



ИНСТИТУТ ЮНЕСКО
ПО ИНФОРМАЦИОННЫМ ТЕХНОЛОГИЯМ
В ОБРАЗОВАНИИ



UNITED NATIONS EDUCATIONAL
SCIENTIFIC AND CULTURAL ORGANIZATION

DIGITAL LIBRARIES IN EDUCATION

Specialized training course

STUDY GUIDE



UNESCO INSTITUTE FOR INFORMATION TECHNOLOGIES
IN EDUCATION

MOSCOW
2007



ОРГАНИЗАЦИЯ ОБЪЕДИНЕННЫХ НАЦИЙ
ПО ВОПРОСАМ ОБРАЗОВАНИЯ, НАУКИ И КУЛЬТУРЫ

ЭЛЕКТРОННЫЕ БИБЛИОТЕКИ В ОБРАЗОВАНИИ

Специализированный учебный курс

ПРАКТИЧЕСКОЕ РУКОВОДСТВО



ИНСТИТУТ ЮНЕСКО ПО ИНФОРМАЦИОННЫМ ТЕХНОЛОГИЯМ
В ОБРАЗОВАНИИ

МОСКВА
2007

ББК 74.58.я.73
И 58

М90 Дэвид Бейнбридж, Иен Виттен, Дэвид Николс
Электронные библиотеки в образовании. Специализированный учебный курс.
Практическое руководство / Авторизированный пер. а англ. – М.: Изд. дом «Обучение-Сервис», 2007, 248 с., 23 илл.

ISBN 978-5-902116-21-9

В сотрудничестве с:

Уэйном Макинтошем – Университет Окленда, Новая Зеландия

Тарикере Б. Раджашекар – Индийский институт науки, Индия

Авторизованный пер. с англ.

Моисеевой М.В.

Специализированный учебный курс Института ЮНЕСКО по информационным технологиям в образовании «Электронные библиотеки в образовании» разработан в рамках межсекторального проекта ЮНЕСКО «Методологии для электронных библиотек», основной задачей которого являлось исследование как существующих, так и перспективных технологий и форм работы с электронными библиотеками с учетом влияния этических, социальных, педагогических, организационных и экономических аспектов их использования на обучение, культурную и научную деятельность.

Этот курс посвящен вопросам использования электронных библиотек в образовании, включая также и другие, быстро развивающиеся сферы их применения. Курс поможет работникам сферы образования научиться создавать собственные коллекции для использования в рамках той предметной дисциплины, которую они преподают. Здесь представлен обзор крупномасштабных национальных и международных электронных библиотек для образования, но основной акцент сделан на использование малобюджетных методов построения и эксплуатации электронных библиотек на уровнях от индивидуального до институционального.

ББК 74.58.я.73
И 58

© Институт ЮНЕСКО по информационным технологиям в образовании, 2007

Мнения, выраженные в данном документе, являются мнениями авторов и не обязательно отражают точку зрения Секретариата ЮНЕСКО.

Дополнительную информацию можно получить по адресу:
Институт ЮНЕСКО по информационным технологиям в образовании
ул. Кедрова, д. 8, корп. 3, Москва, 117292, Российская Федерация
Тел.: 7 495 129 2990 E-mail: info@iite.ru
Факс: 7 495 129 1225 Web: www.iite.ru

ISBN 978-5-902116-21-9 (ЗАО «Издательский дом «Обучение-Сервис»)

СОДЕРЖАНИЕ

Предисловие	6
МОДУЛЬ 1. Концепция электронных библиотек и их роль в образовании.	13
Глава 1.1. Понятие «электронная библиотека»	19
Глава 1.2. Электронные библиотеки в образовании	32
Глава 1.3. Программное обеспечение для создания электронных библиотек Greenstone	44
МОДУЛЬ 2. Формы представления документов	57
Глава 2.1. Документы: исходный материал	63
Глава 2.2. Создание электронной библиотечной коллекции	70
МОДУЛЬ 3. Работа с метаданными.	83
Глава 3.1. Разметка.	90
Глава 3.2. Метаданные	104
Глава 3.3. Образовательные метаданные	120
МОДУЛЬ 4. Мультимедийные электронные библиотеки	139
Глава 4.1. Форматы и стандарты мультимедиа	145
Глава 4.2. Создание гетерогенных коллекций	152
Глава 4.3. Расширение возможности Greenstone	168
МОДУЛЬ 5. Открытые стандарты и примеры из практики	181
Глава 5.1. Стандарты метаданных: METS, MODS, RDF.	188
Глава 5.2. Институционные репозитории и интероперабельность	193
Глава 5.3. Примеры из практики образовательных электронных библиотек	210
ПРИЛОЖЕНИЯ	231
Приложение А. Глоссарий	233
Приложение Б. Библиография, журналы и веб-сайты	238

Предисловие

Уважаемые читатели!

Я рад приветствовать всех вас, кто решил изучить курс «Электронные библиотеки в образовании»!

Электронные библиотеки – это большие организованные коллекции информационных объектов. Качественное программное обеспечение для создания электронных библиотек позволяет даже неспециалистам в этой области придумывать, собирать, грамотно структурировать и распространять коллекции, содержащие новую информацию. И это очень важно, поскольку создание подобных коллекций революционизирует процесс обучения и изменяет форму представления учебных материалов.

Появление Веба, или Всемирной паутины (World Wide Web), меняет взгляды общества на информацию, обеспечивая доступ к беспрецедентным объемам информации для всех и каждого. Конечно, Веб вряд ли можно назвать надежным источником для всеобщего просвещения, а неразборчивое его использование даже опасно и, к сожалению, очень широко распространено. И, тем не менее, надо признать, что Всемирная паутина действительно изобилует легко доступной и высококачественной информацией. Многие образовательные учреждения, международные организации, общественные группы, некоммерческие и благотворительные общества занимаются созданием сайтов, на которых они накапливают и представляют разнообразную информацию.

Обычно рассматриваемая в качестве образовательного ресурса Всемирная паутина имеет целый ряд серьезных недостатков: неравномерный и разный по качеству охват тем, мимолетность и непредсказуемость информации (будет ли она завтра на том же самом месте?). Помимо этого, само использование Всемирной паутины таит в себе определенные опасности (а что если студенты получат доступ к неприемлемой информации?). Но еще большей трагедией является то, что целые сегменты общества вообще лишены права на подобный доступ. В то время как большинство населения в развитых странах мира имеет доступ к Интернету даже из дома, лишь незначительное количество жителей развивающихся стран могут приобщиться к этому информа-

ционному богатству. Электронные библиотеки решают подобные проблемы, предоставляя надежные источники и подходящие материалы. Они дают возможность преподавателям создавать коллекции специально для студентов, с которыми они работают, и включать в них информацию, полученную из разных источников. Они также позволяют использовать альтернативные средства распространения (например, на CD-ROM/DVD, в формате, наиболее часто применяемом при работе с информацией в развивающихся странах).

Этот курс рассказывает о все более расширяющихся сферах использования электронных библиотек в образовании, а также о текущих и перспективных технологиях их создания и распространения. Он показывает педагогам, как построить свои собственные коллекции электронных библиотек для использования в тех курсах, которые они преподают. Курс базируется на примерах больших национальных и международных электронных библиотек, используемых в системе образования, но все же больше ориентирует читателя на низкобюджетные методы построения и поддержки электронных библиотек силами отдельных творческих личностей и самоорганизующихся сообществ педагогов, действующих на различных уровнях – от индивидуального до институционального.

Еще одним ключевым компонентом курса являются широко распространенные международные стандарты. Их значение определяется тем, что они обеспечивают совместимость документов, созданных в различных форматах и полученных из различных источников. Эти стандарты позволяют осуществлять обмен информацией между библиотеками, созданными независимо друг от друга, и обеспечивают основу для координации региональных, национальных и международных стратегий создания и распространения образовательных ресурсов.

В комплект учебных материалов по курсу «Электронные библиотеки в образовании», помимо данного пособия, входят:

- Учебник «Как создать электронную библиотеку»;
- CD-ROM «Электронные библиотеки в образовании» с хрестоматией, содержащей рекомендуемые для дополнительного чтения статьи, сам курс, примеры коллекций и программное обеспечение Greenstone;
- CD-ROM «IMARK» с интерактивным учебным модулем «Оцифровка и электронные библиотеки» (Digitisation and Digital Libraries), который входит в состав учебно-методического комплекса по управлению информацией – Information Management Resource Kit (IMARK), разработан Продовольственной и сельскохозяйственной организацией (ФАО) ООН.

Свободно распространяемое программное обеспечение с открытым кодом является ключевым элементом этой стратегии, и подобная высококачественная программа, поддерживающая создание электронных библиотек, уже существует. В данном курсе используется программа для создания электронных библиотек Greenstone. Именно с помощью этой программы вы научитесь создавать собственные информационные коллекции из имеющегося у вас материала, встраивать в них по желанию материалы, полученные из других электронных библиотек и из Всемирной паутины, а также передавать эту информацию вашим студентам через веб-сайт или самозагружающийся компакт-диск (CD-ROM), или тем и другим способом одновременно.

Также в материалах этого учебного курса даются ссылки на учебник «Как создать электронную библиотеку» (“How to Build a Digital Library”, 2003) Иена Виттена (Ian H. Witten) и Дэвида Бейнбриджа (David Bainbridge). Учебник не входит в комплект учебных материалов по курсу «Электронные библиотеки в образовании», поэтому слушателям данного курса, рекомендуется самостоятельно приобрести эту книгу или взять ее в библиотеке.

Перечисленные ресурсы обеспечивают основной материал для чтения по курсу. Интерактивный диск работает на компьютере с операционной системой Windows. Программа создания электронных библиотек Greenstone, которая используется при проведении практических занятий в рамках курса, может работать в любой из трех операционных систем – Windows, Linux или Macintosh OS/X.

Мы перечислили все то, чем вы будете пользоваться в процессе обучения по данному курсу. Приступая к обучению, вам также будет полезно познакомиться с кратким описанием содержания и целевого назначения курса, структурой учебных материалов и критериями оценки успешности обучения.

Чему вы можете научиться?

Этот курс расскажет вам о том...

- что такое электронные библиотеки,
- как они используются в образовании и других сферах деятельности,
- что такое метаданные, и как с их помощью организовать электронные библиотеки,
- какие бывают форматы представления электронных документов,
- как в электронных библиотеках может использоваться мультимедиа,
- как создавать коллекции электронных библиотек и управлять ими,

- какие стандарты действуют для электронных библиотек и образовательных метаданных,
- какие системы могут использоваться для создания институциональных репозиториев.

Для кого этот курс?

Целевая аудитория курса включает две основные группы:

- работники сферы образования: преподаватели, методисты и тьюторы;
- информационно-библиотечные специалисты, работающие в сфере образования: библиотекари, разработчики электронных библиотек, администраторы информационных систем, руководители соответствующих подразделений образовательных учреждений.

Дополнительной аудиторией могут быть также студенты и научные сотрудники.

Основное внимание в данном курсе уделяется системе образования на уровне средней и высшей школы.

Какой уровень подготовки ожидается?

При разработке курса авторы ориентировались на тех, кто уже имеет некоторый опыт создания учебных материалов для курсов, проводящихся в традиционной форме обучения, а также практический опыт использования компьютера для решения таких задач, как, например, работа с текстом. Помимо этого, мы надеемся, что все слушатели нашего курса являются читателями каких-либо библиотек.

Сколько времени требуется для изучения данного курса?

Нам хотелось, чтобы вы уделите данному курсу около 50 учебных часов, включая время на выполнение заданий, или 70 часов, если английский язык не является вашим родным языком или вы довольно медленно работаете на компьютере. От 10 до 15 часов курса отводятся на выполнение практических заданий по созданию различных электронных библиотечных коллекций и 3 часа на выполнение дополнительных упражнений по выбору обучающихся.

Как организован данный курс?

Курс состоит из пяти модулей, каждый из которых включает два или три раздела. Время, отведенное на изучение модулей, распределяется неравномерно, а примерно так, как показано в приведенной ниже таблице:

Модуль	Номера разделов	Время на изучение
1	3	7 – 10 часов
2	2	7 – 10 часов
3	3	15 – 20 часов
4	3	10 – 15 часов
5	3	10 – 15 часов
Время	3	50 – 70 часов

Какие ресурсы вам потребуются?

Каждый обучающийся должен иметь:

- специализированный учебный курс ИИТО ЮНЕСКО «Электронные библиотеки в образовании»;
- компакт-диск, содержащий дополнительные материалы по курсу «Электронные библиотеки в образовании», включая:
 - программное средство для создания электронных библиотек Greenstone,
 - примеры коллекций, все пособия и т.д.,
 - файлы с примерами для выполнения упражнений,
 - хрестоматию с коллекцией материалов для дополнительного чтения,
 - коллекции материалов, созданных в поддержку учебника «How to Build a Digital Library» (I.H. Witten и D. Bainbridge), (далее везде «Как создать электронную библиотеку»),
 - примеры различных коллекций;
- учебник «Как создать электронную библиотеку»,
- интерактивный компакт-диск: Information Management Resource Kit (IMARK), модуль Digitisation and Digital Libraries,
- доступ к компьютеру (Windows 98 и выше).

Для изучения данного курса доступ к Интернету не требуется, но некоторые вопросы будет проще решать при его наличии.

Процедура оценки

Каждый модуль включает несколько заданий, каждое из которых в свою очередь включает серию вопросов. Основной целью заданий является помощь обучающимся при работе с материалами курса и самопроверка. Если вы изучаете данный курс в рамках учебного семинара, то ваши преподаватели будут использовать вопросы этих заданий для оценки вашей текущей успеваемости. Если вы учитесь самостоятельно на дистанционном курсе, то вам, возможно, придется присылать свои ответы по обычной или электронной почте той организации, которая проводит данный курс.

Авторский коллектив

Разработка данного курса была осуществлена командой специалистов в области теории и практики использования электронных библиотек из Новой Зеландии при поддержке Института ЮНЕСКО по информационным технологиям в образовании (ИИТО). Членами команды являются Дэвид Бейнбридж (David Bainbridge), Шотландия и Новая Зеландия, Дэвид Николс (David Nichols), Англия и Новая Зеландия, а также Иен Виттен (Ian H. Witten), Канада и Новая Зеландия. Мы благодарны профессору Уэйну Макинтошу (Wayne Macintosh), Южная Африка и Новая Зеландия и доктору Тарикере Б. Раджашекару (Dr Tarikere Basappa Rajashekar), Индия, за их помощь на ранних этапах работы по данному проекту. Мы благодарим всех членов команды, работающих по проекту New Zealand Digital Library Project в Университете Вайкато, и особенно – Майкла Дьюснипа (Michael Dewsnip), который подготовил компакт-диск курса, и Шаокун Ву (Shaoqun Wu) за проведение тщательного тестирования материалов курса. Основные идеи были сформулированы в ходе проведения специализированного курса ИИТО ЮНЕСКО «Информационные и коммуникационные технологии в дистанционном обучении», и поэтому мы безмерно благодарны доктору Майклу Дж. Муру (Michael G. Moore) из Пенсильванского университета (Pennsylvania State University), США, который координировал разработку данного курса.

Иен Виттен
Редактор и координатор проекта
New Zealand Digital Library Project

МОДУЛЬ 1

КОНЦЕПЦИЯ ЭЛЕКТРОННЫХ БИБЛИОТЕК И ИХ РОЛЬ В ОБРАЗОВАНИИ

Глава 1.1. Понятие
«электронная библиотека»

Глава 1.2. Электронные
библиотеки в образовании

Глава 1.3. Программное
обеспечение для создания
электронных библиотек
Greenstone



Цель

Сформировать понятие «электронные библиотеки» и дать общее представление о роли, которую играют электронные библиотеки в образовании сегодня и будут играть в перспективе, а также проанализировать примеры из практики применения электронных библиотек.

Задачи

Изучив материалы Модуля 1, вы сможете:

- дать определение понятию «электронная библиотека»;
- убедить директора вашей библиотеки не приступать к полной оцифровке всего имеющегося в наличии библиотечного фонда;
- найти электронные библиотеки во Всемирной паутине;
- объяснить, почему электронные библиотеки особенно важны в развивающихся странах;
- перечислить примеры использования электронных библиотек в образовании и объяснить основные направления их использования;
- написать тезисы выступления на тему о том, как программное обеспечение по созданию электронной библиотеки поможет студентам вашего учебного заведения сформировать необходимые знания и умения;
- объяснить разницу между открытым и коммерческим программным обеспечением;
- установить коллекцию Greenstone на компьютере;
- использовать все возможности читательского интерфейса Greenstone для поиска нужной информации;
- объяснить в контексте данного программного обеспечения значения таких понятий, как документ, коллекция и библиотека;
- назвать три различных способа создания коллекций в Greenstone.

Введение к Модулю 1

Первый модуль объясняет значение понятия «электронная библиотека», показывает направления использования электронных библиотек в образовании и позволяет приобрести опыт работы с электронной библиотечной коллекцией при проведении поиска информации.

Модуль состоит из трех разделов. В первом проводится анализ концепции электронных библиотек, и даются примеры их применения в различных областях, при этом особое внимание уделяется культурным, историческим и гуманитарным аспектам их применения, а также технологии их создания. Второй раздел рассказывает об основных направлениях использования электронных библиотек в образовании и проводит анализ потенциальных возможностей подобных библиотек. Третий раздел представляет программу создания электронных библиотек Greenstone, которая используется как основное программное обеспечение на протяжении всего курса.

Этот модуль, как и другие модули курса, включает задания, которые помогут вам поразмышлять об электронных библиотеках и их роли в образовании. Эти задания также могут быть использованы преподавателем для оценки ваших знаний по курсу.

Модуль 1. Материалы для чтения

Глава 1.1. Понятие «электронная библиотека»

Материал для чтения 1. Учебник «Как создать электронную библиотеку»: Предисловие и все разделы Главы 1 за исключением Раздела 1.4.

Назначение. Этот вводный материал объясняет концепцию электронной библиотеки и рассказывает об эволюции библиотеки на протяжении столетий в историческом контексте. В материал также включено несколько иллюстраций. Показана роль электронных библиотек в развивающихся странах и дан обзор проблем этики и авторского права. При чтении этой главы подумайте о том, как эти определения и сценарии соотносятся с реалиями той библиотеки, которая вам хорошо знакома.

Материал для чтения 2. CD-ROM IMARK: Раздел 1. Обзор концепций (4 урока).

Назначение. Изучить обширный материал для дальнейшего знакомства с электронными библиотеками, включающий множество примеров и детально рассматривающий вопросы практического применения авторского права. Этот материал рекомендуется для дополнительного изучения, поскольку он расширит ваше представление об опыте использования электронных библиотек.

Глава 1.2. Электронные библиотеки в образовании

Материал для чтения 1. CD-ROM «Электронные библиотеки в образовании», Хрестоматия, перевод статьи Гари Марчионини и Германа Море «Значение цифровых библиотек для преподавания и обучения». Marchionini, G. and Maurer, H. (1995) The Roles of Digital Libraries in Teaching and Learning. Communications of the ACM (1995), Vol. 38, No. 4, pp. 67–75.

<http://www.ils.unc.edu/~march/cacm95/cacm.html>

Назначение. Эта статья описывает процесс эволюционного развития электронных библиотек, направленный на удовлетворение существующих у преподавателей и учащихся потребностей, а также намечает пути дальнейшего их совершенствования. В процессе чтения статьи задумайтесь о том, какие виды учебных ресурсов вы используете в настоящее время и что бы вы смогли сделать, если бы вам стали доступны значительно большие объемы подобных ресурсов.

Материал для чтения 2. CD-ROM «Электронные библиотеки в образовании», Хрестоматия, перевод статьи Ганса Роеса «Электронные библиотеки и образование: общие направления и перспективы развития». Roes, H. (2001) Digital Libraries and Education: Trends and Opportunities.

<http://www.dlib.org/dlib/july01/roes/07roes.html>

Назначение. Эта статья знакомит вас с широким спектром стратегических, политических и организационных вопросов, с которыми сталкиваются библиотекари и преподаватели, активно использующие информационные и коммуникационные технологии в высшем образовании.

Глава 1.3. Программное обеспечение для создания электронных библиотек Greenstone

Материал для чтения 1. Учебник «Как создать электронную библиотеку», Раздел 1.4.

Назначение. Этот материал кратко описывает программу создания электронных библиотек Greenstone. Когда вы будете с ним знакомиться, вспомните основные направления применения электронных библиотек в образовании, которые вы изучали при чтении Раздела 1 и решите, в какой степени Greenstone может способствовать развитию этих направлений.

Материал для чтения 2. CD-ROM «Электронные библиотеки в образовании».

Назначение. Данный CD-ROM содержит все материалы для чтения по курсу, включая данный специализированный учебный курс, причем в формате, удобном для проведения полнотекстового поиска, и иллюстрирует другие возможности коллекций, созданных при помощи Greenstone. Коллекция способствует выполнению двух задач одновременно: она позволяет вам приобрести некоторый опыт в просмотре содержимого электронной библиотеки, а также познакомиться с материалами, предназначенными для дополнительного чтения.

Материал для чтения 3. CD-ROM «Электронные библиотеки в образовании»: коллекция «Как создать электронную библиотеку» на компакт-диске с данным курсом.

Назначение. Эта коллекция включает указатель по пособию и цветные версии всех рисунков. Она наглядно демонстрирует необычайное разнообразие коллекций Greenstone и одновременно предоставляет собой полный указатель по этому ресурсу, поэтому вы сможете обращаться к ней за консультацией во время прохождения курса.

Материал для чтения 4. CD-ROM IMARK: Урок 6. Учебник по Greenstone.

Назначение. Этот урок самообучения знакомит вас с различными аспектами работы с программой Greenstone. Изучите его для того, чтобы разобраться с вопросами инсталляции, особенностями интерфейса и вариантами создания коллекций.

Глава 1.1. Понятие «электронная библиотека»

Прежде чем приступать к дальнейшему чтению материалов, поразмышляйте о том, что такое электронная библиотека и чем она отличается от традиционной, с одной стороны, и от Всемирной паутины (World Wide Web), с другой.

Задание 1

Подумайте о библиотеке, находящейся в вашем университете, или о какой-то другой библиотеке, которую вы хорошо знаете.

- С какими проблемами можно столкнуться при создании электронной библиотеки с тем же самым содержательным наполнением?
- Что выиграют и в чем проиграют пользователи электронной библиотеки в сравнении с пользователями традиционной библиотеки?
- При условии ограниченности доступных вам ресурсов, какие именно материалы вы будете оцифровывать в первую очередь?
- Как бы мог выглядеть электронный аналог каталога традиционной библиотеки?
- Какие дополнительные возможности поиска и просмотра может предложить электронная библиотека?
- Чем будет отличаться ваша электронная библиотека от Всемирной паутины?
- Почему людям, имеющим доступ к WWW, также нужны и электронные библиотеки?

Если вы не можете дать уверенный ответ на некоторые из этих вопросов из-за того, что у вас еще нет опыта использования электронных библиотек, не волнуйтесь – мы вернемся к этим вопросам еще раз в конце данного раздела. Итак, продолжим...

Электронные библиотеки для образования, как и любые другие электронные библиотеки, могут существовать в различных масштабах. Существующие примеры подобных библиотек являются лишь первыми нерешительными шагами в этом направлении, поскольку технологии создания электронных библиотек еще только начинают становиться доступными для отдельных людей и организаций. Как пойдут дела

в будущем будет зависеть от отдельных творческих личностей, а также от значительных совместных усилий и командной работы. Став студентом этого курса, вы также будете способствовать формированию и развитию электронных библиотек в образовании: так как вы тоже становитесь ее творцом.

Задание 2

Прочитайте предисловие и начало Главы 1 (с. 1–5) учебника «Как создать электронную библиотеку» и напишите по одному абзацу текста, в котором выразится ваше собственное видение того, какой должна быть:

- крупномасштабная национальная электронная библиотека;
- образовательная электронная библиотека среднего масштаба, созданная на базе одного из факультетов университета;
- небольшая электронная библиотека, самостоятельно созданная в ходе учебного курса одним из студентов.

Существует множество определений понятия «электронная библиотека», и мы приводим их ниже. Когда вы будете знакомиться с этими определениями, задумайтесь о том, что их всех объединяет. Помимо этого, выделите черты сходства и отличия электронных и традиционных библиотек.

Первые два определения основываются на том, что электронная библиотека – это то же самое, что и традиционная библиотека или традиционная система поиска информации, но только представленная в цифровом формате.

- *Библиотека, которая кодирует журналы, книги и информацию в цифровом формате.*
- *Коллекция текстов, рисунков и пр., закодированная с целью хранения, поиска и чтения с помощью компьютера.*

Другие определения делают акцент на функциях традиционной библиотеки, но идут дальше, оговаривая тот факт, что в электронных библиотеках могут быть предложены новые услуги. Второе из приведенных здесь определений предусматривает довольно полный набор услуг, который включает и сохранение информации («защиту») – основную заслугу всех традиционных библиотек, не упомянутую, хоть это и удивительно, ни в одном из других определений.

- *Коллекция цифровых версий информационного содержания, аппаратное, программное обеспечение, персонал, поддерживаю-*

щий функции традиционной библиотеки, также выполняющий интеллектуальные операции, такие, как поиск, просмотр и навигацию.

- *Интегрированный набор услуг по получению, каталогизации, хранению, поиску, защите и извлечению информации.*

В некоторых определениях акцент делается на технических характеристиках: размере, масштабе, распределенной структуре, доступе к сети. Второе определение подчеркивает возможность сохранения динамически меняющейся информации, т.е. того, что не может делать традиционная библиотека.

- *Коллекция, включающая значительное количество цифровых объектов, представляющая все виды материалов и медиасредств, которые хранятся в распределенных информационных репозиториях, доступ к которым осуществляется по национальным компьютерным сетям.*
- *Большая коллекция информации, которая представлена в цифровом формате. Электронная библиотека может содержать документы, рисунки, звуковые обозначения и информацию, полученную в процессе выполнения определенных действий (например, изображение, получаемое с постоянно передающего информацию метеоспутника).*

Другие определения отражают тесную взаимосвязь электронной библиотеки с WWW. Является ли Всемирная паутина электронной библиотекой или нет? — мы вернемся к этому вопросу после выполнения следующего задания.

- *Электронные библиотеки могут включать справочный материал или ресурсы, доступ к которым может осуществляться по Всемирной паутине. Оцифрованные фрагменты библиотечной коллекции или оригинального материала, произведенного непосредственно для Веба, могут также включаться в электронную библиотеку.*

Родственным термином является понятие «виртуальная библиотека», хотя можно было бы считать этот термин синонимом «электронной библиотеки», поскольку в обоих терминах делается акцент на неосязаемую природу хранящегося в библиотеке материала. Это объясняет более частое использование этого термина для обозначения информационного портала, созданного в электронной форме и доступного откуда угодно.

Задание 3

Прочитайте Разделы 1.1 и 1.2 (с. 5–20) учебника «Как создать электронную библиотеку» и напишите эссе, в котором будет изложено ваше мнение о том, смогут ли электронные библиотеки конкурировать, сосуществовать или же дополнять традиционные библиотеки в сфере образования в обозримом будущем – лет, скажем, через 20 (предположительный объем эссе: 500 слов).

Дополнительное упражнение 1

Учебный модуль *IMARK* содержит значительный объем информации об электронных библиотеках. В связи с этим мы рекомендуем вам проработать Уроки 1.1 и 1.2 указанного модуля. Первый определяет концепцию электронных библиотек, обсуждает их преимущества и приводит краткое описание процессов, лежащих в основе их создания и использования. В следующем приводится множество примеров электронных библиотек, существующих в настоящее время. Работая с этим материалом, вы больше узнаете об электронных библиотеках и приобретете опыт их использования. Большинство компьютеров запускают установку программы, которая требуется для работы с учебным модулем на CD-ROM *IMARK*, автоматически. (А если точнее, то они запускают программу автозапуска *AutoPlay*: как правило, она имеется на каждом компьютере). Если же с вашим компьютером ситуация другая, и в тот момент, когда вы вставляете диск, ничего не происходит, то вам нужно открыть папку Мой компьютер (*My Computer*), затем выбрать и открыть дисковод, который определен под CD-ROM (это, как правило, дисковод *D:*), а затем запустить файл *IMARK_VI0Een*, сделав двойной щелчок по его названию.

Как только начнется процесс установки, программа задаст вам серию вопросов, касающихся целей инициализации. Для быстрого запуска программы выполните следующее:

- Щелкните **<START COURSE>**
- Щелкните **<Start Learning>**
- Примите все предложенные условия
- Заполните необходимые данные в персональной анкете учащегося
- Щелкните **<Save>**
- Зарегистрируйтесь, если вы работаете в режиме онлайн (рекомендуется)
- Выберите **<Start course>**
- Откройте Урок 1.1 ...

Итак, при размышлении о том, является ли World Wide Web электронной библиотекой, мы приходим к выводу, что нет. Как объясняется в материалах для чтения к Заданию 3, WWW недостает двух исключительно важных элементов: организации и разборчивости. Библиотека — это организованная коллекция, причем организованная в соответствии с некоторыми определенными принципами, а не просто аналог Всемирной паутины с такими же, как и в ней, запутанными ссылками. Отбирается информация в библиотеке также по определенным критериям, независимо от того, проговорены они заранее или нет. В каком-то смысле библиотеки представляют собой управляемые (или «курируемые кем-то») коллекции, которые буквально «находятся под присмотром» ответственного руководителя, в роли которого может выступать, например, менеджер. За WWW никто вообще не следит.

С самого раннего периода существования Всемирной паутины люди пытались привнести в нее некоторый порядок и сделать более удобной для поиска ресурсов путем развития поисковых систем или как их еще называют — «движков». Сейчас мы представим вам некоторые из наиболее популярных поисковых систем Веб.

Пример из практики: Поисковая система Google —
<http://www.google.com>

За годы своего существования, то есть с момента запуска в 1998 году и до того времени, как мы об этом пишем, поисковая система Google является одним из наиболее быстро растущих культурных феноменов, когда-либо известных в мире, и одним из его наиболее экономически прибыльных предприятий.

Первоначально поисковая система просто создавала копию всей Всемирной паутины, или, по крайней мере, той ее части, которую она могла охватить. (Считается, что система Google создает несколько резервных копий, распределенных по серверам, находящимся в различных частях США.) Эта копия хранилась до момента ее периодического обновления путем повторного копирования из Веба, что происходило, возможно, каждые несколько недель. На основе этого принципа строится указатель (индекс), который записывает каждое слово, появляющееся во всем тексте, и количество веб-страниц, на которых оно появляется. Огромная по своим размерам структура данных позволяет обнаруживать все страницы, содержащие какое-либо конкретное слово или набор слов. Когда пользователь вводит свой запрос в виде определенного набора слов, система выдает в ответ список с «результатами поиска», то есть список страниц, на которых присутствуют все эти слова.

На большую часть сделанных запросов выдается список, включающий колоссальное количество подходящих страниц. Успех Google складывается из двух основных инноваций. Первая – это способ сортировки результатов, выдаваемых пользователю. Система первым делом анализирует структуру связей между веб-страницами для определения их общего количества, а, следовательно, важности или ранга каждой страницы в отдельности. Ранг определяется количеством страниц, которые на нее ссылаются, то есть, другими словами, делают ее популярной, и рангом самих этих страниц. Идея такова: если несколько высоко котирующихся страниц указывают на вашу страницу, то ее категория будет также высокой.

Вторая ключевая инновация Google основывается на наблюдении, что страница чаще всего и лучше всего описана именно на тех страницах, которые на нее указывают. Точно так же, как и ваше мнение о себе самом менее ценно, чем мнение о вас других людей, существует колоссальная разница между тем, что утверждается на самой веб-странице, например, что на ней дается «хорошее» представление какой-то определенной темы, и похожими утверждениями по поводу этой страницы, которые даются на других страницах, указывающих на нее. Google использует текст, ассоциированный со ссылкой, называющийся «якорным текстом», что является важной подсказкой по содержанию той страницы, на которую указывает эта ссылка.

Google постоянно стремится к тому, чтобы сделать свою систему более удобной, особенно для непрофессиональных пользователей, многие из которых воспринимают ее в качестве парадной двери в Интернет. Она предоставляет возможность искать информацию в системе, которая еще во многом продолжает оставаться анархичной.

Хотя полнотекстный поиск, упрощенный поисковыми системами Всемирной паутины, может значительно облегчить его, он менее тщателен и, в целом, менее удобен, чем та организационная система, которая является стандартной для библиотек с их указателями по авторам, названиям, темам, ключевым словам, аннотациям и т.д. Более того, специализированная тематическая коллекция Веб, собранная путем подбора страниц в соответствии с определенными принципами поиска и организации, конечно же, тоже может быть электронной библиотекой. В этом курсе вы научитесь создавать подобные коллекции.

Электронные библиотеки имеют особое значение для развивающихся стран. Одной из причин является то, что доступ к традиционным источникам информации, например, книгам, там зачастую

очень затруднен, а электронные библиотеки дают возможность публиковать данные коллекции в сети. Другой причиной является то, что доступ к сети Интернет в развивающихся странах, как правило, низкоскоростной, что еще больше увеличивает неравенство в получении знаний в развивающихся и развитых странах. Технологии электронных библиотек могут улучшить сложившуюся ситуацию, потому что несмотря на негативное отношение к этому многих людей, распространение информации не должно зависеть только от Интернета.

Развивающиеся страны являются благодарными пользователями электронных библиотек. Но для устойчивого развития этих стран в условиях цифровой революции чрезвычайно важно, чтобы они не превратились в социальные группы, состоящие исключительно из «читателей» уже готовой информации. Одним из путей решения этой проблемы является обеспечение участия этих стран в производстве информационных коллекций. Если вы живете в развивающейся стране, то этот курс поможет вам научиться создавать свои собственные электронные библиотечные коллекции для образования. Вы научитесь делать такие коллекции доступными в Веб на глобальном уровне и размещать их на CD-ROM или DVD, составив, таким образом, свою собственную информационную коллекцию, которую можно было бы просматривать даже на самом скромном по оснащению компьютере.

Задание 4

Прочитайте Раздел 1.3. (с. 20–24) учебника «Как создать электронную библиотеку». Исходя из имеющегося у вас опыта, объясните, как отсутствие доступа к информации может негативно отразиться на образовании, а также выделите несколько различных направлений применения электронных библиотек в сфере образования, наиболее значимых для развивающихся стран.

Вопросы копирайта обычно являются наиболее проблемными для электронных библиотек. Законы авторского права в разных странах разные, и если у вас есть вопросы по практике применения этих законов в конкретных ситуациях, то вам потребуются помощь юриста, специализирующегося в авторском праве.

Дополнительное упражнение 2

CD-ROM IMARK: Урок 1.3. «Авторское право: правовые основы и законодательная база» и 1.4 «Вопросы авторского права и их реализация в условиях библиотек». *Мы рекомендуем вам внимательно поработать над уроком 1.3. Урок 1.4. не включен в данный курс, но если у вас есть возможность, вы должны с ними познакомиться. Эти материалы помогут вам детально разобраться в практических вопросах использования авторского права.*

Задание 5

Прочитайте Раздел 1.5 (с. 28–35) учебника «Как создать электронную библиотеку».

Поисковая система Google (см. рисунок ниже) создает отдельную копию для каждой из обнаруженных ею веб-страниц, что служит основой для последующего индексирования. Когда пользователь щелкает по ссылке в результатах поиска, Google сразу переводит его на соответствующую страницу Веба. Далее пользователю предоставляется возможность щелкнуть по вторичной ссылке, чтобы увидеть частную копию этой страницы, хранящуюся в буфере Google, что особенно удобно в случае, если исходная страница уже больше не существует. Когда Google показывает копию, он предваряет ее четким предупреждением о том, что это не исходная веб-страница, также дает возможность загрузить нужные данные. Например, ниже приведено предупреждение Google по странице www.greenstone.org:

This is Google's cache of <http://www.greenstone.org/> as retrieved on 3 Apr 2005 03:21:26 GMT. Google's cache is the snapshot that we took of the page as we crawled the web. The page may have changed since that time. Click here for the [current page](#) without highlighting. This cached page may reference images which are no longer available. Click here for the [cached text](#) only. To link to or bookmark this page, use the following url:
http://www.google.com/search?q=cache:vsSEdeaNiAJ:www.greenstone.org/+greenstone&hl=en&lr=lang_en

Google is not affiliated with the authors of this page nor responsible for its content.

Прокомментируйте политику Google с этической и юридической точки зрения.

Заключение

Добравшись до этой страницы курса, вы уже изучили значительный объем основополагающей информации об электронных библиотеках. В материалах для чтения к Заданию 2 изложены четыре сценария, цель которых развеять миф о том, что электронные библиотеки являются ничем иным, как простой разновидностью традиционных библиотек, но с байтами вместо книг. В материале для чтения к Заданию 3 обсуждалась концепция электронной библиотеки в историческом контексте эволюционного развития библиотеки на протяжении столетий. Курс, который вы изучаете, ориентирован на применение электронных библиотек в развивающихся странах, поэтому в следующем материале для чтения (Задание 4) рассматривалось использование различных электронных библиотек в этом контексте. Заключительный материал для чтения был посвящен сложному вопросу авторского права и этических проблем, связанных с повторным использованием материала.

Давайте вернемся к вопросам, которые мы поднимали в самом начале, в Задании 1, и дадим краткий комментарий, приведя по этим вопросам нашу собственную точку зрения. Не волнуйтесь, если материал для чтения к этой Главе изменил ваши взгляды — ведь именно для этого образование и существует! Мы предложили подумать о вашей университетской библиотеке (или о какой-либо другой библиотеке, которую вы хорошо знаете), задав такие вопросы:

- *С какими главными проблемами можно столкнуться при создании электронной библиотеки с тем же самым содержательным наполнением?*

Оцифровка сразу всего библиотечного фонда была бы непростительной ошибкой. Одной из наиболее серьезных проблем в этом случае стало бы выполнение нетривиальной задачи по переводу всего библиотечного фонда в цифровой формат путем сканирования материалов и сохранения их на компьютере. Книги проще отсканировать, если их можно разброшюровать, то есть освободить от корешков, но, возможно, есть и какой-то менее разрушительный способ работы с первоисточниками. Работа со старыми и хрупкими материалами займет особенно много времени. Но самая серьезная проблема возникнет в связи с юридической обоснованностью всего этого предприятия. Как правило, авторское право не распространяется лишь на небольшой процент (если вообще такое возможно) библиотечного фонда, и вряд ли библиотеке удастся получить разрешения от всех владельцев авторских прав на оцифровку этих ресурсов. Мы надеемся, что изуче-

ние данной Главы убедило вас в том, что электронные библиотеки не являются просто оцифрованными версиями реально существующих библиотек.

- *Что выиграют и в чем проиграют пользователи электронной библиотеки в сравнении с пользователями традиционной библиотеки?*

Когда электронная библиотека становится доступной в Вебе, она открывает свои фонды для многочисленных новых читателей. Главным преимуществом, которое получают читатели подобной библиотеки, будет удобство удаленного доступа — с рабочего места, из дома (если у них есть доступ к Интернету) или, возможно, через филиал библиотеки. Другим преимуществом будет то, что читатели смогут по-новому проводить поиск информации в библиотеке, например, проводить полнотекстовый поиск. Большинство из нас не любит читать книги в режиме онлайн, поэтому, если коллекция была уничтожена в процессе оцифровки (смотрите выше), читатели библиотеки понесут невосполнимые потери. Более того, высокая стоимость создания электронной библиотеки означает со всей очевидностью то, что ресурсы были перераспределены с других проектов, которые, возможно, были еще более важными. И снова повторим, что вряд ли преимущества онлайн-библиотеки перевесят те колоссальные затраты, которые будут необходимы для ее создания.

- *При условии ограниченности доступных вам ресурсов, какие именно материалы вы будете оцифровывать в первую очередь?*

В Материалах для чтения к этой Главе вы не найдете прямого ответа на данный вопрос. Ответ в значительной степени зависит от решений, принимаемых на местном уровне. Приоритеты должны быть отданы материалу, созданному именно в указанном сообществе, и который должен быть защищен авторским правом, в особенности в отношении тех материалов, которые больше всего пользуются читательским спросом — например, должны быть включены конспекты по учебным курсам или объемный методический материал. Оцифровка может использоваться для дальнейшего продвижения и рекламы, то есть для привлечения внимания к тому, чем более всего сильна данная библиотека, например, уникальными коллекциями первоисточников. Это также может сократить частоту передачи в пользование хрупких оригиналов, особенно если они очень часто запрашиваются. Другими аргументами являются совместное использование фондов в партнерстве с другими библиотеками и совместное освоение доступных источников финансирования.

- *Как мог бы выглядеть электронный аналог каталога существующей традиционной библиотеки?*

Рассматриваемая Глава не исследовала вопросов, связанных с пользовательским интерфейсом напрямую, хотя вы увидите немало таких примеров в последующих разделах. Обратите внимание, что во многих современных библиотеках — и практически во всех библиотеках в развивающихся странах — каталоги в электронной форме уже существуют. Это создает необходимые условия для работы с картотеками, в том числе для осуществления поиска и просмотра источников по заголовку, автору и теме. Как только читатели библиотеки осваивают такую работу, большинство из них начинают отдавать предпочтение именно электронным каталогам. Большим недостатком библиотеки, целиком представленной в электронном формате (хоть на самом деле это происходит не по вине каталога) является потеря интуитивной ориентации читателя из-за отсутствия возможности привычно просмотреть книги на полках. В реальной (физической) библиотеке читатели используют каталог для поиска конкретной книги, а затем зачастую просматривают и те книги, которые соседствуют с ней на книжной полке, что считается очень удобным. В электронной библиотеке это не так то просто сделать.

- *Какие дополнительные возможности поиска и просмотра может предложить электронная библиотека?*

Эта Глава привлекла ваше внимание к полнотекстному поиску, который является основой таких поисковых систем, как Google. Многие электронные библиотеки также предлагают пользователям эту услугу. Существует большое количество других возможностей, которые здесь не были упомянуты, такие, например, как проведение морфологического поиска по документам библиотеки или автоматическое создание списков акронимов и их определений, отобранных из тех же самых документов. Вы встретите эти примеры в Разделе 1.3.

- *Чем будет отличаться ваша электронная библиотека от World Wide Web?*

Вы уже узнали, что электронные библиотеки представляют собой специализированные по теме коллекции ресурсов, которые были тщательно подобраны и хорошо организованы, что отличает эти библиотеки от WWW, в который любой человек может добавить любую информацию, а сама организация совершенно бессистемна. Электронные библиотеки имеют границы, а Всемирная паутина — нет. И как было сказано выше, для описания этого отли-

чия иногда используется слово «курируется», которое может быть применимо к электронным библиотекам, но уж точно не к Вебу. И еще электронные библиотеки могут доставлять информацию тем людям, у которых нет доступа к WWW.

- *Почему людям, имеющим доступ к World Wide Web, нужны электронные библиотеки?*

Может так случиться, что информация, имеющаяся в электронной библиотеке, уже существует во Всемирной паутине — в большинстве названных выше коллекций именно так и происходит. Однако притом, что полнотекстовый поиск является мощным средством обнаружения информации, зачастую в WWW очень сложно найти то, что вам нужно. Более того, очень сложно убедиться в том, что найденное представляет собой авторитетный источник информации. Как и тематические коллекции, составленные из заранее отобранных материалов, электронные библиотеки, как правило, лучше и надежнее с точки зрения достоверности источников информации, чем большинство источников Всемирной паутины.

В заключение хотелось бы дать еще один комментарий по этому Заданию: всегда можно найти более целесообразные направления использования ресурсов, которые предполагается потратить на решение фундаментальной задачи по оцифровке всей библиотеки. В материалах для чтения представлено несколько примеров таких решений, например: описание Библиотеки по развитию гуманитарных наук (Humanity Development Library), история организации библиотеки Катайи (Kataayi) в сельской части Уганды, история создания онлайн-архивов исследовательских материалов по физике, инициативы по сохранению языка и традиций народности зиа пуэбло (Zia Pueblo) в Нью Мехико и созданию библиотеки популярной музыки. В этих материалах приводятся краткие описания коллекций, посвященных борьбе со стихийными бедствиями, а также коллекций, способствующих сохранению культуры коренных народов, собранных на основе информации, почерпнутой из местных источников. В данном Разделе и прилагаемых к нему материалах для чтения особо подчеркивается, что электронные библиотеки заметно улучшают доступ к информации, поскольку перспективные возможности открываются именно тогда, когда формируются потребности.

Первая глава Модуля 1 помогла вам организовать свою работу с разнообразными источниками информации. Подводя итоги проделанной вами работы, отметим, что вы:

- поразмышляли о том, что такое электронная библиотека и в чем ее отличие от традиционной библиотеки и ее оцифрованной версии,
- обсудили некоторые из существующих взглядов на будущее электронных библиотек в образовании,
- подумали над собственным определением понятия «электронная библиотека» и познакомились с несколькими различными определениями, сделанными до вас,
- установили, чем электронные библиотеки отличаются от World Wide Web,
- узнали о поисковых системах Всемирной паутины, об их сильных и слабых сторонах, проявляемых при поиске информации,
- узнали, какую роль играют электронные библиотеки в развивающихся странах,
- рассмотрели этические и юридические основы авторского права.

Теперь вы готовы сделать первый шаг в изучении электронных библиотек и их использования в образовании. Пожалуйста, переходите к Главе 1.2.

Глава 1.2. Электронные библиотеки в образовании

Пожалуйста, начните работу по этой главе с размышлений о том, как и с какими документами вы имеете дело, когда вы учитесь сами или учите других людей.

Задание 1

Подумайте о каком-то конкретном курсе (который вы преподаете или с которым вы хорошо знакомы) с позиции студента:

- Какие информационные ресурсы нужны студентам?
- Как будут развиваться общеучебные навыки студентов, если они получат доступ к специально отобранной в Вебе тематической информации?
- Как изменится жизнь студентов, если они получат доступ к информационным ресурсам, которые вы сами для них подготовили?
- Какие именно преимущества получают студенты, если они смогут создавать собственные коллекции ресурсов по тому или иному курсу?
- Как электронные библиотечные коллекции могут помочь студентам решить проблему использования материала, накопленного ими в период обучения?
- Как бы мог выглядеть стандартный информационный ресурс, созданный в поддержку курса, похожего на «Электронные библиотеки в образовании»?

Электронные библиотеки могут сыграть заметную роль в образовании, если будут использоваться в качестве:

- среды обучения (опыт студента);
- источника достоверной информации (опять же, в поддержку опыта студента);
- ресурса для обучения (разработка курса).

В Задании 1 вам было предложено сосредоточить внимание на опыте студента, что соответствует первым двум пунктам. Выполнение третьего подразумевает привлечение учителей, которые преподают тот или иной предмет студентам, а также сами, вместе со студентами, пользуются услугами одной и той же библиотеки. Задачи, которые

они решают при разработке курсов, сильно отличаются от тех, что стоят перед студентами, обучающимися по этим курсам.

Все библиотеки — традиционные и электронные — являются типичными организациями по обслуживанию потребителей, деятельность которых основывается на общих для всех них фундаментальных требованиях. Изучая направления применения электронных библиотек в образовании, мы должны рассмотреть потребности различных групп пользователей или организаций, а именно:

- студентов,
- учителей,
- структур, принимающих решения, включая правительства и организации, разрабатывающие образовательные стандарты.

Каждая из этих групп предъявляет свои требования к содержанию и форме организации образовательных электронных библиотек. В этой главе рассматриваются требования студентов и преподавателей. Главы 3.2 и 5.3 исследуют вопрос, как стандарты метаданных могут поддержать существующие потребности обучаемых и руководителей образовательных учреждений.

Что отличает электронные библиотеки от традиционных? Назовем несколько отличий:

- актуальное содержание: в электронной библиотеке гораздо проще обновлять материал;
- содержание получено из первичных источников: например, Американское аэрокосмическое агентство NASA распространяет актуальные научные данные о Земле и космосе, которые могут использовать для своих нужд студенты и ученые;
- полнота представления содержания: информация легко достижима как вглубь, так и вширь;
- содержание может быть представлено в различных форматах (изображения, карты, аудио, видео, потоки цифровых данных и т.д.);
- доступ к содержанию упрощен: новые данные можно распространять с помощью компьютерных сетей, что гораздо проще, чем обновление бумажных копий,
- содержание можно напечатать: студенты могут без труда создать свои собственные материалы и опубликовать их в электронной библиотеке.

К этому списку можно было бы добавить еще два преимущества:

- возможность повторного использования ресурсов: учителя могут совместно использовать ресурсы, что гораздо эффективнее, чем использование традиционных печатных изданий;
- простота интеграции: чем больше работы студентам приходится выполнять на компьютере, тем проще им становится включать в создаваемые ими работы материалы, полученные из внешних источников.

Простота интеграции дает возможность сделать курсовую работу более содержательной, но вместе с тем несет определенную угрозу, так как случаи плагиата становятся для студентов совершенно обычным делом. В Главе 3.3 Модуля 3 мы увидим, что вопросы авторского права затрагиваются каждый раз, когда речь заходит об образовательном контенте.

Задание 2

Прочитайте статью «Значение цифровых библиотек для преподавания и обучения» Гари Марчионини и Германа Море. (Вы сможете найти эту статью, наряду с другими, которые предлагаются для изучения в рамках данного курса, в Хрестоматии, включающей материалы для дополнительного чтения по курсу «Электронные библиотеки в образовании»; она также размещена на компакт-диске с материалами курса. Если вы хотите работать с электронной версией статьи, то сразу перейдите к выполнению практических упражнений Главы 1.3 чтобы узнать, как это делается, а затем вновь вернитесь к этому заданию.)

Ответьте на вопросы с позиции учителя, разрабатывающего свои уроки:

- Какие виды ресурсов оказались бы для вас наиболее полезными?
- Как можно проследить за использованием цифровых ресурсов? Является ли плагиатом возможность легко скопировать документ?
- Из каких источников в настоящее время учителя подбирают материалы для уроков?

Материалы из Интернета могут использоваться в образовательных целях, но не стоит забывать о том, что они представлены не в такой жестко контролируемой среде, как материалы электронной библиотеки. Учителям, скорее всего, понравятся ограничения, существующие для студентов, работающих с курируемыми коллекциями, поскольку заложенные в эти коллекции средства могут:

- определять ресурсы, предназначенные для той или иной конкретной учебной деятельности, выполняемой студентом;
- ограничивать доступ к некоторым ресурсам в зависимости от возраста студента или конкретного содержания;
- отслеживать использование ресурсов.

Задание 3

Прочитайте статью «Электронные библиотеки и образование: общие направления и перспективы развития» Ганса Роеса (эта статья также находится в сборнике материалов для дополнительного чтения и на компакт-диске с курсом).

- Как можно более эффективно организовать контент, предназначенный для учителей?
- Что вам больше нравится: небольшие уроки по узкой теме или большие предметно-ориентированные коллекции?

Организация контента для учителей в образовательных электронных библиотеках зависит от элементов метаданных коллекции. Образовательные метаданные создаются на основе вопросов, понятных и удобных для учителей, поскольку метаданные многих больших библиотек предназначены для выхода на более широкую аудиторию, а именно на публику в целом. Категоризация, которая обеспечивает доступ только для одной группы пользователей, может быть неудобной для другой группы, в которой могут оказаться и учителя.

Пример из практики: DLESE – <http://www.dlese.org/>

Электронная библиотека по образованию в сфере наук о Земле (Digital Library for Earth System Education – DLESE) обеспечивает:

- доступ к **высококачественным коллекциям образовательных ресурсов**;
- доступ к **наборам цифровых и графических данных**, включающим инструменты и интерфейсы, которые позволяют эффективно использовать их в рамках учебного процесса;
- **поддержку пользователей**, преподавателей и обучающихся, которые создают, используют или обмениваются образовательными ресурсами.

Коммуникационные сети обеспечивают взаимосвязь и взаимодействие между всеми уровнями образования в сфере наук о Земле.

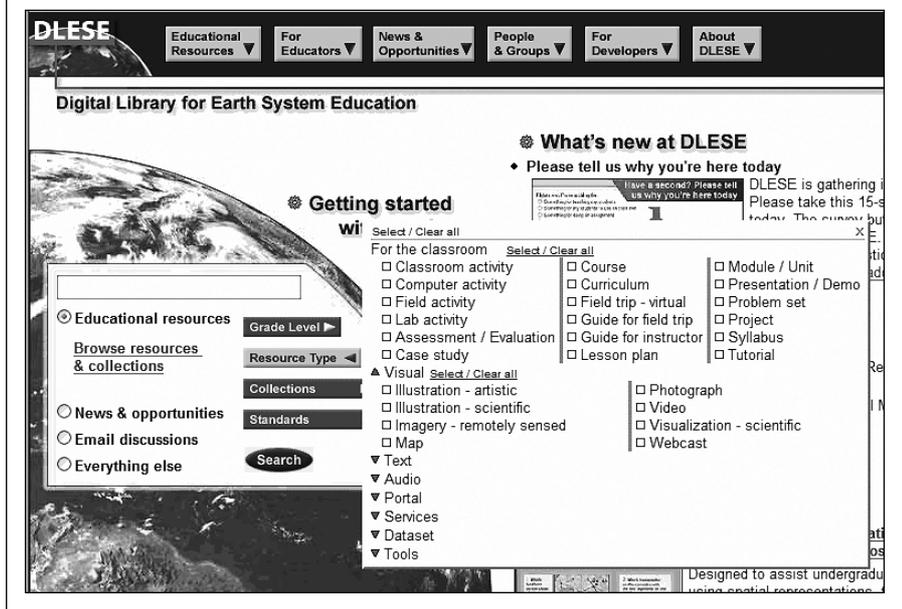
Ресурсы DLESE включают электронные материалы для учителей и учащихся, такие как: планы уроков, карты, изображения, наборы данных, наглядные пособия, контрольные задания, расписания занятий и онлайн-курсы.

Материал сгруппирован по определенным критериям, например:

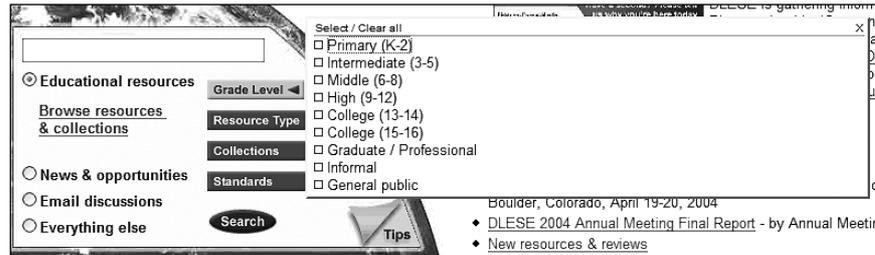
- Класс/уровень обучения
- Тип ресурса
- Вид деятельности (например, экскурсия, работа в компьютерном классе, тьюториал)
- Форма представления информации:
 - Визуальная (карты, видео, фотографии)
 - Текстовая (аннотации, списки литературы, словари)
 - Аудио/звуковая (музыка, аудиозаписи рассказов)

Познакомьтесь с иллюстрациями, показывающими различные организационные схемы, предлагаемые данной системой.

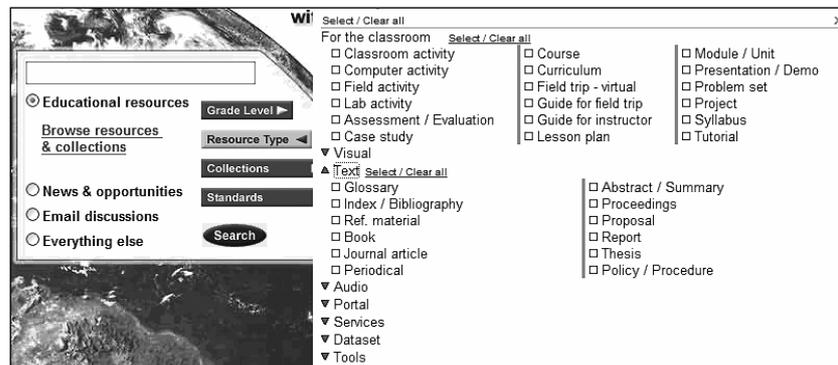
Ресурсы для работы в классе и категории визуальных средств:



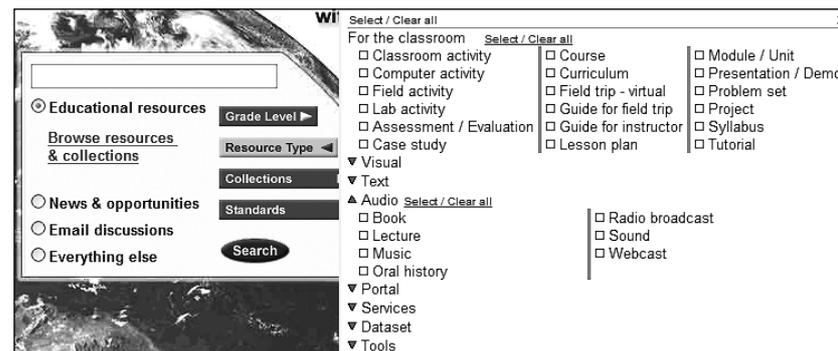
Категории по классу/уровню обучения:



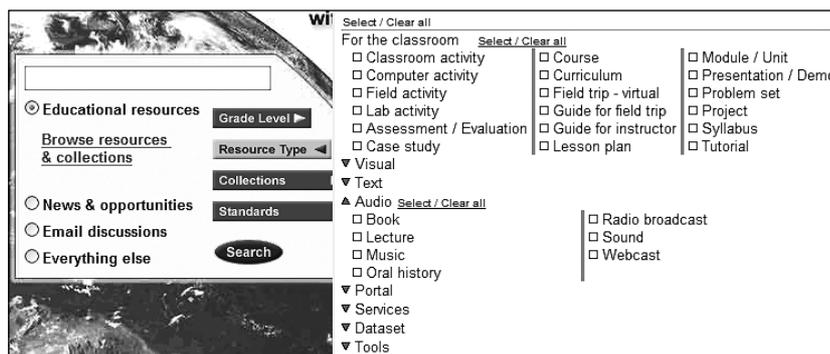
Текстовые категории:



Аудио/звуковые категории:



Категории услуг по поддержке коммуникации и совместной деятельности:



Задание 4

Подумайте о конкретном курсе, который вы преподаете или с которым вы хорошо знакомы. Как бы вы описали:

- Различные виды учебной деятельности студентов?
- Различные виды используемых медиасредств?

А теперь изучите категории, которые представлены в электронной библиотеке DLESE (Earth System Education). Если у вас на компьютере установлен браузер и подключен Интернету, то эти категории вы сможете найти в режиме онлайн на сайте <http://www.dlese.org/>; в качестве альтернативы воспользуйтесь иллюстрациями с примерами из практики DLESE.

Теперь сравните ваши категории с представленными в DLESE.

- Сколько выделенных вами категорий совпадает с категориями DLESE?
- Являются ли категории DLESE специфичными для наук о Земле или они годятся для любого другого предмета?
- Знаете ли вы какие-либо ресурсы или виды деятельности, которые не могут быть описаны с помощью этих категорий?

Пример из практики: NLVM – <http://matti.usu.edu/nlvm>

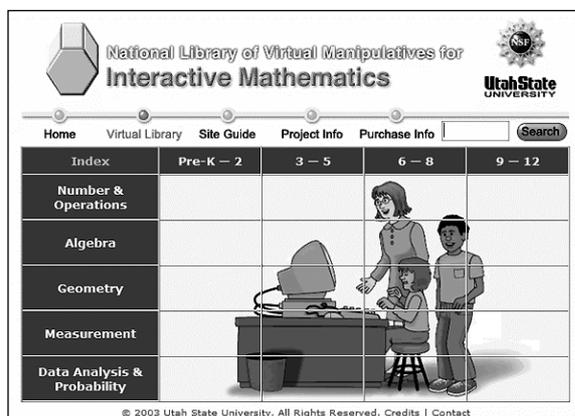
Электронные библиотеки могут содержать не только статичные материалы как текст или изображения. Интерактивное программное обеспечение, как и любой другой вид информации, также можно организовать в коллекции, обеспечив соответствующий доступ к нему пользователей. Национальная библиотека виртуальных манипуляций для интерактивной математики (The National Library of Virtual Manipulatives for Interactive Mathematics – NLVM), созданная в Университете штата Юта (Utah State University), – это коллекция программных средств, например, Java-приложения, которые могут выступать в качестве интерактивных средств в веб-браузере для наглядного объяснения математических понятий, таких как фракталы, системы счисления, вероятность, дроби и диаграммы Венна.

«Этот проект, длящийся три года и поддержанный Национальным научным фондом (National Science Foundation – NSF), направлен на развитие библиотеки, содержащей уникальные интерактивные, основанные на веб-технологии виртуальные лаборатории (или тематические тьюториалы), представленные, главным образом, в виде приложений Java, разработанных специально для обучения математике (в основном, для начальной и средней школы).

Со временем, мы сделаем так, что все материалы будут доступны на нескольких серверах Интернета, что позволит нам создать мощную национальную библиотеку, получив возможность использования которой учителя смогут разнообразить проводимые ими уроки математики. Материалы также будут представлять интерес для системы повышения квалификации и подготовки учителей начальной школы в области математики».

(Источник: <http://matti.usu.edu/nlvm/nav/projinfo.html>)

Ниже проиллюстрированы основные категории NLVM:



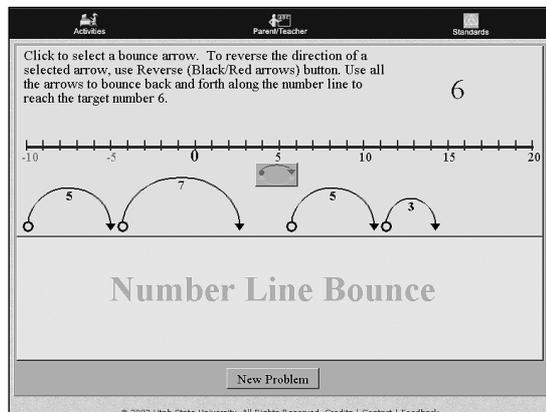
К каждой из этих категорий и каждому вопросу учебной программы есть ссылка на виртуальные инструменты, или манипуляции, которые можно использовать при их изучении:

Measurement (Grades 9 - 12)

Virtual manipulatives related to the NCTM *Measurement* standard for grades 9 - 12.

-  **Converting Units** – Use a simple system for converting units.
-  **Fill and Pour** – Solve puzzles requiring you to fill and pour containers.
-  **Geoboard** – Use geoboards to illustrate area, perimeter, and rational number concepts.
-  **Geoboard - Circular** – Use circular geoboards to illustrate angles and degrees.
-  **Great Circle** – Use a 3D globe to visualize and measure the shortest path between cities.
-  **How High?** – Try your hand at the classic Piagetian conservation of volume test.
-  **Pattern Blocks** – Use six common geometric shapes to build patterns and solve problems.

Пример виртуальной манипуляции для понятия числового луча:



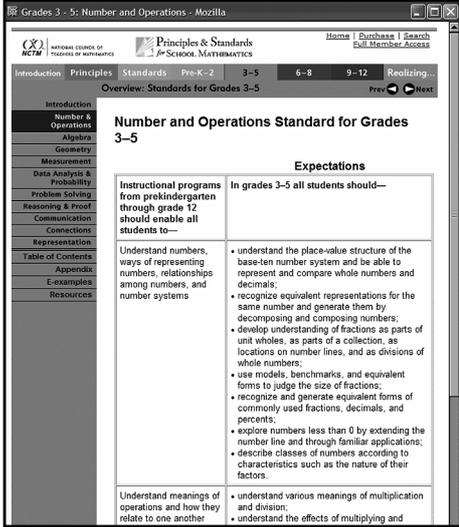
Click to select a bounce arrow. To reverse the direction of a selected arrow, use Reverse (Black/Red arrows) button. Use all the arrows to bounce back and forth along the number line to reach the target number 6.

Number Line Bounce

New Problem

© 2003 Utah State University. All Rights Reserved. Credits | Contact | Feedback

NLVM привязывает учебные задания к стандартам — документам, определяющим требования к знаниям и умениям учащихся разных возрастов. Например, на экране с примером виртуальной манипуляции для понятия «числовой луч» ссылка дается на документ, подготовленный Национальным советом учителей математики (National Council of Teachers of Mathematics), который можно найти по адресу <http://standards.nctm.org/document/appendix/alg.htm>:



The screenshot shows a web browser window displaying the NCTM website. The page title is 'Grades 3-5: Number and Operations - Mozilla'. The main content area is titled 'Number and Operations Standard for Grades 3-5'. A table with two columns is visible, detailing expectations for students in grades 3-5. The first column lists broad instructional goals, and the second column lists specific skills and understandings that students should have.

Expectations	
Instructional programs from prekindergarten through grade 12 should enable all students to—	In grades 3-5 all students should—
Understand numbers, ways of representing numbers, relationships among numbers, and number systems	<ul style="list-style-type: none"> understand the place-value structure of the base-ten number system and be able to represent and compare whole numbers and decimals; recognize equivalent representations for the same number and generate them by decomposing and composing numbers; develop understanding of fractions as parts of unit wholes, as parts of a collection, as locations on number lines, and as divisions of whole numbers; use models, benchmarks, and equivalent forms to judge the size of fractions; recognize and generate equivalent forms of commonly used fractions, decimals, and percents; explore numbers less than 0 by extending the number line and through familiar applications; describe classes of numbers according to characteristics such as the nature of their factors.
Understand meanings of operations and how they relate to one another.	<ul style="list-style-type: none"> understand various meanings of multiplication and division; understand the effects of multiplying and

Данный пример показывает, что деятельность NLVM нацелена на решение конкретных задач образовательного процесса, определенных в официальных документах. А лица, принимающие решения и участвующие в разработке этих документов, отстаивают свои интересы, и, разумеется, несут свою меру ответственности за результаты использования ресурсов в образовательном процессе и повышение эффективности работы учителей и учащихся.

Задание 5

Сравните системы DLESE и NLVM, которые были описаны в приведенных выше примерах из практики.

- Категоризация DLESE гораздо более детальная, чем в NLVM. Насколько это необходимо?
- Сможет ли выиграть NLVM от введения более детальной категоризации? Какие еще механизмы доступа можно было бы ввести для улучшения этой библиотеки?
- Насколько важно в вашей предметной среде учитывать существующие образовательные стандарты при создании учебных материалов, размещаемых в электронных библиотеках? Нужно ли использовать их при организации библиотеки?
- В какой степени глобальные электронные библиотеки должны поддерживать национальные стандарты?

Заключение

Электронные библиотеки представляют работникам сферы образования множество новых замечательных возможностей. Они могут включать в себя колоссальное по разнообразию и виду содержание. При этом, говоря о содержании, мы в первую очередь думаем о текстовых документах, но при ближайшем рассмотрении оказывается, что электронные библиотеки могут содержать любой другой вид информации – например, изображения, карты, аудио, видео и даже виртуальные лаборатории. То же самое можно предлагать читателям традиционных библиотек, но организовать это намного сложнее, поскольку для разных пользователей потребуется различный интерфейс и средства просмотра. Электронные библиотеки просматривают на компьютерах, предназначенных для общих целей и которые могут представлять все виды ресурсов и быть более или менее единообразными. Более того, в них нет ограничений на используемые ресурсы: электронные библиотеки могут включать необработанные данные и даже интерактивные программные модули, как мы это видели на примере NLVM. Другая возможность заключается в наличии глобального доступа: теперь учебными материалами можно обмениваться в таких масштабах, которые предыдущим поколениям трудно было представить.

Этот раздел также продемонстрировал вам, что у электронных библиотек есть разные категории пользователей и заинтересованных лиц. Хотя представители региональных органов управления образованием интересуются (или должны интересоваться) традиционными школьными библиотеками, они вряд ли будут сами пользоваться ими как информационным ресурсом. Традиционные школьные библиотеки спроектированы, прежде всего, для детей, на втором месте – для учителей, в меньшей степени для руководителей системы образования, хотя, в принципе, любая электронная библиотека может служить целям всех трех групп одновременно. Это расширение пользовательской базы отражено в новых видах структур и классификаций, разработанных для обеспечения многоцелевого использования одних и тех же ресурсов разными аудиториями.

Конечно, у всех технологий есть свои недостатки, и нам нужно заранее о них узнать. Одним из недостатков, с которым сегодня приходится сталкиваться во многих учебных ситуациях, является плагиат. Компьютерные сети позволяют студентам быстро обнаружить информацию, нужную им для выполнения заданий и включить ее в свои

ответы. Хотя соблюсти все правила цитирования совсем несложно, студенты предпочитают вообще этого не делать. Коллекции сочинений и рефератов широко распространены в международном масштабе, и уже есть коммерческие веб-сайты, которые за определенную плату выдают студентам ответы, являющиеся плагиатом.

Еще одним важным аспектом является то, что электронные библиотеки сами создают условия для усиления контроля над ними со стороны органов управления образованием на региональном, национальном и даже глобальном уровне. Большинство студентов учатся на примерах из местной практики и лучше усваивают материал, если он имеет прямое отношение к тому месту, где они живут. Большинству учителей нравится сложный процесс разработки собственных учебных курсов и подготовки авторского методического материала. Тенденция к усилению контроля со стороны органов управления образованием над этими процессами может привести к снижению удовлетворенности учителей своей работой и негативно отразиться на авторитете учительской профессии, что скажется на учащих, успеваемость которых зависит от того, насколько яркий и мотивированный преподаватель их учит.

Таким образом, вы получили представление об электронных библиотеках и о направлениях их использования в образовании. Теперь пришло время приобрести некоторый практический опыт работы в такой библиотеке. Пожалуйста, переходите к Главе 1.3, в которой вы познакомитесь с программой создания электронных библиотек Greenstone.

Первоисточники

Мы нашли три принципиально важных первоисточника, которые помогли нам при подготовке данной главы. Различные роли, которые электронные библиотеки могут играть в образовании, определенные нами в начале рассматриваемой главы, обсуждаются в работе Мазулло и Мака (Masullo and Mack, 1996). Черты отличия электронных библиотек от традиционных выделяются и обсуждаются в работах Уоллас, Крайчика и Солоуей (Wallace, Krajcik and Soloway, 1996); дополнительная возможность многократного использования ресурсов освещена в работе Мендела (Mendel, 1999).

Глава 1.3. Программное обеспечение для создания электронных библиотек Greenstone

Эта глава представляет вам специализированный комплект программного обеспечения для создания электронных библиотек. Пожалуйста, начните работу с определения категорий потенциальных пользователей подобного программного обеспечения и требований, которые они предъявляют к подобного рода программным средствам.

Задание 1

Люди, приходящие в обычные библиотеки, выполняют одну из двух главных функций: читателя либо библиотекаря. Представьте, что (в противоположность тому, что было описано в конце Главы 1.1.) ваша местная университетская библиотека, или любая другая, которая вам хорошо известна, была полностью заменена электронной. Используя имеющуюся у вас информацию о библиотеках, напишите один абзац, с изложением требований, предъявляемых к программному обеспечению для электронной библиотеки с точки зрения читателя и библиотекаря. Условие одно – их функции не должны очень сильно отличаться от тех, которые они выполняли до появления этой библиотеки.

Напишите еще один абзац, объясняющий, как может измениться роль библиотекаря в этом новом мире цифровых технологий.

При выполнении практических работ в рамках данного курса вы будете руководствоваться программой создания электронных библиотек Greenstone, которая поможет вам организовать информацию на современном уровне и сделать ее доступной в Интернете или на переносных устройствах типа CD-ROM или DVD.

Задание 2

Прочитайте Часть 1.4 (с. 24–28) учебника «Как создать электронную библиотеку», а затем ответьте на следующие вопросы в зависимости от специфики программы электронной библиотеки Greenstone:

- Что такое «документ»?
- Что такое «коллекция»?
- Что такое «библиотека»?
- По какой лицензии на программное обеспечение выпускается программа Greenstone?

Greenstone является быстро развивающейся системой, с постоянно появляющимися новыми дополнениями к ней. Например, если в начале интерфейс программы был представлен на 12 языках мира, то сегодня это число увеличилось до 32 языков: английский, арабский, армянский, боснийский, вьетнамский, галисийский, грузинский, датский, иврит, индонезийский, испанский, итальянский, казахский, канадский, каталонский, китайский, латвийский, маори, немецкий, португальский, русский, сербский, турецкий, украинский, фарси, финский, французский, тайский, хинди, хорватский, чешский, японский. Это серьезный показатель международного интереса к данной программе.

Задание 3

Greenstone выпускается по лицензии GNU General Public License^{*}, которая официально определяется как «Программное обеспечение с открытым источником» (Open Source Software). Прочитайте определение программы с открытым источником, приведенное ниже, и ответьте, есть ли у вас разрешение на:

- передачу программы Greenstone третьим лицам?
- продажу Greenstone?
- внесение каких-либо изменений в Greenstone?
- создание коллекций на основе Greenstone и их свободное распространение?
- создание коллекций на основе Greenstone и продажу доступа к ним?
- создание CD-ROM/DVD на основе коллекций, созданных в Greenstone, и их продажу?

^{*} У вас будет возможность прочитать эту Лицензию, когда вы проинсталлируете программу Greenstone в ходе работы над Разделом 2.2 Модуля 2 учебника. Если у вас есть доступ к Интернету, то вы можете также найти ее по адресу www.gnu.org.

Программное обеспечение с открытым источником

Наличие открытого источника не означает, что все, что получает пользователь — это доступ к исходному коду программы. Условия распространения программных средств с открытым источником должны отвечать следующим критериям (с сайта <http://www.opensource.org>):

Свободное распространение

Данная лицензия не накладывает ограничений на любую из сторон в отношении продажи или отчуждения данной программы, если она являлась частью программного продукта, составленного из программ, имеющих несколько разных источников. Лицензия не требует выплаты авторского вознаграждения или других выплат, связанных с данным актом продажи.

Код источника

Данная программа должна включать код источника, а также допускать возможность распространения кода источника и компилятора. В тех случаях, когда отдельные формы продукта распространяются без кода источника, должны быть предусмотрены легко доступные средства его получения, предоставляемые по приемлемой стоимости и бесплатно скачиваемые из Интернета. Желательно, чтобы код источника предоставлялся программистам для дальнейшей модификации этой программы. Преднамеренное изменение кода источника недопустимо. Выпуск промежуточных форм, таких как препроцессор или транслятор запрещаются.

Вторичные разработки

Лицензия должна давать право на модификации и вторичные разработки, а также их дальнейшее распространение на основе таких же условий, что и лицензия на исходное программное средство.

Целостность авторского кода источника

Лицензия может накладывать ограничения на распространение кода источника в измененной форме в том случае, если лицензия позволяет распространение «заплаточных файлов» (patch files) с кодом источника с целью дальнейшей модификации программы в процессе создания продукта. Лицензия должна разрешать распространение программных средств, созданных на основе модифицированного кода источника. Лицензия может включать требование того, чтобы вторичным разработкам присваивались другие названия или номера версий по сравнению с исходным программным средством.

Отсутствие дискриминации в отношении личностей или групп

Данная лицензия не должна ущемлять права человека или группы людей.

Отсутствие дискриминации в отношении направления использования и сферы деятельности

Лицензия не должна содержать ограничений на использование данной программы по какому-либо направлению или сфере деятельности. Например, она не должна ограничивать использование программы ни в сфере бизнеса, ни в исследованиях в области генетики.

Распространение лицензии

Права, действующие в отношении данной программы, должны распространяться на третьих лиц, кому данная программа была передана в пользование, а, следовательно, дополнительная лицензия, определяющая отношения между ними, не требуется.

Лицензия не должна распространяться на программный продукт

Права, действующие в отношении программы, не должны зависеть от того, частью какого распространяемого программного продукта она является. Если программа отделяется от дистрибутива и используется или распространяется в рамках данной лицензии, то все стороны, которым эта программа будет передана, имеют те же самые права, что и лица, которые инициировали распространение исходного программного продукта.

Лицензия не должна ограничивать другие программные продукты

Лицензия не должна накладывать ограничения на другое программное обеспечение, которое распространяется вместе с данным лицензированным программным продуктом. Например, в лицензии нельзя настаивать на том, чтобы все другие программы, распространяемые на том же носителе, должны также рассматриваться как программы с открытым источником.

Лицензия должна быть технологически нейтральной

Лицензия не может определять приоритетное использование какой-либо отдельной технологии или стиля интерфейса.

Теперь давайте взглянем на электронную библиотечную коллекцию, построенную с использованием Greenstone. Для этого вы должны воспользоваться читательским интерфейсом Greenstone (Greenstone User Interface, GUI), доступ к которому вы можете получить через веб-браузер. Позже, работая с Модулем 2, вы создадите свои собственные коллекции в Greenstone, для чего вы будете использовать другой интерфейс – интерфейс библиотекаря (Greenstone Librarian Interface, GLI).

Итак, возьмите имеющуюся у вас копию компакт-диска с курсом «Электронные библиотеки в образовании». Обратите внимание: этот диск не то же самое, что комплект материалов на CD-ROM IMARK (Information Management Resource Kit), который также предоставляется слушателям данного учебного курса. Он содержит три заранее подготовленные коллекции Greenstone, которые можно просмотреть на любом компьютере с операционной системой Windows (3.1/3.11, 95/98/ME, NT/2000 и XP). Greenstone также работает на Linux, Unix и Macintosh OS/10, но она записывает на компакт-дисках только те из заранее подготовленных коллекций, которые создавались именно под Windows. Эти коллекции имеют только читательский интерфейс Greenstone, и именно поэтому их можно просматривать даже на компьютерах с Windows (3.1/3.11 и 95/98/ME). Библиотечарский интерфейс (The Librarian Interface), который используется во всех следующих модулях, работает исключительно на современных версиях Windows (NT/2000 и XP). Это объясняется тем, что этот интерфейс моложе, чем сама система Greenstone, поэтому написан на языке программирования Java.

В ходе выполнения следующего упражнения вы проведете установку этих коллекций на ваш компьютер. Большинство модулей данного курса включают несколько упражнений, имеющих практическую направленность. Упражнения отличаются от заданий тем, что они не требуют ответы на поставленные вопросы, но шаг за шагом направляют вас к выполнению конкретных практических действий. Они занимают главное место в данном курсе: *не пропускайте их*.

Упражнение: Установка коллекций

1. Вставьте CD-ROM с курсом «*Электронные библиотеки в образовании*» в компьютер с операционной системой Windows. Если процесс установки не начался сразу же (из-за того, что на вашем компьютере отключена функция автозапуска AutoPlay), определите дисковод, на котором находится ваш CD-ROM/DVD (обычно, это дисковод D:), откройте папку *prebuilt*, и сделайте двойной щелчок по файлу запуска программы *Setup.exe*.
2. В процессе установки вам будет предложено выбрать папку, в которой будет размещена программа: мы рекомендуем выбрать папку, предложенную по умолчанию, а именно C:\GSDL.
3. Вам также предоставляется возможность запустить Greenstone непосредственно с CD-ROM или скопировать весь CD-ROM на ваш компьютер. Мы рекомендуем последнее: пожалуйста, поставьте галочку там, где сказано *Install all collection files*. На копирование всех файлов уйдет около 2-х минут.
4. И, наконец, программа-инсталлятор предлагает вам установить браузер Netscape. Не просите программу сделать это, если на вашем компьютере уже установлен один из браузеров.

Этот процесс инсталляции намного сложнее описать, чем на самом деле выполнить! Он займет у вас всего несколько минут.

5. Для запуска *Greenstone* откройте в Windows меню Пуск, Программы и выберите *Greenstone*, а затем в подменю выберите пункт *Электронные библиотеки в образовании* и войдите в библиотеку, щелкнув по **<Enter Library>**.

В этой Главе мы знакомимся с двумя из трех коллекций, имеющих-ся на CD-ROM: «*Электронные библиотеки в образовании*» и «*Как создать электронную библиотеку*». Третью коллекцию, названную IEEE LOM, мы посмотрим в Главе 3.3 (LOM расшифровывается как “Learning object metadata” – Метаданные учебных объектов).

Упражнение: Коллекция курса

Сначала изучите коллекцию «*Электронные библиотеки в образовании*».

1. Щелкните по пиктограмме данной коллекции. Вы перейдете на домашнюю страницу коллекции, которая часто называется страницей «Об этой коллекции» (about page).

На домашней странице находится область доступа с набором кнопок с названиями поиск (search), содержание (content), авторы от А до Я (authors a-z), модули (modules) и акронимы (acronyms). Эта область доступа является отправной точкой для проведения поиска в любой из коллекций *Greenstone*.

1. Щелкните по **<authors a-z>**. Появится список пиктограмм с изображением книжной полки. Щелкните по одной из пиктограмм, под названием Marchionini, G. чтобы увидеть две статьи Гарри Марчионини (Gary Marchionini).
2. Один из этих материалов представлен в виде PDF файла, а другой – HTML файла. Откройте по очереди каждый из этих документов.
3. Щелкните кнопку **<contents>** в окне доступа к коллекции. После этого вы увидите две книжные полки: одну – с материалами данного руководства по изучению и другую – с материалами для дополнительного чтения по курсу. Выберите любую из них и посмотрите, что в ней находится.
4. Щелчок по уже открытой книжной полке закрывает ее. Закройте только что открытую вами книжную полку и затем выберите другую и изучите ее содержимое.
5. Щелкните по кнопке **<acronyms>** в окне доступа к коллекции и найдите значение акронима “LOM”.
6. Щелкните **<search>** и найдите слово “LOM”. Проверьте разницу в полученных результатах при поиске текста и поиске по заголовкам (используйте выпадающее меню на поисковой странице).
7. Щелкните по пиктограмме коллекции **Электронные библиотеки в образовании**, расположенной вверху слева. Это действие вернет вас на страницу коллекции «Об этой коллекции».

Ниже окна доступа к коллекции, которое расположено на странице «Об этой коллекции», находится окно поиска информации (похожее на то, что обычно появляется на поисковой странице), в котором дается описание этой коллекции и инструкции по поиску представленной в ней информации.

Над окном доступа к коллекции находится пиктограмма коллекции «Электронные библиотеки в образовании», а справа от нее пиктограмма, представляющая раздел about (Об этой коллекции), и три кнопки – home (домашняя страница), help (помощь) и preferences (настройка).

8. Щелкните кнопку **<home>**. Это вернет вас на домашнюю страницу Greenstone.
9. Вернитесь к коллекции (щелкнув по ее пиктограмме) и выберете **<help>**. Вы получите более подробную информацию о том, как получить доступ к коллекции.
10. Щелкните кнопку **<preferences>**. Вы перейдете на страницу, на которой сможете изменить некоторые из установок программы.
11. Теперь исследуйте коллекцию, свободно перемещаясь по ней в разных направлениях. Щелкайте по любым ссылкам: все пиктограммы, которые появятся на экране — активные, и по ним можно перейти на другие страницы. Если вы наведете курсор мыши на пиктограмму и удержите его в этом положении, то, как и в большинстве браузеров, вы скоро увидите появившееся краткое сообщение, которое подскажет, какое действие можно ожидать после щелчка по этой пиктограмме. Экспериментируйте! Используйте для поиска самые обычные слова, такие как артикль “the” или союз “and” на которые вы точно получите какой-либо ответ, и при этом вы точно ничего не испортите. (Обратите внимание: в отличие от многих поисковых систем Greenstone индексирует абсолютно все слова, включая артикли.)

Задание 4

Прочитайте страницу Help коллекции «Электронные библиотеки в образовании», а затем ответьте на следующие вопросы:

- Что находится в данной коллекции?
- Назовите пять способов перехода к искомому документу в данной коллекции.
- Сколько именно документов в данной коллекции написано Эриком Дювалем (Erik Duval)?
- Сравните, сколько раз появлялись в данной коллекции слова «he» (он) и «she» (она).
- Сколько раз слово «метаданные» появлялось в заголовках? А в самом тексте?
- Какова разница между словами some (некоторые) и all (все) при проведении поиска?
- Что стоит за акронимом “MODS”?
- Как вы переключаете интерфейс с английского языка на русский? Остается ли интерфейс на русском, когда вы возвращаетесь на домашнюю страницу Greenstone?

- Найдите поисковый термин, который даст вам при проведении поиска разные результаты в зависимости от того, игнорируете (ignore) ли вы окончания слова или это слово целиком должно совпадать (must match) со словом, определенным на странице Preferences.
- Какая существует разница между графическим (Graphical) и текстовым (Textual) форматами интерфейса (на странице Preferences)?

Вторая из заранее подготовленных коллекций курса, с которой мы сейчас познакомимся, называется «Как создать электронную библиотеку?» (How to build a digital library). Как объясняет само название, основу ее составляет книга, которая была предложена вам для изучения и записана для этого на компакт-диск. Однако читать коллекцию не так просто как саму книгу, поскольку книга уже была опубликована коммерческим издателем, который не дал бы разрешения на ее свободное распространение в электронной форме. Вместо этого, коллекция содержит указатель предложений, имеющихся в книге. Вы можете найти и прочитать отдельные предложения, но не параграфы, разделы или главы. Это объясняется тем, что в Greenstone каждое отдельное предложение может расцениваться как полноправный «документ»! При этом каждый из документов является независимым объектом. Это довольно необычное использование программы, показывающее универсальность Greenstone.

Когда вы получаете предложение (то есть, документ) из данной коллекции, оно сопровождается номером страницы в опубликованной версии данного учебного пособия. Поисковая функция в коллекции ведет себя как усовершенствованный указатель к книге, то есть она является интерактивным указателем, приспособленным для полнотекстового поиска. Так же, как и сам указатель, коллекция включает:

- онлайн-приложения к книге,
- цветные версии всех иллюстраций,
- текстовые версии иллюстраций в XML,
- введение и примеры к Главе 1,
- иерархический браузер фраз (доступ, к которому осуществляется с помощью кнопки *phrases*),
- список акронимов.

Последние два пункта списка — это инструменты Greenstone, которые можно использовать в любой другой коллекции. Система анализирует определенные части текста (в нашем примере все подписи к рисункам) и выделяет иерархию фраз, которые специальный браузер фраз создает с помощью приложения Java. Попробуйте, например, напечатать *XML* в браузере иерархий фраз, для того чтобы увидеть четыре фрагмента текста, в которых данное слово появляется в подписях к рисункам. Система также анализирует полностью все тексты документов и выбирает любые сокращения и их расшифровку.

Задание 5

Используйте коллекцию материалов «Как создать электронную библиотеку» для ответов на следующие вопросы:

- В скольких предложениях коллекции содержится слово образование (education)?
- Какая история из коллекции School Journal («Школьный журнал») описывается в этой книге?
- Сколько сокращений, используемых в книге, начинается со слова Standard (Стандарт)?
- Что означает слово *tapu*?
- Сколько раз появляется слово *library* (библиотека)? Слово *libraries* (библиотеки)?
- Сколько раз слово *Library* (Библиотека) появляется с заглавной буквы?
- Сколько раз появляются слова – производные от слова *appear*?
- Назовите английскую поэму, которая была написана примерно в 1000 году н.э.
- Кто такой Alan Kay (Алан Кэй)?
- На какой странице было впервые упомянуто что-либо, относящееся к китайской культуре?

Дать ответы на большинство из подобных вопросов будет довольно сложно, если вы будете искать их в традиционной книге.

На диске IMARK вы найдете несколько практико-ориентированных уроков по работе с программой для создания электронных библиотек Greenstone.

Задание 6

Проработав Урок 6.1 *Тьюториал Greenstone комплекта IMARK*, вы узнаете:

- основные характеристики программного обеспечения,
- требования и условия инсталляции,
- какие функции интерфейса будут доступны для конечных пользователей,
- три различных подхода к построению коллекций, которые поддерживают это программное обеспечение.

Ответьте на вопросы в конце данного Урока.

Заключение

Вы начали изучать эту главу с размышлений о том, как обеспечить потребности и оказать поддержку двум наиболее важным группам пользователей традиционных библиотек: читателям и библиотекарям. Программа для создания электронных библиотек обязательно должна быть доступной для простых читателей, чтобы они могли пользоваться ею практически без всякого дополнительного обучения. Электронные библиотеки будут обречены на неудачу, если для освоения работы с ними придется изучить целую книгу! Они должны поддерживать универсальный доступ: быть свободно распространяемыми, способными запускаться на любом компьютере, включая самые примитивные; иметь возможность работать, не подключаясь к Интернету; они должны представлять материалы на любом языке и т.д. Только незначительное количество электронных библиотечных систем удовлетворяют всем этим требованиям. В частности, большинство из существующих систем ориентированы на обязательную работу с Интернетом. Минусом в этом случае является необходимость решать серьезную технологическую проблему, т.к. сервер подобной электронной библиотеки с находящейся на нем поисковой системой и другими инструментами обработки информации должен, в том числе, работать и на устаревших системах, иначе такая библиотека не будет обеспечивать универсальный доступ к информации для всех желающих.

В рассмотренной главе вам была представлена программа Greenstone, и вы обсудили, в какой степени она удовлетворяет требованиям универсальной доступности. Лицензия GNU (General Public License), по которой она выпускается, дает большую свободу одновременно и для коммерческого, и для некоммерческого ее использования. Очень

важно, чтобы вы полностью осознали, какая свобода вам предоставлена! Это хорошо проиллюстрировано вопросами, перечисленными в Задании 3 (все ответы на них должны быть «да»). Вам не нужно получать разрешение на распространение программы Greenstone, ее продажу или любую модификацию – создавайте с ее помощью свои коллекции на переносных носителях, раздавайте или продавайте их. По сути, единственным ограничением, которое накладывает на вас эта лицензия, является то, что если вы изменяете данное программное обеспечение, вы обязаны бесплатно предоставлять эту измененную вами версию всем желающим (хотя вы также можете ее продать). Это означает, что вы не можете встроить компоненты Greenstone в свою собственную коммерческую схему.

Вы также попробовали поработать с читательским интерфейсом. Он предоставляет пользователям богатый набор инструментов поиска и просмотра информации в коллекции. Им не сложно пользоваться: в нем есть онлайн-помощь (представленная на всех языках интерфейса), поэтому никакой другой помощи, дальнейших инструкций или дополнительной документации вам не потребуется.

Особые механизмы доступа, которые работают в каждой отдельной коллекции Greenstone, определены человеком, который проектировал и создавал эту коллекцию. В следующем Модуле вы узнаете об огромном разнообразии существующих форматов документов, которые вы, возможно, захотите включить в вашу коллекцию. Вы также узнаете о том, как создать свои коллекции с помощью библиотечарского интерфейса – Greenstone Librarian Interface.

Модуль 1. Концепция электронных библиотек и их роль в образовании

МОДУЛЬ 2

ФОРМЫ ПРЕДСТАВЛЕНИЯ ДОКУМЕНТОВ

Глава 2.1. Документы:
исходный материал

Глава 2.2. Создание
электронной библиотечной
коллекции



Цель

Познакомиться с существующим разнообразием форматов документов, их сильными и слабыми сторонами, а также особенностями представления в них символов на разных языках мира; приобрести опыт создания электронной библиотечной коллекции из документов, созданных в разных форматах (файлов Word и PDF).

Задачи

После завершения работы по Модулю 2, вы сможете:

- привести примеры сложных ситуаций, которые возможны при представлении на компьютере документов на неевропейских языках;
- описать, как используется Unicode для представления различных наборов символов;
- различать такие версии Unicode, как UTF-8, UTF-16 и UTF-32;
- отличать языки описания страницы и форматы документов, подготовленных в текстовом редакторе;
- рассказать об истории языка описания страниц PostScript и его дальнейшем превращении в формат PDF;
- выбрать форматы документов, наиболее приемлемые для вашей электронной библиотеки;
- обсудить недостатки использования фирменных форматов типа Microsoft Word;
- установить программу Greenstone, а также проводить ее последующие обновления, с одновременным обновлением созданных ранее коллекций;
- создавать коллекции из документов Word, RTF, PDF и PostScript;
- добавлять к коллекции Greenstone метаданные;
- спроектировать коллекцию Greenstone и включить в нее поисковые индексы, классификаторы для просмотра и коллекции пиктограмм;
- объяснить различие между метаданными, извлеченными и составленными пользователем;
- определить местонахождение созданных вами коллекций в файловой структуре компьютера.

Введение к Модулю 2

В Модуле 1 вы узнали, что такое «электронные библиотеки» и познакомились с направлениями их использования в образовании. Также вы провели поиск информации в электронной библиотечной коллекции.

Модуль 2 знакомит вас с особенностями создания электронных библиотек. Документы — это исходный материал, на основе которого создаются электронные библиотеки, а символы являются исходным материалом для формирования документов. Хотя мы и стремимся создать электронные библиотеки с наполнением, которое бы «звенело и сверкало» и было рассчитано на высокий интеллектуальный уровень читателей, нам придется вернуться на землю и начать с азов — так, как это делает ребенок, который приступает к изучению алфавита. Алфавиты устроены просто, и именно они демонстрируют удивительное разнообразие языков мира. И документы, из которых они строятся, также служат разным целям! Под каждую цель есть свои форматы документов. Форматы низшего уровня, являющиеся общим знаменателем для всех прочих — это форматы представления документов во Всемирной паутине, языки описания страниц, которые описывают уже созданные документы, форматы представления документов в текстовых редакторах, предназначенных для удобства и быстроты редактирования, форматы обработки документов, которые работают вне сети и контролируют процесс вывода текста на странице. Документы могут быть представлены не только в виде текста: вы узнаете больше о форматах графических изображений, аудио, видео и мультимедиа.

Другой особенностью является то, что понять как на самом деле построена электронная библиотечная коллекция, очень не просто. В этом модуле вы проведете инсталляцию программы создания электронной библиотеки и научитесь ее использовать для объединения документов в единое целое. Вы также подготовите описание метаданных на основе заголовков и имен авторов. Далее вы разработаете проект коллекции, определив ее внешний вид и то, какие инструменты доступа она будет поддерживать. Создадите указатели и разработаете структуры просмотра документов, проведете предварительный просмотр только что созданных новых коллекций с целью проверки проекта и представления его пользователям либо через WWW, либо путем записи коллекции на CD-ROM или DVD. И вновь отметим, что за любым самым высокоинтеллектуальным решением стоит огромная тяжелая и кропотливая черновая работа.

Этот модуль включает две главы, примерно соответствующих по тематике предыдущим. Первая рассматривает вопросы о том, как именно представлены документы в электронных библиотеках. Вторая показывает, как использовать Greenstone для создания коллекции документов. В этих главах вы познакомитесь с четырьмя основными текстовыми форматами: Microsoft Word, RTF (Rich Text Format), PDF (Portable Document Format) и PostScript. Вы узнаете о них в Главе 2.1 и создадите на их основе коллекции в Главе 2.2., которая описывает процедуру инсталляции программы Greenstone и выполнение различных действий по ее поддержке.

Модуль 2. Материалы для чтения

Глава 2.1. Документы: исходный материал

Материал для чтения 1. Учебник «Как создать электронную библиотеку»: Глава 4 (до Раздела 4.4 включительно).

Назначение. В самом начале этого Материала для чтения дается описание формата Unicode, стандарта кодирования символов всех языков мира, а затем предоставляется характеристика существующих языков описания страниц и форматов текстовых процессоров. Проведите параллель между прочитанным и тем, что вы знаете по собственному опыту работы с различными языками и форматами документов. Основное внимание в этом разделе уделяется форматам Word, RTF, PDF и PostScript.

Материал для чтения 2. CD-ROM IMARK: Раздел 2. «Электронные документы и форматы», Уроки 2.2, 2.6, и 2.7.

Назначение. Урок 2.7 дает ценную практическую информацию по языку описания страниц PDF, характеристика которого представлена в Материале для чтения 1.

Уроки 2.2 и 2.6 предлагаются как дополнительные, они не настолько важны для данного раздела, однако показывают различные перспективы по дальнейшему использованию материала, с которым вы здесь знакомитесь, поэтому если вы сможете, то воспользуйтесь нашими рекомендациями и изучите их.

Глава 2.2. Создание электронной библиотечной коллекции

Материал для чтения 1. CD-ROM IMARK: Урок 6.2. «Использование GLI: Сбор и преобразование документов».

Материал для чтения 2. CD-ROM IMARK: Урок 6.3. «Использование GLI: Проектирование и создание коллекции».

Назначение. Эти два раздела, предназначенные для самостоятельного изучения, позволят вам на практике расширить свои знания по установке программы Greenstone и проектированию с помощью библиотечарского интерфейса (Greenstone Librarian Interface, GLI) коллекции, включающей документы в форматах Word, RTF, PDF и PostScript.

Глава 2.1. Документы: исходный материал

В рассматриваемой главе вы узнаете о том, как документы, которые, по сути, являются наиболее важными для формирования электронных библиотек, представлены в компьютере. Существует великое разнообразие таких документов, поэтому, прежде чем взяться за изучение деталей, пожалуйста, уделите время и поразмыслите над ними.

Задание 1

Задумайтесь о возможных видах исходного материала, на основе которого можно было бы создать образовательную электронную библиотеку.

- Перечислите как можно больше известных вам форматов сохранения данных для текстовых документов.

Рассмотрите отдельные символы, которые используются в оформлении текстовых документов.

- Что еще, помимо символов, которые вы можете обнаружить на англоязычной клавиатуре (строчные и прописные буквы, цифры, знаки пунктуации и различные дополнительные символы), представлено в ваших документах?
- Если, помимо английского языка, вы сейчас используете и другие языки, напишите параграф с описанием проблем представления на компьютере текста на этих языках.

Продолжите изучение этой главы с Главы 4 учебника «Как создать электронную библиотеку», в которой рассказывается о различных способах представления документов в машиночитаемой форме.

Вначале мы познакомимся с Unicode, являющимся универсальным способом представления всех существующих в мировых языках символов. Поскольку компьютерная технология родом из Западной Европы и США, то на ранних этапах ее развития преобладало использование западноевропейских языков и, прежде всего, английского. В международной практике необходимо иметь дело и с другими наборами символов и языковыми системами, что порождает немало проблем.

Проблемы интернационализации документов не сводятся исключительно к вопросам представления в документах отдельных символов.

Когда компьютеры выполняют полнотекстную индексацию коллекции документов, они обычно сокращают слова до их корневой основы или, по крайней мере, предлагают самому пользователю произвести этот выбор. Это позволяет сделать запрос невосприимчивым к различным вариантам написания слов, что зачастую очень удобно. Конечно же, такой вид поиска, называемый морфологическим поиском, зависит исключительно от языка и вызывает немало вопросов. Более того, во многих восточных языках слова далеко не всегда отделяются друг от друга пробелами, как это делается в европейских языках, и это очень серьезно влияет на проведение полнотекстного индексирования.

Если вы имеете дело с международными коллекциями, то вы должны иметь представление о подобных проблемах. Современные системы поиска информации неизменно применяют методы морфологического поиска только для английского языка или предоставляют возможность программистам создавать на их основе модули для других языков. Современные системы редко используют приемы фрагментации текста при работе с восточными языками.

Задание 2

Прочитайте начало Главы 4 и Частей 4.1 и 4.2 (с. 131–163) учебника «Как создать электронную библиотеку» и ответьте на следующие вопросы:

- На решение каких проблем нацелен Unicode?
- Что подразумевается под словами «двусторонняя совместимость»?
- Какая существует разница между составным символом и объединенным символом?
- Почему это осложняет проведение поиска?
- Приведите пример полнотекстного запроса и используемого для него контекста, в котором морфологический поиск будет иметь большое значение.
- Приведите пример запроса и контекста, в котором важность морфологического поиска не столь велика.
- Перечислите ограничения на использование информации на русском языке в поисковых системах, спроектированных исключительно для английского языка.

Дополнительное упражнение 1

Урок 2.2. учебного модуля на CD-ROM IMARK предлагает вам познакомиться с другим подходом к решению проблемы кодирования символов, а именно — созданием многоязычных веб-страниц.

Программы создания электронных библиотек, как правило, предусматривают подобные действия. Тем не менее, очень полезно разобраться в том, что именно происходит в вашей системе, поэтому мы рекомендуем вам проработать этот Урок и расширить свои представления о сложных вопросах кодировки символов.

Давайте переключим наше внимание на то, как представлены документы в популярных компьютерных форматах. Мы будем различать языки описания страниц, которые разработаны для представления документов, наборы шрифтов для документов и форматов текстовых процессоров, которые разработаны для поддержки процесса создания документов, но также часто используются и для обработки уже готовых документов. Существенным отличием является то, что старые версии текстовых процессоров сохраняли формат сразу всех страниц документа, включая нумерацию страниц и структуру каждой страницы в отдельности, выравнивание и т.д., в то время как в более поздних версиях эти характеристики менялись в зависимости от типа используемого принтера, размера бумаги, установок по умолчанию границ документа и пр.

Данное учебное пособие дает детальное описание PostScript, исходного языка описания страниц, и объясняет, как в дальнейшем он превратился в PDF, или Portable Document Format. В большинстве случаев, при работе с электронной библиотекой, очень важно знать о формате PostScript, поскольку множество довольно старых документов до сих пор были представлены именно в нем. Однако, если вы делаете постраничную версию документа в наши дни, вы, скорее всего, будете сохранять ее в формате PDF, хотя для некоторых компьютерных платформ такое дополнительное программное обеспечение должно приобретаться на коммерческой основе, в то время как версия PostScript существует везде и всегда, т.к. все еще является именно тем базовым языком, на котором общаются между собой компьютеры и принтеры. Урок 2.7. учебного модуля на CD-ROM IMARK дает детальное практическое представление о языке описания страниц PDF и рассказывает о его использовании, что в целом очень хорошо дополняет материал из вашего учебного пособия. Вы должны познакомиться и с тем, и с другим первоисточником.

Задание 3

Прочитайте Часть 4.3. (с. 163–184) учебника «Как создать электронную библиотеку» и проработайте Урок 2.7. учебного модуля на CD-ROM IMARK. Затем ответьте на следующие вопросы:

- Насколько велика вероятность того, что старые машиночитаемые документы будут представлены в PostScript или в PDF?
- Почему так сложно сканировать файл, сохраненный как PostScript, и извлекать из него только текст?
- Какие проблемы с языком PostScript привели к разработке формата PDF?
- Почему формат PDF выходит за рамки такого определения как «язык описания страниц»?
- Что лучше: PostScript или PDF: (а) для использования в онлайн, (б) для международного использования?

Как уже было отмечено выше, форматы текстовых процессоров были спроектированы для создания документов, а также для коллективного обсуждения документов. Однако подобные документы, созданные с помощью коммерческого продукта могут не читаться на некоторых из существующих программных платформ. Это неудобно не только пользователю, но и тому, кому он посылает документ в формате Microsoft Word, поскольку, возможно, получатель не сможет прочитать этот документ. Ни для кого не секрет, что если вы хотите прочитать тот же самый документ на следующий год или в последующее десятилетие, то вы целиком и полностью полагаетесь на конкретного производителя программного обеспечения, поскольку именно от него зависит, смогут ли будущие версии этой программы прочитать старые файлы. На самом деле, они не смогут этого сделать: сегодняшние версии Microsoft Word не могут декодировать многие старые файлы Word, и эти документы настолько недоступны, как если бы вы сами их зашифровали, а затем выбросили ключ к шифру.

Итак, для работы в настоящее время мы рекомендуем формат PDF, который позволит коллективно обсуждать и сохранять созданные документы. Не используйте форматы текстового процессора, если вы не хотите, чтобы ваши читатели стали вашими соавторами и внесли свою правку в уже заверченный документ. Если вы создаете электронную библиотеку, в которую хотите включить документы Microsoft Word, то сначала переведите их в формат PDF (или HTML). Это позволит

вам лучше проконтролировать форму представления документов читателям, обеспечивая большую мобильность, универсальность доступа и повышая шансы на то, что в будущем вы сможете эти документы восстановить.

И все же, из-за широкого распространения текстовых процессоров будет удобно, если программа создания электронной библиотеки поддерживает работу с подобными документами, извлекая из них текст для полнотекстового индексирования и переводя их в более универсальную форму представления для читателей, не требуя при этом наличия необходимого в подобных ситуациях текстового процессора. В учебнике «Как создать электронную библиотеку» представлена информация об основных форматах текстового процессора Microsoft Word и RTF, или Rich Text Format, который разработан для переноса и использования документов, созданных в текстовом процессоре, в другие различные приложения, а также LaTeX, который часто используется для математических и иных научных документов.

Задание 4

Прочитайте Часть 4.4. (с. 184–194) учебника «Как создать электронную библиотеку» и ответьте на следующие вопросы:

- Как вы думаете, если один и тот же документ сохранить в формате Word и RTF, то какой из двух файлов будет больше по объему?
- Позволяют ли Word или RTF добавлять новые функции форматирования, не нарушая формата документов, созданных ранее? Почему?
- Перечислите несколько способов включения графических иллюстраций в документы, созданные в формате RTF.
- Насколько сложно извлечь простой текст из документов, хранящихся в формате (а) RTF, (б) Word и (в) LaTeX?
- Назовите три способа включения в электронную библиотеку документов, сохраненных в формате Word с помощью быстрого сохранения (Файл/Сохранить). Обсудите все за и против.

Дополнительное упражнение 2

Урок 2.6. учебного модуля на компакт-диске IMARK рассказывает о текстовых процессорах на примере Microsoft Word. Однако этот материал ориентирован, прежде всего, на то, чтобы научить вас пользоваться интерактивными инструментами Word для создания документов, которые не только хорошо смотрятся, но и грамотно структурированы.

Преимущество использования хорошо структурированных документов в электронной библиотеке заключается в том, что это позволяет (при необходимости) автоматизировать намного больше функций при поиске и представлении документов. Для того чтобы извлечь из этого модуля максимальную пользу, вы сначала должны прочитать Урок 2.1, в котором вводится понятие «разметка» и объясняется ее значение.

Заключение

Рассматриваемая глава познакомила вас с основополагающими вопросами, касающимися представления символов и документов.

Некоторые из этих вопросов своими корнями уходят в так называемую интернационализацию. Если бы компьютеризация в равной степени охватила бы все страны мира, то эти проблемы разрешились бы давным-давно. В действительности же, компьютерная технология развивалась на основе английского языка, довольно простого по структуре, и базирующегося на таком же простом латинском алфавите. Но многие современные документы, использующиеся в международной практике, гораздо сложнее по структуре и языку.

Другие вопросы касаются стандартизации. Хотя Unicode является международным стандартом, языки описания страниц являются неформальными стандартами, которые не были приняты в качестве международных. С тех пор, как PostScript был впервые представлен в 1985 году, он претерпел значительные изменения: сначала появились версии PostScript первого, второго и третьего уровней затем в 1992 году – PDF 1.0, превосходящий все предыдущие версии PostScript и далее развившийся до PDF 1.5, который был выпущен в середине 2003 года. Наряду с расширением представлений о том, какими должны быть языки описания страниц, совершенствуется и программное обеспечение. Определенную проблему при работе с этими документами создает существование множества их разных версий и дополнений. Когда PostScript только появился, документы всегда можно было обработать, причем даже те, которые создавались в других странах и на других компьютерах. К сожалению, в процессе развития PostScript многие из этих документов стали считать нечитаемыми, поскольку программное обеспечение на местах не могло их интерпретировать. Появление PDF изменило ситуацию к лучшему, но и с ним уже начинает происходить то же самое, только не так стремительно.

До сих пор приходится решать вопрос коммерциализации. PostScript и PDF поддерживают открытые стандарты: полная инфор-

мация о том, как они представляют документы, доступна всем. И все же, они являются коммерческими продуктами, и их стоимость является серьезным препятствием для широкого распространения PDF, поскольку компании берут деньги за программное обеспечение, которое создает PDF файлы. (В самом начале вообще нужно было платить деньги даже за чтение PDF файлов!) Хорошо знакомый всем пользователям формат Word также не является открытым и не гарантирует (хотя и относится к программному обеспечению семейства Microsoft) надежную преемственность версий документов Word. Идея понятна: прибыль зависит напрямую от повторных продаж более совершенных версий программного продукта, в то время как совместимость разных версий программ является технически непростой задачей, но при этом не дает прибыли. Детальное описание форматов RTF и LaTeX доступно всем без исключения, а для обработки документов LaTeX существует значительное количество программ с открытым источником.

В этой главе мы намеревались подготовить вас к тернистому пути, по которому вам придется пройти, чтобы создать электронные библиотеки. Хотя при решении этих проблем могут потребоваться специальные технические знания, даже навыки программирования, мы верим, что преодоление возникающих проблем и даже допущенные вами ошибки сослужат вам хорошую службу и сформируют необходимые навыки выявления проблем и анализа ситуации, а также позволит вам обсудить с техническими экспертами пути их решения.

В некоторых случаях вы сможете справиться с проблемой за счет грамотного использования существующего программного обеспечения. Например, программы Microsoft имеют больше шансов справиться со сложными документами Word, чем программы создания электронных библиотек с открытым кодом, чьи авторы не вникали во внутренние детали конкретного формата документов Word. Если у вас появились сложности в работе с документом, и у вас установлена программа Word, вам следует открыть в этой программе статьи из Хрестоматии (файлы с расширением .doc), а затем конвертировать их в другой формат, возможно, в PDF. Если вы обнаружите, что версия в PDF также вызывает проблемы, то вы должны попробовать записать документ в HTML. Хотя качество документа будет ниже, но вы все равно сможете включить его в вашу электронную библиотеку.

А теперь вы уже готовы попробовать создать свои собственные электронные библиотечные коллекции из документов и в форматах, о которых вы здесь узнали. Чтобы сделать это, переходите к Главе 2.2.

Глава 2.2. Создание электронной библиотечной коллекции

Эта глава целиком посвящена практической работе. В ходе работы над Главой 1.3 вы установили заранее укомплектованные коллекции и приобрели опыт работы с читательским интерфейсом Greenstone. Далее вы проведете инсталляцию полной версии программы Greenstone, которую установить очень просто, однако потребуются некоторое время для копирования файлов. Кроме этого, вы научитесь использовать для создания коллекций интерфейс библиотекаря. Если читательский интерфейс достаточно прост в использовании и не требует дополнительного освоения, библиотечарский интерфейс более сложен, поскольку он ориентирован на решение более трудных задач.

Большая часть данной главы базируется на учебном модуле компакт-диска IMARK, содержащем блестящее введение в Greenstone. При изучении главы вы будете работать с Уроками 6.2 и 6.3 на CD-ROM IMARK, посвященными использованию интерфейса библиотекаря (GLI – Greenstone Librarian Interface).

Задание 1

Проработайте Урок 6.2. учебного модуля на CD-ROM IMARK и выполните все практические задания, включающие:

- установку Greenstone с помощью IMARK CD-ROM;
- запуск интерфейса библиотекаря (Librarian Interface);
- объединение документов;
- улучшение вида документов путем назначения их метаданных.

Вы найдете Greenstone на компакт-диске IMARK в папке **software_tools** → **Greenstone**. Сделайте двойной щелчок по названию файла **setup.exe** для запуска процесса установки. В этот раз вам нужно инсталлировать больше файлов, поэтому процесс установки займет примерно в два раза больше времени, чем в прошлый раз.

Обратите внимание: мы исходим из того, что программа Greenstone ранее на вашем компьютере ни разу не устанавливалась. Но если это не так, то очень важно удалить старую версию до того, как записывать на компьютер новую. Задание 4, приведенное ниже, описывает, как это сделать.

(При этом вам не нужно удалять подготовленные коллекции, которые вы устанавливали в ходе работы с Главой 1.3.).

Урок 6.2. учебного модуля на компакт-диске IMARK описывает процесс установки. Однако он не касается одной существенной детали инсталляции, которая чрезвычайно важна для достижения поставленных нами целей. Как только программа Greenstone установлена, и вам зададут вопрос о том, хотите ли вы установить ImageMagick: скажите «Да» в ответ на этот вопрос. Для того чтобы установить эту программу, вы должны обладать правами «администратора» в Windows. Далее мы описываем шаги, которые вам нужно предпринять для решения этой проблемы, и, как и раньше, мы рекомендуем вам использовать в основном установки по умолчанию. Вот что вам нужно сделать.

1. "This will install ImageMagick 5.5.7 Q8. Do you wish to continue?"(Будет проведена установка ImageMagick 5.5.7 Q8. Вы хотите продолжить?) **Yes Да**).
 2. "Welcome to the ImageMagick Setup Wizard" Click **<Next>**. (Добро пожаловать в мастер ImageMagick Setup Wizard. Щелкните <Далее>).
 3. "Information: Please read the following ..." Click **<Next>**. ("Информация: пожалуйста, прочитайте следующее ..." Щелкните <Далее>).
 4. "Select Destination Directory ..." Leave at default and click **<Next>**. («Выберите директорию ...» Оставьте директорию по умолчанию и щелкните <Далее>).
 5. "Select Start Menu Folder ..." Leave at default and click **<Next>**. («Выберите меню «Пуск» ...» Оставьте установку по умолчанию и щелкните <Далее>).
 6. "Select Additional Tasks ..." Leave at default and click **<Next>**. («Выберите дополнительные задачи ...» ... Оставьте выбор по умолчанию и щелкните <Далее>).
 7. "Ready to Install". Click **<Install>**. («Готов к установке». Щелкните <Инсталлировать>).
- Файлы копируются.
8. "You have now installed ..." Click **<Next>**. («Вы установили ...» Щелкните <Далее>).
 9. "Setup has finished ...". Deselect "View index.html" and click **<Finish>**. («Установка завершена ...». Откажитесь выполнить команду «Просмотреть указатель в HTML» и щелкните <Готово>).

Обратите внимание: версию Greenstone, которую вы только что установили, можно запустить через Электронную библиотеку Greenstone с помощью меню Пуск→Программы. Коллекции, подготовленные для курса «Электронные библиотеки в образовании», которые вы установили, находятся в том же самом меню под строкой Greenstone, в Главе 1.3.

Теперь ответьте на следующие вопросы:

- Теперь на вашем компьютере установлены две отдельно существующие программы Greenstone – данная и та, что вы установили в ходе работы над Главой 1.3! Как вы находите каждую из них в меню Windows Пуск?
- Назовите пять основных этапов создания коллекций с помощью интерфейса библиотекаря?
- Сколько всего документов находится в созданной вами коллекции?
- Какие из форматов документов, описанных в предыдущей Главе, представлены в данной коллекции?
- Какой из этих документов открывается после того, как вы делаете двойной щелчок по их названиям?
- Какой набор метаданных используется в данной коллекции и из скольких элементов метаданных он состоит?
- Изучите систему помощи в интерфейсе библиотекаря, используя кнопку в верхнем правом углу экрана. Сколько подглав включает глава, названная «Enriching the Collection with Metadata» (Улучшение внешнего вида коллекций с помощью метаданных)?

Задание 1 помогло вам одолеть половину пути, который нужно пройти для создания электронной библиотечной коллекции. Задание 2 завершает этот процесс.

Задание 2

Проработайте Урок 6.3. учебного модуля на компакт-диске IMARK и выполните все необходимые практические задания, что включает:

- принятие решения по выбору инструментов поиска и просмотра, которые должны быть предложены читателям данной коллекции;
- проектирование коллекции;
- ее создание;
- ее предварительный просмотр;
- ее запись на CD-ROM.

Теперь ответьте на следующие вопросы:

- Какой из компонентов Greenstone определяет форматы документов, включенных в коллекцию?
- Какую часть панели Design вы использовали для выбора графического образа коллекции?
- Какие два различных вида поиска предлагает Greenstone (когда в библиотечарском интерфейсе включен режим расширенного поиска (Advanced Search))?

- Сколько поисковых указателей используется сейчас в вашей коллекции? Где они показаны в читательском интерфейсе? Как они называются?
- Сколько там просмотровых классификаторов и где в интерфейсе читателя они показаны?
- Сколько времени потребовалось вашему компьютеру, чтобы создать эту коллекцию? Рассчитайте скорость передачи данных в мегабайтах за единицу времени.
- Сколько места на компакт-диске занимает эта коллекция?
- Сколько документов появляются в списке заголовков a–z?
- Сколько документов вы получите, если проведете поиск документов по такому простому слову как *computer* (компьютер)?
- Сколько документов в данной коллекции написаны самой Салли Джо Каннингэм (Sally Jo Cunningham) и сколько в соавторстве с другими?
- Есть ли в коллекции копии одного и того же документа, записанные в разных форматах?

Эта коллекция включает документы, представленные в разных форматах — Word, RTF, PDF и PostScript. При выполнении Задания 2, вам, возможно, было трудно разобраться в том, что же именно произошло с некоторыми из них. Вы переместили в коллекцию двенадцать документов и, если все прошло хорошо, они все появились в списке результатов поиска, проведенного с помощью такого обычного запроса как *компьютер (computer)*. (Вы можете сказать, все ли прошло хорошо, основываясь на результатах, описанных в материалах для чтения, а именно — на этапе работы с коллекцией, названной *Build*, где говорится о том, сколько именно документов было включено в коллекцию.) Но сколько именно документов появляется в списке названий документов списка *a–z*? Их может быть меньшее количество, потому что этот список формируется на основе *dc.Title* (что означает «метаданные Dublin Core *Title*»), т.е. в него попадают только те документы, которые имеют значение для этого элемента метаданных. Возможно, вы не определили метаданные для всех 12 документов. Например, на этапе *Enrich* вы не смогли просмотреть содержание всех типов файлов — это зависит от программного обеспечения, установленного на вашем компьютере, и от того, как именно оно установлено.

Greenstone конвертирует все документы в формат HTML, поэтому эта коллекция включает как исходные форматы документов (Word, RTF, PDF или PostScript), так и их HTML версии. Выберите не-

сколько документов и сравните обе версии. Вы увидите, что качество версии в HTML ниже, форматирование не очень удачное (например, текст больше не разбивается на страницы), но сам текст остается без изменений, и это именно тот текст, который индексируется для полно-текстового поиска.

Если вы не смогли внести все необходимые метаданные из-за того, что ваш компьютер, скажем, не может показывать файлы PostScript, вы можете сделать это прямо сейчас, проведя в вашей коллекции поиск по какому-нибудь распространенному слову (например, компьютер – *computer*) и найдя название, имя автора и т.д. в HTML версии этого файла.

Дальше давайте посмотрим, как Greenstone может создать из этих же документов простейшую коллекцию. Greenstone может сам извлекать метаданные из файлов документов, хотя, конечно же, он не может этого сделать лучше человека! Используя эти извлеченные метаданные, можно очень просто и быстро создать самую элементарную коллекцию. В Greenstone используются очень тонкие установки по умолчанию, которые позволяют вам работать над созданием коллекции, не уделяя особого внимания вопросам ее дизайна. Прежде, чем вы повторите этапы создания коллекции, вы познакомитесь с тем, как с помощью интерфейса библиотекаря Greenstone сможет создать удобный ярлык для исходных файлов, что очень облегчит дальнейшую работу с коллекцией.

Задание 3

Для создания ярлыка для исходных файлов вернитесь к панели Gather и определите местонахождение той папки на вашем компьютере, в которой находятся исходные документы в форматах Word, RTF, PDF и PostScript (вы использовали эти файлы в ходе выполнения Задания 1). Выберите папку, а затем щелкните по ней правой кнопкой мыши. Следуя инструкциям, создайте ярлык. Закройте все папки в проводнике, и вы увидите, что исходные файлы теперь помечены ярлыком.

Повторите этапы создания коллекции, описанные в Уроках 6.2 и 6.3 учебного модуля на компакт-диске IMARK, используя при этом другое название для данной коллекции. Пропустите этапы работы, связанные с использованием панелей Enrich и Design, другими словами, просто перетащите те же самые 12 файлов и переходите напрямую к панели Create. А теперь ответьте на вопросы:

- Сколько в этой коллекции поисковых указателей? Как они называются?
- Сколько в ней просмотровых классификаторов, и как они называются?
- Что появляется на месте пиктограммы, которую вы установили для данной коллекции в Задании 2?
- Сколько документов появляется в списке заголовков a–z?

Поскольку вы не вводили каких-либо метаданных, то по какому именно принципу сформировался список заголовков a–z? Откуда появились эти заголовки? Создав коллекцию, выберите файл на панели Enrich и просмотрите до самого конца весь список метаданных. Вы увидите значения метаданных, которые начинаются с “ex.” ex.Title задает метаданным значения, используемые в классификаторе a–z. Элемент метаданных ex.Source всегда повторяет название файла, содержащего документ, поэтому классификатор filenames всегда показывает все документы, содержащиеся в коллекции.

Greenstone выбирает подобные метаданные из файлов документов. Помимо заголовка (Title), эти метаданные включают информацию о стандарте кодирования (например, iso_8859_1), формате файла (например, PS for PostScript), размере файла (в байтах), внутреннем идентификаторе документа Greenstone, языке написания документа (например, “en” для английского языка), а также имени приложения, которое его открывает.

- Автоматическое формирование метаданных не может быть абсолютно точным (и никогда таким не станет). Найдите какие-либо ошибки в сформированных заголовках. Как вы думаете, какой из методов использует программа Greenstone для формирования заголовков?
- Две созданные вами коллекции отличаются друг от друга по метаданным, которые они используют для просмотровых классификаторов. Где именно в интерфейсе библиотекаря можно найти об этом информацию? Greenstone выделяет метаданные даже в том случае, когда вы определили метаданные только лишь на этапе работы Enrich. Откройте еще раз коллекцию, которую вы создали, выполняя Задания 1 и 2, и найдите выделенные метаданные, используя панель Enrich. Хотя эти метаданные и выделены, они в этой коллекции не используются.

Задание 3 иллюстрирует два очень важные момента, на которых следует остановиться отдельно. Во-первых, метаданные, сформированные вручную, являются (или должны являться) более качественными, чем извлеченные метаданные. Разумеется, это делает сам процесс их создания более дорогим! Во-вторых, Greenstone имеет сложнастраиваемые установки по умолчанию. Если даже вы не сделали ничего другого, кроме как перенесли в коллекцию извне несколько файлов и щелкнули *Build*, Greenstone постарается сделать из этого что-либо

приемлемое. Эта политика пронизывает весь дизайн Greenstone. Конечно, результаты работы будут напрямую зависеть от того, насколько активно вы поработали на этапах *Enrich* и *Design*.

Следующее упражнение познакомит вас с тем, как управлять процессом установки программы Greenstone. Как мы отмечали в Задании 1, к настоящему времени на вашем компьютере уже проведены две отдельные инсталляции Greenstone: первую установку вы провели, выполняя Задание 1 данной главы, а другую – в Главе 1.3. Предположим, что вы установили программу в директории, предлагаемой инсталлятором по умолчанию. В ходе установки, которую вы проводили в Задании 1, это было сделано на *C:\Program Files\gsdl*, а версия Greenstone, которую вы проинсталлировали в Главе 1.3 – на *C:\GSDL*. Мы также полагаем, что в Главе 1.3 вы последовали нашим рекомендациям и скопировали весь компакт-диск, выбрав строку с командой *Установить все файлы коллекции (Install all collection files)*. (Если вы это не сделали, тогда мы просим вас провести повторную установку вашего компакт-диска с курсором «Электронные библиотеки в образовании» в соответствии с указаниями, данными в Главе 1.3.)

В этом упражнении вы обновите программу Greenstone установкой более свежей версии и перенесете уже созданные коллекции в новую версию с последующим удалением старой. Выполняя эти упражнения, вы научитесь определять местонахождение программы и самих коллекций на вашем компьютере, а также регулярно обновлять их в связи с появлением новых версий программы.

Упражнение: Обновление инсталляции

Перед выполнением этого упражнения убедитесь в том, что на вашем компьютере не запущен библиотечный интерфейс или локальный библиотечный сервер Greenstone. Как правило, завершение работы с GLI также прекращает работу сервера.

Первым делом удалите версию инсталляции Greenstone, которую вы загрузили на свой компьютер в Главе 1.3 и которая находится в директории *C:\GSDL*. Сначала мы перенесем содержащиеся в ней три коллекции – коллекцию «*Электронные библиотеки в образовании*», «*Как создать электронную библиотеку*» и демонстрационную коллекцию LOM на рабочий стол Windows для того, чтобы мы их не удалили заодно с остальным содержимым. Greenstone хранит каждую коллекцию в отдельной поддиректории общей папки. Перенесите всю папку *C:\GSDL\collect* целиком на рабочий стол Windows вашего компьютера. Обратите внимание, что в папке должно быть три вложенные папки – по одной на каждую коллекцию.

А теперь в меню Пуск Windows перейдите в строку с открывающимся меню Greenstone и выберите команду **<Uninstall>** для того, чтобы убрать версию инсталляции Greenstone, которую вы выполнили в Главе 1.3. Это мини-версия Greenstone, которая содержит читательский интерфейс, а также некоторые заранее подготовленные коллекции (в нашем случае, три): все это можно найти на любом из компакт-дисков, который вы создадите с помощью Greenstone. Вы не можете создать коллекцию без этой мини-версии Greenstone.

Затем мы просим вас обновить имеющуюся у вас версию программы Greenstone. Версия, которую вы инсталлировали с компакт-диска *IMARK* — это версия 2.51. Ваш компакт-диск — «*Электронные библиотеки в образовании*» — также содержит отдельную копию полной версии этой программы, отличающуюся от только что удаленной мини-версии тем, что она более высокого уровня и называется версией 2.60. Именно с ней мы и будем работать на протяжении всего курса.

Обратите внимание: прежде чем проводить инсталляцию любой новой версии Greenstone, вам нужно полностью удалить существующую версию.

Во-первых, убедитесь в том, что в данный момент вы не работаете с программой Greenstone. Затем удалите старую версию программы, перейдя на панель управления Windows (выбрав строку Настройки в меню *Пуск*). Щелкните по группе *Установка или удаление программ*, выберите *Greenstone Digital Library Software* и дайте команду *Удалить*. (Для того, чтобы это сделать, вам, возможно, нужно иметь уровень доступа для работы в Windows с правами «Администратора».)

По завершении этой процедуры вам будет задан вопрос, хотите ли вы удалить все имеющиеся у вас коллекции Greenstone: в этом случае ответьте Да поскольку у вас пока еще нет сколько-либо ценных коллекций. Заметьте, что программа-инсталлятор удалит только те файлы, которые были установлены в самом начале, но не удалит ни один из тех файлов, которые вы могли бы создать позднее. Вам нужно очистить вашу систему, проверив папку *C:\Program Files\gsdl*, созданную при инсталляции, и полностью удалить ее, если она пока еще не исчезла.

Теперь проведите установку версии программы Greenstone, имеющуюся на вашем компакт-диске «*Электронные библиотеки в образовании*». Сначала вставьте компакт-диск в дисковод и отмените процесс авто-установки, если он запустится автоматически. Откройте CD-ROM, щелкнув правой кнопкой мыши по соответствующему названию дисковода (обычно, это D:) в окне *Мой компьютер* и выбрав команду **<Открыть>** (**<Open>**) из меню.

Среди файлов, которые вы увидите, двойным щелчком активизируйте тот, что называется *setup.exe*. Затем следуйте инструкциям по инстал-

ляции, выбирая установки по умолчанию. Вам не нужно вновь устанавливать ImageMagick.

Вы заметите несколько незначительных изменений в процедуре установки, поскольку версия 2.60 использует другую программу-инсталлятор. Это как раз то, что вы должны знать наверняка: новая версия Greenstone была размещена в папку *Program Files\greenstone*, в то время как версия 2.51 была размещена в папке *Program Files\gsdl* (обе эти папки были определены по умолчанию, и вы могли их изменить в процессе установки). Данное нововведение унаследуют и все последующие версии Greenstone. В связи с этим, если вы захотите сохранить существующие коллекции, вам придется отдельно перенести содержимое папок со старого места на новое. Будущие версии Greenstone также будут устанавливаться на новое место *Program Files\greenstone*, поэтому с этой проблемой вы больше не столкнетесь.

А теперь обратитесь к обновленной установке Greenstone, перейдя в основное меню Windows Пуск и найдя там название программы *Greenstone Digital Library Software*, а затем, выбрав *Greenstone Digital Library*, начать работу в читательском интерфейсе (Reader's interface), а не в интерфейсе библиотекаря (Librarian interface). Вы увидите, что в дополнение к тем примерам коллекций документов, с которыми вы будете работать в дальнейшем, в верхней части домашней страницы Greenstone есть еще одна коллекция – *Greenstone Demo*.

В той папке, которую в начале выполнения этого задания вы поместили на рабочий стол Windows, находятся еще три коллекции. Перенесите эти три вложенные папки (они называются *course*, *howto* и *lomdemo*) внутрь основной папки, в которую была установлена программа Greenstone, а именно:

C:\Program Files\greenstone\collect

Теперь завершите работу с программой Greenstone (вы можете это сделать, закрыв веб-браузер) и повторно запустите ее по той же самой схеме. Эти три коллекции появятся на домашней странице Greenstone.

И, наконец, запустив библиотечарский интерфейс из меню *Пуск* Windows, вы увидите, что слева от других кнопок появится новая, называемая "Download". Вы научитесь использовать эту новую команду в Главе 3.1.

Интерфейс основной части программы Greenstone существует на четырех языках: английском, французском, испанском и русском. Однако существуют и другие языковые версии этой программы. Если вы хотите пользоваться Greenstone с интерфейсом на другом языке, то вам нужно установить программу Greenstone Language Pack, которая есть на имеющемся у вас компакт-диске с курсом. Чтобы сделать это, найдите папку *Greenstone Language Pack* на CD-ROM и сделайте двойной щелчок по пиктограмме с файлом запуска *setup.exe*.

На этом этапе работы мы надеемся, что вы согласитесь с утверждением, что, используя Greenstone, очень просто создать электронную библиотечную коллекцию, объединив в ней документы самых разных форматов. Но существуют и отдельные подводные камни, появление которых мы предполагали в ходе изучения этих форматов в Главе 2.1. В связи со сложностью существующих форматов, и, в особенности, из-за нечеткости работы самого формата Word, иногда что-то перестает работать.

Greenstone использует программное обеспечение с открытым источником для конвертирования документов, созданных в различных форматах, в стандартный формат внутреннего представления документов, в основе которого лежит HTML. Иногда данная программа неспособна конвертировать некоторые файлы. Например, как вы узнали из Задания 3 (Урок 2.7 учебного модуля на CD-ROM IMARK), документ в PDF формате может состоять не из текста, а из последовательности отсканированных изображений – рисунков или текстов. Хотя он воспринимается как обычный текст, прочесть его машиной невозможно без использования оптического распознавания символов (Optical Character Recognition, OCR), операции, сложность которой зависит от качества изображения, вида текста и от используемых шрифтов, и не всегда полученные результаты будут совершенны. Дополнительное программное средство, используемое приложением Greenstone для прочтения PDF документов, для этого не предназначено. Хотя вы и увидите сам PDF файл, текст в нем не будет проиндексирован. Другой пример – текст из текстового документа Word иногда выделяется некорректно, что обычно происходит при включенном режиме «быстрого сохранения» текста. В этом случае кроме основного текста может появиться и дополнительный, который был удален, но в действительности не убран из быстро сохраненного файла. Третий пример – конвертация может не сработать для определенного типа файлов Word, как правило, для очень старых версий или тех, что были созданы на еще более старых системах Macintosh. Даже Microsoft Word не может читать некоторые файлы, которые были созданы много лет назад.

Подобные проблемы возникают довольно редко, и мы надеемся, что вы с ними не столкнетесь! Но если это и случится, то следующее задание поможет вам с ними справиться. Кто предупрежден, тот вооружен.

Задание 4

Создайте свежую коллекцию Greenstone из двух файлов, которые хранятся в папке `sample_files\difficult_documents`, расположенной на компакт-диске с курсом «Электронные библиотеки в образовании». Эти файлы называются `No extractable text.pdf` и `Weird characters.pdf` – их названия указывают на те проблемы, которые они могут вызвать! Используйте конфигурацию коллекции по умолчанию: то есть, просто объедините эти два файла, создав новую коллекцию.

Теперь проведем предварительный просмотр коллекции.

- Сколько документов появляется в списке заголовков a–z?
- Сколько документов появляется в списке имен файлов?
- Что необычного в документе, который появляется в окне?
- Были ли вам продемонстрированы какие-нибудь предупреждения, в то время как Greenstone создавала коллекцию?
- Как именно выглядел документ в PDF формате, который не вошел в коллекцию?

Списки заголовков и имен файлов показывают только один из документов. В процессе создания коллекции появляется такое сообщение: «Один документ был обработан и включен в коллекцию; один был отклонен» (“One document was processed and included in the collection; one was rejected”).

Библиотечарский интерфейс может работать в разных режимах. До сих пор вы пользовались режимом, называемым «Librarian» (Библиотечарский), который был установлен по умолчанию. Используйте вкладку Preferences в меню File для перехода в режим Эксперта (Expert) и затем заново создайте коллекцию. Панель Create выглядит в режиме Эксперт по-другому, так как она предлагает большее количество команд: в нижней части окна найдите кнопку Build Collection и щелкните по ней. Теперь вы увидите сообщение, в котором говорится, что файл не может быть обработан и объясняется, почему именно.

- Какие еще существуют режимы работы в Greenstone, помимо Библиотечаря и Эксперта?

Аномалия с символами на китайском языке, с которой вы столкнулись в Задании 4, является проблемой, наглядно показывающей некоторую несовместимость символов в Windows, проявившуюся в этом конкретном документе. На самом деле, эта проблема появляется только на компьютерах с операционной системой Windows, а на других компьютерах документы ведут себя абсолютно нормально. Решение этой проблемы очевидно: определите для этого файла метаданные

Title вручную. Документ, который был отвергнут в Задании 4, является PDF файлом, состоящим из последовательности нескольких включенных изображений, и в нем вообще отсутствует машиночитаемый текст. Greenstone не может справиться с таким видом PDF файла. Другим вариантом решения могло бы стать создание машиночитаемой версии документа путем выполнения оптического распознавания символов (optical character recognition, OCR) данного файла или его конвертация из PDF формата в последовательность графических файлов и последующая их обработка с помощью соответствующего приложения (вы будете создавать коллекцию, содержащую последовательность постраничных изображений в Главе 5.3).

Другие проблемы периодически возникают с использованием файлов Word, в особенности, созданных в устаревших версиях Word. При появлении ошибок в программе, которые могут возникнуть в процессе выполнения типичного для Word «быстрого сохранения» отключите данную установку и повторно сохраните файл. В других случаях прочитайте документ в Word и используйте команду «Сохранить как HTML» для перевода его в формат, который Greenstone сможет обработать. Когда вы занимаетесь созданием электронных библиотек, вам нередко приходится быть очень предприимчивым в поиске вариантов выхода из подобных сложных ситуаций.

Заключение

В ходе работы вы устанавливали Greenstone и практиковались в создании электронной библиотечной коллекции, состоящей из документов, представленных в различных форматах.

Вы увидели пять этапов создания коллекции: подбор документов, их обогащение метаданными, проектирование коллекции путем определения инструментов поиска и просмотра, предлагаемых пользователю, построение коллекции и предварительный просмотр результата. (В Главе 3.1 вы узнаете еще об одном возможном этапе: выгрузке документов из WWW.) Обычно, этот процесс проходит несколько итераций: вы прокручиваете по несколько раз этапы проектирования, построения и предварительного просмотра до тех пор, пока не будете удовлетворены полученным результатом. Скорее всего, вы начинали с того, что использовали низкокачественное автоматическое извлечение метаданных и переделывали на основе этого дизайн коллекции. Вы начали работу с небольшого количества документов, но как только вы определили, что удовлетворены общим дизайном коллекции, вы стали

добавлять остальные документы, и перешли к трудоемкому процессу обогащения их метаданными.

По нашему опыту, самую простую заготовку для коллекции вполне можно создать за один или два часа. Последующий процесс проектирования коллекции может занять от дня до недели. И, наконец, добавление всех необходимых метаданных может растянуться на месяцы, и даже годы в зависимости от масштаба проекта.

Вы также попрактиковались в проведении некоторых операций по обслуживанию Greenstone. Вы знаете, где это программное обеспечение размещается на вашем компьютере и где находятся сами коллекции. Домашняя страница Greenstone автоматически показывает коллекции в папке collect (хотя в программе присутствует дополнительный механизм для того, чтобы при желании их можно было скрыть). Вы можете поэкспериментировать, перемещая коллекции, встраивая их в уже существующие коллекции или возвращая их обратно: возможно, вам понадобится перезапустить Greenstone, чтобы увидеть результат произведенных действий. Мы не описывали другие директории, созданные в процессе инсталляции Greenstone, поскольку вы сами можете с ними познакомиться. Главным в программах с открытым источником является то, что они всегда открыты для исследований.

Вы также провели дополнительную установку Greenstone. Обратите внимание, что это никак не отразилось на находящихся на вашем компьютере коллекциях. Вообще, идея обновления программы Greenstone путем установки новых версий, появляющихся на сайте <http://www.greenstone.org> примерно два раза в год, очень хороша. Будет неплохо, если вы присоединитесь к дискуссионной группе, упомянутой в Главе 4.3. В новых версиях решаются «баги», которые, возможно, вас беспокоили, и добавляются описания новых возможностей, которые могут оказаться очень полезными. Совместимость с предыдущими версиями программы является для разработчиков Greenstone основным приоритетом: вы не будете испытывать трудностей при обновлении своей программы. И к тому же, это все будет для вас совершенно бесплатно!

Теперь пришло время узнать больше о метаданных вообще и образовательных метаданных в частности. Вы будете изучать метаданные в Модуле 3. Вам предстоит продолжение знакомства с Greenstone, и вы выполните упражнения, которые расширят ваши знания, и шаг за шагом превратят вас в опытного пользователя.

МОДУЛЬ 3

РАБОТА С МЕТАДАННЫМИ

Глава 3.1. Разметка

Глава 3.2. Метаданные

Глава 3.3. Образовательные
метаданные



Цель

Понять роль метаданных в организации таких электронных ресурсов, как электронные библиотеки и системы управления знаниями (LMS – Learning Management Systems); познакомиться с наиболее распространенными стандартами метаданных, включая те из них, которые специально предназначены для образовательных метаданных; оценить роль и значение языка разметки XML и приобрести опыт определения и использования метаданных.

Задачи

По завершении работы над Модулем 3 вы сможете:

- объяснить значение термина «метаданные»;
- понять различие между встроенными и внешними метаданными;
- рассказать историю развития HTML, SGML и XML;
- выделить различия в организации HTML, SGML и XML;
- объяснить целевое назначение XHTML и CSS;
- объяснить, что подразумевается под DTD и объяснить, какое отношение это имеет к документам XML;
- разобраться в лабиринте X-стандартов: XSL, XLink, XPath, XML Schema, XSLT;
- описать стандарт метаданных Dublin Core;
- обсудить, как метаданные MARC могут быть включены в Dublin Core;
- выделять черты сходства и различия библиографических форматов BibTeX и Refer;
- назвать наиболее популярные стандарты метаданных для изображений и мультимедиа;
- выбрать соответствующие форматы для метаданных вашей собственной электронной библиотеки;
- спроектировать и построить разнообразные электронные библиотечные коллекции Greenstone;
- объяснить, в чем отличие целей LOM и SCORM;
- построить коллекции Greenstone на основе метаданных LOM;
- объяснить взаимосвязь, существующую между LOM и SCORM.

Введение к Модулю 3

Документы и метаданные являются двумя основными компонентами любой библиотеки. Документы формируют контент, а метаданные – поддерживают его организацию. В Модуле 2 вы детально изучили документы и их форматы, а в Модуле 3 вы познакомитесь с тонкостями работы с метаданными и особенностями их форматов. Как и раньше, все это будет детально проиллюстрировано в ходе выполнения упражнения по работе с программой электронных библиотек Greenstone.

Также как и документы, метаданные – часто описываемые как «данные о данных» – являются исходным материалом для электронных библиотек. Они составляют основу организации как электронных, так и традиционных библиотек. Связанный с ними термин «разметка» (markup) исторически относится к процессу аннотирования документов с помощью дополнительно составленной текстовой информации. В электронной среде это соответствует аннотированию документов с помощью команд форматирования. Позднее это нашло более широкое применение при аннотировании документов, содержащих структурную информацию, включающую также и метаданные. Глава 3.1 посвящена языку разметки и представляет такие языки разметки как HTML, XML и соответствующие им стандарты. Язык XML больше похож на схему для описания языков разметки, нежели на один из языков разметки. Глава 3.2. посвящена метаданным. В нем проводится сравнение метаданных, используемых в электронных библиотеках, с содержанием традиционных библиотечных каталогов и демонстрируется прием извлечения метаданных из исходного текста самих документов.

В Модуле 2 вы научились создавать простые электронные библиотечные коллекции, используя Greenstone. Метаданные являются ключом к организации данных в электронной библиотеке, а Главы 3.1 и 3.2 содержат серии разноуровневых упражнений, которые позволят вам гораздо лучше разобраться в средствах организации документов, имеющихся в Greenstone. Особые механизмы доступа пользователей к любой из созданных коллекций основываются на использовании метаданных, которые определяются внутри самой коллекции. Вы узнаете, как использовать библиотечарский интерфейс для проектирования различных коллекций. Подобным практическим упражнениям в этих главах отводится значительное место, поэтому вы должны быть готовы уделить им немалую часть своего времени.

Особое место в данном курсе занимает вопрос о метаданных образовательных ресурсов, что является основной темой Главы 3.3.

Для большинства используемых учебных материалов исключительно важными являются такие их свойства как: уровень образования, направление подготовки, уровень сложности, среднее время, затрачиваемое на работу с этими материалами и т.д. Сегодня стало очень популярным разбивать весь учебный материал на отдельные части, называемые «учебными объектами» (learning objects), и использовать метаданные для установления их взаимосвязи и дальнейшего совместного использования в рамках того или иного учебного курса при выполнении типовых заданий. Педагогам были предложены различные стандарты образовательных метаданных для быстрого обнаружения, обмена или совместного использования информации, полученной в образовательных электронных библиотеках. В Главе 3.3 вы узнаете о двух наиболее значимых стандартах представления образовательного контента: LOM (Learning Object Metadata), стандарте метаданных для учебных объектов и SCORM (Sharable Content Object Reference Model), стандарте, состоящем из множества компонентов, описывающих использование целого ряда существующих стандартов, включая LOM. Вы узнаете о том, как построить коллекции образовательных метаданных Greenstone, используя LOM. LOM построен на основе XML и является примером практического воплощения концепции, представленной в Главе 3.1, в рамках которой XML может использоваться для описания других языков разметки.

Модуль 3. Материалы для чтения

Глава 3.1. Разметка

Материал для чтения 1. Учебник «Как создать электронную библиотеку»: Глава 5. Введение, Разделы 5.1, 5.2 и 5.3.

Назначение. Познакомиться с такими понятиями как разметка и метаданные, а затем с описанием языка разметки гипертекста (Hypertext Markup Language – HTML) и расширенным языком разметки (Extensible Markup Language – XML), представлением и форматированием документов, размеченных в XML.

Материал для чтения 2. CD-ROM IMARK: Раздел 2. «Электронные документы и форматы», Уроки 2.4 и 2.5.

Назначение. Урок 2.4 не входит в обязательную часть данного курса, но мы рекомендуем использовать его в качестве дополнительного материала для более глубокого понимания HTML. Урок 2.5 представляет XML и дает дальнейшее развитие идей, заложенных в тексте учебного пособия.

Глава 3.2. Метаданные

Материал для чтения 1. Учебник «Как создать электронную библиотеку»: Глава 2, Раздел 2.2 и Глава 5, Разделы 5.4 и 5.6.

Назначение. В Главе 2 приводятся фрагменты описаний основных элементов, из которых складываются традиционные библиографические системы. Раздел 5.4 описывает два основных стандарта представления метаданных документа: MARC и Dublin Core. В Разделе 5.6 разъясняется, как можно автоматически извлекать метаданные из документа.

Материал для чтения 2. CD-ROM IMARK: Раздел 3 «Стандарты метаданных и предметное индексирование», Уроки 3.1, 3.2 и 3.3.

Назначение. Урок 3.1 представляет описательные метаданные и объясняет их назначение. Урок 3.2 дает объяснения по отдельным элементам стандарта метаданных Dublin Core, а Урок 3.3 описывает существующие расширения к данному стандарту.

Глава 3.3. Образовательные метаданные

Материал для чтения 1. CD-ROM «Электронные библиотеки в образовании», Хрестоматия, перевод статьи Эрика Дюваля «Для чего нужна стандартизация в области образовательных технологий».

Duval, E. (2004) Learning Technology Standardization: Making Sense of it All. International Journal on Computer Science and Information Systems, 2004, No. 1, pp. 33–43. <http://www.comsis.fon.bg.ac.yu/ComSISpdf/Volume01/InvitedPapers/ErikDuval.pdf>

Назначение. Получить полное представление о существующей ситуации со стандартами образовательных метаданных.

Материал для чтения 2. CD-ROM «Электронные библиотеки в образовании», Хрестоматия, перевод статьи Нормы Фрайзена, Джона Мэйсона, Найджела Уарда «Построение прикладных профилей метаданных в образовании».

Friesen, N. Mason, J. and Ward, N. (2003) Building Educational Metadata Application Profiles.

Proceedings of the International Conference on Dublin Core and Metadata for e-Communities, 2003, pp. 63–69. <http://www.bncf.net/dc2002/program/ft/paper7.pdf>

Назначение. Познакомиться с примерами из практики по истории развития двух прикладных профилей метаданных.

Материал для чтения 3. CD-ROM «Электронные библиотеки в образовании», Хрестоматия, Интерактивный тематический веб-сайт – An Introduction to ADL and the SCORM (Введение в ADL и SCORM), Academic ADL Co-Lab, 2003. <http://projects.aadlcolab.org/scourse/latestgreatest/viewer.htm>

Это интерактивный тьюториал по стандарту SCORM, который представлен на компакт-диске с курсом.

Назначение. Получить представление о том, как работают системы управления знаниями (LMS) и как именно стандарт SCORM регулирует обмен содержательными ресурсами между подобными системами.

Материал для чтения 4. CD-ROM «Электронные библиотеки в образовании», Хрестоматия, перевод статьи Эрика Дюваля, Уэйна Ходгинса «Основные направления исследований в области метаданных учебных объектов (LOM)»

Duval, E. and Hodgins, W. (2003) A LOM Research Agenda, Proceedings of the twelfth international conference on World Wide Web, 2003, edited by G. Hencsey, B. White, Y. Chen, L. Kovacs and S. Lawrence, pp. 1–9. <http://www2003.org/cdrom/papers/alternate/P659/p659-duval.html>

Назначение. Обсудить будущее стандартов образовательных метаданных.

Глава 3.1. Разметка

Эта глава объясняет, что такое метаданные и как они влияют на организационную инфраструктуру электронной библиотеки. Метаданные бывают двух типов: встроенные метаданные, или разметка, которые предназначены для оказания помощи читателям в навигации в пределах отдельно взятых документов, и внешние метаданные, которые представляют собой описательную информацию о документе в целом. Далее мы увидим как это различие, хоть оно и кажется на первый взгляд незначительным, приносит немалую путаницу в работу по созданию электронной библиотеки.

В обычных документах встроенные для навигации по документу метаданные отображаются типографским способом с помощью различных стилей заголовков, списков, сопроводительных надписей, сносок и т.д. Информация, требуемая для навигации по документу, становится очень запутанной и смешивается с информацией, требуемой для его представления. Работники типографий используют термин «разметка» для описания структуры отдельных документов и осуществления контроля за их внешним видом в момент представления пользователю. Разметка – это разновидность метаданных, и именно она находится в центре внимания данной главы. Множество других форматов для представления внешних метаданных является предметом изучения следующей главы.

В начале мы просим вас поразмышлять на тему организации традиционной библиотеки и о принципах организации (или отсутствии таковой) Всемирной паутины.

Задание 1

При изучении Модуля 1 мы пришли к выводу, что Всемирная паутина (World Wide Web) не является электронной библиотекой. Одной из главных причин этого стало отсутствие ее организации и неразборчивость. В обычной библиотеке используются два основных приема доступа к документам – поиск в библиотечном каталоге и поиск на книжных полках, где книги хранятся в специально организованном порядке. Доступ к веб-страницам осуществляется за счет гиперссылок с одной страницы на связанные с ней другие страницы, а также путем полнотекстового поиска с использованием поисковых систем.

Выделите черты сходства и различия в этих четырех механизмах доступа [Обратите внимание: длина списка не должна превышать десять пунктов]. Предположим, что у вас были необходимые полномочия и ресурсы для проектирования и внедрения библиотечного каталога в WWW. Какие характеристики документов перечисляются в вашем местном библиотечном каталоге? Какие из этих характеристик, а также из дополнительно описанных, вы бы перечислили в каталоге для Веба?

Мы ожидаем, что один из вопросов, который может у вас возникнуть при выполнении Задания 1, будет касаться степени подробности описания метаданных: применительно к отдельным веб-страницам, группам страниц или веб-сайтам в целом? Указатели, существующие в обычных библиотеках, как правило, разработаны на уровне отдельных книг или журнальных серий. Они редко описывают то, что находится внутри книги, а именно: отдельные главы или статьи. Что же является аналогом отдельной книги или журнальной серии во Всемирной паутине? На это нет четкого ответа.

В основе современной системы WWW лежит язык HTML, или язык разметки гипертекста (Hypertext Markup Language). И поскольку большинство электронных библиотек используют Веб для распространения информации — хотя это и не единственный механизм распространения — HTML чаще используется как основной формат для интерактивного просмотра документов в электронной библиотеке, причем даже тех документов, которые доступны и в других форматах. Мы могли бы обсудить HTML еще раньше, в Модуле 2, однако мы сделали это именно сейчас, поскольку HTML является сердцевинной языка разметки. Более того, он обладает еще одним важным качеством. Слово «гипертекст» относит его к средству связи между документами, которое лежит в основе организации Всемирной паутины. Это ставит перед нами непростую задачу: насколько подробно следует описывать метаданные для Веба?

Задание 2

Прочитайте начало Главы 5 и Разделов 5.1 и 5.2 (с. 221–237) учебника «Как создать электронную библиотеку» и ответьте на следующие вопросы:

- Перечислите три основных отличия между целями и философской основой HTML и XML.
- Для чего лучше всего подходит HTML?

- Как метаданные включаются в HTML?
- Как метаданные включаются в XML?
- Что такое DTD?
- Как могут быть представлены в HTML (и XML) случайные не-ASCII Unicode символы?
- Как может быть представлен в XML (и HTML) целый не-ASCII Unicode документ?
- Как из HTML файла извлечь текст?
- Какие вы знаете два вида разметки?

Дополнительное упражнение 1

Учебный модуль на CD-ROM IMARK в Уроке 2.4 представляет HTML, обобщая его черты и демонстрируя процесс создания простого HTML документа. В рамках данного учебного курса вам не нужно создавать HTML документы, хотя вы будете пользоваться подобными документами, подготовленными другими людьми, при построении коллекций собственной образовательной электронной библиотеки. Тем не менее, это полезное умение, относящееся к данной теме, и мы рекомендуем вам проработать этот Урок, для того, чтобы лучше разобраться в вопросах кодировки символов.

Разметка — это разновидность метаданных. Структурная разметка делает структуру документа достаточно подробной и очевидной, позволяя перемещаться в пределах документа. Презентационная разметка, по сути, не относится к категории метаданных, но часто тесно переплетается со структурной разметкой. Когда HTML еще только появился, оба вида разметки было сложно отделить друг от друга. Важной тенденцией за последнее десятилетие стала попытка выделить структуру из презентации и создать средства для осуществления структурной разметки, которые могли бы переводить содержание документа в различные стили визуального представления.

Задание 3

Прочитайте Раздел 5.3 (с. 237–253) учебника «Как создать электронную библиотеку» и проработайте Урок 2.5 учебного модуля на компакт-диске IMARK. Затем ответьте на следующие вопросы:

- Объясните, как CSS помогает усилить различие между структурной и презентационной разметкой?

- Каковы три компонента XSL, и какой из них ближе всего к CSS?
- Что вы будете использовать для перевода документа XML в другой формат и обратно?
- Предположим, что у вас есть книга, представленная в формате XML. Какой из языков семейства XML позволит вам обратиться к отдельной главе или части текста?
- Каково назначение пространства имен?

Упражнение, которое следует ниже, предложит вам создать коллекцию Greenstone из относительно большого количества HTML файлов. Мы выбрали несколько документов, посвященных периоду английской истории, относящемуся к правлению Тюдоров. Вы можете найти эти документы на компакт-диске с курсом «*Электронные библиотеки в образовании*» в папке *sample_files\tudor*. Вам нужно будет представить, что вы — учитель английской истории, готовящий учебный материал для своих студентов на основе документов, предварительно выгруженных из WWW.

На основе этих файлов мы сформируем основную коллекцию. В зависимости от мощности вашего компьютера, это может занять несколько минут, и если у вас мало свободного времени, то вы можете сразу открыть одну из папок, например *Monarchs*, и воспользоваться ее содержанием. Затем мы посмотрим на то, какими разными по виду изображения могут быть файлы, размещенные на панели Gather и Enrich, что очень пригодится вам при работе с файлами определенного типа.

Упражнение: HTML коллекция «Тюдоры»

1. Вызовите библиотечарский интерфейс Greenstone (Librarian Interface) (из меню Пуск Windows) и запустите новую коллекцию, которая называется **tudor** (используйте меню File). Введите необходимые данные во всплывающее диалоговое меню и оставьте **Dublin Core**, который вам предлагается по умолчанию в качестве набора метаданных.
2. На панели **Gather** откройте папку *sample_files* в коллекции *tudor*.
3. Перетащите *englishhistory.net* слева направо для включения ее в коллекцию **tudor**.
4. Переключитесь на панель **Create** и щелкните по команде **<Build Collection>**.

5. Когда процесс создания коллекции завершен, просмотрите коллекцию в режиме **preview**.
6. Заметьте, что средства просмотра в этой коллекции (*titles a–z* и *file names*) основываются целиком и полностью на метаданных. Вернитесь к библиотечарскому интерфейсу (Librarian Interface) и познакомьтесь с метаданными, которые были извлечены для некоторых файлов.

Возможно, вы заметили, что коллекция включает несколько случайных графических файлов, равно как и документы HTML. Это ошибка. Дело в том, что многие документы HTML содержат изображения и, хотя Greenstone и пытается определить, какие из них относятся к страницам HTML, а какие изображения нужно включить в коллекцию, в данном случае это было выполнено не очень удачно. (Это произошло из-за того, что веб-сайт, с которого выгружались данные файлы, случайно не был задействован в ходе обычного обновления файловой структуры.)

7. Переключитесь на панель **Design** и выберите раздел **Document Plugins**. Помимо **plugin HTMLPlug** вы увидите *smart_block*. Это программа, которая пытается распознать изображения, включенные в страницы HTML, и предотвратить их включение в коллекцию, что в нашем случае сработало очень неудачно! Выберите строку **plugin HTMLPlug** и щелкните по **<Configure Plugin>**. Откроется окно. Прокрутите страницу, пока вы не увидите команду **smart_block**, а, увидев ее, отключите ее действие. Щелкните **<OK>**.
8. Переключитесь на панель **Create** и создайте (**build**) коллекцию, а затем просмотрите ее (**preview**). Данная коллекция точно такая же, как и предыдущая, за исключением того, что в ней появление этих случайных изображений запрещено. Подключаемая программа действует как «трубопровод»: файлы один за другим передаются по очереди и обрабатываются ею до тех пор, пока она не обнаружит тот файл, который она способна обработать. По умолчанию (т.е. без *smart_block*) приложение HTML блокирует *все* изображения.

Теперь рассмотрим различные способы отображения файлов:

9. Переключитесь на панель **Gather** и откройте справа *englishhistory.net→tudor*.
10. Измените в расположенном справа меню **Show Files** команду **All Files** на **HTM & HTML**. Заметьте, что файлы, которые там изображены, были отфильтрованы так, что демонстрируются только те, которые принадлежат к данному типу.

11. Измените в этом меню команду **Show Files на Images**. И вновь эти файлы вернутся на свое место и будут помещены над всеми остальными.
12. Теперь опять вернитесь к установкам меню **Show Files** и выберите **All Files**, иначе позднее вы можете с ними запутаться. Помните, что если вам покажется, что на панелях **Gather** или **Enrich** явно фигурируют не все ваши файлы, то именно эти установки и есть причина возникшей проблемы.

Одним из поводов, почему вам, как учителю истории, пришлось бы взяться за создание подобной коллекции, могло быть желание записать ее на компакт-диск для распространения среди тех студентов, у которых затруднен доступ в Интернет. В следующем упражнении мы создадим файлы, которые необходимы для выполнения этой задачи. Когда они будут готовы, вам уже не придется использовать программные ресурсы вашего собственного компьютера для того, чтобы создать этот CD-ROM. По желанию вы можете экспортировать сразу несколько коллекций. Возможно, вы не захотите что-либо записывать на CD-ROM до тех пор, пока вы не завершите обучение на этом курсе, после чего вы сможете собрать вместе все созданные вами коллекции и записать их на диск, что будет свидетельствовать о выполненной вами работе.

Упражнение: Экспортирование на CD-ROM

13. Запустите библиотечарский интерфейс Greenstone (Librarian Interface), если он еще не запущен на вашем компьютере.
14. Выберите **File→Write CD/DVD image**, и затем в открывшемся окне выберите для экспортирования коллекцию **tudor**. Вы можете выбрать какое-нибудь свое название для CD-ROM, иначе по умолчанию будет использоваться название “collections”. Сделайте это прямо сейчас, введя “Tudor collection” в поле **CD/DVD name**; после чего щелкните **<Write CD/DVD image>**.

Нужные вам для экспортирования файлы записаны в папке:

C:\Program Files\greenstone\tmp\exported_Tudorcollection

Для записи этих файлов на компакт-диск вам нужно обратиться к имеющемуся на вашем компьютере программному обеспечению. В операционной системе Windows XP эта возможность встроена на уровне самой

системы: если у вас есть программа записи на CD-ROM или DVD, то вы просто вставляете чистый диск в дисковод и перетаскиваете содержимое из папки *exported_Tudorcollection* в папку, которая представляет диск.

В результате вы получаете самозапускающийся компакт-диск Greenstone, который начинает процесс инсталляции сразу же при помещении в дисковод. Этот процесс вы наблюдали при первичном использовании компакт-диска с материалами вашего курса, в Главе 1.3.

Документы, из которых вы составляли коллекцию Тюдоров (Tudor), были скачены из сети в то время, когда CD-ROM по данному курсу еще только готовился. Если читатели вашей библиотеки подключены к Интернету, вы можете представить им оригинальные версии документов, а не их копии, хранящиеся в электронной библиотеке. После того, как читатели щелкнут по имеющейся в документе ссылке, программа электронной библиотеки загрузит оригинальный адрес ресурса (URL) в адресной строке браузера, который приведет читателя непосредственно на нужную страницу. Так, подключенный к сети сервер электронной библиотеки, может обеспечить условия для поиска и просмотра материала в онлайн. В других ситуациях вполне достаточно разместить адреса первоисточников документов в самой коллекции (как это сделано в коллекции Тюдоров), чтобы пользователи могли прочитать их, не подключаясь к сети. Документы останутся доступными даже тогда, когда веб-сайт перестанет работать или изменится его адрес; а коллекция может быть записана на автозапускаемый компакт-диск.

Для онлайн-библиотеки, которая индексирует материал в Вебе, важно, чтобы это индексирование осуществлялось в тот момент, когда пользователь в процессе запроса нужного ему документа переходит по ссылке к исходному документу в WWW. В этом случае электронная библиотека будет действовать как поисковая система WWW. Greenstone поддерживает эту функцию, поэтому теперь мы попробуем изменить имеющуюся у нас коллекцию Тюдоров. Если у вас нет доступа к Интернету, то вы можете пропустить или поверхностно ознакомиться со следующим упражнением.

Упражнение: Коллекция веб-страниц

1. Откройте имеющуюся у вас коллекцию **tudor** и с помощью панели **Gather** проверьте файлы, которые вы в нее перетащили. Первая папка называется *englishhistory.net*, и именно в ней при открытии обнаруживается **tudor** и так далее. Файлы представляют в полном объеме все то, что находится на страницах (и сопровождающих их изображениях) и что составляет основу раздела Tudor на веб-сайте *englishhistory.net*. Они были выгружены из Веба так, что при этом была сохранена структура исходного веб-сайта. Это позволяет восстановить исходный URL любой страницы из находящихся в папке.
2. На панели **Design** выберите раздел **Document Plugins**, а затем выберите там строку **plugin HTMLPlug** и щелкните по **<Configure Plugin>**. Появится всплывающее окно. Найдите команду **file_is_url** (она расположена ниже первой группы команд) и активизируйте ее. Щелкните **<OK>**.

Установка этой команды для HTMLPlug означает то, что Greenstone задает дополнительный набор метаданных для каждого документа, названного URL, в котором дается его исходный URL.

Очень важно то, что файлы, собранные в этой коллекции, начинаются с доменного имени веб-ресурса (в данном случае *englishhistory.net*). Процесс выполнения команды прервется, если вы перетащите ее поверх папки **tudor**, потому что в этом случае метаданные URL будут выглядеть примерно так:

`http://tudor/englishhistory.net/tudor/...`,

а не так, как нужно

`http://englishhistory.net/tudor/...`

Если вы уже ранее скопировали папку **tudor**, то удалите ее и сделайте свежую копию. Перетащите папку **tudor** в правую часть панели **Gather**, где в нижнем правом углу находится мусорная корзина. Затем сделайте свежую копию файлов, начав с папки *englishhistory.net*, открыв **tudor** на левой панели и перетащив ее содержимое в противоположную сторону.

3. Для использования новых метаданных URL, мы должны изменить ссылку, прописанную для ярлыка таким образом, чтобы она вела на исходный URL, а не на его копию, размещенную в электронной библиотеке. Перейдите на панель **Design**, выберите раздел **Format Features** и отредактируйте содержание формата **Vlist**, заменив

`[link][icon][/link]`

на

`[weblink][webicon][/weblink]`

Щелкните **<Replace Format>** для сохранения внесенных изменений.

4. Перейдите на панель **Create** и, используя команды **build** и **preview**, создайте и просмотрите созданную вами коллекцию. Коллекция ведет себя без изменений, за исключением того, что когда вы щелкаете по ярлыку документа, ваш веб-браузер отображает исходный документ, размещенный непосредственно в Вебе (принимая во внимание, что в то время, когда вы выполняете это упражнение, документа там еще существует!). Если вы не работаете в сети, вы не сможете получить и просмотреть этот документ.

В этом упражнении вы обеспечили пользователей сетевой версией каждого документа. В качестве альтернативы, в зависимости от ожидаемого способа использования, имеет смысл задать ссылки как на локальную копию, так и сетевую версию документа: это можно сделать, используя более сложный оператор форматирования. Также, если онлайн-ресурс довольно часто изменяется, веб-сайт должен регулярно обновляться и повторно сканироваться, а коллекция переделываться, иначе проиндексированная информация устареет.

Та же самая технология может быть применима и к целому множеству веб-сайтов. Это похоже на следующий сценарий: учитель может создать коллекцию полезных ресурсов по теме «Англия времен Тюдоров», обнаружив относящиеся к теме веб-сайты (и части сайтов), выгрузив их и создав электронную библиотечную коллекцию. Напоследок давайте вернемся к коллекции **tudor**, но на этот раз повторно выгрузим ее из Всемирной паутины вместо того, чтобы использовать заранее подготовленную версию, которая поставляется на компакт-диске с данным курсом. Обратите внимание: для завершения этого упражнения вам нужно было установить Greenstone с компакт-диска «Электронные библиотеки в образовании». Компакт-диск «Оцифровка и электронные библиотеки» с учебным модулем IMARK, названный содержит старую версию Greenstone, в которой отсутствует опция выгрузки файлов из Интернета.

Упражнение: Скачивание документов из Веба

Панель **Download** интерфейса библиотекаря Greenstone позволяет вам скачивать из Веба как отдельные файлы, так и части веб-сайтов или даже веб-сайты целиком.

1. Запустите новую коллекцию **webtudor**, в основу которой будет положена коллекция **tudor**.

2. В веб-браузере загрузите <http://englishhistory.net>, пойдите по ссылке *Tudor England* и щелкните **<Enter>**. Вы окажетесь на странице с адресом

<http://englishhistory.net/tudor/contents.html>

Именно отсюда мы скачивали файлы, которые использовали для наполнения коллекции **tudor**.

3. Вы можете сделать то же самое, скопировав адрес этой страницы в веб-браузере, перенеся его на панель **Download** и щелкнув по кнопке **<Download>**. Поскольку выгружаемые файлы могут быть объемом в несколько мегабайт, потребуются дополнительные ресурсы памяти вашего компьютера, равно как и ваше терпение! Для того чтобы ускорить ход событий при выполнении данного упражнения, мы сосредоточим внимание на небольшом разделе данного сайта. Находясь на панели **Download** в окне **Source URL** введите следующий адрес

<http://englishhistory.net/tudor/monarchs/edward6.html>

Существует несколько вариантов управления процессом скачивания файлов. Для того чтобы скопировать раздел сайта *monarchs*, выберите команду **Only mirror files below this URL**. Если вы этого не сделаете, в процессе выгрузки будут также задействованы ссылки на другие разделы сайта *englishhistory.net*.

4. Теперь щелкните **<Download>**. В нижней части панели появится окно, отражающее ход выполнения скачивания файлов.

Более детальную информацию можно получить, щелкнув по **<View Log>**. Процесс скачивания можно остановить и при необходимости возобновить, а также прервать, щелкнув по **<Close>**. Скачивание может оказаться довольно длительным процессом, поскольку в него вовлечено множество сайтов, и поэтому Greenstone устанавливает очередность выгрузки файлов. Когда в окно с адресом ресурса (Source URL) помещается следующий адрес, и вы щелкаете **<Download>**, открывается новое окно статуса скачивания, которое добавляется к тем аналогичным окнам, которые уже выведены на панель. Как только одно из окон покажет, что процесс скачивания завершен, автоматически открывается следующее окно, и начинается скачивание новой страницы.

5. Скачиваемые файлы размещаются в папке **Downloaded Files**, расположенной сверху, слева от панели **Gather**. Возможно, вам не понадобятся все выгруженные файлы, а те, которые нужны, вы выберете, перетаскивая через весь экран отобранные файлы из этой папки

в правую сторону экрана вашей коллекции, точно так же, как мы это делали ранее, когда отбирали данные из папки *sample_files*. В этом примере мы включим в коллекцию абсолютно все, что было выгружено.

Выберите папку **englishhistory.net** в папке скаченных файлов (**Downloaded Files**) и перетащите ее через экран в область показа коллекции.

6. Переключитесь на панель **Create** для того, чтобы создать (**build**) и просмотреть (**preview**) коллекцию. Она получится меньше по объему, чем предыдущая коллекция, поскольку мы включили в нее только те файлы, которые относятся к *monarchs*. Именно эти файлы представляют самые последние версии этих документов. Поскольку в основу вашей коллекции **webtudor** была положена коллекция **tudor**, то она включает измененный формат `[weblink][webicon][weblink]`, что приводит к тому, что в новой коллекции теперь присутствуют ссылки на исходные веб-документы.

За долгие годы своего существования HTML накопил много расширений и дополнений, и сейчас он представляет собой довольно запутанный формат, с которым непросто справиться. HTML-приложение Greenstone читает файлы HTML и анализирует их содержание. Это процесс не всегда протекает гладко. Если веб-страницы используют определенные интерактивные возможности, то документы в этой коллекции не обязательно будут точно соответствовать тому, что было в оригинале. В наиболее сложных случаях документ может появиться в виде пустой страницы или вообще не появиться. В других случаях внешний вид страницы может претерпеть значительные изменения: может быть потерян фоновый рисунок или какая-нибудь интересная выполняемая функция.

Вот те проблемы, о которых все уже знают.

Фреймы. Документы, которые используют фреймы (frames) HTML, разделяются на отдельные компоненты, которые воспринимаются системой как отдельные, не зависящие друг от друга, документы. HTMLPlug не может обработать веб-страницы, которые включают фреймы.

Комментарии. Приложения HTML иногда используют комментарии `<!-- ... -->` для сообщения дополнительной информации, используемой при обработке документа. HTMLPlug игнорирует комментарии.

Тэг <head>. HTMLPlug сбрасывает информацию в тэге заголовка документа (<head>), что может привести к потере файлов стиля или функций JavaScript. (Переключатель *keep_head* удерживает эту информацию от сброса, но его нельзя делать установкой по умолчанию, поскольку это может привести к появлению других проблем.)

JavaScript, язык программирования, который может быть встроен в HTML для управления содержанием и структурой документа, например, он может загрузить новое изображение во время работы с программой. HTMLPlug его игнорирует, поскольку на этапе создания коллекции он должен установить все внешние ресурсы (изображения и т.д.), которые нужны для воссоздания документа в целом.

Встроенные технологии. Страницы HTML могут включать еще более расширенный набор дополнительных веб-технологий. В приложении HTMLPlug есть специальные функции, которые помогают с ними справиться, но обращение к ним требует специальных знаний.

Основным общим правилом является то, что установки по умолчанию в HTMLPlug хорошо работают только тогда, когда страницы, включаемые в коллекцию, совпадают по форме с базовым синтаксисом HTML, то есть, когда они не используют JavaScript, встроенные веб-технологии, действия с DOM или не имеют устаревшей фрейм-овой структуры. В Вебе есть множество примеров довольно привлекательных страниц, которые отвечают этим требованиям, а подход «чем проще, тем лучше» также гарантирует вам то, что ваши веб-страницы будут показывать именно то, что вы хотели, причем с помощью разных веб-браузеров и в различных операционных системах. Использование новых оригинальных инструментов только снижает шансы на то, что страница будет представлена корректно, а иногда это будет приводить к сбоям в работе браузера, имеющегося у пользователя.

Заключение

Рассматриваемый модуль рассказал вам о том, как работать с метаданными. Метаданные — это своеобразный клей, который скрепляет в единое целое все документы электронной библиотеки. С точки зрения читателя, это именно то, что и придает библиотеке ценность единой коллекции, в отличие от набора отдельных, не связанных друг с другом, документов. Глава 3.1 была посвящена языку разметки. Исторически разметка развивалась как один из процессов подготовки книги к изданию, и, как это задумывалось изначально, она в большей степени предназначалась для наглядного представления результатов работы, нежели для представления метаданных. Именно этим объясняется включение данной Главы в Модуль 2, который посвящен представлению документов. И на самом деле, HTML является языком представления документов.

Три фактора осложняют изучение этого вопроса. Один из них связан с определением понятия индивидуальной «работы». Это более-менее ясно в случае с печатным изданием — книгой (хотя даже здесь возникает много вопросов с разными изданиями, версиями, переводами, аудио записями для слепых или фильмами, в основу которых положен сюжет книги, то есть, что является составной частью одной и той же «работы», а что — нет). Понятие «работа» еще более туманно в ситуации с гипертекстом, однако, веб-страницы не могут рассматриваться в качестве индивидуальных объектов, существующих сами по себе, поскольку они являются лишь частью массивной, связанной бесчисленными нитями структуры, которую мы называем Вебом. Второй осложняющий фактор — это тенденция к разделению формы презентации и ее структуры, а также кодирование каждой из них по отдельности. Это было проиллюстрировано на примере перехода от исходного и неструктурированного, ориентированного на представление данных, языка HTML к тщательно структурированному XML, делегирующему все презентационные решения отдельным таблицам стилей CSS. Третьим фактором является то, что изначально HTML отдельно кодировал метаданные *Title*, а также давал возможность размещать прочие метаданные в теге `<meta>`. Более того, даже без явного выставления тегов, некоторые метаданные с большей или меньшей точностью можно извлечь непосредственно из текста документа. Мы узнаем о том, как это делается, в следующей главе (Задание 3).

Наряду с тем, что вы узнали о HTML, XML и подобных стандартах, эта глава продемонстрировал вам процесс создания коллекции HTML документов с помощью Greenstone. Вы узнали о том, как поместить коллекцию Greenstone на компакт-диск и как использовать Greenstone для загрузки файлов из Веба с последующим встраиванием их в существующие коллекции.

В мировой практике Greenstone очень часто используется именно для создания и организации коллекций документов в HTML. Как вы уже видели, библиотечарский интерфейс (Librarian Interface) имеет панель *Download*, которая позволяет загружать отдельные веб-страницы, части веб-сайтов или даже целые веб-сайты в специальную область памяти на вашем компьютере таким образом, чтобы вы могли в дальнейшем использовать их для создания своих коллекций. Это дает вам в руки такой мощный инструмент, что мы чувствуем себя обязанными напомнить вам об одном из уроков Главы 1.1, в котором мы говорили о том, что создание и распространение информационных коллекций налагает на вас определенные обязательства, которые вы должны узнать прежде, чем вы начнете этим заниматься. Существуют законы, регулирующие авторское право: получение доступа к документу не означает то, что вы можете передавать его другим людям. Существуют и социальные ограничения: коллекции должны уважать традиции сообществ, создавших эти документы. Есть также этические нормы и правила в отношениях между людьми. Перо гораздо сильнее по своему действию, чем меч! — не забывайте про силу информации и мудро ею пользуйтесь.

Глава 3.2. Метаданные

В этой главе вы продолжите изучение метаданных. Мы сосредоточим наше внимание на внешних метаданных, являющихся типичными для содержания любого библиотечного каталога. То, что описывается с помощью метаданных, может быть сохранено вне документа. Однако, многие системы разметки документов, а XML здесь является хорошим примером, часто допускают возможность встраивания «внешних» метаданных в документ. То же самое применимо и для обычных, не электронных, документов: название, имя автора, дата публикации и информация об издателе книги — обычно размещаются непосредственно в самой книге. Внешние метаданные — это на самом деле те метаданные, которые могут быть отделены, вынесены из документа, и, тем не менее, они будут выполнять свои функции и нести заложенную в них информацию. В отличие от них, встроенные метаданные вне самого документа существовать не могут.

Задание 1

Прочитайте Раздел 2.2 Главы 2, (с. 46–55) учебника «Как создать электронную библиотеку» и проработайте Урок 3.1 учебного модуля на компакт-диске IMARK. Затем выполните следующие задания:

- В учебнике говорится, что библиографическая система поддерживает пять различных видов деятельности пользователя: обнаружение местонахождения ресурса, определение, отбор, запрос и управление информацией. Подумайте и обсудите эти виды деятельности в контексте работы с коллекцией Tudor, которую вы создали, работая над Главой 3.1.
- Используя стандартную систему классификации (ту, что используется в вашей библиотеке или в предметном указателе Библиотеки Конгресса США (Library of Congress Subject Headings), дайте определения некоторым терминам, применяемым к документам, находящимся в коллекции Tudor.

Задание 2

Прочитайте Раздел 5.4 Главы 5 (с. 253–261) учебника «Как создать электронную библиотеку» и проработайте Уроки 3.2 и 3.3 учебного модуля на CD-ROM IMARK.

В программе Greenstone есть файл с названием *marctodc.txt*, который управляет процессом преобразования метаданных формата MARC в Dublin Core. Найти этот файл можно следующим образом: если вы установили программу в соответствии с принятыми стандартами, то вы найдете этот файл на *C:\Program Files\greenstone\etc\marctodc.txt*. Изучите его и попробуйте понять, как типовая запись из каталога MARC, например такая, как показана в Таблице 5.2 (с. 255) учебника, переводится в формат Dublin Core.

В следующем упражнении вам предлагается создать коллекцию изображений. Коллекции изображений строятся на основе внешних метаданных, использующихся для выполнения всех действий по поиску и просмотру, в то время как в коллекциях текстовых документов заголовки обычно извлекаются из самих документов, и поиск этих документов может проводиться даже в тех случаях, когда метаданные отсутствуют. В этой коллекции вы сами определите метаданные и добавите средства поиска и просмотра, которые будут действовать на их основе. Вы можете использовать свои собственные рисунки или те, что размещены в папке *sample_files\images* на компакт-диске с курсом.

Обратите внимание: вы не сможете приступить к созданию коллекции образов до тех пор, пока на вашем компьютере не будет установлен ImageMagick, как это было описано в Задании 1 Главы 2.2.

Упражнение: Коллекция изображений

1. Создайте новую коллекцию (File→New) и назовите ее **backdrop**. Заполните все поля, введя необходимую информацию. В поле **Base this collection on** выберите в меню вариант **Simple image collection (image-e)**.

Вам не предлагается определять используемый набор метаданных, поскольку новая коллекция автоматически наследует все те установки, которые были определены для коллекции, взятой за основу.

2. Скопируйте изображения, которые находятся в папке *images*, в только что созданную вами новую коллекцию.
3. Перейдите на панель **Create** и создайте (**build**) коллекцию.

4. Просмотрите (**Preview**) полученный результат.
5. Щелкните **<browse>** в окне навигации и просмотрите список фотографий, отсортированных по именам файлов и представленных в виде эскизов, которые сопровождаются некоторой базовой информацией. Структура этой коллекции точно такая же, как в **Simple image collection (image-e)**, но содержание другое.
6. Переключитесь на панель **Enrich** и просмотрите извлеченные метаданные для *Ascent.jpg*.

А теперь мы добавим наши собственные метаданные и используем их для задания нового режима просмотра коллекции. Мы используем набор метаданных Dublin Core.

7. Коллекция (*image-e*), на которой базируется **backdrop**, использует исключительно извлеченные метаданные. Для добавления еще одного набора метаданных перейдите на панель **Design** в библиотекарском интерфейсе (Librarian Interface) и щелкните по **<Metadata Sets>** в списке, расположенном слева (последний пункт списка). Затем щелкните **<Add Metadata Set>** (нижняя левая кнопка).
8. В открывшемся окне выберите **dublin.mds** и щелкните **<Add Metadata Set>**.
9. Теперь переключитесь на панель **Enrich**, щелкнув по ней. Метаданные по каждому файлу теперь показывают поля Dublin core *dc*. Так же, как и поля *extracted ex*.
10. Мы имеем дело только с первыми тремя файлами (*Ascent.jpg*, *Autumn.jpg* и *Azul.jpg*), чтобы лучше почувствовать, как это все работает на самом деле. Для начала напишите в поле **dc.Title** каждого из файлов названия файлов (точно такие же, но без расширений файлов).
11. Чтобы поля метаданных стали вам доступны, щелкните по **Ascent.jpg**, а затем щелкните по полю **dc.Title**, находящемуся по правую руку от вас. Щелкните по текстовому окну **Value**, введите **Ascent** и щелкните по **<Append>**.

Окно **All Previous Values** будет более пригодным для использования по мере добавления в него новых данных.

12. Повторите те же самые действия для **Autumn.jpg** и **Azul.jpg**.

Теперь мы настроили внешний вид коллекции. На этом этапе процесса создания и предварительного просмотра коллекции мы не увидели ничего нового. Это объясняется тем, что мы не настолько сильно изменили дизайн коллекции, чтобы увидеть все преимущества новых метаданных.

13. Перейдите на панель **Design** (щелкнув по ней) и выберите **Format Features** из списка, расположенного слева. Оставьте **Editing Controls** с их установками по умолчанию так, чтобы поле **Choose Feature** осталось неактивным, а **Vlist** выберите как **Affected Component**. В строке **HTML Format String** отредактируйте текст в соответствии с приведенным ниже примером:

Измените “_ImageName_:” на “Title:”

Измените “[Image]” на “[dc.Title]”

Имена метаданных в Greenstone чувствительны к регистру символов, поэтому очень важно напечатать слово “Title” с заглавной буквы (и не сделать заглавными “dc”).

14. Затем щелкните по команде **<Replace Format>**. Первый из отредактированных вами текстов изменит текстовый фрагмент, который появляется справа от эскизного изображения, а второй изменит метаданные, за ним следующие.
15. Перейдите на панель **Create** и щелкните по **<Build Collection>**. Теперь просмотрите (**preview**) эту коллекцию. Когда вы щелкаете по команде **browse** на навигационной панели, то вид представления документа в ней меняется на “Title: Ascent” и т.д.

Из-за того, что мы определили метаданные только для первых трех объектов, место, отведенное для заголовка, останется пустым. Это объясняется тем, что в этих объектах нет метаданных dc.Title. Чтобы получить все, что нужно, введите все метаданные целиком.

При использовании некоторых параметров дизайна вам нужно каждый раз заново создавать коллекцию, иначе эффект не будет виден. Однако, изменения, производимые операторами форматирования, осуществляются незамедлительно, и вы сразу же можете увидеть результат произведенных изменений, щелкнув по кнопке **Reload** веб-браузера. Ранее мы предлагали вам сначала создать коллекцию, а затем ее просмотреть, что мы делали исключительно в целях упрощения объяснения.

Теперь мы изменим размер эскиза, уменьшив его.

16. Изображения эскизов создаются встроенным приложением *ImagePlug*, поэтому нам нужно обратиться к его исходным параметрам. Для этого переключитесь на панель **Design** и выберите **Document Plugins** из списка, расположенного слева. Двойной щелчок по **plugin ImagePlug** откроет окно с параметрами. (Также вы можете выбрать одним щелчком **ImagePlug**, а затем щелкнуть по **<Configure Plugin>** в нижней части экрана).

Сейчас все команды отключены, поэтому используются только те параметры, которые заданы по умолчанию. Выберите **thumbnailsize**, установите значение параметров на **50**, а затем щелкните **<OK>**.

17. Создайте (**Build**) и просмотрите (**preview**) коллекцию.
18. Как только вы увидите результат произведенных изменений, вернитесь на панель **Design**, выберите параметры конфигурации для *ImagePlug*, а затем отключите установку размера эскиза с тем, чтобы его размер вернулся к исходному, т.е. его нормальному размеру, каким он был при повторном создании коллекции.

Теперь добавьте метаданные, которые опишут фотографии, включенные в данную коллекцию. И вновь для иллюстрации мы сосредоточим наше внимание на первых трех изображениях (*Ascent.jpg*, *Autumn.jpg* and *Azul.jpg*).

19. Переключитесь на панель **Enrich** и выберите *Ascent.jpg*. Мы поместим наше описание в элемент метаданных **dc.Description**, для чего его нужно выбрать на панели, расположенной справа.

Какое описание нам нужно ввести? Для того чтобы вспомнить содержание того или иного файла, вам надо их открыть, сделав по ним двойной щелчок, что предусмотрено библиотечарским интерфейсом. Двойной щелчок запускает соответствующее приложение, ориентируясь на то, что написано в расширении файла Word для файлов с расширением *.doc, Acrobat для файлов с расширением *.pdf и так далее. Сделайте двойной щелчок по файлу *Ascent.jpg*: изображение будет загружено и показано вам в редакторе Microsoft Photo Editor (хотя это зависит от установок вашего компьютера).

20. Вернувшись назад в библиотечарский интерфейс, введите текст **Moon rising over mountain landscape** как значение в поле **dc.Description** и щелкните **<Append>** для его сохранения.

21. Повторите это действие для *Autumn.jpg* и *Azul.jpg*.

22. Заново создайте коллекцию и внедрите в нее новые метаданные.

23. Теперь обновите оператор форматирования для использования новых метаданных **dc.Description**. Перейдите на панель **Design** и войдите в раздел **Format Features**, выбрав его из списка названий в левой части экрана и убедившись в том, что **VList** активизирован. Поместите курсор мыши в строке **HTML Format String** сразу же после текста

```
[dc.Title]<br>
```

и добавьте следующее:

```
Description: [dc.Description]<br>
```

Затем щелкните **<Replace Format>**.

24. Просмотрите (**Preview**) результат. Вам не нужно создавать коллекцию как вы это делали, выполняя действия в п. 22 для включения метаданных, поскольку изменения, произведенные операторами форматирования, начинают действовать немедленно. Каждый заголовок будет теперь сопровождаться не только эскизом, но и описанием каждого из изображений.

Теперь мы добавим еще один новый параметр просмотра, в основе которого будут лежать эти описания.

25. Переключитесь на панель **Design** и выберите **Browsing Classifiers** в левом списке. Выберите **AZList** в строке меню **Select classifier to add**, а затем щелкните **<Add Classifier>**.

26. Появится окно, с помощью которого можно управлять параметрами классификатора. Выберите в меню метаданных строку **dc.Description** и щелкните **<OK>**. Теперь переключитесь на панель **Create**, создайте (**build**) коллекцию и просмотрите (**preview**) результат. Выберите ссылки на новые описания (**descriptions**), которые появятся в окне навигации.

В списке показываются только три объекта, поскольку только они совпадают с заданными метаданными (в этом случае с **dc.Description**). Исходный список для просмотра включал все фотографии, имеющиеся в коллекции, так как он основывался на извлеченных метаданных **ex.Image**, включающих имя файла, которое задается для всех изображений из данной коллекции.

И, наконец, мы создаем поисковый указатель, в основу которого также положены метаданные **dc.Description**.

27. Перейдите на панель **Design** и выберите **Search Indexes** в левом списке. Введите текст “**descriptions**” для названия указателя **Index Name**, выберите **dc.Description** и щелкните по **<Add Index>**.

28. Перейдите на панель **Create**, создайте (**build**) коллекцию, а затем просмотрите (**preview**) ее. Для примера проведите поиск по слову “**autumn**” (осень).

Greenstone позволяет добавлять метаданные практически к любой коллекции. И эти метаданные могут использоваться по-разному — в качестве основы для полнотекстовых указателей, для инструментов просмотра или демонстрации. Greenstone не ограничивает выбор дизайна и стиля создаваемых коллекций. Она лишь задает тот каркас, структуру, которую библиотекари используют для быстрого формирования стиля коллекции, которая отвечает их нуждам с учетом особенностей тех исходных документов, с которыми они работают. Plusom

является то, что все это легко меняется, но есть и минус, который заключается в том, что для эффективного проведения таких изменений нужно сначала научиться это делать. Существует много способов настройки стиля дизайна коллекции, и именно этим объясняется включение такого большого количества упражнений в данном курсе.

В следующем упражнении мы вернемся к первичной коллекции Tudor, которую мы создали при выполнении первого упражнения в Главе 3.1, и усовершенствуем ее тремя разными способами. Вначале мы выделим небольшое количество метаданных, которые позволят нам разделить документы по четырем категориям, и добавим специальный инструмент, который позволит читателям просматривать документы по этим категориям. Затем мы поделим полнотекстовый указатель на четыре соответствующие части. В данном упражнении мы сделаем это на основе значений метаданных, которые мы только что определили, хотя деление на категории может осуществляться на основе любых других значений метаданных или на основе имен файлов, представленных в коллекции документов. Затем мы добавим иерархический указатель фраз. И, наконец, мы посмотрим, как можно контролировать процесс создания коллекции, что очень важно при работе с коллекциями, включающими большое количество документов.

Как вы узнали из последнего задания Главы 2.2, библиотечарский интерфейс обеспечивает для пользователей четыре уровня доступа. Пока мы действовали в режиме **библиотекаря** (Librarian mode), но для выполнения этого упражнения мы должны переключиться на режим **специалиста библиотечных систем** (Library Systems Specialist) или на более высокий уровень. Причиной этого является необходимость определить «обычные выражения». Обычные выражения — это мощный инструмент подбора текста, непростой и требующий времени на освоение, поэтому мы не будем давать более подробные объяснения, т.к. для выполнения задания они не понадобятся. И именно это объясняет, почему этот инструмент недоступен в режиме библиотекаря.

Упражнение: Усовершенствованная коллекция Tudor

Мы возвращаемся к коллекции Tudor, чтобы добавить метаданные, отражающие предметную иерархию. После этого мы создадим на их основе классификатор, дающий возможность читателям просматривать документы по отдельным темам, например, только по монархам (Monarchs), родственникам (Relatives), жителям страны (Citizens) и др.

1. Откройте свою коллекцию **tudor** (исходную версию, а не ту, что называется **webtudor**), перейдите на панель **Enrich** и выберите папку *monarchs* (являющуюся вложенной папкой в *tudor*). Установите в метаданных **dc.Subject and Keywords** значение **Tudor period|Monarchs**. (Для краткости в дальнейшем мы будем называть этот элемент метаданных просто как **dc.Subject**.) Вертикальная полоска (“|”) — это маркер иерархии. Выбор папки и использование кнопки **Append** для установки метаданных действуют таким образом, что значения метаданных годятся для всех файлов, которые находятся в этой папке, во всех вложенных в нее папках и т.д. Вы увидите на экране специальное сообщение, предупреждающее вас об этом.
2. Повторите то же самое для папок *relative* и *citizens*, установив в метаданных **dc.Subject** значения, соответственно, **Tudor period|Relatives** и **Tudor period|Citizens**. Обратите внимание, что иерархический список появится в области **All Previous Values**.
3. И, наконец, выберите все оставшиеся файлы — те, которые не попали в папки *monarchs*, *relative* или *citizens* — выделив первый из файлов, а далее, удерживая клавишу Shift, выделите все остальные. (Это обычный прием одновременного выделения нескольких объектов.) Задайте установки в метаданных **dc.Subject** как **Tudor period|Others**: это делается в одно действие, но выполняется оно с небольшой задержкой.
4. Перейдите на панель **Design** и выберите в левом списке **Browsing Classifiers**. Установите классификатор **Hierarchy** в меню **Select classifier to add**, а затем щелкните **<Add Classifier>**.
5. Появится окно, с помощью которого можно управлять установками классификатора. Измените метаданные на *dc.Subject*, а затем щелкните **<OK>**.
6. Для того, чтобы все было в полном порядке, вам нужно убрать (**remove**) классификатор (**classifier**) для метаданных **Source** (включенный по умолчанию) из списка заданных и действующих в настоящее время классификаторов, поскольку он мало на что влияет при работе с данной коллекцией.
7. Теперь перейдите к панели **Create**, создайте (**build**) коллекцию и просмотрите (**preview**) ее. Выберите новые предметные (**subjects**) ссылки, которые появляются на панели навигации и щелкните по книжным полкам для того, чтобы пройтись по четырем созданным вами структурам.

Далее мы разделим полнотекстовый указатель на четыре отдельные части. Для того, чтобы это сделать, мы сначала выделим четыре вложенные коллекции, полученные путем «фильтрации» документов по определенным критериям, заложенным в их метаданные *dc.Subject*. Далее мы назначим указатель для каждой из вложенных коллекций.

8. Перейдите на панель **Design** и щелкните по **<Partition Indexes>**. Эта функция отключена, поскольку вы работаете в режиме библиотекаря (*Librarian Mode*), что обозначено в окне заголовка в верхней части окна.
9. Перейдите теперь в режим специалиста библиотечных систем (*Library Systems Specialist*). Вы можете сделать это на вкладке **Preferences** (в меню *File*), щелкнув по **<Mode>**. Прочитайте также названия других режимов работы в программе. Обратите внимание, что название выбранного вами режима работы появляется в окне заголовка.
10. Вернитесь в раздел **Partition Indexes** на панели **Design**. Убедитесь в том, что по умолчанию выбрана функция **Define Filters**. Определите фильтр для вложенной коллекции с названием *monarchs*, который будет отличаться от **dc.Subject and Keywords**, и напечатайте **Monarchs** как обычное выражение, соответствующее этой группе. Щелкните по **<Add Filter>**. Этот фильтр выявит любые файлы, в **dc.Subject** метаданных которых присутствует слово *monarchs*.
11. Определите и другой фильтр, **relatives**, по которому слово **Relatives** будет выделяться в **dc.Subject**. Определите третий и четвертый, **citizens** и **others**, которые, соответственно, выделяют их из **Citizens** и **Others**.
12. Определившись с вложенными коллекциями, мы разделим указатель на соответствующие части. Щелкните по кнопке **<Assign Partitions>**. Выберите первую вложенную коллекцию и назовите ее **monarchs**, затем щелкните **<Add Partition>**. Повторите это действие и для других трех вложенных коллекций, называя их части **relatives**, **citizens** и **others**. Создайте (**build**) и просмотрите (**preview**) коллекцию.
13. На поисковой странице находится открывающееся меню, которое позволяет вам выбрать одну из этих частей для проведения поиска. Например, попробуйте провести поиск в части *relatives* по ключевому слову *mary*, а затем поискать в той же самой части документы по слову *monarchs*.
14. Вернитесь в режим библиотекаря (*Librarian mode*), используя **Preferences** (в меню **File**).

Теперь мы добавим в коллекцию интерактивный иерархический указатель фраз, который в программе Greenstone получил название PHIND.

15. Перейдите на панель **Design** и в левом списке выберите пункт **Browsing Classifiers**.
16. Выберите **Phind** в меню **Select classifier to add**. Щелкните **<Add Classifier>**. Появится окно, в котором вам будет предложено выбрать конфигурацию для этого классификатора: оставьте все, как есть, воспользовавшись установками по умолчанию (указатель фраз при этом будет работать с полнотекстными документами), и щелкните **<OK>**.
17. Еще раз создайте (**build**) коллекцию, просмотрите (**preview**) ее и проверьте, как работает новая функция **phrases**, появившаяся на панели навигации. Вы получите интересные результаты, если попытаете поискать с помощью PHIND документы по слову **king**.

И, наконец, мы посмотрим, как можно проконтролировать процесс создания библиотеки. Разработка новой коллекции обычно происходит путем циклического повторения таких этапов работы, как создание, предварительный просмотр и добавление новых функций для придания коллекции улучшенного вида. Вам будет проще работать, если на этапе проектирования коллекции количество документов, содержащихся в ней, будет уменьшено. Это можно сделать, используя в процессе создания коллекции параметр “maxdocs”.

18. Перейдите на панель **Create** и посмотрите, какие из команд, показанных в верхней части экрана, вам доступны. Выберите maxdocs и установите его счетчик цифр на **3**. Теперь приступайте к созданию (**build**) коллекции. Вы обнаружите, что теперь коллекция состоит из пяти документов (а не трех, как вы это определили: по техническим причинам то значение цифрового показателя, которое вы ввели в **maxdocs**, является лишь примерным значением).
19. Просмотрите страницу **titles a–z** в коллекции, которую вы только что перестроили. Ранее этот список растягивался на десятки страниц по каждой букве алфавита, но сейчас в нем присутствуют всего три документа — первые три файла, с которыми вы встречались в процессе создания коллекции.

Упражнение, приведенное выше, познакомило вас с несколькими полезными характеристиками Greenstone. Коллекция Tudor самая значительная из тех, что мы создали на данный момент, но она является небольшой с точки зрения заложенных в нее возможностей. Настоящие коллекции часто содержат тысячи или сотни тысяч, а может быть, и миллионы документов. Процесс разработки коллекции можно значительно ускорить путем использования на первых этапах работы, скажем, только пятидесяти из них, что было проиллюстрировано в последней части упражнения.

Теперь мы создадим коллекцию, которая будет базироваться исключительно на метаданных: в данном случае — на файле с записями MARC. Greenstone содержит и другие приложения для метаданных, например, для BibTeX и Refer (упоминаемые в Задании 4 Материалов для чтения), а также для CDS/ISIS и PROCite, которые также являются широко распространенными библиографическими форматами. Вы можете создать коллекции из документов, представленных и в этих форматах. Коллекция записей MARC из Библиотеки Конгресса США, в которой собраны материалы о рок-группе «Битлз», находится на компакт-диске с курсом в папке *sample_files\marc*.

Упражнение: Библиографическая коллекция

1. Запустите новую коллекцию, которая называется “**Beatles**” **Bibliography**. Введите необходимую информацию а за основу возьмите “New Collection”. Вам не потребуется включать какие-либо наборы метаданных, поскольку метаданные уже извлечены из записей MARC, и, следовательно, появятся в коллекции в виде извлеченных метаданных.
2. На панели **Gather** откройте папку *marc*, перетащите **locbeatles50.marc** на правую панель и оставьте ее там. Открывшееся окно задаст вам вопрос, хотите ли вы добавить **MARCPlug** в коллекцию для того, чтобы обработать этот файл. Щелкните **<Add Plugin>**, поскольку это приложение потребуется вам для обработки записей MARC.
3. Переместите приложения **TextPlug** в **PSPlug** (**ZIPPlug**, **GAPlug** и **MARCPlug** остаются на месте). Строго говоря, в переносе этих лишних приложений никакой необходимости нет, но практика показывает, что во избежание неожиданных побочных эффектов лучше включать только те приложения, которые вам нужны.

4. Теперь на панели **Design** выберите **Browsing Classifiers** и удалите (**remove**) классификатор **Source**, по умолчанию установленный для метаданных. В этой коллекции все записи взяты из одного и того же файла, поэтому метаданные **Source**, которые настраиваются по имени файла, не особенно для нас интересны.
5. Переключитесь на панель **Create**, постройте (**build**) и просмотрите (**preview**) коллекцию. Пролистайте заголовки **titles a–z** и просмотрите одну или две записи. Попробуйте провести поиск, например, по документам, в которых есть слова **George Martin**.
6. Добавьте метаданные **Subject** к классификатору **AZCompactList**. Выберите этот пункт в соответствующем меню раздела **Browsing Classifiers** на панели **Design** и щелкните **<Add Classifier>**. В открывшемся окне выберите **ex.Subject** в качестве значения метаданных, активизируйте команду **mingroup** и установите в поле ее значение как **1**.

Список **AZCompactList** похож на **AZList**, за исключением того, что термины, которые неоднократно появляются в иерархии, автоматически объединяются как одна новая группа, показываемая отдельной пиктограммой на книжной полке. Установка **1** для минигруппы (**mingroup**) означает, что книжная полка появляется даже тогда, когда на ней есть всего один объект, тем самым, делая более общим процесс демонстрации объектов коллекции.

7. Создайте (**build**) коллекцию и просмотрите (**preview**) полученный результат.
8. Сделайте так, чтобы для каждой группы объектов, находящихся на книжной полке, было показано, сколько именно объектов в ней находится. Для этого проведите необходимые изменения для оператора форматирования **VList** на панели **Design** в **Format Features: {If}{[numleafdocs],<td><i>([numleafdocs])</i></td>}**
9. Щелкните по **<Replace Format>**, переключитесь на панель **Create**, а затем щелкните по **<Preview Коллекция>** (здесь вам не нужно заново создавать коллекцию.)

А теперь проведите поиск по добавленным полям.

10. Находясь на панели **Design**, выберите команду **Search Types** из списка, расположенного слева, и активизируйте команду **Enable Advanced Searches**.

11. Еще раз постройте (**build**) коллекцию и просмотрите (**preview**) результаты. Обратите внимание, на то, что домашняя страница коллекции больше не показывает окно со списком запросов. Это происходит из-за того, что поисковая форма слишком велика, чтобы уместиться здесь полностью. Для проведения поиска вам нужно щелкнуть по **search** в окне навигации. Заметьте, что страница *Preferences* изменилась, и теперь, используя ее установки, можно контролировать процесс расширенного поиска.

Чтобы закончить с этой коллекцией, пометьте ее каким-нибудь изображением, которое будет использоваться для представления этой коллекции на странице Greenstone, а также появляться в верхней части каждой из страниц коллекции:

12. Находясь в общем разделе (**General**) панели **Design**, щелкните по кнопке **<Browse>**, расположенной рядом с иконкой **URL to 'about page'**, и, используя появившийся вслед за этим браузер файлов, найдите папку `sample_files\marc`, расположенную на компакт-диске с курсором. Выберите `beatleslogo.jpg` и щелкните по **<Open>**.

Greenstone скопирует изображения в область, где расположена ваша коллекция, так что коллекция будет работать даже тогда, когда вы уберете CD-ROM из дисковода.

13. Повторите эти действия для **URL to 'home page' icon**, выбрав то же самое изображение.

14. Теперь создайте (**build**) коллекцию и просмотрите (**preview**) ее.

Задание 3

Прочитайте Раздел 5.6. Главы 5 (с. 266–280) учебника «Как создать электронную библиотеку». Greenstone содержит инструменты для распознавания языка, расшифровки акронимов, извлечения ключевых фраз и создания иерархий фраз. Вы встретитесь с браузером иерархий фраз в расширенном варианте коллекции Tudor, а также, когда вы будете изучать записанный на компакт-диске с курсором «Электронные библиотеки в образовании» учебник «Как создать электронную библиотеку» в Задании 5 Главы 1.3. Выберите один из следующих проектов.

- Текст документа появляется как извлеченные метаданные и может просматриваться через панель Enrich. Исследуйте существующие ограничения по использованию данного инструмента путем создания нескольких многоязычных документов и импортирования их в существующую коллекцию.

- Расшифровка акронимов включается на панели конфигурации отдельно для каждой исполняемой программы. Найдите какие-либо документы, которые содержат акронимы и их определения, и познакомьтесь со схемой извлечения данных.
- Извлечение ключевых фраз запускается точно так же. Проанализируйте качество ключевых фраз из выбранных вами документов. Напишите небольшой отчет по результатам вашего исследования.

Заключение

В этой главе вы узнали много нового о метаданных. Вы получили представление о роли, которую библиографическая система играет в поддержке работы пользователей библиотечной системы, а также разобрались в целевом назначении метаданных.

Вы детально познакомились с двумя отдельными стандартами: MARC и Dublin Core (ограниченным и полным). Существует бесчисленное множество других стандартов, но в большинстве случаев они очень специфичны и сильно отличаются друг от друга. MARC является традиционным стандартом, ориентированным, прежде всего, на библиотекарей. При этом он часто подвергается критике как сложный и чрезмерно объемный, в особенности, когда он используется для описания ресурсов, более недолговечных, чем печатные издания, типичные для библиотечных фондов. Dublin Core ориентирован на пользователей и предназначен для работы с новыми источниками в Вебе, но его также критикуют за поверхностный и любительский характер. Полный вариант Dublin Core более других развивает структуру документа. Промежуточные стандарты включают MARCXML, являющийся разновидностью XML для представления записей MARC и MODS для «схемы описания объектов метаданных» (“metadata object description schema”), метод, не так давно разработанный в Библиотеке Конгресса США для представления библиографической информации, которая по средствам выражения находится где-то между MARC и Dublin Core. Мы изучили MODS в Главе 5.1.

Упражнения по работе с Greenstone составляют значительную часть рассматриваемой главы. Как и прежде, они преследуют двойную цель: дают возможность попробовать на практике материал, изложенный в данной главе, продолжают формировать навыки по использованию Greenstone для создания образовательных электронных библиотечных коллекций. Вы встретили несколько новых функций, таких как разделение полнотекстового каталога на части и добавление иерархического

классификатора, использующего структуру просмотра, выстроенную на основе специально предписанных метаданных. Нам хотелось бы, чтобы вы потренировались на своих собственных коллекциях и детально разобрались в тонкостях использования этих новых инструментов.

Метаданные играют в Greenstone две отчетливые роли: их можно интерактивно добавлять в документы, включенные в коллекции, или на их основе создавать коллекции. В этой главе вы на практике познакомились и с тем, и с другим. Вы создали коллекцию изображений, доступ к которой можно было получить только с помощью заданных вами метаданных; пользователи могут проводить поиск по этим метаданным и просматривать их. Простейшая коллекция изображений не содержит полнотекстных документов или внутренних гиперссылок, поэтому метаданные являются здесь единственным механизмом доступа. Вы также построили коллекцию записей MARC (точно также это могли бы быть записи CDS/ISIS, PROCite, BibTeX, или Refer) и снабдили ее интерфейсом, удобным для проведения тематического поиска и очень похожим на множество существующих систем библиотечных каталогов.

Двойственная роль метаданных в Greenstone отражает полярные точки зрения, упомянутые ранее. Метаданные, которые добавляются к документам в библиотечном интерфейсе, не очень значительны по объему, поскольку составляются на основе таких стандартов как Dublin Core или его расширенных версий, созданных на местах. Этот интерфейс не является полноценным редактором метаданных. Он предназначен для полнотекстных коллекций, которые включают меньшее количество документов, чем коллекции, состоящие исключительно из метаданных. Метаданные, которые вы описываете в данном случае, преследуют только одну цель — помочь пользователям электронной библиотеки свободнее ориентироваться в коллекции. Если у вас есть хорошо подготовленные и полные метаданные, созданные в таком же формате, как, например, записи MARC, то вам нужен специализированный редактор для работы с ними, а вовсе не библиотечный интерфейс. Однако, как вы уже видели, вы *можете* создавать коллекции Greenstone из тех записей, которые собраны в библиотечных каталогах.

Эти два направления использования метаданных в Greenstone функционируют на разных уровнях. Библиотечный интерфейс разработан для коллекций, включающих десятки тысяч документов, каждый из которых сам по себе может быть очень объемным, но он не

годится для работы с сотнями тысяч индивидуальных записей метаданных. Тестирование показало, что для загрузки в библиотечарский интерфейс на любом современном компьютере коллекции, состоящей из 10 тысяч небольших документов, содержащих индивидуальные записи метаданных, потребуется всего около 20 секунд. Greenstone может справиться и с очень большими коллекциями метаданных (например, она была протестирована на коллекциях, имеющих более 10 миллионов записей MARC), если только они не разбиваются на индивидуальные записи метаданных в библиотечарском интерфейсе.

Глава 3.3. Образовательные метаданные

Как было показано в предыдущих двух главах этого модуля, использование метаданных напрямую влияет на эффективность организации электронной библиотеки. Метаданные обеспечивают пользователей возможностью находить нужный им материал с помощью инструментов поиска и просмотра. В данной главе мы продолжаем развивать ваши знания о метаданных, сосредоточив внимание на том, какие именно метаданные и как они используются для описания образовательных ресурсов. Мы увидим, какое применение находят образовательные метаданные в существующих электронных библиотеках, и выполним серию упражнений, в ходе которых поработаем с их объектами.

Вспомните рассмотренные нами ранее примеры из практики образовательных электронных библиотек, с которыми мы знакомились в Главе 1.2 Модуля 1: *Электронную библиотеку по образованию в сфере наук о Земле (DLESE)* и *Национальную библиотеку виртуальных манипуляций для интерактивной математики (NLVM)*. Разработчики этих электронных библиотек организовали ресурсы таким образом, чтобы в наиболее полной степени удовлетворить потребности пользователей. При проектировании каждой из библиотек разработчики преодолевали проблемы, связанные с присутствием или отсутствием того или иного рода соответствующих образовательных метаданных.

Подготовка образовательных метаданных для ресурсов, имеющих в фондах крупных библиотек и используемых миллионами преподавателей и студентов, является одной из самых сложных организационных проблем. Руководители систем образования и правительства уже вполне отчетливо осознали, что эти проблемы не могут быть решены на уровне преподавателя. Существование взаимосогласованных стандартов метаданных определяет успешность дальнейшего использования образовательных электронных библиотек на национальном и международном уровнях.

Такой стандарт метаданных как Dublin Core задает рамки для представления электронных ресурсов в Интернете в целом. Что должно быть отражено в стандарте метаданных, если он предназначается именно для образовательных ресурсов? Исследователи работали над этой проблемой немало лет, и на данный момент ими предложено несколько вариантов таких стандартов.

На сегодняшний день наиболее значительными образовательными стандартами метаданных являются стандарты для учебных объектов, называемые LOM и SCORM, именно с ними мы будем детально знакомиться в данной главе. Возможно, эти названия кажутся вам на первый взгляд непонятными акронимами, но мы переходим к изучению такой темы, в которой без подобных акронимов не обойтись, и, выполняя Задание 1, вы в этом сами убедитесь.

Стандарт LOM

Стандарт LOM включает средства описания контента учебного объекта (*Learning Object*), в котором объединяется в единое целое сам ресурс и заложенная в него методика применения в условиях учебного процесса. Стандарт LOM для метаданных положен в основу очень многих из существующих образовательных электронных библиотек.

Полное название этого стандарта пишется так: *IEEE LTSC LOM*, где:

IEEE – Институт инженеров, работающих в сфере электросвязи и электроники (Institute of Electrical and Electronics Engineers); профессиональной организации, деятельность которой направлена на определение стандартов в различных областях, включая компьютерные технологии.

<http://ieee.org/>

LTSC – Комитет по стандартам образовательных технологий (Learning Technology Standards Committee); отдельной группы, организованной IEEE для разработки стандартов в области образовательных технологий.

<http://ltsc.ieee.org/>

LOM – Стандарт метаданных для учебных объектов, разработанный комитетом LTSC; сокращение от Learning Object Metadata.

В соответствии с документами LTSC Стандарт LOM описывает синтаксис и семантику метаданных учебных объектов (Learning Object Metadata), определяемых как характеристики, требующиеся для полного и адекватного описания какого-либо учебного объекта. В данном случае под учебными объектами подразумевается любой объект в цифровом или нецифровом формате, который может быть использован отдельно или совместно с другими объектами, а также на который могут быть даны ссылки с других объектов в процессе электронного обучения. Примерами такого электронного обучения могут

служить компьютерные обучающие системы, интерактивные учебные среды, системы искусственного интеллекта, системы дистанционного обучения и инструменты и среды, поддерживающие групповую учебную деятельность.

<http://ltsc.ieee.org/wg12/>

Задание 1

Прочитайте статью Эрика Дюваля «Для чего нужна стандартизация в области образовательных технологий», а затем ответьте на следующие вопросы:

- Какое значение имеют открытые стандарты для образования?
- Кто должен привлекаться к определению стандартов метаданных для образовательных ресурсов?
- Объясните разницу между учебными объектами (Learning Objects), метаданными учебного объекта (Learning Object Metadata, LOM) и системами управления знаниями (Learning Management Systems).
- Правомерно ли высказывание, что учебный объект, представленный в цифровом формате, всегда сопровождается записью LOM?
- Допустимо ли для какого-либо учебного объекта наличие более одной записи LOM?

Дискуссии по поводу стандартов могут быть довольно абстрактными. Для того чтобы придать им более конкретный характер, в таблице, приведенной ниже, мы показываем полную запись LOM, сделанную с помощью языка XML. Мы будем использовать точно такой же формат записи при работе с Greenstone в Задании 3. После первых нескольких строк запись для удобства чтения размещается в таблице в три столбца, хотя на самом деле это один длинный файл. Некоторые части скрипта показаны жирным шрифтом для того, чтобы вам было легче ориентироваться в комментариях, приведенных ниже; это выделение жирным шрифтом не является частью записи LOM.

Начальный тег **<lom>** включает некоторые определения XML, относящиеся к пространству имен. Далее запись включает шесть разделов верхнего уровня: *classification* (классификация), *general* (общее), *lifecycle* (жизненный цикл), *educational* (образовательные), *rights* (права), and *technical* (технические), в которые объединяются метаданные, представленные в данном источнике. Стандарт LOM включает следующие три категории верхнего уровня: *meta-metadata* (мета-метаданные), *relation* (отношение) и *annotation* (аннотация). Ни одна из этих категорий не является обязательной.

Ресурс, который описывают приведенные здесь метаданные, называется *A Teddy Bear Picnic (Пикник медвежонка)*: это утверждается в соответствующем разделе `<title>` тега `<general>`. В разделе `<classification>` можно прочитать, что предметом данного ресурса является Искусство (*Arts*). Стандарт LOM позволяет при желании вводить информацию на нескольких языках, если только в теге `<language>` не дана установка на использование английского языка как единственно возможного. Эта проблема легко решается с помощью отдельных тегов `<langstring>`, в которых можно установить любой язык, но если такой дополнительной установки сделано не было, то вся информация будет только на английском. Далее, после раздела `<general>` дается описание ресурса: *This film is about a teddy bear trying to enjoy his picnic. Until his sandwich runs away ... now he has to get it back! (Этот фильм о медвежонке, который пытается развлечься, устроив пикник. Это ему удастся до тех пор, пока от него не убежит его сандвич ... и теперь ему нужно этот сандвич вернуть!)*

Пример описания стандарта LOM

<pre><lom xmlns="http://linux2.common.ualgary.ca/rob_text/ims1_2_1" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://linux2.common.ualgary.ca/rob_text/ims1_2_1.xsd"></pre>		
<pre><classification> <keyword> <langstring> Arts </langstring> </keyword> <Objective> <langstring> Discipline </langstring> </Objective> </classification></pre>	<pre><lifecycle> <version> </version> <status> <langstring> Finished </langstring> </status> <contribute> <centity> <vcard> Dalvin Chung Trieu </vcard> </centity> <date> <datetime></pre>	<pre><educational> <context> <value> <langstring> VFSdept </langstring> </value> <source> <langstring> 3D Animation/Maya </langstring> </source> </context> <language> English </language> </educational></pre>

<pre> <langstring> A Teddy Bear Picnic </langstring> </title> <keyword> <langstring> teddy bear </langstring> </keyword> <keyword> <langstring> picnic </langstring> </keyword> <keyword> <langstring> run </langstring> </keyword> <keyword> <langstring> sandwich </langstring> </keyword> <keyword> <langstring> run away </langstring> </keyword> <coverage> </coverage> <description> <langstring> This film is about a teddy bear trying to enjoy his picnic. Until his sandwich runs away ... now he has to get it back! </langstring> </description> </general> </pre>	<pre> 2002-02-01 </datetime> <description> <langstring> Born and raised in B.C. Dalvin grew up having a knack and interest in art since grade 3. He went to study Fine Arts at Kwantlan College for two years before enrolling and completing the classical animation program at VFS. He was one of the top students in his class and got the certificate of excellence. Then he went on to further his studies to complete the Maya program at VFS. His future endeavor is to get into the video gaming industry. </langstring> </description> </date> <role> <langstring> Director/ Creator </langstring> </role> </contribute> </lifecycle> </pre>	<pre> <rights> <description> <langstring> Copyright 2002 Vancouver Film School </langstring> </description> <cost> </cost> <copyrightand otherrestrictions> <langstring> yes </langstring> </copyrightand otherrestrictions> </rights> <technical> <size> 2053707 </size> <duration> 1:38 </duration> <location> http://tree.vfs.com:8080/aloha/uploads/rbrayer/DalvinT.mov </location> <format> video/quicktime </format> </technical></lom> </pre>
---	---	---

Ресурс, который именно эта запись LOM описывает как фильм, на самом деле, характеризуется как объект в машиночитаемой форме в разделе <technical>. Здесь мы получаем информацию о том, что данный объект – это видеозапись в формате QuickTime продолжительностью в 1 минуту 38 секунд и требующая 2053707 байт (или примерно 2 Мб) дискового пространства. LOM допускает включение URL для указания местонахождения онлайн-версии данного ресурса. Разделы <lifecycle>, <educational> и <rights> в данном примере содержат такие детали, как имя автора ресурса, Далвина Чанг Триу (*Dalvin Chung Trieu*), название организации, которой принадлежат авторские права на ресурс – Ванкуверской школе кинематографии (*Vancouver Film School*), и текущий статус работы как завершенной (*Finished*).

Стандарт LOM разработан для описания широкого спектра всевозможных ситуаций. Вы также можете использовать не сам LOM, а любой стандарт из подмножества стандарта LOM, но при этом вы все равно останетесь им довольны. Такие подмножества называются *прикладные профили (application profiles)*. Хотя в самом стандарте LOM и нет указания на то, какие элементы обязательны для описания образовательного ресурса, а какие – нет, в прикладных профилях эти требования, как правило, предусмотрены.

Задание 2

Прочитайте статью «Построение прикладных профилей метаданных в образовании» Норма Фрайзена, Джона Мэйсона и Найджела Уарда и ответьте на следующие вопросы:

- Когда был принят стандарт IEEE LTSC LOM?
- Назовите главные отличия между двумя профилями, которые вы увидели при изучении примеров из практики?
- В чем их основное сходство?
- Что из рассмотренного в примерах из практики напоминает вам о проблемах, решаемых в той организации, которую вы представляете?

Одним из проектов в области электронного обучения, в рамках которого родилась инициатива CanCore, описанная в статье, предлагаемой для чтения к Заданию 2, называется CAREO (Campus Alberta Repository of Educational Objects). Также об этом проекте упоминается в статьях к Главе 4, включенных в Хрестоматию. Речь идет о некой разновидности образовательного репозитория, доступного для

пользователей Интернета. Целью данного репозитория является создание веб-коллекции учебных материалов по разным предметам со встроенной системой поиска, предназначенной для преподавателей, работающих в сельской местности и других местах.

Источник: <http://www.careo.org>

В репозитории представлено более 4000 единиц хранения по искусству, бизнесу, образованию, инженерному делу, юриспруденции и науке. В проекте задействована распределенная сеть экспертов, прошедших процедуру онлайн-регистрации. Для представления метаданных репозиторий использует стандарт LOM. Любой желающий может работать с алфавитным каталогом, но только зарегистрированные участники проекта могут получить доступ к материалам и использовать полный спектр инструментов данной системы. Вы сами также можете пройти процедуру онлайн-регистрации.

В проекте CAREO был создан веб-интерфейс, поддерживающий простой и расширенный поиск, а также просмотр ресурсов. В этом интерфейсе создана рабочая область, в которой участники могут просматривать и редактировать свои персональные данные, размещать на хранение интересующие их объекты, найденные в репозитории, и добавлять новые ресурсы.

В качестве альтернативного можно рассматривать решение по использованию стандартного программного обеспечения для электронной библиотеки, поэтому сейчас мы протестируем небольшую коллекцию Greenstone, включающую записи LOM.

Эта небольшая выдержка из коллекции дает представление об образовательных ресурсах, собранных в репозитории учебных объектов Университета Калгари (University of Calgary's Learning Commons Educational Object Repository). Она состоит из 40 единиц хранения, относящихся к искусству и науке, которые были экспортированы из данного репозитория в формате IEEE LOM и после определенной обработки помещены в коллекцию Greenstone.

Источник: *About this Collection*

на домашней странице этой коллекции – <http://www.careo.org>

Репозитории учебных объектов позволяют пользователям проводить поиск и просматривать записи метаданных. Но здесь мы видим серьезное продвижение вперед, поскольку Greenstone обеспечивает полнотекстовое индексирование всех без исключения онлайн-ресурсов, в которых есть текст. Пользователи могут просматривать объекты коллекции по какой-то определенной предметной теме, ис-

пользуя сортировку объектов по их названиям в алфавитном или хронологическом порядке. Можно также искать текст или заголовки объектов, имеющихся в коллекции, ограничив поиск категориями «Искусство» и «Наука». Когда пользователь получает искомый объект, он может использовать разные приемы для просмотра. Вы можете начать с записи LOM, представленной в табличной форме, разделенной на части, которые можно открыть или свернуть для получения большего или меньшего объема информации. Табуляторы в верхней части экрана меняют вид просматриваемого учебного объекта: при использовании одного из них — «Metadata XML» будут показаны метаданные из исходного формата LOM данного объекта. Если в учебном объекте имеются ссылки на подходящий вид онлайн-ресурсов, то появится третий табулятор, который покажет источник документа в Вебе.

Задание 3

Находясь в веб-браузере, перейдите на домашнюю страницу установленной вами коллекции Greenstone и познакомьтесь с коллекцией «LOM Demonstration», которая помечена пиктограммой IEEE LOM, поскольку использует разработанные им средства поиска и просмотра. После этого ответьте на следующие вопросы:

- Сколько всего объектов находится в данной коллекции и сколько по отдельности находится в категориях Arts и Science?
- Означает ли это, что в данной коллекции есть объекты, для которых отсутствует описание заголовка (General.Title на языке LOM)?
- В скольких объектах не определяется информация о дате (date)?

Стандарт SCORM

Теперь мы познакомимся с другим замечательным стандартом для учебных объектов, а именно — Sharable Content Object Reference Model (SCORM). Он был создан в рамках программы Инициатива по передовому распределенному обучению (Advanced Distributed Learning Initiative), являющейся результатом коллективных усилий правительства, индустрии и науки США. Целью было создание принципиально новой распределенной среды обучения, которая обеспечивает функциональное взаимодействие средств обучения и содержания курса на глобальном уровне.

SCORM нацелен на поддержку разработки многократно используемого учебного контента в виде обучающих объектов (instructional objects), организованных с помощью единой технической инфраструктуры компьютерного или Интернет-обучения.

Источник: <http://www.adlnet.org/index.cfm?fuseaction=scormabt>

Таким образом, SCORM действует как суперобложка, под которой собраны вместе различные компоненты, описывающие существующие стандарты, одним из которых, в частности, является профиль стандарта LOM.

Для лучшего понимания взаимосвязи, существующей между LOM и SCORM, прочитайте приведенный ниже отрывок, который имеет отношение к CanCore. Как вы уже знаете по работе над Заданием 2, CanCore — это прикладной профиль LOM.

Проследить связь между CanCore (LOM) и Sharable Content Object Reference Model (SCORM) путем прямого сравнения их друг с другом невозможно. SCORM по его же собственным характеристикам «является справочной моделью, которая включает набор взаимосвязанных технических описаний и руководств, обеспечивающих соответствие высоким требованиям (Министерства обороны США — Department of Defence), предъявляемым к размещаемому в сети учебному контенту» (SCORM Version 1.1). С другой стороны, CanCore дает справки только для одного стандарта — стандарта LOM, созданного для удовлетворения потребностей работников сферы образования и разработчиков учебных материалов из Канады и других стран. Дополнительные технические описания, на которые ссылается SCORM, включают также IMS Content Packaging Specification, AICC (Aviation Industry CBT Committee) CMI Data Model и AICC CM1001 Interoperability Guidelines. CanCore и SCORM взаимосвязаны и, следовательно, могут сравниваться только в отношении использования стандарта LOM.

Источник: <http://www.cancore.ca/scorm.html>

Это объяснение просто пронизано акронимами! Вы разберетесь в их значениях, выполняя следующее Задание.

Задание 4

На компакт-диске с материалами курса «Электронные библиотеки в образовании» есть тьюториал по SCORM. Для его просмотра щелкните по файлу SCORM_tutorial/viewer.htm. Проработайте представленный в тьюториале материал и ответьте на следующие вопросы:

- Назовите три основных компонента системы поддержки исполнения программ, и за что именно они отвечают.
- Из каких трех компонентов складывается модель сборки контента, и за что именно они отвечают?
- Назовите три элемента модели содержания, и за что именно они отвечают.
- Какова цель структуры содержания в модели сборки контента?
- Какие категории присутствуют в информационной модели метаданных SCORM? Прокомментируйте их взаимосвязь с LOM.
- Тьюториал по SCORM является примером онлайн-обучающего ресурса. Вы только что играли роль студента, знакомящегося с этим стандартом. Насколько этот тьюториал был для вас полезен? Что было более всего адекватно вашему уровню подготовки?
- Используя шаблон, приведенный ниже, в основу которого положена веб-форма, используемая для регистрации метаданных LOM ресурса в репозитории проекта Careopository, сформулируйте и запишите метаданные, которые описывают данный тьюториал по SCORM. Возможно, вам не потребуется заполнять все поля. Вернитесь к примеру описания LOM для Teddy Bear Picnic чтобы вспомнить, как используется тот или иной термин.

Форма для заполнения метаданных LOM для Careo

Общие данные	
Название	
Описание	
Язык	
Упаковка	
Ключевое слово	
Состав	
Структура	

Модуль 3. Работа с метаданными

Образовательные данные	
Язык	
Описание	
Обычное время, выделенное на изучение	(описание)
	(время)
Сложность	(значение)
	(источник)
Возрастная группа	(значение)
	(источник)
Роль пользователя	(значение)
	(источник)
Семантическая плотность	(значение)
	(источник)
Уровень интерактивности	
Вид учебного ресурса	(значение)
	(источник)
Вид интерактивности	(значение)
	(источник)
Контекст	(значение)
	(источник)
Классификация	
Цель	
Ключевое слово	
Описание	
Жизненный цикл	
Версия	
Статус	
Владелец	(роль)
	(данные)
	(дата)
	(описание)
Технические данные	
Размещение	

Размер	
Формат	
Продолжительность	
Требования	(тип)
	(название)
	(источник)
	(макс. версия)
	(мин. версия)
Другие требования по платформе	
Примечания по инсталляции	
Отношение	
Ресурс	(данные каталога)
	(описание)
	(идентификатор)
Вид	(источник)
	(значение)
Права	
Описание	
Копирайт (<i>да или нет</i>)	
Оплата (<i>да или нет</i>)	

Тьюториал по SCORM является далеко не единственным примером онлайн-обучающего ресурса. Отметим также, что и сам тьюториал, упакованный с использованием SCORM (!) как учебный объект, предназначенный для обмена (Sharable Content Object (SCO)), может импортироваться в любые другие системы управления контентом, поддерживающие данный стандарт. Например, в использующийся отдельно от других систем плеер Reload SCORM (www.reload.ac.uk) или в учебную Интернет-среду Moodle (www.moodle.org), в которой пользователи работают с учебным материалом онлайн (обратите внимание, что оба упомянутых нами ресурса относятся к программным средствам с открытым кодом). В тьюториалах, созданных в подобных средах, реализуются богатые функциональные возможности SCORM. Например, когда для перехода от одного учебного модуля к другому студенты должны выполнить определенные задания

или когда они развивают навыки общения в сети, задавая в онлайн вопросы преподавателю, контролирующему процесс обучения.

У нас была возможность посмотреть тьюториал без помощи подобного программного обеспечения, поскольку «упаковка» SCO позволяет создать такой набор веб-страниц, что их можно просматривать обычным браузером, получая при этом доступ к основному содержанию образовательного ресурса. Мы изъяли эти файлы из SCO и поместили их на компакт-диск с курсом, определив домашнюю страницу (*viewer.html*) как стартовую для выполнения следующего задания. Если сказать точнее, то данный тьюториал опубликован в форме SCO путем разархивации определенного набора файлов, что является обменной процедурой SCORM, типичной при онлайн-обмене учебными объектами. На компакт-диске мы разархивировали *полную* версию SCO для данного тьюториала, поскольку для более тщательного изучения стандарта SCORM вам потребуются все имеющиеся в нем файлы. Особый интерес для нас представляет файл *imsmanifest.xml*, сокращенная версия которого показана ниже.

```
<manifest identifier="SCourse-RELEASE-2004-9-24"
version="1.5"

xsi:schemaLocation="http://www.imsproject.org/xsd/
imscp_rootv1p1p2
imscp_rootv1p1p2.xsd
http://www.msglobal.org/xsd/imsmd_rootv1p2p1_imsmd_
rootv1p2p1.xsd
http://www.adlnet.org/xsd/adlcp_rootv1p2_adlcp_
rootv1p2.xsd">
<metadata>
<schema>ADL SCORM</schema>
<schemaversion>1.2</schemaversion>
<lom>
<general>
<title>
<langstring xml:lang="en-US">Academic ADL Co-Lab
SCourse
</langstring>
</title>
<description>
<langstring xml:lang="en-US">
```

```
    The Academic ADL Co-Lab's SCourse project is a
    Collection of Sharable Content Objects (SCOs)
    intended to help people and institutions learn
    about ADL's Sharable Content Object Reference
    Model.
    </langstring>
  </description>
  <keyword>
    <langstring xml:lang="en-US">SCORM course</
    langstring>
  </keyword>
<!-- additional keywords defined -->
  <general>
  <rights>
  <cost>
    <source>
      <langstring xml:lang="x-none">LOMv1.0
      </langstring>
    </source>
    <value>
      <langstring xml:lang="x-none">no</langstring>
    </value>
  </cost>
<!-- ... -->
  </rights>
</lom>
</metadata>
<organizations default="TOC1">
  <organization identifier="TOC1">
    <title>Academic ADL Co-Lab SCourse</title>
    <item identifier="ITEM1" identifierref="SC01_1_1">
      <title>
        Overview of the Advanced Distributed Learning
        Initiative
      </title>
    </item>
    <item identifier="ITEM2" identifierref="SC01_2_3_a">
      <title>
        Specifications, Standards, and the SCORM®
      </title>
    </item>
    <item identifier="ITEM3" identifierref="SC01_2_3_b">
      <title>
        Standards Evolution
```

```
    </title>
  </item>
<!-- items up to ITEM13 defined ... -->
</organization>
</organizations>
<resources>
  <resource identifier="SCO1_1_1" adlcp:scormtype="sco"
    type="webcontent" href="1_1_1/index.htm">
    <metadata>
      <schema>ADL SCORM</schema>
      <schemaversion>1.2</schemaversion>
      <adlcp:location>1_1_1/1_1_1.xml</adlcp:location>
    </metadata>
    <file href="1_1_1/images/academiccolab.gif"/>
    <file href="1_1_1/images/ADLlogo_med.gif"/>
    <file href="1_1_1/images/colabimage.gif"/>
    <!-- further images -->
    <file href="1_1_1/LICENSE/index.html"/>
    <file href="1_1_1/media/ADLObjectives.swf"/>
    <file href="1_1_1/scripts/sco_info.js"/>
    <file href="1_1_1/1_1_1_P1.htm"/>
    <file href="1_1_1/1_1_1_P2.htm"/>
    <!-- further html pages -->
    <file href="1_1_1/index.htm"/>
    <dependency identifierref="common"/>
  </resource>
  <resource identifier="SCO1_2_3_a" adlcp:
    scormtype="sco"
    type="webcontent" href="1_2_3_a/index.htm">
    <!-- continues in a similar vein -->
  </resource>
  <!-- ... -->
</resources>
</manifest>
```

Этот файл несет в себе описание всех отдельных файлов, из которых состоит данный ресурс, а также механизм их объединения в единый учебный объект. Как сказано в тьюториале по SCORM, описания объединены в три основные группы (которые прослеживаются от последней строчки текста файла по направлению вверх): *ресурсы*, лежащие в основе учебных объектов, *элементы*, описывающие как сам ресурс, так и его организационную структуру, и *метаданные* объекта, выраженные с помощью стандарта IEEE LOM. В этом конкретном файле — *imsmanifest.xml* — связь между элементами и ресурсами довольно проста, что облегчает их изучение. Элементы, как правило, представлены в определенном порядке, а ресурсы могут иметь ссылки на другие объекты SCO. Также допустимо, и то, что названные файлы могут включать ссылки на другие файлы (хотя в данном примере скрипта мы этого не видим).

Самостоятельно разработав метаданные LOM для этого ресурса, вы, возможно, захотите сравнить их с теми метаданными, которые описываются в файле *imsmanifest.xml*, поскольку очень любопытно посмотреть, что о ресурсе написали его авторы.

Инструментальные программные средства существуют для того, чтобы помочь пользователям сформировать контент для SCORM, например, упомянутая нами ранее команда разработчиков Reload (Reload — это сокращение от Reusable eLearning Object Authoring and Delivery) также занимается созданием метаданных, формированием контента и разработкой программы-редактора учебных средств.

Наверное, вам интересно узнать, каким именно образом работает тьюториал по SCORM, если он запускается вне системы управления обучением. Например, в некоторых тестах нужно перемещать объекты из одной области экрана в другую и, если вы это сделали верно, то после завершения действия вы услышите аплодисменты. Это стало возможным благодаря тому, что разработчики данного ресурса использовали подключаемое приложение ShockWave. Это приложение также должно быть включено в используемом вами браузере, иначе работа с подобными веб-страницами не получится так, как надо. Эти страницы создавались с использованием синтаксиса HTML, который активизирует файл приложения ShockWave, который и создает такой яркий эффект. Если вы перегружаете страницу с тестом, выполнив, например, только половину заданий, то тест также перегрузится и обновится. Однако, используя возможности SCORM, при разработке ресурса можно было бы задать установку на то, что при преждевременном завершении или прерывании теста,

система управления обучением фиксировала бы результаты прохождения теста, и с ними можно было бы познакомиться.

Рекомендации на будущее

Существующее разнообразие материалов, которые могут быть использованы для образовательных целей, означает, что, несмотря на сложное происхождение и состав существующих стандартов, все еще требуется немало сделать для того, чтобы стандарты типа LOM могли быть названы надежными. Стандарты образовательных метаданных развиваются намного быстрее, чем HTML или электронная почта. Однако они изменяются лишь в пределах версий или пространств имен, которые способствуют тому, чтобы программное обеспечение смогло справиться с записями, почерпнутыми из разных версий стандартов. В статьях, которые предлагаются для чтения к следующему заданию, формулируются некоторые идеи по проведению исследований в области учебных объектов и их использованию в обучении с акцентом на вопросах, имеющих отношение к метаданным. То есть, в этих статьях обсуждается будущее стандартов образовательных метаданных.

Задание 5

Прочитайте статью Эрика Дюваля и Уэйна Ходгинса «Основные направления исследований в области метаданных учебных объектов (LOM)», а затем ответьте на следующие вопросы:

- Какие из исследовательских задач, отмеченных в статье, являются наиболее важными?
- Какое влияние оказывают записи метаданных на интероперабельность LOM?
- Нужно ли вам приобретать специализированное программное обеспечение, чтобы работать с данными LOM в вашей организации?

Заключение

Этот модуль продемонстрировал возможности форматов метаданных для структуризации информации. Стандарты метаданных, представленные в Главе 3.2 — особенно MARC и Dublin Core, — предназначены для описания общих ресурсов. В данной главе вам были представлены стандарты метаданных LOM и SCORM, которые специально спроектированы для передачи содержательных аспектов описываемых ими образовательных ресурсов.

Содержание образовательных ресурсов отличает широкий спектр затрагиваемых тем и разнообразие форм, от книжки с картинками для начальной школы до видеофильма, объясняющего законы квантовой физики. Следовательно, стандарты образовательных метаданных являются сложно устроенными стандартами, сравнимыми с Dublin Core. Совместными усилиями различных заинтересованных организаций были разработаны единые стандарты, отвечающие потребностям всех заинтересованных сторон. Поскольку метаданные представлены в открытом формате XML, оказывается, что для обработки записей таких метаданных очень просто создавать новое программное обеспечение. В Greenstone, например, встроенное приложение LOM создано на основе существующей технологии обработки файлов XML.

Некоторые организации отреагировали на сложность LOM тем, что создали такое описание профиля приложения как дополнительный набор LOM. Эти дополнения стали постоянными для данного стандарта, что является хорошим знаком и предвещает то, что все последующие разработки в данной области также будут базироваться на открытых стандартах. Это позволит любой программе для создания электронной библиотеки импортировать образовательные метаданные и даст возможность пользователям проводить осмысленный просмотр и поиск документов по образовательным категориям.

Большинство пользователей компьютеров не ценят в должной мере значения стандартов. Единственное, что их на самом деле волнует, это, чтобы различные приложения, которыми они используют, могли работать вместе. Но программы, такие как почтовые клиенты и веб-браузеры, зависят исключительно от существования общих стандартов. То же самое справедливо и для традиционных библиотечных каталогов, и даже вдвойне — для электронных библиотек. Чем богаче структура подобных систем (например DLESE), предлагаемых пользователям, тем сложнее инфраструктура стандартов, положенная в их основу. Понимание этих стандартов является важным шагом в направлении создания образовательных электронных библиотек, полностью соответствующих запросам пользователей.

МОДУЛЬ 4

МУЛЬТИ- МЕДИЙНЫЕ ЭЛЕКТРОННЫЕ БИБЛИОТЕКИ

Глава 4.1. Форматы
и стандарты мультимедиа

Глава 4.2. Создание
гетерогенных коллекций

Глава 4.3. Расширение
возможности Greenstone



Цель

Понять принципы, положенные в основу наиболее распространенных форматов мультимедиа, научиться работать с мультимедийными документами и создавать из них электронные библиотечные коллекции; быть в состоянии провести усложненную пользовательскую настройку коллекций и разобраться в основных механизмах действия программы создания электронных библиотек Greenstone.

Задачи

После завершения работы по Модулю 4 вы сможете:

- описать различные способы хранения цифровых изображений;
- объяснить на примерах, что происходит при переводе изображения из одного формата в другие;
- объяснить, в чем отличие сжатия с потерями и без потерь;
- выделить черты сходства и отличия существующих аудио- и видео-форматов;
- порекомендовать формат для графических изображений, аудио- и видео-файлов, которые при различных обстоятельствах можно было бы использовать в электронных библиотеках;
- объяснить, в чем особенность электронной библиотеки Open Video;
- обсудить различные способы просмотра видеоматериалов;
- создать коллекцию гетерогенных мультимедийных документов в Greenstone;
- провести настройку коллекции Greenstone, используя для этого различные приемы;
- прокомментировать примеры коллекций документов Greenstone;
- объяснить, как Greenstone импортирует документы и создает коллекции;
- ориентироваться в файловой структуре Greenstone;
- объяснить назначение встроенных приложений, классификаторов и операторов форматирования;
- объяснить, как Greenstone генерирует веб-страницы;
- определить, что именно отражено в информационной базе данных коллекции Greenstone и в протоколах транзакций, которые она поддерживает.

Введение к Модулю 4

Образовательные электронные библиотеки часто содержат мультимедийные документы. Это объясняется тем, что учебный материал, как правило, создается с использованием элементов мультимедиа. Это могут быть учебники, диаграммы и рисунки, слайды презентаций, онлайн-презентации, учебное видео- и аудиовизуальные записи лекций. И действительно, именно преобладание в их составе мультимедийных документов отличает образовательные библиотеки от других научных библиотек.

Модуль 1 (Глава 1.1.) представляет вам несколько альтернативных концепций электронной библиотеки, включая и определение из учебника, которое начинается так: «Специально подобранная коллекция цифровых объектов, включающая текст, видео- и аудио...». Главной проблемой в ближайшем будущем будет интеграция объектов, представляющих различные виды медиа, в электронные библиотеки таким образом, чтобы каждый из этих объектов стал равноправным членом библиотечного общества. Сегодня в электронных библиотеках присутствуют, главным образом, текстовые документы, другие же типы медиасредств, такие как файлы мультимедиа, ассоциированные с текстовыми метаданными, обеспечивающими основу для поиска и просмотра, поддерживаются во вторую очередь. Но уже в ближайшее время электронные библиотеки смогут решить проблему использования мультимедийных объектов путем внедрения средств поиска и индексирования всех типов документов, средств автоматического аннотирования мультимедийных объектов и классификации мультимедийного контента, а также эффективных средств просмотра документов в мультимедийных коллекциях.

Но мы не будем ждать завтрашнего дня: мы хотим, чтобы вы начали создавать свои образовательные электронные библиотеки уже сегодня! Первая глава этого модуля дает вам значительный объем практической информации о том, как представлять рисунки, аудио и видео в компьютерной форме. Вторая глава целиком состоит из заданий упражнения по работе с программой Greenstone. Вы создадите коллекцию, которая включает элементы мультимедиа — звуки и рисунки, а также текстовые и библиографические документы. В третьей главе речь идет не только о том, как работать с мультимедийными документами, но и о некоторых дополнительных возможностях Greenstone, которые помогут вам создать различные виды коллекций и понять, что лежит в основе того или иного действия, которое вы выполняете в этой программе.

Модуль 4. Материалы для чтения

Глава 4.1. Форматы и стандарты мультимедиа

Материал для чтения 1. CD-ROM IMARK: Раздел 2 «Электронные документы и форматы», Урок 2.3. «Форматы электронных картинок».

Назначение. Знакомство с цифровыми изображениями, а именно: их основными характеристиками, различием между растровым и векторным форматами и проблемами конвертирования между различными форматами.

Материал для чтения 2. Учебник «Как создать электронную библиотеку»: Разделы 4.5, 4.6 и 5.5.

Назначение. Раздел 4.5 включает детальные описания наиболее популярных графических форматов, с которыми вы встретитесь в процессе создания электронной библиотеки. Раздел 4.6. посвящен различным форматам, существующим для видео- и аудиозаписей. Раздел 5.5. рассказывает о стандартах метаданных для мультимедиа.

Материал для чтения 3. CD-ROM «Электронные библиотеки в образовании», Хрестоматия, перевод статьи Гари Марчионини и Гари Гейслера «Открытая библиотека цифрового видео»,

Marchionini, G. And Geisler, G. (2002) The Open Video Digital Library, D-Lib Magazine, December 2002, Vol. 8, No. 12.

www.dlib.org/dlib/december02/marchionini/12marchionini.html

Назначение. Познакомиться с Open Video – электронной библиотекой, созданной в рамках американского исследовательского проекта по созданию видеобиблиотек, которая в полном объеме доступна для всех пользователей.

Глава 4.2. Создание гетерогенных коллекций

Материалы для чтения отсутствуют.

Глава 4.3. Расширенные возможности Greenstone

Материал для чтения 1. CD-ROM IMARK: Раздел 6, Урок 4.

Назначение. Познакомиться с описанием файла конфигурации, который является основой всех коллекций, а также узнать о приемах пользовательской настройки коллекций Greenstone.

Материал для чтения 2. Учебник «Как создать электронную библиотеку»: Разделы 6.1 и 6.4–6.7.

Назначение. В дальнейшем все, что вы будете делать в Greenstone, вы будете делать в библиотечарском интерфейсе. Чтобы понять его работу, нужно разобраться с основополагающими функциями системы, поэтому в этих разделах вы узнаете об основных используемых в программе файлах и директориях, процессе импортирования и создания коллекций, а также о том, как организуются документы внутри коллекции.

Материал для чтения 3. Учебник «Как создать электронную библиотеку»: Разделы 7.1, 7.4, рекомендованные для дополнительного чтения.

Назначение. Узнать еще о внутреннем устройстве Greenstone, в частности, о том, что происходит, когда читатели взаимодействуют с системой.

Глава 4.1. Форматы и стандарты мультимедиа

Библиотеки всегда связаны с литературой. Наш словарь определяет литературу как «все, что записано обществом в прозе или в стихах» и далее добавляет «в широком смысле, литература включает все типы художественных и нехудожественных произведений, предназначенных для публикации». Получается, что все это находится в строгой зависимости от письменности. Но в Новой Зеландии, откуда родом сама программа Greenstone, наиболее ранние культурные традиции связаны с маори, чья литература представлена историями, сказками, стихами и мифами, которые передавались через устное народное творчество. Только тогда, когда из Европы прибыли первые миссионеры, эти первоисточники были записаны и стали «литературой» в том смысле, как это определяется словарями. Нам есть чему поучиться у ранних культур. Например, одним из направлений в современном литературном творчестве, особенно в австралийском и североамериканском, является запись устных народных сказаний коренных народов, проживающих на данной территории.

Задание 1

Подумайте о том курсе, который вы преподаете или с которым вы хорошо знакомы:

- Используются ли в нем какие-либо мультимедийные информационные ресурсы?
- Чему новому могут научиться студенты, если у них будет доступ к источникам мультимедийной информации?
- На наших глазах происходит «тихая революция» в связи с появлением новой и недорогой для пользователей технологии перехвата статичных и движущихся изображений. Предположим, что преподаватель имеет все необходимые ресурсы для создания мультимедийных учебных материалов и готовит на их основе и в соответствии с поставленными целями обучения мультимедийные информационные ресурсы. Как это может повлиять на формирование знаний и умений учащихся?
- Нужно ли учащимся заниматься созданием своих собственных коллекций мультимедийных ресурсов по теме изучаемого курса?
- Как вы себе представляете единый мультимедийный информационный ресурс, созданный в поддержку учебного курса?

Для работы с мультимедийными документами вам нужно кое-что узнать о том, как средства мультимедиа представлены на вашем компьютере и каковы их наиболее распространенные форматы. В целом, для представления мультимедийных документов существует гораздо меньше возможностей, чем для текстовых документов. Но, в связи с недостаточно разработанной квазианалоговой природой этих медиасредств, размер файлов очень быстро увеличивается, становясь непомерно большим, и поэтому процедуры сжатия стали включать непосредственно в форматы этих средств. Решая простую задачу по уменьшению размера файла, программы сжатия неизбежно приводят к некоторым негативным явлениям, которые нужно учитывать при разработке электронных библиотек.

В следующем далее Задании вы узнаете о форматах GIF, PNG и JPEG. Первые два формата годятся для представления изображений, созданных на компьютере, таких как текст, рисунки и логотипы. Формат JPEG создан для передачи полутоновых изображений, к которым относятся фотопортреты и пейзажи.

Задание 2

Проработайте Урок 2.3. учебного модуля на компакт-диске IMARK и ответьте на вопросы в конце этого урока. Затем прочитайте Раздел 4.5. (стр. 194–206) учебника «Как создать электронную библиотеку» и ответьте на следующие вопросы:

- В чем разница между сжатием изображений с потерями и сжатием без потерь?
- Почему стандарт PNG, созданный для сжатия изображений без потерь является более продвинутым по сравнению с GIF?
- Не принимая во внимание технические детали, скажите, какой из форматов больше подходит для электронных библиотек с открытым кодом: GIF или PNG? Почему?
- Когда вы конвертируете изображение из формата GIF или PNG в формат JPEG, что происходит не так, как хотелось бы? Почему? Что происходит, когда вы конвертируете файл обратно из JPEG?
- Во многих электронных библиотеках, хранящих фотографии, цифровые ресурсы представлены в одном из двух форматов: с потерями и без потерь данных. Почему? Каково, по вашему мнению, назначение каждого из них? Какая из версий должна быть показана пользователю на веб-страницах? Какой именно стандарт для изображений должен использоваться в этом случае?

Мультимедиа охватывает и видео-, и аудиоформаты. Следующее Задание предлагает познакомиться со стандартом MPEG, который включает схему MP3, широко используемую для прослушивания музыки. В нем также упоминаются форматы мультимедиа для Apple – QuickTime, и для Microsoft – AVI, а также WAV, AIFF и AU для аудио.

Задание 3

Прочитайте Раздел 4.6 (стр. 206–216) учебника «Как создать электронную библиотеку» и затем ответьте на следующие вопросы:

- Какой вариант стандарта MPEG был разработан для использования в WWW – MPEG-1, MPEG-2 или MPEG-4?
- Какой из форматов вы бы порекомендовали для мультимедийной информации, предназначенной для открытой электронной библиотечной системы: MPEG, QuickTime или AVI?
- Каково основное преимущество RealVideo и RealAudio по сравнению с более ранними схемами запуска и просмотра мультимедийных объектов?
- Как между собой связаны стандарты MP3 и MPEG?
- Назовите два главных преимущества MP3 по сравнению с WAV, AIFF и AU, используемыми для аудиоинформации.

Ни один из упомянутых выше стандартов не является полностью открытым. Как правило, декодеры распространяются свободно, однако работа кодировщиков должна быть оплачена. Например, люди, распространяющие музыку на коммерческой основе и использующие для этого формат MP3, должны платить за лицензию. В ответ на это, сообщество разработчиков программного обеспечения с открытым кодом создает альтернативные стандарты для мультимедиа, такие как аудиоформат Ogg Vorbis и видеоформат Ogg Theora, оба из которых используют общую упаковку метаданных.

Некоторые стандарты метаданных разработаны специально для мультимедийного материала. Формат TIFF является широко используемой файловой структурой, которая приспособлена для работы с большим количеством различных форматов для изображений и включает также обеспечение для описательных метаданных. Формат MPEG-7 является специальным стандартом для описания мультимедийных документов. Вы узнаете о каждом из них, выполняя следующее Задание.

Задание 4

Прочитайте Раздел 5.5. (стр. 261–266) учебника «Как создать электронную библиотеку» и затем ответьте на следующие вопросы:

- Почему программа, которая читает изображения в формате TIFF, может не распознать некоторые определенные файлы TIFF?
- Можете ли вы использовать описание XML в качестве значения в поле метаданных TIFF? Если да, то в каком именно поле или полях?
- Какой стандарт является более сложным и понятным – TIFF или MPEG-7?
- На основе материала, изложенного в учебнике, ответьте, для какой сферы потенциального применения стандарта MPEG-7 он не подходит: образования, журналистики, туризма, культуры и отдыха, развлечений, геоинформационных систем, систем удаленного наблюдения и слежения, биомедицины, торговли, архитектуры, недвижимости, дизайна интерьеров, кино-, видео- и радиоархивов, брачных агентств?

Существует не так много электронных библиотечных систем, которые посвящены видео. Одной из таких систем является электронная библиотека Open Video, описание которой приводится в следующем задании. Это задание будет вдвойне интересно, поскольку оно поможет не только определить существующие проблемы, но и найти решения по вопросам, касающимся использования видео, а также потому, что видео является потенциальным источником материала для образовательных электронных библиотек, которые вы будете создавать сами.

Задание 5

Прочитайте статью Гари Марчионини и Гари Гейслера «Открытая библиотека цифрового видео», а затем ответьте на следующие вопросы:

- Какие особые проблемы возникают при включении видео в электронные библиотеки?
- Что получают образовательные электронные библиотеки от участия в проекте Open Video?
- В каком формате или форматах в нем представлен материал?
- Что такое «ключевое фреймовое извлечение», и какое отношение оно имеет к видео библиотекам?
- Назовите три разных способа представления пользователям обобщенной информации о видео.

Заключение

С мультимедийными объектами обычно работать намного сложнее, чем с текстом. На их создание и редактирование уходит много времени, и этот процесс требует специальных аппаратных средств для перехвата аудио- и видеоизображения, а также их обработки. Более того, создание высококачественных мультимедийных ресурсов требует наличия специальных умений по режиссуре, операторской работе, съемке, звукорежиссуре и т.д. Составление метаданных и, в особенности, кратких описаний документов для мультимедиа является намного более сложным делом, чем для текста. Мультимедийная информация занимает намного больше дискового пространства, чем текст. При организации хранения рисунков надо учитывать, что они занимают столько же места, сколько тысячи, а может быть, и миллионы слов. Процесс сжатия здесь является очень важным моментом, а наиболее эффективные методы компрессии очень дороги. Это поднимает вопрос о качестве изображения или звука. Зачастую необходимо содержать одновременно разные версии одного и того же файла — для хранения и для демонстрации, а возможно, и третью версию — для того, чтобы она служила в качестве ярлыка изображения. Из-за большого объема данных, с которыми приходится работать, создание этих версий на компьютере может занять немало времени.

Другой проблемой, возможно, даже еще более серьезной, является то, что авторское право на мультимедийные объекты во Всемирной паутине находится под большим контролем, чем на текстовые материалы. Мы не рекомендуем вам создавать электронные библиотечные коллекции из текстовых документов, которые вы выгрузили случайно из Веба, пока вы не изучите внимательнейшим образом всю ситуацию с авторским правом. Однако если вы все же этого не сделали, то в худшем случае владелец прав, узнав о факте их нарушения, гневно потребует от вас незамедлительно убрать из WWW размещенную вами информацию. Если же вы все-таки создаете электронную библиотеку из видео- или аудиоматериала, скаченного из Веба, вы невольно подвергаете себя опасности быть оштрафованным или жестоко преследуемым за «пиратство», так ваши действия определяются, например, Американской ассоциацией динамичных изображений (Motion Picture

Association of America). Вас даже могут назвать «террористом».* К счастью, существует движение в поддержку создания мультимедийного контента с открытым кодом и его свободного распространения, примером чего выступает электронная библиотека Open Video.

С другой стороны, нравится нам это или нет, современная культура постепенно смещается от текста в сторону представления информации в мультимедийной форме. Как подчеркивается в Задании 1, с появлением дешевой потребительской технологии для перехвата изображения, в нашей жизни произошла тихая революция — вспомните о телефонах со встроенной видеокамерой и фотоаппаратом! Теперь любой желающий может завести фотоальбом на собственном домашнем компьютере или снимать на видео, а затем, используя усовершенствованные приемы редактирования, создать фильм профессионального качества или записать CD-ROM с собственной музыкой (или музыкой, созданной другими). После появления широкоэкранный телевидения наше общество, привыкшее к доминированию текста, постепенно стало все более и более благосклонно воспринимать такое средство, как движущиеся изображения. Книголюбые могут сетовать по поводу отрицания мира печатных слов, приводя постоянный аргумент о превосходстве страниц над виртуальностью и превознося мощь написанного слова, «развивающего воображение и образность мышления» в отличие от любого телевизора. Мы хотим, чтобы наши дети проводили больше времени за чтением книг, а не за видеоиграми. Но мы не можем жить вне этого мира, а он сильно меняется.

Электронные мультимедийные библиотеки откроют доступ к информации людям вне зависимости от их образовательного уровня. Со временем это внесет свой вклад в решение проблемы с различными «цифровыми неравенствами», которые пронизывают наш мир — социальным неравенством между информационно богатыми и информационно бедными в каждой из наших стран, демократическим неравенством между теми, кто использует и не использует все разнообразие электронных ресурсов для участия в жизни общества, а также глобальным неравенством, которое отражает колоссальную диспропорцию в доступе к информации людей из промышленно

* «Мы ведем нашу собственную борьбу с терроризмом», — сказал Джек Валенти, президент Motion Picture Association of America (МРАА), в связи с антипиратской кампанией, которая проводилась в виде нескольких рейдов по кампусам колледжей. Amy Harmon, “Black Hawk Download: Moving Beyond Music, Pirates Use New Tools to Turn the Net into an Illicit Video Club,” New York Times, 17 January 2002.

развитых и развивающихся стран. Мультимедиа может оказаться довольно сложным делом, но это стоит того!

Положительным фактором является то, что мультимедийные объекты всех видов могут включаться в коллекции, созданные с помощью таких доступных программных средств формирования электронных библиотек, как Greenstone. Для того, чтобы понять, как именно это можно сделать, переходите к Главе 4.2.!

Глава 4.2. Создание гетерогенных коллекций

Все электронные библиотечные коллекции, которые мы до недавнего времени создавали, включали документы, принадлежащие к одному и тому же типу. В этой Главе, целиком состоящей из большого по объему упражнения, выполняемого с использованием программы Greenstone, вы создадите сложную гетерогенную коллекцию, которая включает мультимедийные элементы — звуки и рисунки — наряду с текстовыми и библиографическими документами.

Представьте, что вы — учитель музыки, и вы хотите подготовить для своих учеников учебный материал в виде развернутой коллекции, иллюстрирующей определенное направление в музыке. Скорее всего, вам потребуется включить в нее аудиозаписи в формате MP3, несколько музыкальных примеров, которые вы сами наиграли и записали на компьютере с помощью клавиатуры MIDI, дискографию в форме HTML-файлов, дополнительный материал (такой как, например, музыкальные обзоры в форматах PDF и Word), библиотечные записи по теме в формате MARC, нотные станы с нотами и гитарными аккордами в виде текстовых файлов, изображения альбомных обложек в формате JPEG и т.д.

Для примера на компакт-диск с курсом помещена небольшая электронная библиотечная коллекция с информацией по рок-группе «Битлз».

Упражнение – этап 1: посмотрите, что можно сделать!

(Разделение на этапы введено переводчиком)

1. Скопируйте всю папку целиком
sample_files→beatles→advbeat_large
(со всем ее содержимым) в вашу папку *collect* в Greenstone. Если вы провели стандартную установку Greenstone, то вы ее найдете в обычном месте, а это:
My Computer→Local Disk (C:)→Program Files→greenstone→collect
Поместите туда *advbeat_large*.
2. Если у вас на компьютере уже запущен локальный библиотечный сервер электронной библиотеки Greenstone, то вам следует его перезапустить, щелкнув по пиктограмме CD на панели задач, а затем нажать кнопку *RestartLibrary*. Если сервера нет, то запустите его, выбрав *Greenstone Digital Library* в меню *Start*.

3. Изучите коллекцию «Beatles». Обратите внимание на то, что кнопка *browse* делит материал на семь разных типов. Внутри каждой из этих категорий документы помечены соответствующими пиктограммами. Некоторые документы имеют пиктограмму аудиофайла: когда вы щелкаете по ним клавишей мыши, вы слышите музыку (в случае, если на вашем компьютере установлена нужная программа для проигрывания записей). Другие документы имеют изображения рисунков: когда вы щелкаете по ним, вы видите изображения.
4. Посмотрите на окно браузера *titles a–z*. Каждый заголовок сопровождается книжной полкой, которая может содержать несколько объектов, связанных одной темой. Например, *Hey Jude* включает обложку альбома, аудиофайл MP3 и MIDI-версии, слова песни и дискографию.
5. Обратите внимание на низкое качество представленных метаданных. Например, у четырех объектов под названием *A HARD DAY'S NIGHT* (под “Н” в браузере *titles a–z*) оказались разные заголовки. Коллекцию гораздо проще создать, если предварительно все было бы проверено и вручную скорректировано, но это очень большая работа. Только совсем небольшое количество метаданных было введено вручную — в менее чем десяти объектах. Все исходные метаданные остались нетронутыми, и с помощью инструментов Greenstone их можно автоматически вычистить. Далее вы узнаете: как это делается, и, что все не так просто.
6. В окне просмотра файлов Windows найдите файлы, которые включаются в данную коллекцию, а именно папку:
`sample_files→beatles→advbeat_large→import`
 В семи вложенных папках верхнего уровня оказалось более 450 файлов. Организация — минимальная, и она наглядно отражает то, что эти файлы набирались в разное время и разными способами. Например, *html_lyrics* и *discography* были взяты с веб-сайтов, а *cover_images* включает альбомные обложки в формате JPEG. Пройдитесь по всей структуре файлов и посмотрите примеры файлов по каждому из типов, по очереди открывая каждый из них.

Отсутствие порядка в исходной информации типично для материалов, собранных из различных источников. Вам, как теперь уже почти готовому специалисту по созданию электронных библиотек, предстоит решить задачу превращения этого агломерата неорганизованной информации в привлекательную, удобную для использования электронную библиотечную коллекцию. Вы можете начать с создания полного

набора метаданных для каждого объекта коллекции. Это большая работа, а коллекция будет излишне объемной и потребует избыточных затрат времени от учителя музыки, который всего лишь хотел обеспечить своих учащихся полезным учебным материалом. В действительности, очень многое можно сделать, затратив минимум усилий, но предварительно хорошо разобравшись в структуре исходной информации. Именно этот подход мы и будем использовать в данной Главе.

Мы продолжим реорганизацию заготовки коллекции «Beatles», которую вы только что увидели. Мы разработаем коллекцию, используя небольшой набор материалов исключительно с целью ускорить процесс реорганизации имеющейся в нашем распоряжении коллекции.

Упражнение – этап 2: Создание основы коллекции

7. Запустите новую коллекцию (File→New) называющуюся **small_beatles**, используя установки по умолчанию “New Collection”. Использование в качестве основы существующей коллекции Advanced Beatles сделало бы вашу жизнь намного проще, но мы хотим, чтобы вы научились строить коллекцию с азав! Заполните поля необходимой информацией. Используйте набор метаданных Dublin Core, установленный по умолчанию.
8. Скопируйте файлы, находящиеся в папке `sample_files→beatles→advbeat_small` в вашу новую коллекцию. Сделайте это, открыв папку `advbeat_small` и, выбрав в ней восемь объектов (с `cover_images` до `beatles_midi.zip`), перетащите их на другую сторону поля. Поскольку некоторые из этих файлов являются файлами в формате MP3, то вам будет задан вопрос о том, хотите ли вы включить в свою коллекцию **MP3 Plugin**. Щелкните **<Add Plugin>**.
9. Перейдите на панель **Enrich** и просмотрите файлы. Метаданные в них пока отсутствуют. Вспомните, что для просмотра файлов вам нужно делать двойные щелчки по их названиям.
(В этой коллекции нет MIDI-файлов: такие файлы требуют гораздо более серьезной настройки, поскольку для них нет специального встроенного приложения (MIDI plugin). Мы вернемся к ним позже.)
10. Перейдите на панель **Create** и создайте (**build**) коллекцию.
11. Просмотрите (**preview**) результат.

Использование установок по умолчанию довольно удобно, но мы можем улучшить организацию и представление коллекции, поработав над ее ди-

зайном. Именно это и будет в центре нашего внимания при выполнении оставшихся упражнений.

Упражнение – этап 3: Корректировка метаданных вручную

12. Скорее всего, вы захотите подправить некоторые метаданные – например, опечатки в заголовках “MAGICAL MISTERY TOUR.” Эти документы находятся в разделе с дискографией, а в названиях их файлов содержатся одни и те же опечатки. Найдите один из них на панели **Enrich**. Обратите внимание на то, что в элементе извлеченных метаданных **ex.Title** присутствуют необходимые данные, но они неверно написаны. Вы не можете скорректировать этот элемент из-за того, что данные извлекаются непосредственно из файла и обновляются каждый раз, когда эта коллекция перестраивается.
13. Сделайте вместо этого следующее. Добавьте для этих двух файлов (Magical Mystery Tour) метаданные **dc.Title**. Перейдите на панель **Enrich**, откройте папку с дискографией и просмотрите ее до конца списка, где расположены отдельные несистематизированные файлы. Установите значение **dc.Title** для двух «неправильных» файлов.

У нас появилась новая проблема. Метаданные **dc.Title** более не появляются в списке titles a–z, поскольку установки классификатора сориентированы на использование **ex.Title**. Изменение установки классификатора на использование **dc.Title** приведет к потере всех извлеченных заголовков. К счастью, есть другая возможность справиться с этой проблемой за счет определения списка имен метаданных в самом классификаторе.

14. Перейдите на панель **Design** и выберите раздел **Browsing Classifiers**. Сделайте двойной щелчок по классификатору **Title** (он идет первым) и измените установки конфигурации.

- Напечатайте “dc.Title,” перед ex.Title в окне метаданных – т.е. получится так:

dc.Title,ex.Title

Снова создайте (**build**) коллекцию и просмотрите (**preview**) ее.

Извлеченные метаданные не отличаются точностью, но ведь они ничего не стоят. С другой стороны, метаданные, написанные вручную, очень надежны, но дороги. Следующее упражнение покажет вам, как найти альтернативу за счет использования извлеченных метаданных, но с последующей корректировкой их вручную в случае обнаружения в них ошибок. Если это, возможно, и не устроит профессионального библиотекаря, то уж точно подойдет учителю музыки, который хочет сформировать свои коллекции при минимальной затрате сил.

В следующем упражнении мы наведем порядок в коллекции и добавим в нее несколько дополнительных возможностей.

Упражнение – этап 4: Простая настройка

15. Сначала давайте уберем из классификатора **AZList** имена файлов, которые нам не очень нужны, и организуем вместо этого новую структуру для просмотра файлов, при которой документы группируются по категориям (дискография, слова песен, аудио и т.д.). Категории описываются с помощью метаданных, введенных вручную.

- Перейдите на панель **Enrich**, выберите папку *cover_images* и измените значение метаданных **dc.Format** на “Images”. Установка этого значения на уровне папки означает, что все файлы, содержащиеся в ней, также его унаследуют.
- Повторите это действие. Установите “Discography” для папки *discography*, “Lyrics” для *html_lyrics*, “MARC” для *marc*, “Audio” для *mp3*, “Tablature” для *tablature_txt* и “Supplementary” для *wordpdf*.
- Перейдите на панель **Design** и выберите раздел **Browsing Classifiers**.
- Удалите классификатор **ex.Source** (второй в списке).
- Введите классификатор **AZCompactList**. Выберите **dc.Format** в поле метаданных и “Category” для **buttonname**.

Снова создайте (**build**) коллекцию и просмотрите (**preview**) ее.

16. В Greenstone нет заранее подготовленной кнопки для “Category”, поэтому она появляется на панели навигации просто как текст. При этом в программе есть готовая кнопка для *browse* (она используется в коллекции «Beatles», с которой вы работаете).

- Вернитесь к классификатору **AZCompactList** за метаданными **dc.Format**. Щелчком выберите пункт меню **sort** и оставьте **Title** в расположенном рядом текстовом окне: это приведет к тому, что классификатор будет выводить документы в алфавитном порядке по их названиям. Также определите “Browse” в качестве **buttonname**.

Вам придется еще раз создать коллекцию для того, чтобы увидеть, как это все работает.

17. Наряду с аудиофайлами вы видите пиктограмму MP3 файла, после щелчка по которой запускается аудиозапись, а также текстовый документ, содержащий информацию, не представляющую никакой ценности и не предназначенную для просмотра. Чтобы это не появлялось, вам придется поработать с оператором форматирования.

- Перейдите на панель **Design** и выберите раздел **Format Features**.
- Убедитесь в том, что выбран именно **VList** и проведите изменения в полях, отмеченных в примере маркером. Вам нужно ввести три строки, вместо первой строки, и удалить вторую строку.

Измените:

```
<td valign=top>[link] [icon] [/link]</td>
<td  valign=top>[srclink] {Or} { [thumbicon] , [srcicon] } [/srclink]
</td>
<td valign=top>[highlight]
{Or} { [dls.Title] , [dc.Title] , [Title] , Untitled }
[/highlight] {If} { [ex.Source] , <br><i>([ex.Source])</i></td>
на:
<td valign=top>
{If} { [dc.Format] eq 'Audio' ,
      [srclink] [srcicon] [/srclink] ,
      [link] [icon] [/link] }</td>
<td  valign=top>[highlight] {Or} { [dls.Title] , [dc.Title] ,
[Title] , Untitled } [/highlight] {If} { [Source] , <br><i>([Source])</i></td>
```

- После чего щелкните **<Replace Format>**:

Для выполнения этого задания мы заранее подготовили для вас простой текстовый файл с новым текстом. Откройте в WordPad следующий файл:

sample_files→beatles→format_tweaks→audio_tweak.txt

(Не перепутайте и используйте именно WordPad, а не Notepad, поскольку Notepad не сможет правильно передать прерывание строк). Поместите текст в буфер обмена, выделив его в WordPad и выбрав команды Edit→Сору. Теперь вернитесь к интерфейсу библиотекаря, выделите весь текст, который относится к текущему оператору форматирования Vlist, и, используя команды Edit→Paste, замените старый оператор на новый. Не забудьте нажать **<Replace Format>**, когда вы завершите эту работу.

Посмотрите (**preview**) полученный результат. Если вы используете локальный библиотечный сервер Greenstone, то перейдите на панель **Create** и щелкните **<Preview Collection>**, что даст команду этому серверу провести повторное сканирование операторов форматирования. Вам не нужно повторно создавать коллекцию, поскольку операторы форматирования используются только в процессе работы системы.

Щелкните по кнопке *Reload* в окне вашего браузера для перезагрузки страницы.

18. Пока мы там находимся, давайте уберем имя файла первоисточника в том месте, где он появляется после каждого документа.

- В описании формата VList удалите выделенный маркером текст:

```
<td valign=top>
{If}{[dc.Format] eq 'Audio',
[srclink][srcicon][srclink],
[link][icon][link]}</td>
<td valign=top>[highlight] {Or}{[dls.Title],
[dc.Title],[Title],Untitled} [/highlight] {If}{[Source],
<br><i>([Source])</i></td>
```

Не забудьте щелкнуть **<Replace Format>** после завершения всей работы! Просмотрите (**preview**) полученный результат (вам не нужно заново строить эту коллекцию.)

19. Иногда бывает так, что несколько документов имеют один и тот же заголовок. Например, *All My Loving* появляется одновременно как заголовок в словах песни и в нотах (под заголовком *ALL MY LOVING*). В браузере *titles a-z* это можно отрегулировать, сгруппировав документы вместе под пиктограммой книжной полки. Это можно выполнить с помощью списка AZCompactList.

- Перейдите на панель **Design** и выберите раздел **Browsing Classifiers**.
- Удалите классификатор **Title** (находящийся сверху).
- Добавьте классификатор **AZCompactList** и введите в качестве метаданных **dc.Title,ex.Title**.
- Активизируйте **min_group** и установите значение на 1. Этим вы зададите универсальный способ представления документов путем создания книжной полки под каждый заголовок.
- Завершите действие, нажав **<OK>**.
- Перенесите новый классификатор в самое начало списка (кнопка *Move Up*).

Снова создайте (**build**) коллекцию и просмотрите (**preview**) ее. Теперь оба объекта с заголовками *All My Loving* появляются под одной и той же книжной полкой. И все же, многие документы не удастся объединить в связи с тем, что у них нет унифицированных названий: например, *A Hard Day's Night* появляется в четырех разных вариантах. Далее мы еще научимся соединять подобные документы.

20. Сделайте так, чтобы книжные полки отражали точное количество документов, которое они включают. Для этого введите соответствующую строку в оператор форматирования VList на панели **Design**:

```
<td valign=top>
{If}{[dc.Format] eq 'Audio',
[srclink][srcicon][/srclink],
[link][icon][/link]}</td>
<td>{If}{[numleafdocs],[numleafdocs]}</td>
<td valign=top>[highlight]{Or}{[dls.Title],
[dc.Title],[Title],Untitled}[/highlight]</td>
```

Вы найдете этот текст в *format_tweaks→show_num_docs.txt* и сможете его скопировать и вставить в документ точно так же, как делали раньше. Не забудьте щелкнуть **<Replace Format>**.

Посмотрите (**Preview**) результат (вам не нужно создавать коллекцию.)

21. Теперь займитесь изображениями. Здесь тоже оказались лишние документы. Сначала перейдите на панель **Enrich**, откройте папку *cover_images* и добавьте метаданные **dc.Title**, в которых для каждого из десяти представленных документов прописывается название альбома, к которому они относятся. Не забудьте, чтобы увидеть файл, вы должны сделать по нему двойной щелчок.
22. Для подавления лишних документов вновь измените оператор форматирования **VList** на панели **Design**, добавив две выделенные подсветкой строки и закрыв фигурную скобку:

```
<td valign=top>
{If}{[dc.Format] eq 'Audio',
[srclink][srcicon][/srclink],
{If}{[dc.Format] eq 'Images',
[srclink][thumbicon][/srclink],
[link][icon][/link]}</td>
<td>{If}{[numleafdocs],[numleafdocs]}</td>
<td valign=top>[highlight]{Or}{[dls.Title],
[dc.Title],[Title],Untitled}[/highlight]</td>
```

23. Добавьте классификатор просмотра **Phind**, который выбирает информацию и строит высказывания на основе заголовка (Title) и текста (text) документа (установки по умолчанию).

24. Для завершения процесса создания коллекции используйте кнопку просмотра **URL to 'about page' icon** в разделе **General** на панели **Design** для выбора следующего изображения: *advbeatles_large→images→flick4.gif*.

Снова создайте (**build**) и просмотрите (**preview**) коллекцию.

Обратите внимание на то, как в предыдущем упражнении мы подготовили метаданные *dc.Format* для всех документов коллекции, затратив на это минимум сил. Мы сделали это за счет выделения файловой структуры исходной информации. Хотя мы и сетовали ранее на то, что структура папок очень запутанная, но все же вы смогли ею воспользоваться при составлении метаданных.

В следующем упражнении мы встроим в коллекцию файлы MIDI. В программе Greenstone пока еще нет встроенного приложения MIDI, но это не означает, что вы не сможете использовать файлы MIDI. Мы также почистим браузер *titles a–z*.

Для того чтобы это сделать, нам нужно перевести библиотечарский интерфейс в другой режим. Интерфейс поддерживает четыре уровня работы пользователей: помощников библиотекаря (Library Assistants), которые могут добавлять документы и метаданные в существующие коллекции и создавать новые коллекции, структура которых в точности повторяет уже существующие; библиотекарей (Librarians), которые могут, помимо перечисленного выше, разрабатывать дизайн новых коллекций, но которым недоступны специальные технические вещи (например, регулярные выражения); специалистов библиотечных систем (Library Systems Specialists), которым доступны все возможности по дизайну коллекций, но которые не могут выполнять действия по устранению возникающих проблем (например, интерпретировать ошибки, появляющиеся в связи с использованием программ, написанных на языке Perl), и эксперты (Experts), которые могут выполнять все действия без исключения.

Вы уже кратко сталкивались с режимами эксперта (Expert mode) в Задании 4 Главы 2.2, и специалиста библиотечной системы (Library System Specialist mode) при выполнении упражнений Главы 3.2 усовершенствованной версии коллекции Tudor. До настоящего времени, вы действовали, находясь в режиме работы библиотекаря (Librarian mode). Для выполнения следующего упражнения, мы перейдем в режим работы специалиста библиотечных систем (Library Systems Specialist).

Упражнение – этап 5: Улучшенная настройка

Для переключения режимов работы щелкните File→Preferences→Mode и выберите режим специалиста библиотечных систем (Library Systems Specialist). Обратите внимание, что в появившемся описании вам нужно сформулировать регулярные выражения для того, чтобы извлечь максимум из этого режима работы. Этим мы и займемся.

25. **UnknownPlug** является очень удобным и типичным для этого вида встроенным приложением. Оно ничего не знает о каких-либо заданных форматах, но его можно пристраивать к ним для удобства обработки особых типов документов, похожих на MIDI, с помощью расширений имен файлов и набора основных метаданных.

- Добавьте **UnknownPlug**;
- Активизируйте его поле *process_extension* и установите в нем значение *mid* для распознавания файлов с расширением *.mid;
- Установите *file_format* на “MIDI” и *mime_type* на “audio/midi”.

В этой коллекции все файлы MIDI собраны в одном заархивированном файле *beatles_midi.zip*. **ZIPPlug** (который уже есть в списке встроенных приложений, установленных по умолчанию). Он используется для разархивирования файлов и перемещения их вниз по списку встроенных приложений до тех пор, пока они не достигнут **UnknownPlug**.

26. Создайте (**build**) коллекцию и просмотрите (**preview**) ее. К сожалению, файлы MIDI не показываются как Audio под кнопкой browse. Это объясняется тем, что для этих файлов не были прописаны метаданные **dc.Format**.

- Вернувшись опять на панель **Enrich**, щелкните по файлу *beatles_midi.zip* и установите значение формата **dc.Format** на “Audio” — сделайте это, щелкнув по “Audio” в списке **All Previous Values**. Все файлы, извлеченные из архивного Zip-файла, также унаследуют эти установки.

27. Теперь мы снова вернемся к нашему браузеру *titles a-z* и почистим его. Цель — объединение вариантов заголовков путем разбора чужеродного текста. Например, нам нужно произвести такие же действия в отношении “ANTHOLOGY 1”, “ANTHOLOGY 2” и “ANTHOLOGY 3”, как и при группировке по назначению. Для того чтобы это сделать:

- Перейдите в список заголовков **AZCompactList**, находящихся под списком **Browsing Classifiers** на панели **Design**;

- Активизируйте `removesuffix` и установите его значение на:
`(?i) (\\s+\\d+) | (\\s+[[:punct:]]).*`

Создайте (**build**) коллекцию и просмотрите (**preview**) результат. Проследите за тем, сколько заголовков удалось объединить на той или иной книжной полке. Проверьте, насколько хорошо вы поняли смысл регулярных выражений, сделав попытку усовершенствовать эти слияния. Обратите внимание: `[:punct:]` используется для любого знака пунктуации. Пиктограммы к документам Word и PDF даны неверно, но мы это исправим в следующем операторе форматирования.

Вы выполняли предыдущее упражнение, находясь в режиме специалиста библиотечных систем (Librarian Systems Specialist), так как вы должны были использовать «регулярные выражения», а это то, что библиотекари, как правило, не обучены делать. Изучение вопросов использования регулярных выражений не является задачей этого курса.

Одним из очень мощных по возможностям направлений использования регулярных выражений, продемонстрированных в этом упражнении, было наведение порядка в браузере *titles a–z*. Возможно, более правильным было бы просто иметь аккуратно написанные метаданные заголовков. Метаданные, извлеченные из HTML-файлов, являются запутанными и неполными, и мы это как раз и увидели в браузере *titles a–z*. Корректировка метаданных для заголовков будет более простым, с точки зрения логики действий, но очень трудоемким решением. Вместо этого, мы бы предпочли использовать регулярные выражения в классификаторе *AZCompactList* для исправления метаданных заголовков. Это очень сложно понять и также довольно сложно сделать, но если вас это не испугает, то вы освоите быстрый способ корректировки извлеченных метаданных и избежите ввода огромного объема метаданных. Если (как мы ожидаем) вы прошли профессиональную подготовку как преподаватель или библиотекарь, а не компьютерщик, то вы вряд ли справитесь с этим сами, зато вы будете понимать весь процесс, и сможете сформулировать свою просьбу специалисту, которого позовете на помощь, если она вам понадобится.

Следующее упражнение связано с экспертной настройкой. Эта часть курса не является обязательной для изучения, и вы можете попытаться сделать это самостоятельно, если вы имеете к этому склонность и если у вас для этого есть время.

Дополнительное упражнение – этап 6: Экспертная настройка

28. Последним штрихом в создании нашей коллекции станет внесение в нее несколько декоративных деталей.

- Используя окно просмотра файлов Windows вне Greenstone, найдите папку

sample_files→beatles→advbeat_large

Скопируйте находящиеся там папки *images* и *macros* в корневую папку вашей коллекции. (Вы можете спокойно заменить существующую папку *images*: изображение, которое в ней хранится, также включено в папку, которую вы копируете.) В папке *images* содержатся некоторые полезные пиктограммы, а в папке *macros* содержатся описания макросов для названий этих пиктограмм. Для того чтобы увидеть эти описания, откройте в текстовом редакторе находящийся в папке с макросами файл *extra.dm*.

- Повторно отредактируйте свой оператор форматирования **Vlist**, чтобы он выглядел следующим образом

```
<td valign=top>
{If}{[numleafdocs],[link][icon][link]}
{If}{[dc.Format] eq 'Lyrics',[link]_iconlyrics_[link]}
{If}{[dc.Format] eq 'Discography',[link]_icondisc_[link]}
{If}{[dc.Format] eq 'Tablature',[link]_icontab_[link]}
{If}{[dc.Format] eq 'MARC',[link]_iconmarc_[link]}
{If}{[dc.Format] eq 'Images',[srclink][thumbicon][srclink]}
{If}{[dc.Format] eq 'Supplementary',[srclink][srcicon]
[/srclink]}
{If}{[dc.Format] eq 'Audio',[srclink]{If}{[FileFormat] eq
'MIDI',_iconmidi_,_iconmp3_[srclink]}
</td>
<td>
{If}{[numleafdocs],([numleafdocs])}
</td>
<td valign=top>
[highlight]
{Or}{[dc.Title],[Title],Untitled}
[/highlight]
</td>
```

Целиком это выражение находится в файле *format_tweaks→multi_icons.txt*.

Просмотрите (**preview**) свою коллекцию так же, как вы это делали ранее. Вы увидите, что теперь для дискографии, текстов песен, нот и метаданных MARC используются разные пиктограммы. Различаются даже аудиофайлы MP3 и MIDI. Если вы наведете курсор мыши поверх любой из этих пиктограмм, то появится подсказка, объясняющая на том языке, который в настоящее время используется в интерфейсе программы, какой тип файла представляет данная пиктограмма. Обратите внимание: *extra.dm* определяется только на английском или французском языках.

Теперь мы изменим изображение, которое используется как фоновое в данной коллекции, на этот раз больше внимания уделяя работе с файлами. Откройте папку *macros* в вашей коллекции и найдите в ней файл *extra.dm*. Щелкните по нему правой кнопкой мыши. Если вслед за этим будет предложено на выбор несколько программ, то выберите **WordPad** и откройте этот файл (**open it with**).

29. Содержание этого файла незначительно по объему, поскольку в нем кратко описывается, что нужно отменить из функций, установленных по умолчанию для этой коллекции. В WordPad, практически в начале текста, вы должны увидеть:

```
_httpiconchalk_ {_httpcimages_/beat_margin.gif}
_widthchalk_ {1800}
_heightchalk_ {68}.
```

Используйте прием копирования и вставки для того, чтобы изменить эти три строчки следующим образом:

```
# Original statements
#_httpiconchalk_ {_httpcimages_/beat_margin.gif}
#_widthchalk_ {1800}
#_heightchalk_ {68}
_httpiconchalk_ {_httpcimages_/tile.jpg}
_widthchalk_ {22}
_heightchalk_ {22}
```

Значок решетки (**#**) в начале строки указывает на то, что это — комментарий, поэтому Greenstone проигнорирует следующий за ней текст. Мы использовали это, чтобы прокомментировать три исходных выражения и заменить в них некоторые из модифицированных строк. Сохраните первоначальную версию на случай, если позднее нам потребуется восстановить исходные строки. Эти три строки относятся к изображению, которое используется в качестве основного фона коллекции. Новое изображение *title.gif* также было в той самой папке *images*, которую мы скопировали ранее.

31. Находясь в **WordPad**, сохраните *extra.dm*.
32. Просмотрите (**preview**) домашнюю страницу коллекции. Теперь в качестве фонового изображения используется совсем другой рисунок.

Редактируя макрофайлы, можно изменять и другие характеристики коллекции, например, колонтитулы, используемые на каждой странице, или стиль выделения, используемый при поиске терминов (определен цвет, способ написания слов (жирный шрифт) и т.д.).
33. При желании вы можете вернуться к последним изменениям, которые вы сделали, написав комментарий в трех добавленных вами новых строках (добавив **#**) или не комментируя исходные три строки (удалив символ **#**). Не забудьте сохранить этот файл. Для отмены всех пользовательских настроек, которые вы выполнили, просто удалите содержание в папках *macros* и *images*.
34. Теперь, для завершения работы нам нужно создать расширенную версию этой коллекции. Для этого:
 - Закройте текущую коллекцию.
 - Запустите новую коллекцию, названную *advbeat_large*.
 - Основой для этой новой коллекции сделайте *small_beatles*.
 - Скопируйте содержание *sample_files*→*beatles*→*advbeat_large*→*import* в только что созданную коллекцию. Поскольку в этом коллекционном наборе содержится намного больше файлов, то процесс их копирования также займет больше времени.
 - Создайте (**build**) коллекцию и посмотрите полученный результат. (Если вы хотите, чтобы эта коллекция была представлена отдельной пиктограммой, вам надо добавить такую пиктограмму на панели **Design**.)

Независимо от того, справились вы или нет с предыдущим дополнительным упражнением, которое мы сознательно постарались сделать как можно более сложным для выполнения, мы хотим, чтобы вы завершили его на мажорной ноте, проведя заключительное усовершенствование вашей коллекции, добавив несколько полезных инструментов поиска. В коллекции «Beatles» много художественных фотографий и альбомов, для которых мы зададим привлекательную и очень подходящую для просмотра таких вещей форму визуального просмотра, называемую *коллаж* (*collaging*). Изображения, выбираемые в коллекции одно за другим в случайном порядке, будут также

выводиться на экран. Через какое-то время они начнут постепенно исчезать и будут заменяться другими изображениями, появляющимися поверх первых, возможно затеняя предыдущее изображение при появлении. Это обеспечивает динамичный, притягивающий взгляд, а может быть даже и гипнотический эффект, похожий на заставку экрана компьютера, которая строится на использовании изображений, имеющихся в самой коллекции. Но этот эффект заметно отличается от простой заставки экрана, поскольку, если пользователь заинтересуется тем или иным изображением из числа предъявленных и щелкнет по нему мышью, на экран будет выведен документ, хранящийся в электронной коллекции и содержащий это изображение.

Упражнение – этап 7: Необычная настройка

35. Перейдите на панель **Design** и выберите раздел **Browsing Classifiers**. Раскройте меню **select classifier to add** и выберите **Collage**. Щелкните **<Add Classifier>**. Здесь не нужно проводить дополнительных настроек, поэтому просто щелкните **<OK>** в нижней строке меню.
36. Теперь смените панель на **Create**, постройте (**build**) и просмотрите (**preview**) коллекцию.

Мы уже включили интерактивный иерархический индекс фраз, с которым мы впервые столкнулись в Главе 3.2 (при описании усовершенствованной коллекции Tudor). Вам будет интересно использовать его для знакомства с текстами песен «Битлз».

Заключение

Если вам удалось добиться успеха в выполнении упражнений данной главы, то примите наши поздравления! А если вы дошли до конца дополнительного упражнения по выбору, посвященного экспертной настройке, то поздравляем вас вдвойне — вы идете по правильному пути становления продвинутого пользователя Greenstone. Даже если не все выполненные вами упражнения были в равной степени успешны, то все равно вы узнали очень многое о том, как создаются электронные библиотечные коллекции, содержащие гетерогенный материал.

Greenstone — это быстро развивающаяся система. Она далека от совершенства, и для нее постоянно разрабатывается что-то новое, поэтому в самом конце Главы 2.3 мы уделили столь значительное вни-

мание вопросам установки новых версий программного обеспечения и рекомендовали вам регулярно обновлять используемую вами версию программы, присоединившись к дискуссионной группе Greenstone (см. следующий раздел), где помещаются объявления обо всех новых выпусках, или проверяя сайт <http://www.greenstone.org> на предмет появления новых версий. Например, библиотечарский интерфейс является относительно недавней инновацией, и до его появления пользоваться Greenstone было намного сложнее.

Следовательно, если при выполнении упражнений у вас возникли трудности с пониманием технических деталей, не отчаивайтесь! Будут созданы новые программы, которые обязательно упростят вашу работу. Например, операторы форматирования очень трудны для понимания (вы узнаете о них более подробно в следующей Главе), поэтому обязательно появится интерактивный редактор для операторов форматирования, который сделает процесс форматирования материала в Greenstone намного проще, так как он будет работать как современный текстовый процессор. В последнем задании вы выполняли трудную работу, используя различные пиктограммы и цвета основного фона, для изменения вашей коллекции — но и это в итоге тоже будет делаться с помощью интерактивного графического редактора.

Между тем, лучшей стратегией создания продвинутых по техническому исполнению коллекций является подбор уже существующей коллекции с той структурой, которая вам нужна, или обращение за помощью к эксперту для построения таковой, а затем использование этой готовой структуры для создания вашей собственной новой коллекции в библиотечарском интерфейсе с помощью функции “Base this collection on”, использованной вами при выполнении самого первого упражнения Главы 3.2.

Глава 4.3. Расширение возможностей Greenstone

Все более углубляясь в материал данного курса, мы посвятили немало времени и сил для того чтобы научиться работать в библиотечарском интерфейсе. Как подсказывает само название, именно понятие *интерфейс* является ключевым для программы Greenstone, с его помощью *контролируются* все действия, которые мы выполняли в программе в рамках предложенных вам упражнений. В данном разделе мы познакомимся более детально с программной частью Greenstone, чтобы глубже проникнуть в суть ее работы.

Все упражнения, которые вы выполняли или еще будете выполнять, были специально подготовлены для вас авторами курса, однако когда вы начнете работать со своими документами и создавать собственные коллекции, не все из выполненного может пойти гладко. Вы вовлекаетесь теперь в такой творческий процесс, в котором вам придется принимать и сложные решения и преодолевать всевозможные препятствия. Но вы получите удовлетворение: всегда приятно видеть, как нечто, созданное твоими собственными руками, развивается и растет. Однако будет еще лучше, если, имея необходимые знания, вы сможете судить о том, что именно происходит внутри созданного вами продукта.

В продолжение этого детального ознакомления данная глава переведет вас из категории владельца автомобиля (который имеет лишь смутное представление о работе двигателя) в категорию механика. И вы, как настоящий механик, делающий заключение о том, что происходит с неработающей машиной на основе отдельных наблюдаемых вами симптомов, сможете объяснить, почему только что созданная вами коллекция работает не так, как вы ожидали, и что еще нужно проверить. Какие есть ограничения на доступ к той или иной папке, есть ли у вас все нужные файлы и — вообще — что именно находится в файле? Вы будете достаточно подготовлены, чтобы указать нужные части системы и вскрыть нестандартные компоненты. Также, как и при выполнении любого упражнения на обнаружение ошибки, вы всегда должны помнить о том, что проблемы часто возникают не из-за единичной причины, а из-за сочетания различных обстоятельств.

Главы 6 и 7 учебника «Как создать электронную библиотеку» целиком посвящены описанию системы Greenstone и используемого программного обеспечения. В этой Главе нашего пособия мы вместе проработаем наиболее важные части изучаемого материала.

Сначала мы разберемся в том, какова структура коллекций документов, доступ к которым вы получили при инсталляции Greenstone. Этот материал будет объяснен с точки зрения исходного программного обеспечения Greenstone, а не библиотечарского интерфейса. При работе с коллекцией все те команды, которые вы выбираете на панели дизайна библиотечарского интерфейса (Librarian Interface's design panel), записываются в файл конфигурации коллекции (*collection configuration file*). Этот файл сохраняется на диске и используется исходным программным обеспечением при активизации кнопки, которая дает команду на создание коллекции. Данное действие запускает процесс импортирования и создания документа, то есть, две фазы в работе исходной программы. При этом происходит обращение к файлу конфигурации для получения инструкций о том, какие структуры и указатели должны быть сформированы для данной целевой коллекции.

В процессе изучения примеров коллекций документов вы увидите исходное содержимое файлов конфигурации этих коллекций. Используя библиотечарский интерфейс, вы научитесь распознавать (читать) отдельные части файлов конфигурации. Операторы форматирования почти дословно передают суть включенных в них значений, и текст «об этой коллекции» “about this collection text” будет обозначаться именно этим сочетанием ключевых слов:

```
collectionmeta collectionextra "..."
```

Задание 1

Проработайте Раздел 6 Урока 4 учебного модуля на компакт-диске IMARK «Примеры коллекций документов Greenstone». Затем сделайте следующее:

- Определите какие (если вообще таковые имеются) коллекции документов работают с исходными документами, представленными в интересующем вас формате.
- Найдите любую коллекцию документов, имеющую характеристики, которые вы хотели бы задать в своей собственной электронной библиотечной коллекции.
- Сравните два списка, которые вы видели выше. Насколько они совпадают?

Далее мы прочитаем первую из двух глав, посвященных Greenstone, в книге «Как создать электронную библиотеку» (Глава 6). Первая часть главы предназначена главным образом для повторения. Мы пропустим следующую часть, в которой приводится описание «Коллекционера» (“The Collector”), подсистемы Greenstone, которая собирает детали о самой коллекции и о документах, которые должны быть в нее включены. Используя для этого последовательность веб-форм, Collector создает или изменяет коллекцию. Collector работает под управлением одной из версий приложения Java, использующейся в библиотекарском интерфейсе, и похожей на него, за исключением того, что он еще и загружает файлы, предназначенные для создания библиотеки на сервер Greenstone. Более детальную информацию об этом можно найти в разделе Greenstone FAQ (Часто задаваемые вопросы и ответы).

Мы выбрали самое главное в Разделе 6.4, которое представляет два основных этапа в создании любой коллекции и структуры файлов и директорий Greenstone. Раздел 6.5 в деталях описывает архивный формат Greenstone, промежуточный формат на основе XML, использующийся для хранения документов и метаданных в промежутке между этапами импортирования и создания библиотеки. Раздел 6.6 обобщает информацию о структуре файлов конфигурации коллекции, а Раздел 6.7 раскрывает технические подробности по встроенным приложениям, классификаторам и операторам форматирования.

Задание 2

Прочитайте Разделы 6.1 и с 6.4 по 6.7 учебника «Как создать электронную библиотеку». Для просмотра всего этого материала в Greenstone используйте имеющуюся у вас программу управления файлами (C:\Program Files\greenstone\ в стандартной инсталляции в Windows) и «попытайтесь открыть» некоторых из этих файлов. Обычно каждый файл, находящийся в данной файловой области, должен читаться или копироваться, а открыть его вы можете любым текстовым редактором, например, таким как NotePad, WordPad или Emacs. Некоторые файлы являются бинарными, и вы также можете открыть их в текстовом редакторе, но работать с ними будет невозможно. Если вы случайно открыли такой файл или какой-либо другой, то завершите работу в текстовом редакторе, не сохраняя файл для предотвращения изменения системы.

Теперь ответьте на следующие вопросы:

- Объясните в общих чертах, что происходит на этих двух этапах создания коллекции.
- Опишите подробно ваши действия на каждом из этапов создания коллекции.
- Где в файловой структуре располагаются коллекции?
- Какие директории создаются при появлении новой коллекции? Опишите назначение каждой из них.
- Объясните, что означает понятие «устойчивого идентификатора документа».
- Что является конечным результатом импортирования файлов?
- Что по умолчанию используется в Greenstone в качестве инструмента индексирования?
- Определите с помощью рисунка 6.10, что является метаданными для заголовка верхнего уровня этого документа? Собранные вручную содержание документа, показанное на рисунке 6.10 (Раздел 1, Раздел 1.1. и т.д.), демонстрирует заголовки метаданных, связанные с каждым из разделов, а также с номером раздела.
- Объясните назначение встроенных приложений Greenstone.
- Объясните назначение классификаторов.
- Объясните назначение операторов форматирования.

В проектах, связанных с открытыми системами, все очень быстро меняется. В 2003 году, когда создавался учебник, библиотечарский интерфейс был ни чем иным, как первоначальным вариантом прототипа программы. В конце Главы 6 вы, познакомились с кратким комментарием по созданию графических коллекций (Раздел 6.8.). Инструментальное программное средство, о котором там шла речь, превратилось в библиотечарский интерфейс, который используется в данном курсе. Интуитивно понять и освоить этот интерфейс оказалось так легко, что именно с ним в представлении многих пользователей при первом упоминании ассоциируется Greenstone.

Работу в Greenstone можно разделить на две логические части: создание коллекции и ее использование. Рекомендуемый вам материал для дополнительного чтения поможет проанализировать процесс создания коллекции. Однако, понимание этих двух основных компонентов и того, какая роль отводится им в процессе формирования коллекции, еще не даст вам полной картины. Проще говоря, операторы форматирования – это элементы работающей системы, и поэтому все

изменения, которым вы их подвергаете, сразу же проявляют себя, но они столь часто используются, что им посвящена отдельная Глава 6. Мы кратко на этом остановимся. В данной главе даны общие вопросы презентации коллекций такие, как включение в коллекцию специальных пиктограмм, создающихся в процессе работы с коллекцией и управляющихся через библиотечарский интерфейс. Вся информация об использовании и создании коллекции хранится в одном и том же файле конфигурации, и многие части системы игнорируют объекты, которые не содержатся в этом файле. Основа, используемая для создания коллекций Greenstone, написана на языке программирования Perl, а управляющая система на C++.

В Главе 7 учебника «Как создать электронную библиотеку» внимание сосредоточено на работе программы Greenstone. Она предназначена для тех, кто хочет выйти за пределы библиотечарского интерфейса и напрямую поработать с элементами файловой структуры, возможно, даже внести изменения в код источника и прописать его заново, чтобы добиться радикальных изменений в этом программном средстве. Мы узнаем, как это можно сделать в следующем Задании.

С помощью операторов форматирования при желании вы сможете выполнить лишь незначительную отладку коллекции. Это сравнимо с тем, как можно поменять цвет машины или усилить двигатель. Техническая информация, представленная в Главе 7, вооружит вас знаниями для коренной модификации двигателя и изменения вида машины в целом. Владая определенными приемами, вы можете сделать из своей машины трактор, настолько это мощный по возможностям инструмент. Электронная библиотека Kids, описанная в Главе 7, относится к этому уровню программ. Она настолько расширяет возможности Greenstone, что включает персональные странички учащихся, которые впервые начали с ней работу, специальный модуль для учителей, оснащенный дополнительными инструментами, рабочее место для студентов, где они пишут и публикуют свои рассказы, а также доска объявлений для проведения общих дискуссий. Коллекции управляются с двух разных сайтов, но без сомнения, воспринимаются пользователями как единый и унифицированный ресурс.

Эта глава объясняет отношения между двумя основными компонентами системы поддержки исполнения программ: клиентом и сервером. Работа системы обеспечивается протоколом, который соединяет их в единое целое и обеспечивает примерно такое их функционирование, которое вы видели, знакомясь с электронной библиотекой Kids. Разделение на клиента и сервер позволяет перенести вопросы пре-

зентации данных на клиента, а вопросы работы с контентом коллекции, доступа к базе данных и индексирования файлов — решать на уровне серверной программы. Макрофайлы обеспечивают механизм представления данных, который независим от используемого языка. В материалах главы также описывается, что именно хранится в файлах базы данных, как работает механизм «действия», что называется журналами регистрации и какова их конфигурация на уровне сайта в целом.

Задание 3

Прочитайте следующие выдержки из Главы 7 учебника «Как создать электронную библиотеку»: введение, Разделы 7.1, 7.2 и 7.4. Раздел 7.3 (по приемам работы) не является обязательным для чтения и оставлен нами как дополнительное задание.

- Объясните, что такое нулевой протокол в Greenstone.
- Напишите макро определения для следующего:
a macro called `_poweredby_` that displays the message “powered by Greenstone” by default, “actionn par Greenstone” if the user interface language is French, and “accionado por Greenstone” if Spanish.
- Какой тип информации хранится в информационной базе данных коллекции?
- Какая информация хранится в журнале регистрации всего сайта?
- Где искать информацию о текущих ошибках в программе Greenstone?

Для многих пользователей операторы форматирования — это самое неприятное в Greenstone. Они содержат небольшие выдержки из исходного кода HTML, совмещенные со специальными, довольно запутанными командами самого Greenstone. Если хоть что-нибудь как, например, единичный специальный символ `/` в закрывающем теге HTML или фигурная скобка в выражении `{If}` — будет потеряна или поставлена не на том месте, извлеченный отрывок может быть показан неверно, а в исключительных случаях и вообще не показан. Приведем несколько полезных операторов форматирования, которые помогут вам самостоятельно сформировать некоторые из наиболее часто употребляемых конструкций:

Общий контроль за документами в VList:

```
<td>[link] [icon] [/link]</td><td><i>[Title]</i></td>
```

Добавление ссылки на первоисточник в документах VList:

```
<td>[srclink] [srcicon] [/srclink]</td>
```

```
<td>[link] [icon] [/link]</td><td><i>[Title]</i></td>
```

Добавление рисунка эскиза в VList:

```
<td>[link]<img  
    src='_httpcollimg_/[assocfilepath]/[Thumb]'  
    width=[ThumbWidth] height=[ThumbHeight]>  
[/link]</td>  
<td valign=middle><i>[Title]</i></td>
```

Добавление группы информации для иерархического классификатора:

```
<td>[link] [icon] [/link]</td>  
<td>{If}{[numleafdocs],[Title] <i>([numleafdocs])  
</i>,[Title]}  
</td>
```

Со всеми этими возможностями вы в той или иной мере участвуете в выполнении практических заданий по данному курсу. Здесь мы просто сгруппировали их и для удобства представили вам в одном списке.

Сообщество пользователей Greenstone

Из Главы 1.3 вы узнали о том, что Greenstone является программным продуктом с открытым кодом и что любой желающий может увидеть исходный код Greenstone. Также, как и в ходе многих подобных проектов, связанных с программными продуктами с открытым кодом, вокруг Greenstone сформировалось международное сообщество разработчиков, которые вносят свой вклад в создание новых возможностей и устранение неполадок. Также существует и сообщество пользователей Greenstone, которые помогают друг другу в решении возникающих проблем и вносят новые предложения по использованию этой программы. Такие сообщества являются прекрасным ресурсом при поиске ответов на вопросы, которые у вас могут возникнуть при работе с Greenstone. И все же мы рекомендуем вам лучше познакомиться с теми источниками информации, описание которых приводится ниже.

Система помощи программы. В библиотечарском интерфейсе (The Librarian Interface) имеется встроенная система помощи, содержащая много полезной информации, доступ к которой можно получить в меню Help. Также, когда вы задерживаете курсор мыши над большинством кнопок и текстовых полей рабочего окна программы, вы получаете тек-

стовые сообщения, подсказки, которые позволяют вам лучше сориентироваться в том, что происходит в данный момент работы.

Руководства. На компакт-диске с курсом размещено несколько различных руководств по Greenstone.

- *Руководство по установке (Installer's Guide)*

Содержит детальное описание процесса установки Greenstone. Обратите внимание на то, что составители инструкций в этом руководстве исходят из того, что Greenstone устанавливается на компьютер пользователя непосредственно с распространяемого в рамках курса CD-ROM. Поэтому очевидно, что при установке путем выгрузки программы из Веба эти инструкции нужно дополнительно адаптировать.

- *Руководство пользователя (User's Guide)*

Общие сведения об использовании коллекций Greenstone, библиотечарского интерфейса и функций администрирования.

- *Руководство разработчика (Developer's Guide)*

Более детальное описание процесса создания коллекции Greenstone, включая создание коллекций с помощью командной строки или команды в DOS. Также включает описание структуры управления файлами Greenstone.

- *От бумаги к коллекции (From Paper to Collection)*

Этот документ описывает весь процесс создания электронной библиотечной коллекции из печатных документов на бумаге, включая сканирование и процесс OCR (оптического распознавания символов).

Прочая документация

- *Внутри коллекций Greenstone (Inside Greenstone Collections)*

Одним из самых непростых вопросов использования Greenstone является файл конфигурации вашей коллекции. Для лучшего понимания того, как именно это надо делать, в данный документ включены объяснения по файлам конфигурации нескольких реально существующих коллекций Greenstone, а также дается пример по настройке интерфейса Greenstone под конкретного пользователя.

- *Примеры документов коллекций (Documented Example Collections)*

Эта подборка включает одиннадцать документов из коллекций Greenstone, которые содержат странички «about», описывающие как именно эти коллекции были созданы. Они представлены на английском, французском, испанском и русском и являются исключительным ресурсом для тех, кто хочет понять, как можно создать наиболее распространенные виды коллекций. Также, если вы воспользуетесь библиотечарским интерфейсом для создания новой коллекции на основе какой-либо из представленных в этой подборке коллекций, то вы унаследуете стиль и особенности форматирования этой коллекции.

Учебник, который предлагается в качестве основного источника информации по курсу, также содержит массу полезных сведений о Greenstone:

«*How to Build a Digital Library*», by Ian H. Witten and David Bainbridge. Morgan Kaufmann, San Francisco, California, 2003

Наиболее часто задаваемые вопросы и ответы на них (FAQ – Frequently Asked Questions). Разработчики Greenstone составили список вопросов, которые часто задают пользователи этой программы, а заодно они составили и список своих ответов на эти вопросы! Увидеть все вопросы и ответы вы можете на сайте <http://greenstone.org> and click the tab marked faq.

Архивы списков рассылки. На протяжении нескольких лет члены сообщества пользователей Greenstone вели список рассылки для обсуждения проблем и совместного поиска решений. Архивы, содержащие эту переписку, являются хорошим источником для получения ответов на те вопросы, которые до вас уже кто-то задавал (но не так часто, чтобы их включили в список FAQ). Архивы, организованные также в виде коллекций Greenstone, доступны в сети по адресу:

<http://www.nzdl.org/gsarchives>

Списки рассылки. Если вам так и не удалось получить ответ на вопрос, попробуйте послать его по спискам рассылки (mailing lists). Существует два таких списка – один для пользователей, другой для разработчиков (рекомендуем сначала изучить первый из них):

- *Список для пользователей Greenstone (Greenstone User's List)*

Этот список используется для общих дискуссий по Greenstone. Для того, чтобы послать сообщение на этот список, вы должны направить его по адресу

greenstone-users@list.scms.waikato.ac.nz.

- *Список для разработчиков Greenstone (Greenstone Developer's List)*

Этот список предназначен для обсуждения технических подробностей теми людьми, кто разрабатывает или вносит изменения в Greenstone. Для того, чтобы послать сообщение на этот список, вы должны направить его по адресу `greenstone-devel@list.scms.waikato.ac.nz`. Обратите внимание: прежде, чем опубликовать свое письмо в списке рассылки, вам надо оформить подписку на этот список.

Вступление в члены сообщества. Сообщество Greenstone с радостью принимает всех вновь прибывших, независимо от того пользователи они или разработчики, и поддерживает их стремление поделиться своими знаниями и помочь другим людям. Люди приносят свой вклад в зависимости от имеющихся у них опыта и интересов. Кто-то помогает отвечать на вопросы пользователей, другие могут дать совет по более сложной проблеме, а кто-то пишет новые программные дополнения (расширения).

Перевод на другие языки. Greenstone поддерживает интерфейсы и документы на разных языках. Команда разработчиков вполне компетентна в вопросах программирования, но вот в различных языках народов мира они разбираются не очень хорошо. И здесь они полагаются на других членов сообщества, которые обеспечивают перевод интерфейса на другие языки. Помощь в переводе Greenstone на другие языки не требует много времени и технических знаний, но является бесценным вкладом в развитие этого сообщества.

Если вы захотите помочь в переводе интерфейса на язык, который пока не поддерживается этой программой, пожалуйста, направьте свое письмо команде Greenstone по адресу `greenstone@cs.waikato.ac.nz`.

Коммерческие приложения. Если вам нужна специальная помощь по работе с Greenstone, включая ручную настройку коллекции, вы можете обратиться в службу коммерческой поддержки по адресу:

- DL Consulting, Hamilton, New Zealand
<http://www.dlconsulting.co.nz/>
contact@dlconsulting.co.nz

Заключение

Хотя основной темой данного Модуля были мультимедийные коллекции, эта глава несколько выходит за рамки этой темы. Мультимедийные коллекции расширяют границы программного обеспечения и заставляют вас глубже погрузиться в тему и разобраться в устройстве и функционировании самой системы.

В данной Главе мы познакомились с тем, что лежит в основе Greenstone, и это оказалось для нас довольно сложным делом. Разработчики прилагают немалые усилия, чтобы сделать библиотечарский интерфейс простым и интуитивно понятным в использовании, а также довести до нужного уровня абстракции некоторые сложные основополагающие детали. Однако, несмотря на все это, Greenstone является сложной системой. Как и в случае с подобными системами в реальном мире — автомобилями, самолетами, электростанциями — для решения возникающих проблем от пользователя требуется хотя бы минимальное знание предмета. Это знание позволяет вам подняться на уровень выше в достижении ваших целей. Эта глава дает вам некоторые знания, хоть и в небольшом объеме, но мы надеемся, что вы получите такое же удовлетворение от работы с Greenstone, как если бы вы прокатились на хорошо отрегулированном автомобиле по современному прямому шоссе.

Одним из наиболее важных моментов, характеризующих любое программное обеспечение с открытым кодом, является наличие пользователей этой программы. Существование сообщества пользователей, которые помогают и поддерживают друг друга — это одна из наиболее сильных черт открытых систем, и именно этим они больше всего отличаются от коммерческого программного обеспечения. Международное сообщество является бесценным ресурсом: владеющим информацией, опытом, отзывчивым и доступным круглосуточно, и, помимо всего, бесплатным! Эта глава познакомила вас с сообществом Greenstone и рассказала о том, как стать его членом. Мы надеемся, что вы вступите в его ряды, будете учиться, получать удовольствие от своей работы и вносить посильный вклад в общее дело.

Не каждый готов стать механиком. И мы признаем, что если технически подкованные люди, как и разработчики Greenstone, пытаются

вникнуть в детали, то наш читатель, скорее всего, преподаватель или библиотекарь — скажет, что уровень изложения деталей в данном разделе является избыточным и слишком сложным для понимания. Если это так, то просто успокойте себя тем, что самое худшее для вас уже позади. Больше мы не будем так углубляться в дебри Greenstone. В следующем модуле мы переходим к другим стандартам, различным системам и только что появившимся приложениям.

Модуль 4. Мультимедийные электронные библиотеки

МОДУЛЬ 5

ОТКРЫТЫЕ СТАНДАРТЫ И ПРИМЕРЫ ИЗ ПРАКТИКИ

Глава 5.1. Стандарты
метаданных: METS, MODS,
RDF

Глава 5.2. Институционные
репозитории
и интероперабельность

Глава 5.3. Примеры
из практики образовательных
электронных библиотек



Цель

Научиться применять полученные на этом курсе знания и умения в реальных условиях современных электронных библиотек, институциональных репозиториях и образовательных электронных библиотек, использующих современные стандарты метаданных.

Задачи

После завершения работы по Модулю 5, вы сможете:

- отличать понятия местонахождение (location) и имя (name) и понимать, как это различие сказывается при поиске документов в Вебе;
- находить пространства имен в документе XML;
- распознавать, когда в документе XML используется Xlink, и объяснить его назначение;
- различать схемы XML и XML DTD;
- рассказать о формате Open eBook;
- объяснять назначение RDF, MODS и METS;
- знать различные компоненты, описанные в спецификации METS, и объяснить их значение;
- описывать функционал институционального репозитория;
- назвать и описать протокол, используемый при передаче метаданных в институциональном репозитории;
- объяснять целевое назначение протокола Z39.50 и в общих чертах рассказать о том, как он используется для передачи информации;
- рассказать о принципиальном отличии DSpace и Greenstone;
- встроить в коллекцию Greenstone документы, экспортированные из институционального репозитория DSpace;
- объединять метаданные и документы, представленные на нескольких серверах OAI, в единую коллекцию Greenstone;
- объяснять, как можно использовать Greenstone для распространения коллекции в рамках OAI;
- описывать сервисы, предоставляемые пользователям в мультимедийном репозитории Merlot;
- использовать Merlot для поиска других образовательных репозиторий;

- описывать возможность создания коллекций газетных иллюстраций в Greenstone;
- использовать Greenstone для создания коллекции отсканированных изображений;
- принимать решение по проведению анализа поведения пользователей в электронной библиотеке на основе протоколов выполняемых ими действий;
- рассказать о том, что случилось с проектом BBC Domesday, и сравнить его по результатам с одноименным проектом Вильгельма Завоевателя (William the Conqueror).

Введение к Модулю 5

Этот заключительный модуль позволит вам более широко взглянуть на проблему управления электронными библиотеками и организацию взаимодействия между ними. Мы познакомимся с системами институциональных репозиториях и существующими образовательными электронными библиотеками, а также другими коллекциями, имеющими большой потенциал для использования в образовании.

Модуль состоит из трех глав. В первой описаны стандарты для представления метаданных, то есть, документы, содержащие различные компоненты, связанные друг с другом гиперссылками, а также целые коллекции. Для работы с этими стандартами нам, прежде всего, потребуется познакомиться с XML, который лежит в их основе. Вторая глава исследует требования, предъявляемые к институциональным репозиториям, и приводит описания протоколов для обмена данными между ними. В третьей мы увидим некоторые действующие образовательные электронные библиотеки и узнаем о том, какие услуги они предлагают пользователям. Мы также обсудим методы изучения их поведения в электронной библиотеке.

В процессе работы над модулем мы продолжим наше знакомство с Greenstone; мы узнаем, можно ли использовать эту программу в качестве инфраструктуры для различных видов электронных библиотечных систем.

Модуль 5. Материалы для чтения

Глава 5.1. Стандарты метаданных: METS, MODS, RDF

Материал для чтения 1. Учебник «Как создать электронную библиотеку»: все материалы из Главы 8, за исключением Раздела 8.4.

Назначение. Обеспечить более глубокое понимание XML и связанных с ним стандартов.

Материал для чтения 2. CD-ROM «Электронные библиотеки в образовании»: Хрестоматия, перевод статьи Ричарда Гартнера «METS: Стандарт кодирования и передачи метаданных».

Gartner, R. (2002) METS: Metadata Encoding and Transmission Standard

Oxford University Library Services, JISC, 2002.

http://www.jisc.ac.uk/uploaded_documents/tsw_02-05.pdf

CD-ROM «Электронные библиотеки в образовании». Хрестоматия: перевод статьи Ричарда Гартнера «MODS: схема описания объекта метаданных»

Gartner, R. (2003) MODS: Metadata Object Description Schema

Oxford University Services, JISC, 2003.

http://www.jisc.ac.uk/uploaded_documents/tsw_03-06.pdf

Назначение. Получить общее представление о стандартах METS и MODS.

Глава 5.2. Институциональные репозитории и интероперабельность

Материал для чтения 1. Учебник «Как создать электронную библиотеку»: избранные части Главы 8.

Назначение. Узнать больше о том, как с помощью коммуникационных протоколов осуществляется интероперабельность между электронными библиотеками.

Материал для чтения 2. CD-ROM «Электронные библиотеки в образовании»: Хрестоматия, перевод статьи Мириам А. Дрейк «Скрытые сокровища институциональных репозиториях»

Drake, M.A. (2004) "Institutional Repositories: Hidden Treasures", Searcher, May 2004, Vol. 12, No. 5.

<http://www.infoday.com/searcher/may04/drake.shtml>

CD-ROM «Электронные библиотеки в образовании»: Хрестоматия, перевод статьи Уильяма Никсона «Проект DAEDALUS: предварительный опыт работы с программными приложениями EPrints и DSpace в Университете Глазго»

Nixon, W.J. (2003) "DAEDALUS: Initial experiences with EPrints and DSpace at the University of Glasgow,"

Ariadne, October 2003, Vol. 37.

<http://www.ariadne.ac.uk/issue37/nixon/>

Назначение. Познакомиться с институциональным репозиторием в целом, а затем выполнить проект по созданию примера одного из таких хранилищ, сосредоточив внимание на отдельных вопросах.

Материал для чтения 3. CD-ROM «Электронные библиотеки в образовании»: Хрестоматия, перевод статьи Иена Виттена, Дэвида Бейнбриджа, Роберта Тэнсли, Чи-Ю Хуанга и Кэтрин Дон «StoneD: мост между Гринстоун и DSpace»,

Ian H. Witten, David Bainbridge, Robert Tansley, Chi-Yu Huang, and Katherine J. Don *StoneD A Bridge between Greenstone and DSpace*, Working Paper 2005/02, Department of Computer Science, University of Waikato, New Zealand.

Назначение. Познакомиться с программным средством создания институционального репозитория на основе открытого кода – DSpace, определить черты сходства и различия этой программы и программы Greenstone и понять, как они могут использоваться вместе.

Глава 5.3. Примеры из практики образовательных электронных библиотек

Материал для чтения 1. CD-ROM «Электронные библиотеки в образовании»: Хрестоматия, перевод статьи Марка Эпперли, Те Така Кигана, Сэлли Джо Каннинггема и Иена Виттена «Доступ к газетам на языке маори через Интернет»,

Delivering the Maori-Language Newspapers on the Internet, Mark Apperley, Te Taka Keegan, Sally Jo Cunningham and Ian H. Witten In *Rere atu, taku manu! Discovering history, language & politics in the Maori-language newspapers*, edited by Jennifer Curnow, Ngapare Hopa & Jane McRae. Auckland University Press, 2002, 211-232

Назначение. В этой публикации рассказывается история создания электронной библиотеки *Niupera*.

Материал для чтения 2. CD-ROM «Электронные библиотеки в образовании»: Хрестоматия, перевод статьи Джехада Наджара, Стефана Терниера и Эрика Дюваля «Поведение пользователей при работе с хранилищами учебных объектов: эмпирический анализ»,

Najjar, J. Ternier, S. and Duval, E. (2004) User Behaviour in Learning Objects Repositories: an Empirical Analysis, Proceedings of the ED-MEDIA 2004 World Conference on Educational Multimedia, Hypermedia and Telecommunications, edited by L. Cantoni and C. McLoughlin, pp. 4373-4379, 2004,
<http://www.cs.kuleuven.ac.be/~najjar/papers/edmedia2004.pdf>

Назначение. В статье описано как дизайн образовательной электронной библиотеки определяет характер работы пользователей.

Материал для чтения 3. CD-ROM «Электронные библиотеки в образовании»: Хрестоматия, перевод статьи Дугласа Брауна «Потерянный в киберпространстве»,

Brown, D.(2003) *Lost in Cyberspace: The BBC Domesday Project and the Challenge of Digital Preservation*, June 2003, Cambridge Scientific Abstracts.

<http://www.csa.com/hottopics/cyber/oview.html>

Назначение. Увлекательная история, которая подчеркивает важность и одновременно сложность сохранения цифровой информации.

Глава 5.1. Стандарты метаданных: METS, MODS, RDF

Стандарты METS и MODS, которые мы будем изучать в данной Главе, являются современными стандартами метаданных, базирующимися на XML и управляемыми Библиотекой Конгресса США. Они были разработаны специально для использования в электронных библиотеках. METS расшифровывается как Metadata Encoding and Transmission Standard, а MODS – Metadata Object Description Schema. Также мы познакомимся со стандартом RDF, или Resource Description Framework, который был разработан для управления интероперабельностью метаданных, особенно в гипертекстовой среде Веба. Все эти три стандарта имеют прямое отношение к электронным библиотекам образовательных ресурсов.

Для того, чтобы подготовиться к использованию этих стандартов, нам нужно более детально познакомиться с XML и подобными технологиями. Это и является целью первого Задания.

Задание 1

Прочитайте Главу 8 учебника до Части 8.3 включительно и ответьте на следующие вопросы:

- Расшифруйте акронимы URL, URN и URI?
- Какое отношение они имеют друг к другу?
- Каково назначение пространства имен в XML?
- Объясните разницу между глобальными и ограниченными пространствами имен. Приведите пример небольшого отрывка XML, в котором используются теги обоих видов – и из глобальных, и из ограниченных пространств имен.
- Сколько пространств имен могут использоваться в документах XML?
- Назовите черты сходства и отличия по установлению ссылок между документами в XLink и HTML. (Возможно, для этого вам будет удобно нарисовать схемы, иллюстрирующие разные формы связывания документов.)
- Назовите черты сходства и отличия XML и DTD по работе со структурой и типом информации.
- Постройте схему RDF (с линиями, похожими на те, что изображены на рис. 8.6 Части 8.2 учебника), которая опишет данное учебное пособие.
- Дополнительное задание. Используя текстовый редактор или редактор XML опишите ту же самую информацию в XML

Теперь пришло время представить MODS и METS. Вспомните, что мы говорили в самом начале Главы 3.1 о различиях между встроенными метаданными, задача которых помочь читателям в навигации по документам, и внешними метаданными, которые являются описательной информацией о документах. Стандарт MODS является стандартом для внешних метаданных, разработанных по тем же принципам, как и те метаданные, с которыми вы уже работали. METS подходит для всех форм метаданных — и встроенных, и внешних, и может рассматриваться как некоторая рамочная основа, к которой постоянно привязывается содержание документа, его структура, а также традиционные внешние метаданные.

По многим параметрам стандарт MODS может быть представлен как современная форма библиотечных метаданных MARC, выраженная в XML. На самом деле, MARC Standards Office использовался как инструмент для определения MODS. Метаданные MARC можно довольно легко конвертировать в MODS. И все же между ними есть определенная разница: их появление разделяет более 30 лет, поэтому нет ничего удивительного в том, что более новый стандарт использует некоторые идеи из более старого. Существует также и отдельный стандарт MARCXML, который в точности совпадает со стандартом MARC, но разработан под XML. Стандарт MARCXML также администрируется офисом MARC Standards Office, находящимся в Библиотеке Конгресса США.

В Задании 2 Главы 4.3 мы познакомились с архивным форматом Greenstone, основанным на XML схеме, который используется для хранения документов и метаданных в промежутке между этапами импортирования и создания библиотеки. Используя такие теги, как `<Section>`, `<Content>` и `<Metadata name="Title">`, он определяет точный, совместимый с XML, синтаксис, формирующий иерархическую структуру сложно устроенного документа и сопровождающих его метаданных на любом из уровней иерархии.

METS выполняет роль схожую с ролью архивного формата Greenstone, но он был создан относительно недавно и поэтому является более универсальным. Наиболее важным является то, что он позволяет описывать параллельные иерархические структуры. Например, документ может состоять из иерархии глав, разделов, подразделов и параграфов, а по вторичной независимой иерархии — из верхних колонтитулов, основного текста и нижних колонтитулов. Это невозможно сделать в архивном формате Greenstone, поскольку в нем есть ограничение на проведение только одной структурной иерархической декомпозиции. В дополнение отметим, что любой стандарт метаданных, который создан на основе XML, может быть встроен в METS, включая Dublin Core, XMLMARC и (как мы еще увидим) MODS.

Задание 2

Прочитайте две статьи Ричарда Гартнера «METS: стандарт кодирования и передачи метаданных» и «MODS: схема описания объекта метаданных». Теперь ответьте на следующие вопросы.

- Какие существуют разделы самого верхнего уровня в METS? Напишите по одному краткому параграфу, объясняющему их назначение.
- Какие двойные требования пытается согласовать друг с другом стандарт MODS?
- Сколько элементов верхнего уровня и подчиненных элементов имеет стандарт MODS?
- Приведите выдержку из формата XML, которая показывает, как можно использовать вложенные элементы MODS внутри элементов более высокого уровня для оценки представленных данных.
- Какие из элементов верхнего уровня являются обязательными?
- Перечислите пять элементов верхнего уровня, и напишите по одному параграфу, представив краткую характеристику каждому из них.

Greenstone может импортировать и экспортировать документы METS. Уточним это положение. Принимая во внимание то, что METS – это не столько формат, сколько структура описания документов и метаданных, он может импортировать и экспортировать документы, которые подпадают под определенный профиль формата METS. Профиль METS описывает класс документов METS в таких деталях, без которых не могут работать авторы этих документов и программисты, поскольку они включают все необходимые данные, как по созданию, так и по последующей обработке документов.*

Документы METS можно импортировать в Greenstone с помощью встроенного приложения (METS plug-in), в то время как в обратном направлении любая из коллекций Greenstone может быть экспортирована в METS за счет использования команды *Export* из меню *File* библиотечарского интерфейса.

Greenstone может также использовать стандарт METS внутри системы как альтернативу более старому архивному формату Greenstone (Greenstone Archive Format). Это делается путем установления в METS специальной функции *saveas*, которая включает процесс импортирования в Greenstone, о котором вы узнали в Главе 4.3 (Задание 2). Эта функция дает Greenstone команду конвертировать документы, обработанные

* Более подробную информацию по профилям METS можно найти по адресу http://www.loc.gov/standards/mets/profile_docs/METS.profile.requirements.rtf

на этапе импортирования, в формат METS. Файлы, как и ранее, помещаются в папке *archives*. На этапе создания коллекции приложение METSPlug обнаруживает и обрабатывает эти файлы (которые обязательно включают ту же самую информацию, что и архивные файлы Greenstone), а затем создает точно такую же коллекцию, как и предыдущая.

Дополнительное упражнение: Экспорт в формате METS

1. В GLI откройте коллекцию Tudor.
Для замены METSPlug на GAPlug вам нужно находиться в режиме эксперта (Expert).
2. Откройте *File→Preferences→Mode* и переключите режим работы на Expert.
3. Перейдите на панель **Design** и выберите **Document Plugins**. Удалите **GAPlug** из списка приложений и добавьте **METSPLug**.
4. Теперь перейдите на панель **Create**, определите параметры процесса импортирования и установите *saveas* для *METS*. Если вы не работаете в режиме *Expert*, функция импортирования будет вам недоступна.
5. Перестройте коллекцию.
6. В программе-проводнике Windows найдите папку *archives*, в которой находится коллекция Tudor. Для каждого документа в этой коллекции Greenstone создал по два файла: *docmets.xml*, содержащий основное описание METS, и *doctxt.xml*, вспомогательный файл. (Обратите внимание: пока вы не подключитесь к Интернету, вы не сможете посмотреть *doctxt.xml* в используемом вами веб-браузере, поскольку он является удаленным ресурсом.) В зависимости от формы исходных документов они могут сопровождаться дополнительными файлами, например, такими как файлы графических изображений, используемые для оформления веб-страницы. Одной из особенностей MET является его способность связывать информацию во внешних XML файлах. Greenstone использует эту функцию для связывания содержания документа, который находится во внешнем XML файле *doctxt.xml* с его иерархической структурой, которая описана в основном файле METS *docmets.xml*.

Заключение

Если вы будете работать с электронными библиотеками в ближайшие несколько лет (или десятилетий), то поддержка стандартов станет вашей постоянной необходимостью. Есть старая поговорка о том, что «самое замечательное в стандартах — это то, что их существует такое огромное количество, что всегда есть из чего выбрать», и мы могли бы дополнить ее: «и к тому же, они все время меняются». Ни один из стандартов, описанных в этой главе, десять лет назад еще даже не существовал (включая сам XML, техническое описание которого было впервые опубликовано в феврале 1998 года), а все, существующие сегодня, по прошествии еще десяти лет покажутся устаревшими. В учебном пособии по такому же точно курсу, опубликованному в 2003, не были упомянуты ни стандарт METS, ни MODS.

По существующей тенденции стандарты становятся все более комплексными, отражающими всю сложность информации и ее применения. Стандарты начинают в значительной степени основываться на других стандартах. По этим двум причинам, разработчики, действующие в вечном цейтноте и на ограниченной ресурсной базе, в отчаянии разводят руками и начинают описывать дополнительные наборы стандартов, с которыми они смогли справиться. В METS эти дополнительные наборы были официально определены как профили (profiles) METS. В результате оказалось, что если какие-либо две системы могут независимо друг от друга использовать один и тот же стандарт, то не факт, что они смогут взаимодействовать друг с другом так, как нужно. И здесь вам придется прочитать первоисточник.

Все эти стандарты представлены в XML. Для перевода с одного профиля на другой нужно переводить документ из формата XML в другую форму. Кое-что при этом может быть неизбежно потеряно, так как два профиля не несут абсолютно одинаковой информации, но в любом случае вероятность того, что основное содержание документа при его изменении с помощью XML будет сохранено, достаточно велика. Вполне возможно, что XSLT, язык перекодировки XSL, описание которого вы встречали в Разделе 3.1, (Задание 3), учебника «Как создать электронную библиотеку» будет играть главную роль в практической реализации взаимодействия различных вариантов этих стандартов при их внедрении.

Глава 5.2. Институционные репозитории и интероперабельность

В этой главе вы познакомитесь с институционными репозиториями, являющимися разновидностью электронных библиотек, а также с проблемами, связанными с их интероперабельностью. Интероперабельность — это способность электронной библиотеки обмениваться своими ресурсами и сервисами с другими подобными системами. Ресурсами при этом, как правило, являются документы, метаданные и их всевозможные комбинации. Сервисы обладают широким спектром функциональных возможностей — от поиска до извлечения документа на основе его идентификатора и т.п. В Главе 5.1 мы будем изучать два стандарта MODS и METS, которые отражают структуру и содержание документов и метаданных. Этим обеспечивается интероперабельность ресурсов. Для поддержки интероперабельности сервисов доступ к ним должен осуществляться с помощью стандартных протоколов.

В Задании 1 мы познакомимся с двумя стандартными протоколами для электронных библиотечных сервисов: Z39.50 и Open Archive Initiative Protocol for Metadata Harvesting (OAI-PMH). Затем в Задании 2 мы узнаем о DSpace и EPrints, двух программных системах с открытым кодом, которые специально разработаны для поддержки институционных репозиториях. Далее мы изучим разные способы обеспечения интероперабельности между DSpace и Greenstone. Завершают эту Главу несколько упражнений, которые проиллюстрируют интероперабельность между DSpace и Greenstone в обоих направлениях. Затем мы проиллюстрируем сервисы, обеспечивающие интероперабельность в процессе создания коллекции Greenstone из метаданных и соответствующих им документов, которые были экспортированы из внешней электронной библиотеки с помощью OAI-PMH. О программном обеспечении, которое использовалось для поддержки внешней библиотеки, ничего не известно за исключением того, что она совместима с протоколом OAI.

Задание 1

Прочитайте Раздел 8.5 в конце Главы 8 учебника «Как создать электронную библиотеку». (Раздел 8.4 является дополнительным материалом, который вы при желании можете изучить.) А теперь ответьте на следующие вопросы.

- Кто администрирует протокол Z39.50? Когда появились первые упоминания этого протокола, и в каком году была принята первая версия данного стандарта?
- С какой целью создан регистратор Z39.50, и какие элементы стандарта он включает? Сколько категорий включает данный протокол, и что называется минимальным (базовым) исполнением?
- Что послужило предпосылкой проекта Open Archives Initiative, запуск которого был осуществлен в 1999. Первая версия стандарта была выпущена в январе 2001, (версия 1.1) с небольшими изменениями в июле 2001 и новая версия (2.0) в июне 2002. Сравните эту модель развития с Z39.50 и дайте свои комментарии по поводу тех или иных обнаруженных вами различий.
- Выделите черты сходства и отличия в подходах, реализованных в Z39.50 и OAI-PMH.
- Подумайте о проекте создания электронной библиотеки, связанной с вашей местной университетской библиотекой (или какой-либо другой библиотекой, с которой вы хорошо знакомы), требующей интероперабельности на уровне обслуживания. Какой из этих двух протоколов вы бы порекомендовали и почему?

После опубликования учебника «Как создать электронную библиотеку» (How to Build a Digital Library) в 2003 году была ратифицирована новая версия стандарта Z39.50.

А теперь мы переходим от протоколов к институциональному репозиторию.

Задание 2

Прочитайте статью Мириэм А. Дрейк «Скрытые сокровища институциональных репозиториев», в которой дается общее представление об институциональном репозитории, а также приводятся ответы на следующие вопросы:

- Какова главная цель создания институционального репозитория и в каких типах организаций создано большинство из существующих репозиториев?

- Базируясь на материалах данной статьи, назовите десять категорий документов, которые типичны для любого университетского репозитория? Насколько это соответствует реальной ситуации с репозиторием в том университете, где вы работаете?
- Назовите два главных фактора, от которых зависит устойчивое развитие любого институционного репозитория.

Теперь прочитайте статью Уильяма Никсона «Проект DAEDALUS: предварительный опыт работы с программными приложениями EPrints и DSpace в университете Глазго», в которой описывается опыт одного из университетов Великобритании по созданию и поддержке двух известных систем институционных репозитория с открытым кодом, и ответьте на следующие вопросы:

- В DAEDALUS описание уровня записи (метаданные) было расширено за счет включения типа и статуса документа для того, чтобы автоматически направлять этот документ в ту или иную коллекцию. Назовите пять категорий данной коллекции? Как они соотносятся с перечнем десяти типов документов, о которых говорилось в статье Дрейка (Drake)?
- В статье описываются три ключевые области развития и поддержки институционных репозитория: конфигурация, подчинение и администрирование. Кто в вашей организации обладает необходимой профессиональной компетентностью для выполнения работ в каждой из этих областей? Если сейчас у вас нет ни одного человека, кто смог бы выполнить работу в какой-нибудь из этих областей, тогда определите кто может стать самым подходящим кандидатом для проведения обучения работе в этой области?

Как уже было сказано в начале данного раздела, электронные библиотечные системы могут быть связаны друг с другом путем обмена ресурсами (документами и метаданными) или совместным использованием сервисов, принадлежащих участникам такого обмена. Следующий далее материал для чтения позволит вам сосредоточить внимание на двух больших электронных библиотечных системах с открытым кодом. В представленной статье обсуждаются различные способы организации взаимодействия подобных систем.

Задание 3

Прочитайте статью Иена Виттена и др. «StoneD: мост между Гринстоун и DSpace» и ответьте на следующие вопросы:

- Назовите шесть наиболее важных отличительных черт, характерных одновременно для систем Dspace и Greenstone.
- В статье перечислено несколько отличий между DSpace и Greenstone. Назовите два из них и напишите по небольшому абзацу, описывающему каждый из них.
- Перечислите четыре направления построения взаимосвязей между двумя системами. Для каждого из них отметьте, является ли он примером интероперабельности на уровне источника или интероперабельности на уровне сервиса.

В статье, которую вы прочитали, перед выполнением Задание 3, обсуждается создание коллекции Greenstone из файлов, которые были экспортированы из DSpace. Теперь мы создадим именно такую коллекцию для того, чтобы увидеть, как осуществляется на практике интероперабельность на уровне источника. В этом курсе вы не инсталлировали DSpace (что требует серьезной технической подготовки, а также значительного времени и, помимо всего прочего, эта программа не работает в Windows). Поэтому Задание начинается непосредственно с того, что все документы уже были заранее экспортированы из коллекции DSpace.

В упражнении используются файлы PDF, Word и MP3, которые являются смесью различных статей, интервью и протоколов интервью. В отличие от других упражнений по данному курсу эти документы не имеют первоисточников, существующих в реальном мире, так как они были созданы в лаборатории, чтобы показать вам интересный пример документов выдуманного музыкального отделения. Не забудьте про это, когда вы будете просматривать только что созданную коллекцию!

Упражнение: От DSpace к Greenstone

1. Запустите библиотечный интерфейс (Greenstone Library Interface), если вы с ним не работаете в данный момент.
2. Перейдите в режим работы системного специалиста (Library System Specialist), используя команды File→Preferences, так как далее вам потребуется изменить порядок встроенных приложений на панели **Design**.

3. Откройте новую коллекцию (*new collection*) **StoneD** и заполните поля, введя все необходимые данные. Оставьте набор метаданных в **Dublin Core** по умолчанию.
4. Перейдите на панель **Design** и выберите слева раздел **Document Plugins**. Уберите (**Remove**) **TEXTPlug**, **EMailPlug** и **HTMLPlug**. Реальной необходимости убирать эти встроенные приложения нет, однако этим действием вы снимете риск возникновения лишних помех.
5. Теперь добавьте **DSpacePlug**. Оставьте установки по умолчанию без изменений и нажмите **<OK>**.
6. Используя стрелки вверх и вниз, измените (**Move**) местоположение **DspacePlug**, поместив его между **GAPlug** и **PDFPlug**.
7. Теперь добавьте **MP3Plug** с его установками по умолчанию. Установленное положение на потоковую подачу объектов менять не нужно.
8. С помощью панели **Gather** найдите на основном CD-ROM с материалами курса папку **sample_files\dspace\exported_docs**. Она содержит пять примеров объектов, экспортированных из институционного репозитория DSpace. Скопируйте их в свою коллекцию, перетащив их в правую часть панели.
9. Создайте (**build**) коллекцию и просмотрите (**preview**) ее, обратив внимание на установки по умолчанию, действующие для коллекции Dspace.

Если вы будете вести поиск по заголовкам a–z, то вы обнаружите в списке семь документов, хотя только пять из них были экспортированы из DSpace. Два исходных объекта имели альтернативную форму представления файлов в папке. Установки встроенного приложения DSpace контролируют подобные ситуации и в подобных случаях рассматривают такие объекты, как отдельные документы Greenstone.

Ниже мы приводим пример установок встроенного приложения (**first_inorder_ext**) для того, чтобы показать, как могут работать вместе альтернативные формы. Эта установка позволяет воспринимать документы, которые имеют одно и то же название, но разные расширения как единый объект коллекции. Один из файлов просматривается как первичный документ, он индексируется, и из него (по возможности) будут извлекаться метаданные, в то время как остальные файлы будут рассматриваться просто как ассоциированные с ним.

В установке *first_inorder_ext* в качестве аргумента используется список расширений файлов (отделенных друг от друга запятыми): первичным документом становится тот, расширение которого окажется самым верхним в списке.

10. Выберите **DSpacePlug** и щелкните **<Configure Plugin>**. Включите режим установки параметров *first_inorder_ext*. Установите значения *pdf, doc, mp3* в появившемся окне и нажмите **<OK>**.

11. Создайте (**build**) и просмотрите (**preview**) коллекцию.

Теперь у вас осталось только пять документов, поскольку в список включается только по одной версии каждого документа, т.е. только первичные версии документов.

DSpace экспортирует файлы, содержащие метаданные Dublin Core по заголовку и автору (наряду с прочими). Далее мы дооснастим коллекцию соответствующими инструментами индексирования и просмотра.

12. Находясь на панели **Design**, выберите **Search Indexes**. Удалите индексы *ex.Title* и *ex.Source*, а затем добавьте индекс для **dc.Title**, называемый «заголовки» (*titles*) и для **dc.Contributor**, называемый «авторы» (“authors”).

13. Оставаясь на той же панели **Design**, выберите **Browsing Classifiers** и удалите (**delete**) оба классификатора **AZList** (*ex.Title* и *ex.Source*). Добавьте классификатор **AZList** для **dc.Title** и еще один для **dc.Contributor**.

14. Теперь выберите раздел **Format Features** на панели **Design** и замените оператор форматирования **VList** следующим текстом:

```
<td valign=top>
  [srclink]{or}{[thumbicon],[srcicon]}/[srclink]
</td>
<td valign=top>
[highlight]{or}{[dls.Title],[dc.Title],[ex.Title],Untitled}/[
highlight]
  {If}{[ex.Source],<br>
  <i>([ex.Source])</i>}{If}{[equivlink],<br>
  Also available as:[equivlink]}
</td>
```

Вы найдете этот текст в файле *format_tweak.txt*, находящемся в папке с примерами *sample_files* в *dspace*. Вы можете скопировать и вставить этот текст так же, как вы это делали ранее (в Разделе 4.2). Не забудьте, завершив работу, нажать кнопку **<Replace Format>**.

15. Создайте (**build**) коллекцию снова и просмотрите (**preview**) ее.

У вас в списке по-прежнему осталось только пять документов, но напротив некоторых из объектов, например, *Interview with Bob Dylan*, появилась строка “Also available as” (Так же доступный, как...), сопровождаемая пиктограммами, которые ведут к альтернативным формам представления данного документа.

В упражнении, следующем ниже, вы экспортируете коллекцию Greenstone в форме, приемлемой для DSpace. Это можно будет сделать с помощью команд меню File библиотечарского интерфейса, в котором есть команда Export..., позволяющая вам экспортировать коллекции в различной форме. Однако для более глубокого понимания Greenstone мы выполняем эту работу, запрашивая программу в командной строке Windows. Это требует наличия некоторых технических навыков и поэтому данное упражнение обозначено как дополнительное. Если вы не привыкли работать в режиме командной строки, то рекомендуем вам пропустить это упражнение.

Дополнительное упражнение: от Greenstone к DSpace

16. Откройте окно DOS для перехода в режим командной строки. Вы сможете сделать это в меню Пуск→Программы→Стандартные, но детали могут отличаться в зависимости от того, какую версию системы Windows вы используете. Если вы не можете найти командную строку с помощью основного меню, то выберите Пуск→Выполнить... и введите cmd в открывшееся окно.

17. В окне DOS переместитесь в домашнюю директорию, в которую вы установили Greenstone. Это должно выглядеть примерно так:

```
cd C:\Program Files\greenstone
```

18. Напечатайте:

```
setup.bat
```

что позволит вам перейти к программному управлению Greenstone.

19. Замените название директории, в которую вы поместили созданную вами коллекцию при выполнении последнего упражнения:

```
cd collect\stoned
```

Обратите внимание: даже если в названии коллекции используются заглавные и строчные буквы, названия директорий в библиотечарском интерфейсе пишутся только прописными буквами.

20. Введите следующую команду для экспортирования коллекции с использованием формата DSpace (import/export format):

```
perl -S export.pl -saveas DSpace -removeold stoned
```

Экспортирование в Greenstone является дополнительным процессом. Если вы еще раз введете команду export.pl, то к тем файлам, которые уже существуют в папке export, будут добавлены новые, экспортированные, файлы, но с новыми именами папок. Для того, чтобы лучше проследить за всеми изменениями, что, собственно, и является нашей основной целью при выполнении этого упражнения, мы можем давать эту команду несколько раз. Команда – removeold удалит все файлы, которые были вами экспортированы.

21. Эта команда создаст новую вложенную папку collect\stoned\export. Познакомьтесь с ней, используя имеющийся у вас проводник. Там вы найдете файлы, которые нужны для систематизации этого набора документов в DSpace. Они очень похожи на те, с которых мы начали предыдущее упражнение.

Вы точно также можете использовать команду export.pl при работе с любыми другими коллекциями Greenstone, например, в коллекции материалов для чтения по данному курсу, а также при переносе результатов произведенной инсталляции DSpace с помощью программных средств импортирования данных.

Следующее упражнение позволит вам познакомиться с интероперабельностью на уровне сервисов, что осуществляется за счет использования протокола Open Archive Initiative Protocol for Metadata Harvesting (OAI-PMH). Упражнения, которые вы выполняете в данном курсе, рассчитаны на работу за изолированным компьютером, поэтому мы не будем соединяться с каким-либо реальным внешним сервером, работающим как поставщик данных. Вместо этого, мы подготовили для вас некоторый набор файлов, который помещен на переданном вам компакт-диске с материалами курса. Эти файлы имеют вид XML-записей, которые вырабатывает протокол OAI-PMH. В следующем упражнении описывается, как можно выполнить такое условное соединение.

Мы уже встречались с подготовленной заранее и построенной на OAI коллекцией Greenstone в Задании 1 Главы 4.3, когда изучали

пример коллекции документов, полученной при инсталляции Greenstone. В следующем упражнении вы опять будете с ней работать: вы пройдете все шаги, которые необходимо проделать для восстановления этой коллекции. Коллекция состоит из изображений, но вы не сможете ее создать, если на вашем компьютере не установлена программа ImageMagick, как это было описано в Задании 1 Главы 2.2. Если хотите взглянуть на то, что у вас должно получиться в результате выполнения этого упражнения, то можете прямо сейчас посмотреть пример коллекции документов.

Упражнение: Коллекция OAI

1. Запустите новую коллекцию под названием **OAI Service Provider**. Введите на всех полях нужную информацию. Вы можете оставить установку метаданных по умолчанию Dublin Core, хотя мы и не будем ею пользоваться.
2. На панели **Gather** найдите папку *sample_small* в *sample_files/oai*. Перетащите эту папку в коллекцию и сбросьте ее там.
3. В процессе выполнения операции по копированию появляется окно с вопросом о том, хотите ли вы включить **OAIPlug** в список встроенных приложений, использованных в этой коллекции, поскольку Библиотечарский интерфейс (Librarian Interface) не нашел подходящие для обработки этого типа файла запускаемые программы. Нажмите кнопку **<Add Plugin>** для добавления этой программы в существующий список.

Когда файлы копируются таким способом, библиотечарский интерфейс обрабатывает каждый из них и использует расширение этих файлов для проверки наличия в данной коллекции соответствующей программы. До недавнего времени ответом на этот вопрос всегда было «да», поэтому все переданные файлы были незамедлительно обработаны. В этот раз, однако, ни одна из программ в имеющемся списке не в состоянии обработать файловые записи OAI, которые мы скопировали (их расширение — .oai).

В некоторых случаях существует несколько запускаемых программ, которые могут обработать один и тот же файл, например, расширение *.xml используется во многих XML-форматах. Появляющееся в этом случае окно предлагает выбрать подходящую для данного расширения запускаемую программу из предложенных вариантов. Как правило, сделать выбор очень просто. Если вы хотите, вы можете вообще проигнорировать вопрос (щелкнув **<Don't Add Plugin>**), поскольку программу можно добавить позднее, в разделе Document Plugins панели Design.

4. Вам нужно сконфигурировать программу Image. На панели **Design** выберите раздел **Document Plugins**, затем строку **plugin ImagePlug** и щелкните **<Configure Plugin>**. В открывшемся вслед за этим окне найдите установку параметра **screenviewsize**, включите его и напечатайте цифру 300 в окне, расположенном за ним, что позволит вам задать размер окна просмотра изображения в 300 пикселей. Щелкните **<OK>**.

5. Теперь переключитесь на панель **Create** для того, чтобы создать (**build**) и просмотреть (**preview**) коллекцию.

Как и при создании других коллекций, мы возьмем за основу установки по умолчанию, принятые в Greenstone, а результат, который мы получаем, вполне нам подходит, хотя его еще можно усовершенствовать. Выполняя следующие шаги, мы улучшим нашу коллекцию, используя метаданные, которые были собраны с помощью OAI-PMH в файлы с расширением .oai.

6. В разделе **Browsing Classifiers** панели **Design**, удалите два классификатора **AZList** (*ex.Title* и *ex.Source*).

7. Добавьте классификатор **AZCompactList**, который будет основываться на метаданных **ex.Subject**.

8. Теперь добавьте классификатор **AZCompactList**, который использует метаданные **ex.Description**. На панели с его установками выберите **mincompact = 1**, **maxcompact = 10** и **buttonname = Captions**.

9. В разделе **Search Indexes** на панели **Design** удалите все индексы и добавьте новый, который называется "captions" и который основывается на метаданных *ex.Description*.

10. Создайте (**build**) коллекцию и просмотрите (**preview**) ее.

Можно провести и другие усовершенствования для улучшения общей презентабельности коллекции.

11. На панели **Design** выберите **Format Features**. Сначала замените оператор форматирования **VList** на следующую запись:

```
<td>
  {If}{[numleafdocs],[link][icon][link],[link][thumbicon]
    [link]}
</td>
<td valign=middle>
  {If}{[numleafdocs],[Title],<i>[Description]</i>}
</td>
```

Вы найдете этот текст в файле *vlist_tweak.txt* в папке oai, расположенной в *sample_files*. Не забудьте нажать **<Replace Format>** после завершения работы.

Этот оператор форматирования определяет в дальнейшем внешний вид вертикальных списков, таких, как списки результатов поиска и перехвата при показе пиктограмм эскизов, сопровождаемых метаданными `Description`. Установкой по умолчанию в Greenstone является использование извлеченных метаданных, поэтому `[Description]` – это то же самое, что и `[ex.Description]`.

Затем, в открывающемся списке **Choose Feature** выберите **DocumentHeading** и сделайте так, чтобы в его операторе форматирования (который пока пуст) можно было прочитать

```
<h3>[Subject]</h3>
```

Когда вы добираетесь непосредственно до документа, представленного в коллекции, над кнопками `detach` и по `highlighting` появляется заголовок документа. По умолчанию `DocumentHeading` демонстрирует метаданные документа `ex.Title`. В данном конкретном примере экспортированных записей ОАИ заголовками являются имена JPEG файлов, причем эти имена файлов совершенно не информативны (например, `01dla14`). Вы можете увидеть их на панели **Enrich**, если выберите какое-либо изображение в `sample_small→oai→JCDLPICS→srcdocs`, а затем проверите соответствующее ему имя файла и метаданные `ex.Title`. В показанном выше операторе форматирования вместо этого представляются метаданные `ex.Subject`.

И, наконец, вы, наверное, уже заметили, что в том месте, где, собственно, и должен был появиться документ, вместо него вы видите предупреждение `This document has no text` (Этот документ не содержит текста). Чтобы снять эту проблему, выберите **DocumentText** в открывающемся списке **Choose Feature** и используйте приведенный ниже отрывок текста в качестве оператора форматирования, который до этого действия был не заполнен. Этот текст находится в файле `doctxt_tweak.txt` упоминавшейся ранее папки `format_tweaks`.

```
<center><table width=_pagewidth_ border=1>
<tr><td colspan=2 align=center>
<a href=[OrigURL]>[screenicon]</a></td></tr>
<tr><td>Caption:</td><td> <i>[Description]</i> <br>
(<a href=[OrigURL]>original [ImageWidth]x[ImageHeight]
[ImageType] available</a>)
</td></tr>
<tr><td>Subject:</td><td> [Subject]</td></tr>
<tr><td>Publisher:</td><td> [Publisher]</td></tr>
<tr><td>Rights:<td> [Rights]</td></tr>
</table></center>
```

Этот оператор форматирования меняет способ представления документа. Документ будет выведен на весь экран, и он будет также связан с оригиналом документа, хранящимся в полном объеме на Вебе. Кроме того, будут представлены количественные данные о самом изображении такие, как ширина, высота и тип файла.

12. Вспомните о том, что эти операторы форматирования обрабатываются системой поддержки исполнения программ (Глава 4.3), поэтому для того, чтобы увидеть, что получилось в результате ваших действий, вам не нужно перестраивать коллекцию. Перейдите на панель **Design** и нажмите **<Preview Collection>**, чтобы увидеть результаты.

С целью ускорения процесса создания коллекции в нее было включено меньшее количество источников документов, чем в заранее подготовленных коллекциях, устанавливаемых в процессе инсталляции Greenstone. Однако, несмотря на все произведенные изменения, функциональность этой коллекции осталась прежней.

В выполненных вами упражнениях вы не получали данных с какого-либо внешнего сервера OAI-PMH. Для выполнения этой операции потребуется запустить программу, управляемую из командной строки (как и в процессе экспортирования данных, описанном в предыдущем дополнительном упражнении). Ваш компьютер должен быть напрямую подключен к Интернету, поскольку использование сетевой защитной системы может повлиять на возможность программы выгрузить на компьютер нужную вам информацию. Если ваше рабочее место не отвечает этому требованию, то пропустите это дополнительное упражнение и переходите к следующему заданию.

Дополнительное упражнение: Скачивание файлов через OAI

13. **Сохраните (Save)** свою коллекцию. Удостоверьтесь в том, что название директории записано как `oaiservi` (оно появляется в окне заголовка библиотечарского интерфейса), после чего выйдите (**quit**) из GLI.
14. Повторите еще раз все действия, описанные в первых пяти пунктах предыдущего дополнительного упражнения: откройте командное окно, найдите директорию, в которую была установлена Greenstone, запустите файл `setup.bat` и еще раз поменяйте директорию на `collect\oaiservi` — папку, содержащую коллекцию провайдера сервиса OAI, которую вы создали при выполнении последнего упражнения.

15. С помощью текстового редактора откройте файл конфигурации коллекции `oaiservi\etc\collect.cfg`. Добавьте в него следующую строку (все это должно быть написано в одну строку):

```
acquire OAI -src rocky.dlib.vt.edu/~jcdlpix/  
cgi-bin/OAI1.1/jcdlpix.pl -getdoc
```

Хотя конкретное местоположение данной строки не так уж и критично, мы рекомендуем вам поместить ее поближе к началу файла после строк, содержащих общую информацию и информацию об авторе, но выше строки с указателем. Сохраните файл и выходите из программы редактирования.

16. Удалите содержимое коллекционной папки *import*. Она включает отсканированные версии коллекционных файлов, помещенные туда во время выполнения последнего упражнения. Далее вы увидите, как получить новые данные с внешнего сервера OAI.

17. Вернувшись назад, к командной строке DOS, запустите `perl -S importfrom.pl oaiservi`

Greenstone тут же начнет свою работу и произведет поток диагностической информации. Программа `importfrom.pl` установит соединение с провайдером данных OAI, который значится в файле конфигурации коллекции (это делается для каждой отдельной строки “acquire” в данном файле), и экспортирует все имеющиеся на данном сайте записи.

18. Выгруженные файлы сохраняются в папке импорта. Как только команда будет выполнена, все окажется на месте, и вы сможете приступить к созданию коллекции. Подтвердите успешную загрузку нужных вам записей OAI, повторно создав коллекцию.

В OAI-PMH отсутствуют специальные средства для описания содержания документа, поскольку в самом стандарте четко объясняется, что при его создании такая цель поставлена не была. Однако существует и негласное соглашение о том, что в поле метаданных *Identifier* может включаться URL того документа, который он описывает. В Greenstone есть команда *getdoc*, которая использовалась в предыдущем дополнительном упражнении. Команда проверяет поле *Identifier* в каждой загружаемой записи, чтобы увидеть, есть ли в ней URL, и если да, то этот URL загружается в файл и связывает его с этой записью.

Опуская детали, отметим, что программа импортирования OAI, используемая в Greenstone (*importfrom.pl*), взаимодействует с внешним сервером по протоколу HTTP, который положен в основу Веба.

Если записи метаданных (и ассоциированные с ними документы) уже были скачены ранее, то их следует скачать вновь, если копия на сервере новее, чем существующая на компьютере пользователя.

В следующем Задании вы используете только что приобретенные знания об OAI и Greenstone. Мы предполагаем, что даже если вы и не справились с предыдущим дополнительным упражнением, то немало об этом прочитали и проработали все инструкции.

Задание 4

Как бы вы использовали Greenstone для создания виртуальной электронной библиотеки, которая будет составлена из данных, полученных из разных репозиториях OAI, каждый из которых поддерживает отдельная организация, но при условии, что доступ к этим репозиториям будет осуществляться из одной точки и будет универсальным?

Мы научились создавать коллекции Greenstone из материала, позаимствованного из распределенного репозитория OAI. А что делать в том случае, если вам захочется открыть другим пользователям доступ к своим коллекциям по протоколу OAI? В программу Greenstone заложена такая возможность. Убедитесь в том, что серверная программа “OAI server”, называемая *oaiserver*, активизирована. В этом случае любой желающий может получить доступ к вашим коллекциям Greenstone через протокол OAI, используя тот же самый адрес (URL), который использовался для того, чтобы получить доступ к вашей инсталлированной коллекции Greenstone, но при этом вместо *library* должно быть написано *oaiserver*. Например, если к вашим коллекциям Greenstone можно получить доступ по адресу:

<http://127.0.0.1/cgi-bin/library>,

то ваш сервер OAI будет доступен по адресу:

<http://127.0.0.1/cgi-bin/oaiserver>

(в вашем случае 127.0.0.1 будет заменен на адрес вашего компьютера).

Вы можете заранее определить, какие из ваших коллекций могут быть открыты для внешнего доступа по OAI. Установки параметров OAI определены в файле *oai.cfg*, находящемся в директории *etc* Greenstone. Этот файл предоставляет общую информацию о репозитории и перечисляет коллекции, которые должны быть открыты. По умолчанию все коллекции являются закрытыми для внешнего доступа: для того, чтобы снять эту установку, включите название коллекции в список *oaicollection*. Те коллекции, которые были созданы

с помощью версий Greenstone 2.52 и более ранних, должны быть обновлены прежде, чем они будут использоваться через OAI. Сервер OAI поддерживает только один стандарт метаданных — Dublin Core. Для коллекций, использующих другие наборы метаданных, должны действовать правила по переводу существующих метаданных в формат Dublin Core, описанные в файле *oai.cfg*.

Как мы упоминали ранее, ваша коллекция будет доступна для других только в том случае, если вы активизируете программу *oaiserver*. При установке Greenstone (проведенной вами в Задании 4 Раздела 2.2) по умолчанию использовался локальный библиотечный сервер Windows. Для запуска сервера OAI вы должны запускать не его, а библиотечную веб-версию Greenstone. Для этого вам нужно установить веб-сервер (в нашем случае — Apache). Инструкции по установке включены в Руководство по установке Greenstone (Greenstone Installer's Guide), имеющееся на вашем компакт-диске с курсом: если вы не справитесь с этим сами — обратитесь за помощью к техническому специалисту. Вам нужно осуществить установку веб-сервера, а затем повторно установить Greenstone (также как вы выполняли это в Задании 4 Главы 2.2), но на этот раз, выбрав *web library* вместо *local library*. Когда вы используете библиотечную версию Greenstone, программа *oaiserver* запускается автоматически. (Она запускается как программа “CGI”, библиотечная программа Greenstone, поскольку находится в директории командных файлов Greenstone *cgi-bin*.)

Заключение

Эта глава была посвящена рассмотрению технических вопросов гораздо в большей степени, нежели общечеловеческих, организационных или политических — хотя последние, конечно же, являются самыми важными факторами, влияющими на разработку и адаптацию технологии электронных библиотек. Сферы применения образовательных электронных библиотек так разнообразны, что сложно что-либо выделить. Некоторые важные позиции мы рассмотрим в контексте институциональных репозиториев.

Институциональные репозитории обеспечивают инфраструктуру для сбора всех продуктов интеллектуального труда организации, их размещения и хранения, а также обеспечения к ним широкого доступа с целью их использования. При их организации особое внимание нужно уделить главным заинтересованным лицам. Например, в университетской среде можно выделить следующие группы таких лиц:

- сотрудники факультетов, чья грантовая и научно-исследовательская деятельность предусматривает накопление и распространение результатов интеллектуального труда,
- сотрудники библиотек, в круг основных обязанностей которых входит освоение, использование и поддержка репозитория и связанных с ним сервисов,
- персонал университета, участвующий в проведении долгосрочного планирования, и лица, принимающие решения, поскольку создание такого подразделения требует регулярных финансовых вложений,
- широкие круги университетского сообщества, от чьего интеллектуального и политического влияния будет зависеть поддержка данного проекта,
- участники пилотных апробаций.

Для того, чтобы реализация проекта стала успешной, о нем необходимо проинформировать широкие круги общественности и членов профессионального сообщества. Руководители подразделений образовательных учреждений, ведущие специалисты и известные преподаватели должны убедиться в том, что без их заинтересованного участия такой проект не сможет развернуться в полной мере. Можно задать им массу вопросов. Например, можно ли удалять старые версии статей? Можно ли что-то убирать из репозитория и если да, то, при каких обстоятельствах? Может ли рядовой сотрудник университета разместить в репозитории огромные по размерам объекты, например, наборы данных объемом в сотни гигабайт?

Первопроходцы подобных систем играют очень важную роль в тестировании различных инструментов репозитория и механизмов осуществления обратной связи с разработчиками. Они способствуют распространению новых технологий в образовательном учреждении. Они определяют процедуры работы с информацией в репозитории на уровне пользовательского интерфейса и привлекают внимание специалистов к решению возникающих проблем. Они могут продемонстрировать свои достижения еще до того, как репозиторий начнут использовать другие люди, после чего мгновенно распространят об этом информацию по всему миру, поэтому их имена следует упоминать в рекламных материалах и официальных публикациях.

Создание институционального репозитория — это длительный и никогда не прекращающийся процесс. Сколько будет стоить его запуск, поддержка, обслуживание и обновление через десятилетия? Одним из самых серьезных, но до конца непроработанных вопросов является

вопрос стоимости хранения данных, поскольку очень сложно оценить стоимость осуществления такого проекта и стоимость перемещения огромных по объемам данных. Управление системами хранения подобных объектов требует тщательного финансового планирования. Перспективные расчеты по вместимости репозитория выходят на первый план, поскольку именно это будет определять объем финансовых затрат в будущем. И, наконец, институциональные репозитории зачастую создаются на базе библиотек, которые и так уже страдают от серьезного недостатка финансирования.

Организации, занимающиеся внедрением репозитория, очень быстро понимают, что распространение информации о репозиториях и выявление потребностей их потенциальных пользователей, предваряющие технологические решения по созданию таких систем, помогают более четко определить спектр ожидаемых услуг и политику в отношении использования репозитория. Организациям приходится определять структуру управления репозиторием и виды собираемой информации, в которую могут быть включены: результаты исследовательских проектов, диссертации, метаданные или административная информация. Также должны быть разработаны план и стратегии сохранения и восстановления репозитория в случае чрезвычайной ситуации. И, наконец, должна быть продумана структура управленческого аппарата, персонал, наблюдательный совет, которые в наибольшей степени соответствуют традициям данной организации.

Глава 5.3. Примеры из практики образовательных электронных библиотек

Нижеследующая глава представляет примеры из практики действующих систем, которые наглядно иллюстрируют различные способы распространения информации через образовательные электронные библиотеки. В процессе знакомства с ними подумайте, о том наследии, которое каждая подобная система оставит для будущих поколений, желающих ее использовать в совершенно новой компьютерной среде. Здесь кроется самый сильный аргумент из всех имеющихся в пользу открытых стандартов и программного обеспечения по созданию электронных библиотек с открытым кодом.

Пример из практики MERLOT: Мультимедийный образовательный ресурс для онлайн-обучения

Наш первый пример представляет собой каталог материалов, размещенных в сети, не содержащий учебных объектов, как таковых. Как и каталог обычной библиотеки, он содержит информацию о материалах, которые расположены на той или иной книжной полке. Это шлюз, исключительно для сбора метаданных, в то время как электронные библиотеки обычно включают и содержательную составляющую. Вы уже видели это в коллекции, созданной с помощью записей LOM, с которой вы познакомились в Главе 3.3.

MERLOT (The Multimedia Educational Resource for Learning and On-Line Teaching) – это мультимедийный образовательный ресурс для онлайн-обучения. Он является высококачественной коллекцией интерактивных учебных материалов, задачников, обзоров и данных по персоналиям. Также MERLOT – это национальная сеть предметных онлайн-сообществ, которые отбирают и реферируют учебные материалы по той или иной отдельной учебной дисциплине. MERLOT является национальным порталом на котором собираются имеющиеся в WWW реферированные учебные материалы.

Источник: <http://www.merlot.org/help/FAQ.po>

Этот богатый по своим возможностям мультидисциплинарный проект, включает так же процесс реферирования и отбора учебных материалов. MERLOT использует отзывы и комментарии в системных средствах поиска и просмотра. Такой канал обратной связи с пользователями помогает сформировать сообщество преподавателей, которое в свою очередь обеспечивает поддержку и дальнейшее развитие каталога образовательных материалов.

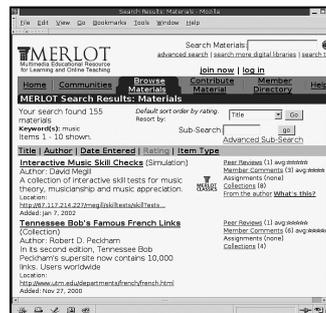
Приведенные ниже иллюстрации дают представление об этой системе. Начиная с домашней страницы, пользователь вводит поисковое слово “music” (его не видно на рисунке) и нажимает кнопку “go” для активизации поиска. Используемое в запросе поисковое слово является достаточно общим, поскольку пользователь пытается получить общее представление о том, что содержится в данном репозитории. Такой тип использования запроса очень часто фиксируется при анализе протоколов записей работы пользователей в репозиториях, причем даже в тех случаях, когда, как и в MERLOT, существует инструмент просмотра по категориям, который можно было бы использовать с большой эффективностью. По запросу было найдено 155 соответствий, и пользователь продолжает свое исследование, выбрав тот документ, который находится в верхней части списка – *Interactive Music Skill Checks* Давида Мегилля (David Megill). Форма начального представления этого учебного объекта объединяет все метаданные для данного ресурса и включает ссылки на экспертные отзывы. Редакционный (экспертный) совет, отвечающий за данную предметную область, присвоил данному документу определение “Merlot classic” (классика MERLOT).

Щелчок в поле *location* открывает сам исходный ресурс. Он состоит из набора тестов по *теории (theory)* и *прослушиванию (aural)*, и далее подразделяющихся по таким категориям как высота (*pitch*), ритм (*rhythm*) и т.д. Как только пользователь выбирает *один из тестов*, в *новом окне* запускается приложение Java, предлагающее пользователю ответить на вопрос. После ввода ответа, пользователь может выбрать команду *check* для проверки правильности выполнения задания.

Модуль 5. Открытые стандарты и примеры из практики



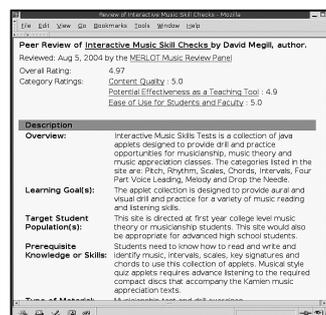
Домашняя страница MERLOT



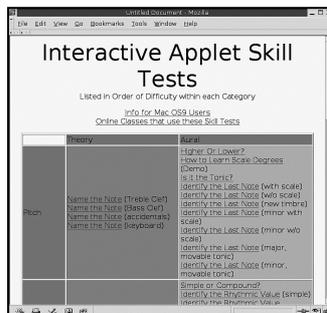
Результаты поиска по слову "music"



Документ, возглавляющий список соответствий



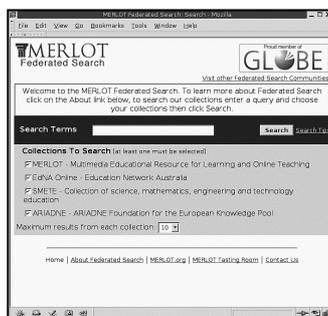
Метаданные данного документа



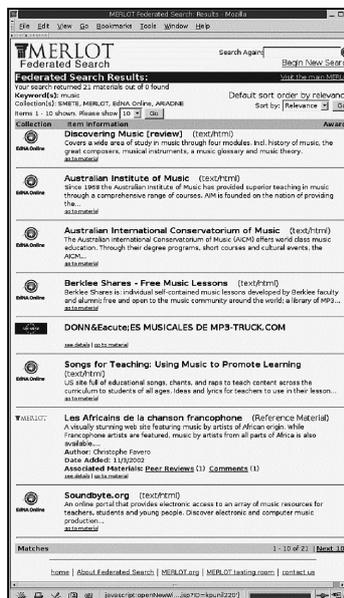
Выбор проверочного теста

MERLOT поддерживает объединенный поиск с помощью кнопки search more libraries на домашней странице. В примере, приведенном ниже, пользователь повторил в своем запросе слово music. Установки по умолчанию для объединенного поиска ограничивают количество объектов, запрашиваемых с каждого сайта, до десяти десятков, поэтому был представлен только 21 подходящий запросу документ: десять из MERLOT, десять из EdNA и один из Agiadne (информация об этих библиотеках будет дана ниже). Сайты, задействованные при объединенном поиске, не отличаются единообразием представления материала, поэтому информация, воспроизведенная в результатах, также различна по форме. EdNA позволяет напрямую перейти к материалу, Agiadne приводит краткие описания ресурсов, а Merlot приводит такие детали, как информация об авторе и дате создания источника, а также ссылки на референтные отзывы.

Глава 5.3. Примеры из практики образовательных электронных библиотек



Объединенный поиск в Merlot



Результаты поиска по слову “music”

Также как и в коллекции Cereo (описанной в Главе 3.3), большая часть функций MERLOT доступна всем пользователям, но только зарегистрированные пользователи могут выполнять такие операции, как размещение объектов в персональной зоне или написание отзывов. Следующее задание предлагает вам посмотреть на эту систему, но вы сможете это сделать только в том случае, если у вас есть доступ в Интернет.

Задание 1 (только в Интернете)

Если у вас есть доступ к сети, найдите каталог MERLOT на сайте <http://www.merlot.org>

- Можете ли вы найти материал, который можно использовать хотя бы в одном из ваших курсов?
- Удобно ли выделены категории для поиска?
- Насколько полезны отзывы и комментарии пользователей?

Рассмотрите проблему поддержки работоспособности каталога, состоящего только из метаданных.

- Как обновляется каталог при изменении ресурсов?
- Как каталог будет работать, если ресурс появляется одновременно в нескольких каталогах?

Познакомьтесь с объединенным поиском в Merlot на сайте:

<http://fedsearch.merlot.org/main/search.jsp>

Используя эту схему, вы можете одновременно проводить поиск по четырем репозиториям учебных материалов:

- MERLOT;
- EdNA – онлайн-система, которая создана для поддержки и распространения информации о преимуществах Интернета для обучения и образования в Австралии;
- SMETE – шлюз к ресурсам, которые способствуют развитию, усилению и совершенствованию обучения и преподавания по естественным наукам, математике, в сфере инженерного и технологического образования в США;
- ARIADNE – создана для расширения кругозора европейцев в области существующих возможностей обучения на основе информационных и коммуникационных технологий.

Повторите первый вопрос данного задания, используя объединенный каталог. Заметили ли вы какую-нибудь разницу?

Вы можете прочитать об онлайн-сообществе MERLOT в материалах их конференций, опубликованных на сайте:

<http://conference.merlot.org/>

Пример из практики Ниурера: газеты на маори

Маори – это коренные жители Новой Зеландии, а Ниурера – это коллекция исторических газет, которые были напечатаны преимущественно для маорийской аудитории с 1842 по 1932 год. Они представляют особый интерес, так как охватывает период европейской колонизации (Новая Зеландия, будучи весьма удаленной от Европы, была открыта европейцами довольно поздно). Газеты можно искать с помощью полнотекстового поиска и просматривать сериями или по отдельным датам выпуска. Данная коллекция стала доступной для пользователей благодаря проекту по созданию электронной библиотеки, выполненному новозеландским Университетом Вайкато (University of Waikato).

Коллекция Ниурера содержит более 17000 газетных страниц, взятых из 34-х отдельных периодических изданий. Некоторые из них издавались на деньги правительства Новой Зеландии, другие при финансовой поддержке самих маори, а остальные издавались религиозными группами.

В основу данной коллекции была положена коллекция микрофишей, собранная Библиотекой Александра Тернбулла (Alexander Turnbull

Library) в Новой Зеландии. Большая часть этой коллекции написана исключительно на языке маори (70%), некоторые публикации на двух языках (27%), а относительно небольшая часть — только на английском (3%).

Коллекция включает четыре основных компонента:

- факсимильные изображения исходных страниц,
- тексты, взятые из газет (для осуществления поиска),
- библиографические комментарии по каждому газетному заголовку,
- аннотации на английском языке по каждому выпуску.

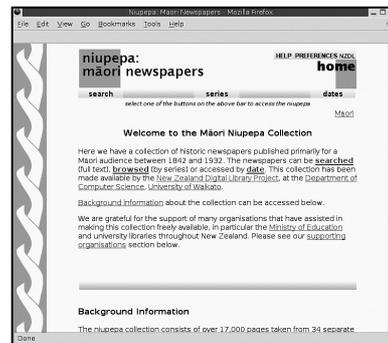
В отличие от других электронных библиотек и репозиториев, о которых было рассказано в данном курсе (а именно, DLESE, NLVM и MERLOT), Niurera не имеет особых образовательных метаданных. Образовательная ценность этого ресурса объясняется повышением уровня его доступности для пользователей. Проект позволил коренным образом изменить доступ к архивам газет на языке маори от ограниченного и доступного лишь читателям, работающим с микрофишами в особых библиотеках, до открытого, практически с любого терминала Интернет. Добавим, что не менее важным является и то, что в рамках проекта были разработаны новые механизмы доступа к информации такие, как, например, полнотекстный поиск. Эти изменения создали основу для дальнейших исследований этого материала, предпринимаемых преподавателями, историками и учеными. На сайте проекта видно, что уже сделан первый шаг в направлении создания полноценной образовательной электронной библиотеки и в ней, наряду с исходными архивными материалами, уже содержатся некоторые методические материалы.

Далее показаны иллюстрации с примерами использования данной коллекции. Пользователь начинает с просмотра домашней страницы, чтобы разобраться в содержании и структуре данной коллекции — сначала на языке маори, а затем на английском. Затем он проводит поиск полнотекстных документов из газет, в которых используется слово *waka* (что на языке маори означает каное, *canoe*). Далее производится просмотр списка подходящих документов и выбирается тот из них, в котором слово *waka* использовано в заголовке. Щелчок по этому слову позволяет вывести на экран весь документ целиком. В этом месте пользователю становятся доступны различные режимы просмотра документа, включая факсимильное изображение первой страницы газеты. На следующей иллюстрации показано как пользователь просматривает серию газетных заголовков, затем серии газетных выпусков, а затем переходит на книжную полку, для того чтобы взять там *Anglo Maori Warder* и уточнить конкретный вопрос.

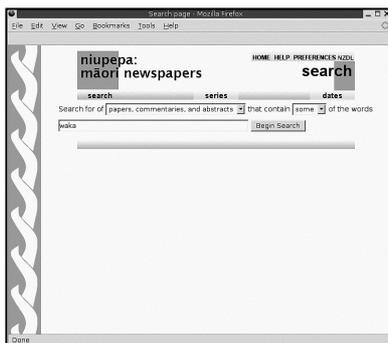
Модуль 5. Открытые стандарты и примеры из практики



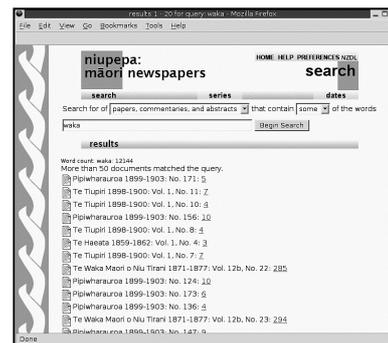
Домашняя страница Niupepa



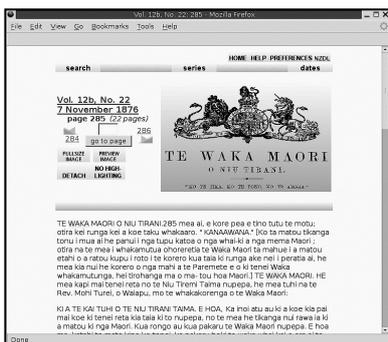
Домашняя страница на английском языке



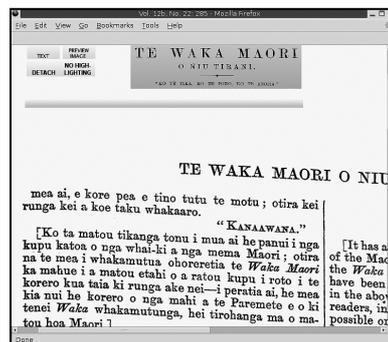
Поиск в коллекции по слову «waka»



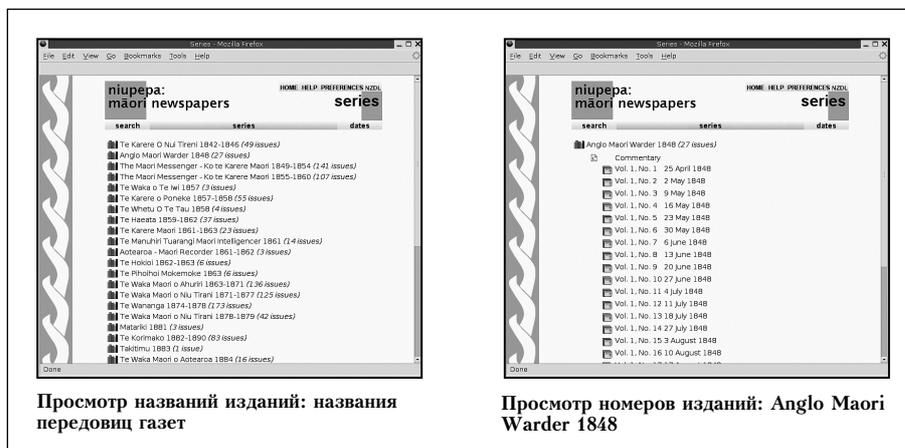
Результаты поиска



Просмотр документа: только текст



Просмотр документа: отсканированное изображение



Задание 2

Прочитайте статью Марка Эпперли и др. «Доступ к газетам на языке маори через Интернет» для того, чтобы больше узнать о проекте Niupepa, а затем ответьте на следующие вопросы.

- Можете ли вы привести примеры хорошо знакомых вам коллекций документов, которые могли бы значительно выиграть от реализации подобного подхода?
- Перечислите некоторые из дидактических возможностей работы с подобным материалом, предоставленным электронной библиотекой.
- Какие вы можете предложить пути исправления ошибок, связанных с реализацией процесса OCR(оптического распознавания символов). Расскажите об этом?

Вы можете сами познакомиться с Niupepa онлайн по адресу:
<http://www.nzdl.org/niupepa>

Упражнение, которое вы сейчас выполните, предложит вам использовать Greenstone для создания небольшой версии Niupepa с помощью пяти газет из двух газетных серий. Вы сможете провести поиск по полнотекстным документам, а также просмотреть документы по названию и дате. Когда вы будете просматривать газету, то вы сначала увидите изображение газеты в режиме предварительного просмотра, а затем, слайд за слайдом, сам текст, снабженный инструментом навигации *goto page* в верхней части страницы.

Коллекция включает собрание разнообразных исполняемых команд, классификаторов и операторов форматирования. Основная часть работы осуществляется с помощью приложения *PagedImgPlug*, которое было специально создано для обработки данных, подобных тем, что мы видим в данном примере. Для каждого документа подготовлен свой файл объекта коллекции, в котором определяется список графических файлов, составляющих документ, и дается информация о номере страницы, на которой они находятся. Также этот файл может (по выбору) сопровождаться текстовым файлом, содержащим машиночитаемую версию данного изображения, которая используется для полнотекстового поиска. Три газеты в нашей коллекции (все три из серии *Te Whetu o Te Tau*) имеют такие текстовые версии, а две из них (из серии *Te Waka o Te Iwi*) представлены только в виде изображений. Файлы объектов также могут содержать метаданные. В нашем примере газетная серия записана как *ex.Title*, а время ее публикации как *ex.Date*. Эти метаданные извлекаются в процессе создания коллекции.

Упражнение: Коллекция отсканированных изображений

1. Запустите новую коллекцию, которая называется **Paged Images**, и заполните поля соответствующей информацией: эта коллекция строится на документах, заимствованных из фрагмента коллекции Niupera; единственными метаданными здесь являются название статьи и дата, причем они извлечены из файлов объекта, которые включены в исходные документы, поэтому необходимость специально прописывать метаданные в этом случае отсутствует.
2. Включите **PagedImgPlug** и установите параметры его *screenview* (режима просмотра) на *on*. Исходные изображения, с которыми мы работаем, были отсканированы с высоким разрешением и поэтому они слишком велики по объему, чтобы загружаться через обычный браузер. После создания коллекции режим просмотра (*screenview*) изображения для каждой страницы будет генерироваться с меньшим разрешением.
3. На панели **Gather** в папке *sample_files* откройте *niupepa/sample_items* и перетащите все ее содержимое в вашу коллекцию, расположенную в правой части экрана.

4. Некоторые из файлов, которые вы только что перетащили, являются текстовыми, содержащими текст, извлеченный из графических файлов этих страниц. Нам нужно их обработать с помощью **PagedImgPlug**, а не **TEXTPlug**. Переключитесь на панель **Design** и удалите **TEXTPlug**. Пока вы там находитесь, вы можете навести порядок и с другими объектами, удалив также **HTMLPlug**, **EMAILPlug**, **PDFPlug**, **RTFPlug**, **WordPlug** и **PSPlug**, поскольку ни один из них вы не будете использовать в дальнейшем.
5. Теперь переходите на панель **Create**, создайте (**build**) коллекцию и просмотрите (**preview**) полученный результат. Проведите поиск по ключевому слову *waka* и посмотрите один из документов, заголовки которых будут представлены в виде списка по завершению поиска (все три появятся под названием *Te Whetu o Te Tau*). Пролистайте заголовки *titles a–z* и откройте один из документов под заголовком *Te Waka o Te Iwi*.

Эта коллекция была построена в Greenstone при действующих установках по умолчанию. Вы можете найти в ней интересующие вас материалы, но информация здесь представлена менее четко и привлекательно с точки зрения дизайнера коллекции по сравнению с тем, как это сделано в основной коллекции Niupera, о чем мы говорили выше. Например, в списке документов *titles a–z* показаны документы, принадлежащие к одним и тем же сериям, поэтому эти документы повторяются, не различаясь по таким важным параметрам, как дата их выпуска. Будет лучше, если вы сгруппируете их по заголовкам серий и покажете даты внутри каждой из этих групп. С помощью классификатора *AZCompactList* это можно выполнить гораздо более тщательно, чем с помощью *AZList*, а затем уже использовать оператор форматирования *VList* для придания коллекции улучшенного вида.

Другая неожиданность вас поджидает, когда вы доберетесь до самой газеты. Вы увидите, что на экране демонстрируется лишь связанный с нею текст. Открывая любую из газет серии *Te Waka o Te Iwi*, в окне просмотра документа вы видите сообщение *This document has no text (В этом документе нет текста)*. Не показывается и сама информация по этому отсканированному изображению (данные по разрешению просматриваемого изображения или что-то в этом роде), хотя она и была подготовлена заранее и размещена в коллекции вместе с самим документом. Эту проблему можно решить с помощью оператора форматирования, который меняет установки по умолчанию, определенные для *DocumentText*.

Упражнения, следующие ниже, продемонстрируют вам приемы дальнейшей настройки и усовершенствования дизайна вашей коллекции.

Упражнение: Усовершенствуем коллекцию изображений

1. На панели **Design** и непосредственно под разделом **Browsing Classifiers** из списка **AZList** удалите классификаторы для *ex.Source* и *ex.Title*.

2. Теперь добавьте **AZCompactList** для *ex.Title* и **DateList** для *ex.Date*.

3. **Измените (Modify)** оператор форматирования для **VList**. Найдите ту часть выражения по умолчанию, в котором говорится

```
{If}{[ex.Source],<br><i>([ex.Source])</i>}
```

и измените его на

```
{If}{[ex.Date],: [ex.Date]}
```

Это повлечет за собой изменение формы представления извлеченной информации, если таковая имеется.

4. В конце этого оператора форматирования, где говорится:

```
</td>
```

допишите

```
{If}{[numleafdocs],<td>([numleafdocs] items)</td>}
```

Следствием использования классификатора **AZCompactList** будет то, что при просмотре заголовков на экране появятся пиктограммы книжной полки. Это обновление оператора форматирования также выразится в том, что в квадратных скобках будет показано количество документов, размещенных на конкретной книжной полке. Это действие становится возможным потому, что метаданные в данном случае определяются **[numleafdocs]** исключительно для пиктограмм книжных полок.

5. Оставаясь в разделе **Format Features** на панели **Design**, выберите из списка команд “Choose Feature” команду **DocumentText**. Строка форматирования HTML, находящаяся в ней, пустая, что означает возможность демонстрировать документ по умолчанию в виде простого текста или, если текста там нет, выводить сообщение “This document has no text” («В этом документе нет текста»). Измените это на:

```
<center>
  <table width=_pagewidth_>
    <tr>
      <td valign=top>[srclink] [screenicon] [/srclink]</td>
      <td>[Text]</td>
    </tr>
  </table>
</center>
```

(доступный текст находится в *niupepa\doc_tweak.txt*)

Использование пиктограммы для экрана [screenicon] будет влиять на размер связанного с нею изображения, что будет сгенерировано путем переключения команды определения размера экрана в приложении PagedImgPlug. Эта пиктограмма будет иметь ссылку на исходное изображение с помощью конструктора [srclink]...[/srclink].

6. Переключитесь на панель **Create**; создайте (**build**) и просмотрите (**preview**) созданную вами коллекцию.
7. По желанию, вы можете добавить лого и изменить цвет фона, как вы это уже делали, выполняя дополнительное упражнение почти в самом конце Раздела 4.2. Вы найдете подходящую пиктограмму в файле piurera\images, который можно запустить с помощью *macros\extra.dm*.

В коллекции, которую вы только что создали, газеты группируются по названиям серий и датам, которые позволяют отличить разные выпуски газет, входящих в одну и ту же серию. Пользователи могут просматривать их в хронологическом порядке, по дате, причем слева в режиме предварительного просмотра будет выводиться исходная страница в версии с высоким разрешением, а справа – простая текстовая версия этой же газеты (если это возможно).

Как и любое другое упражнение, включенное в этот курс, целью данного упражнения было проиллюстрировать шаг за шагом все возможности программы Greenstone. Становясь все более и более опытным, вы сами не заметите, как начнете со всем легко справляться. Не забывайте, что если вы строите свою коллекцию на основе уже существующей, то она автоматически наследует все параметры этой коллекции. Если вы выполнили это упражнение, то при желании вы можете создать совсем новую коллекцию, состоящую из изображений отсканированных текстов, но в основу ее положить именно эту, созданную вами в ходе выполнения упражнения, коллекцию.

Пример из практики: Как работают пользователи электронной библиотеки?

Об интересах читателей традиционной библиотеки можно догадаться по тем книгам, которые они там берут. В электронных каталогах или электронных библиотеках об этом свидетельствуют запросы пользователей и те документы, к которым они обращаются, причем и то, и другое можно записать или запротоколировать (logged). Эти записи, называемые записями транзакций (transaction logs), могут помочь вам определить, каким именно образом используется данный электронный ресурс. Очень часто оказывается, что на практике эти ресурсы используются совершенно не так, как это было задумано разработчиками.

Записи транзакций в электронной библиотеке помогут вам проследить и оценить влияние, которое оказывают такие ресурсы на сообщества учащихся и учителей. При разработке системы очень важно определить, будут ли иметь сотрудники этой библиотеки доступ к такой полезной информации. Необходимо также разобраться в том, как действующее на местах законодательство регулирует деятельность по сбору и использованию информации о работе индивидуальных пользователей в библиотеке. Обычно ценную информацию можно получить без персональной идентификации пользователей.

На рисунке, показанном ниже, приведен небольшой фрагмент протокола работы пользователей с Greenstone, в котором записано то, что происходило в апреле 2003 года. Этот протокол взят с сайта Новозеландской электронной библиотеки (New Zealand Digital Library) – <http://www.nzdl.org>, но он был обработан таким образом, что в протоколе была удалена вся персонифицированная информация о пользователях. Запись этой транзакции осуществляется каждый раз при генерации веб-страницы. По выдержкам из записи транзакций отслеживают деятельность каждого отдельного пользователя, а именно по тем запросам веб-страниц, к которым пользователь обращается во время прохождения курса.

В протоколе записи транзакции записывается доменное имя компьютера, запрашивающего веб-страницу, или, если оно отсутствует, его IP (Internet Protocol) адрес, последовательность цифр, которая идентифицирует данный компьютер. Это именно та часть, которая была преобразована в анонимную выборку как последовательность символов 79ec8da3e7dc19b0. Далее в алфавитном порядке следует список всех аргументов, которые могут быть использованы в указателях URL библиотеки Greenstone. Их очень много – более 100! – но для каждого отдельного запроса страниц, из показанных в примере, большинство не используется, и поэтому эти аргументы пустые.

Первая запись в протоколе транзакции (которая занимает более десятка строк) показывает их все, но после нее идут только постоянные аргументы. В этих записях протокол действия пользователей представляется не в последовательном порядке, а попеременно с действиями других пользователей, которые в это же самое время в разных уголках мира работали с Greenstone.

Формат записи транзакции Greenstone задан главным образом под автоматическую, а не ручную обработку данных. Программа, выполняющая анализ действий пользователя, может генерировать обобщенную статистику, такую, как пропорция транзакций, по которым были сделаны запросы, доступ к документам, запросы помощи и т.д. Хотя эта запись и выглядит как шифр, проявив небольшое терпение, вы вполне можете извлечь из него нужную вам информацию.

Эта последовательность действий показывает, как пользователь входит в Библиотеку развития гуманитарных наук (Humanity Development Library) (с=hdl: где с — это обобщенное название коллекции, а hdl — сокращенное название именно этой коллекции), начиная со страницы Об этой коллекции (About this collection) (р=about). Язык интерфейса — английский (l=en), и у пользователя запущен веб-браузер под систему Windows NT версия 5.1. В следующей записи транзакции (которая приходится примерно на 15-ю строку из числа показанных в данном окне, поскольку первая запись очень длинная) они переходят к поиску с помощью классификатора (classifier) (cl=CL1, а затем в следующей записи — CL1.4) перед тем, как получить документ (d=HASH0157382f65cca62098138be4), расположенный в пределах данной части классификатора. Вы можете увидеть, сколько именно времени пользователь провел, получив доступ к тому или иному ресурсу. Перескочив на самую последнюю запись, показанную в данном протоколе (Wed Apr 02 14:10:47 +1200 2003), вы сможете найти там запрос (a=q обозначает, что Greenstone выполняет запрос — query) по тексту dig well.

Пример из практики записи транзакции

```
/cgi-bin/library 79ec8da3e7dc19b0 [Wed Apr 02 13:57:08 +1200
2003] (a=p, b=0, b1=0, b2=0, bclaboutdesc=, bclcfgchanged=0,
bclclone=0, bclclonechanged=0, bclclonecol=,
bclcontactemail=, bcldirname=, bcldelete=0, bcleconf=0,
bclesrce=0, bclfromsrce=0, bclfullname=, bclinfochanged=0,
bclinput=, bclinputnum=3, bclinputtype=, bcltmp=, bcp=,
beu=, bft=, bl=english, bnu=, bp=, bt=0, c=hdl, cc=, ccp=0,
```

```
ccs=0, cfgfile=, cl=, cm=, cq2=, ct=0, d=, de=, debc=0,
ds=, dsbc=0, e=, el=prompt, er=, f=0, fc=1, fqa=0, fqc=,
fqf=, fqk=, fqn=4, fqs=, fqv=, g=Document, gc=0, gt=0, h=,
h2=, hd=0, hl=1, hp=, hs=0, il=1, j=, j2=, k=1, ky=, l=en,
m=50, n=, n2=, nl=, o=20, p=about, pc=, pfd=0, pfe=0, pfl=0,
pld=10, ple=10, pll=10, ppnum=0, pptext=, pw=, pxml=0, q=,
q2=, qb=0, qt=0, r=1, rd=0, s=0, st=1, t=1, tlng=french,
u=0, ua=, uan=, ug=, uma=listusers, umc=, umnpw1=,
umnpw2=, umpw=, umug=, umun=, umus=, un=, us=invalid, v=0,
w=utf-8, x=0, z=084bf801ce71d63f) "Mozilla/4.0 (compatible;
MSIE 6.0; Windows NT 5.1
; .NET CLR 1.0.3705)"

/cgi-bin/library 79ec8da3e7dc19b0 [Wed Apr 02 13:57:20 +1200
2003] (a=d, ... c=hdl, ... cl=CL1, ... z=084bf801ce71d63f)
"Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; .NET
CLR 1.0.3705)"
/cgi-bin/library 79ec8da3e7dc19b0 [Wed Apr 02 13:57:48 +1200
2003] (a=d, ... c=hdl, ... cl=CL1.4, ... z=084bf801ce71d63f)
"Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; .NET
CLR 1.0.3705)"

/cgi-bin/library 79ec8da3e7dc19b0 [Wed Apr 02 13:58:10
+1200 2003] (a=d, ... c=hdl, ... cl=CL1.4, ... d=HASH0157
382f65cca62098138be4, ... z=084bf801ce71d63f) "Mozilla/4.0
(compatible; MSIE 6.0; Windows NT 5.1; .NET CLR 1.0.3705)"
...
/cgi-bin/library 79ec8da3e7dc19b0 [Wed Apr 02 14:10:47
+1200 2003] (a=q, ... c=hdl, ... q=dig well, ...
z=084bf801ce71d63f) "Mozilla/4.0 (compatible; MSIE 6.0;
Windows NT 5.1; .NET CLR 1.0.3705)"
```

Режим записи протокола в Greenstone можно включать и выключать, и, как правило, он всегда отключен во время установки программы. Однако, в той версии Greenstone, которая поставляется на компакт-диске с данным курсом, эта функция включена по умолчанию, поэтому, когда вы до нее доберетесь, вы можете посмотреть протокол записи всего того, что вы до настоящего времени выполняли в рамках данного курса. Этот файл с протоколом находится в папке `greenstone\etc\usage.txt`

Запись транзакций контролируется с помощью информации, заносимой в файл конфигурации Greenstone – *main.cfg*, который постоянно находится в директории *etc*. В файле есть инструкции о том, как установить нужный режим записи протокола: установить *logcgiargs* на true для того чтобы запись протокола была активизирована и вы получали нужную вам информацию.

Задание 3

Прочитайте статью Джехада Наджара и др. «Поведение пользователей при работе с хранилищами учебных объектов: эмпирический анализ» для того, чтобы узнать, как можно выполнить анализ записи транзакций, произведенной в образовательной электронной библиотеке ARIADNE, упомянутой в Задании 1.

- Считаете ли вы, что этот материал пригодится вам при проведении оценки какой-либо образовательной электронной библиотеки?
- На какие интересные вопросы вы могли бы ответить с его помощью?
- На какие вопросы о пользователях вы не сможете найти ответы в процессе изучения записи транзакций?
- Можете ли вы предложить свои собственные решения, позволяющие организовать автоматическую запись протоколов работы пользователей в электронной библиотеке?
- Поможет ли эта функция пользователям при поиске нужных им ресурсов?

Пример из практики: Проект Domesday

Мы завершаем эту Главу небольшим предупреждением. Создание электронных библиотек может быть очень занимательным и полезным делом, но сам факт создания электронной коллекции еще не гарантирует ее длительного существования. В работе над проектом *Domesday* (призносится как *Dooms-day*), осуществляемой компанией BBC (British Broadcasting Corporation) при поддержке правительства Великобритании и Европейского сообщества, был создан великолепный информационный ресурс, но не была определена долгосрочная перспектива.

Проект *Domesday* был знаменательным мультимедийным ресурсом, созданным в преддверии празднования 900-й годовщины появления оригинала книги *Domesday*. Учащиеся школ и исследователи из разных частей страны собрали значительный массив материалов, который был записан на двух специальных видео дисках.

Источник: <http://www.si.umich.edu/CAMILEON/domesday/what.html>

Задание 4

Прочитайте статью Дугласа Брауна «Потерянный в киберпространстве» для получения более подробной информации о проекте Domesday.

- Ко всем ли данным, полученным вами в ходе выполнения предыдущих проектов, вы можете и сегодня получить доступ?
- Какие шаги вы можете предпринять для того, чтобы обеспечить доступ к полученным вами данным в будущем?

В газетной статье, написанной в 2002, с иронией подмечается, что доступ к материалам, написанным в одиннадцатом столетии, получить гораздо проще, чем к тем, что были созданы в 1986 году:

«Все дело в том, что специальные компьютеры, которые были созданы для проигрывания 12-ти дюймовых видеодисков, содержащих исторические тексты, фотографии, карты и архивные видеозаписи о жизни Британии, безнадежно устарели.

Как результат – никто не может получить доступа к значительному объему информации, собранной в рамках данного проекта (а это по объему равно нескольким томам энциклопедий) и отражающей в наиболее полной мере все стороны жизни британской нации в 1986 году. И в противоположность этому, оригинал книги – *Domesday Book* – инновационная разработка Англии XI столетия, была составлена в 1086 году норманскими монахами и до сих пор хранится в прекрасном состоянии в Офисе гражданских записей в Кью Кец (Public Record Office, Kew) и может быть прочитана любым человеком, умеющим читать и имеющим право работать с этой книгой.»

From “Digital Domesday Book lasts 15 years not 1000”
by Robin McKie and Vanessa Thorpe,
The Observer, Sunday March 3, 2002.

<http://books.guardian.co.uk/news/articles/0,6109,661585,00.html>

Предполагалось, что в рамках проекта *Domesday* в 1986 году будет создан ресурс, который впоследствии станет широко использоваться для образовательных целей всеми последующими поколениями. И, несмотря на все усилия участников проекта по созданию этого ресурса, ему грозит серьезная опасность исчезнуть навсегда.

В проблеме сохранения проекта *Domesday* есть две стороны: техническая и юридическая. С технической точки зрения, одной из стратегий сохранения ресурса может быть создание программы, которая

позволит современным компьютерам эмулировать аппаратное обеспечение, на котором *Domesday* сможет работать, то есть сможет запускать исходную программу для получения нужных документов, их демонстрации и повторного использования на основе оригинального интерфейса. Проект, получивший название CAMiLEON*, тестирует техническую осуществимость и эффективность подобной эмуляции для сохранения интеллектуального контента и внешнего вида этого материала.

Юридические проблемы более серьезны. Многие владельцы авторских прав на материалы, собранные в рамках проекта *Domesday*, разрешили использовать их именно для записи на диски данного проекта, а именно:

- профессиональные фотографии,
- фотографии с национальных конкурсов,
- текст, опубликованный в таких первоисточниках, как газеты и журналы,
- серии теленовостей и обзоров спортивных соревнований,
- карты, лицензированные военной инспекцией Великобритании (UK Ordnance Survey).

Но не было сделано ни одной записи, указывающей, кто именно передал эти материалы и на каких условиях эти данные могут быть использованы. Вопросы защиты интеллектуальной собственности в отношении различных аспектов этой работы являются камнем преткновения в связи с возвращением *Domesday* в число доступных всем ресурсов. Большинство (если не все) компонентов, включая программное обеспечение и сам контент, все еще находятся под защитой авторского права. Одним из направлений работ по восстановлению данного ресурса будет определение владельцев, в отношении которых действуют данные права и получение их разрешений на копирование, изменение или эмулирование этих компонентов. Эта проблема далека от тривиальной: было установлено, что в этом проекте приняло участие более миллиона человек, хотя их участие и может отличаться по размеру вклада!

В решении общей проблемы сохранения электронных ресурсов нет простых решений. Однако, с технической точки зрения программы с открытым источником, которые базируются на открытых стандартах, могут дать гораздо больше шансов для обеспечения долгой жизни вашей информации, чем закрытое коммерческое программное обеспе-

* <http://www.si.umich.edu/CAMILEON/domesday/domesday.html>

чение. У многих пользователей есть старые файлы с документами, которые больше невозможно прочитать, поскольку программы, с помощью которых они были созданы, устарели и перестали использоваться. С расшифровкой подобных файлов может справиться только профессиональный шифровальщик. А при использовании программ с открытым источником, по крайней мере, всегда можно получить доступ к этому источнику, являющемуся ключом к шифру материалов, которые в ином случае просто оказались бы бессмысленными битами нечитаемой информации.

Заключение

Мы прошли вместе с вами длинный путь изучения использования электронных библиотек в образовании. В этом заключительном обращении к вам мы еще раз хотим подчеркнуть, что этот курс посвящен не изучению прошлого, а построению будущего, лучшего будущего для учащихся и учителей.

Электронные библиотеки являются организованными коллекциями, содержащими информацию. Наш опыт работы с Всемирной паутиной, World Wide Web, показывает, что она полна жизни, но при этом бессистемна и никем не контролируется, а, следовательно, с каждым днем она все более способствует появлению еще больших объемов такой же неорганизованной информации. В то же самое время опыт использования открытых для всех каталогов онлайн-публичных библиотек, доступных непосредственно с рабочего стола компьютера пользователя, свидетельствует о следующем. Работа с безусловно сделанной, но очень жестко регламентированной библиотечной системой, которая не в состоянии приблизить пользователя к реальным документам, до которых можно было бы дотронуться рукой, только усиливает разочарование пользователей. Это разочарование становится еще большим в связи с отсутствием правильной организации процесса доступа к информации, интуитивно понятного и доступного всем пользователям.

Электронные библиотеки допускают вас к своим ресурсам по-разному. Технология исключительно нова — ей не больше десяти лет, а идея, что люди могут создавать библиотеки, используя клиентское программное обеспечение, еще новее. Вы можете посмотреть на них с более высокого уровня, и тогда вы увидите электронные библиотеки как грандиозный национальный или интернациональный ресурс. Опустившись ниже, вы увидите их как рядовые, ничем не примечательные

коллекции, созданные индивидуально или группами разработчиков в соответствии с потребностями местных сообществ. Безусловно, и те, и другие библиотеки сосуществуют в одном пространстве. Стандартные протоколы регулируют удаленный доступ ко многим локальным коллекциям, формируя большие по объему ресурсы из тех небольших частичек, что создаются индивидуально, и с любовью и заботой поддерживаются отдельными людьми. Другие протоколы позволяют получить материал из больших централизованных электронных библиотек и соединить их с имеющимися у вас документами для создания именно той коллекции, которая вам нужна.

Не теряя из виду главное стратегическое направление, этот курс сосредоточил основное внимание на проблемах создания библиотек локального уровня: на том, как вы, преподаватель, сможете создать свою собственную электронную библиотечную коллекцию и использовать ее в курсе, который вы преподаете и, если вы хотите, предоставить также и вашим студентам возможность создавать их собственные коллекции. Когда наши учителя сами учились в школе и университете, а это было много лет назад, учебники стали их лучшими друзьями. Они потратили немало лет на чтение этих учебников, их изучение и аннотирование. Многие до сих пор хранят эти учебники на книжных полках – без сомнения, они устаревают, но, несмотря на это, до сих пор являются бесценной частью интеллектуального багажа. Несмотря на все прошедшие годы, они время от времени продолжают обращаться к этим старым книгам. Они знают их так хорошо, что когда нужно срочно вспомнить какой-то забытый факт или формулу, то они незамедлительно находят в них то, что нужно.

Нам, авторам этого курса, очень повезло: мы могли позволить себе покупать книги и хранить их. Сегодняшним студентам везет меньше, и не из-за того, что они уже не могут позволить себе купить нужные книги, а потому что большая часть материала, который их учителя предлагают им почитать, находится в WWW. Будет ли он там через несколько десятилетий, если студентам захочется вновь обратиться к этим ресурсам за подсказкой, и будет ли доступ к этим ресурсам удобен? Вряд ли. А многим другим студентам не везет, поскольку они не только не могут позволить себе купить нужные книги, но и потому что у них нет доступа к Вебу. Есть мнение, что электронные библиотеки будут способствовать выравниванию сложившейся ситуации с доступом к информации. Что получится, если каждый студент, завершивший обучение в школе или университете, будет иметь на руках электронную библиотечную коллекцию, включающую весь материал, который он когда-либо изучал, все книги, все записи по каждому курсу, причем в удобной, компактной, надежной и долговечной форме?

Модуль 5. Открытые стандарты и примеры из практики

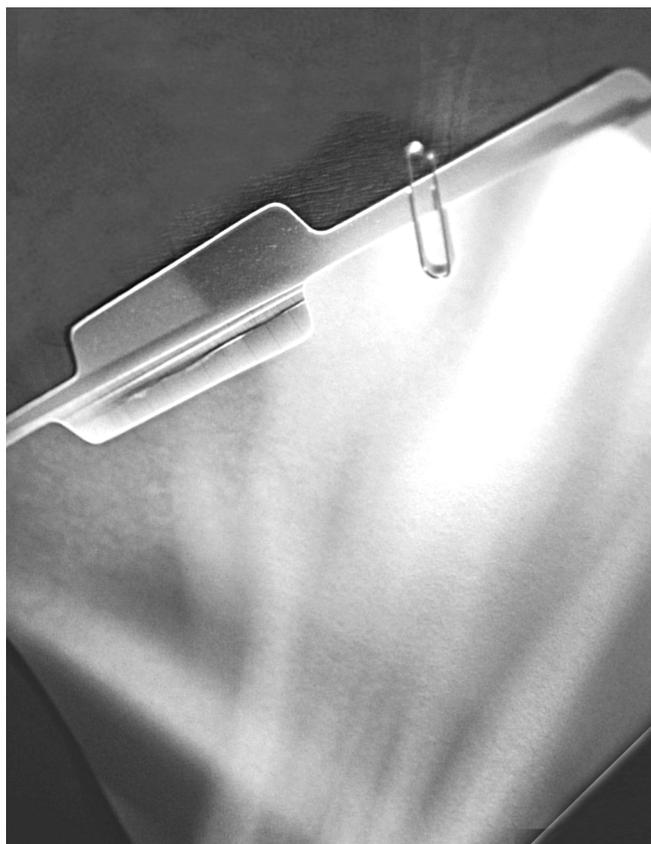
Вот те новые идеи, над которыми стоит задуматься. В действительности, в таких главах курса, как эта, мы стараемся вдохновить вас примером того, что уже сделано другими, хотя найти такие примеры, которые проиллюстрировали бы сказанное нами, было нелегко. По правде говоря, было очень трудно вообще найти хотя бы какие-нибудь примеры удобных в использовании и интересных для изучения образовательных электронных библиотек.

Такова проблема будущего, которое вы будете создавать своими руками, и, которое теперь зависит и от вас.

ПРИЛОЖЕНИЯ

Приложение А.
Глоссарий

Приложение Б.
Библиография, журналы
и веб-сайты



Приложение А

Глоссарий

ASCII	Американский стандартный код обмена информацией — ASCII (сокращ. от American Standard Code for Information Exchange), стандартный 7-ми битный код, появившийся в 1968 году для представления символов латинского алфавита, цифр и специальных символов
Boolean query	Логический (булев) запрос; логический информационный запрос в поисковой системе, который может включать операторы AND, OR и NOT
Browsing	Просмотр коллекции путем сканирования организованного списка значений метаданных, ассоциированных с документами (такими, как имя автора, заголовок, дата, ключевые слова)
Classifier	Классификатор, или кодовый модуль Greenstone, формирующий просмотрный каталог на основе метаданных, имеющихся в документе
Collection	Коллекция; группа документов, объединенных единой системой поиска и просмотра документов
CSS	Наборы многоуровневых стилей (сокращ. от Cascading Style Sheets); способ управления формой представления документов формата HTML и XML
Digital library	Электронная библиотека; коллекция цифровых объектов (текст, аудио, видео), включая также комплекс средств и технологий доступа, получения и поиска информации, а также средств организации и поддержки коллекции
Document	Документ; основной компонент, из которых составляется электронная библиотечная коллекция; может быть представлен в форме текста, графики, аудио или видео и т.д.
DTD	Определение типа документа (сокращ. от Document Type Definition); описание, используемое в XML (и также в SGML) для определения структуры особых наборов документов
Dublin Core	Международный стандарт минимального набора данных, созданный для описания метаданных веб-ресурсов

Encapsulated PostScript	Инкапсулированный вариант формата PostScript, разработанный для отражения графических изображений на одной странице или ее части с целью последующего включения в те или иные документы
Format string	Форматирующая строка; строка, описывающая вид документов и каталогов, представляемых в Greenstone
GIF	Формат представления данных (сокращ. от Graphics Image Format) – популярный способ сжатия графических изображений без потерь, созданный в 1987 г.
GNU Public License	Общественная лицензия GNU; лицензия на использование программного обеспечения, которая разрешает пользователям копировать, свободно распространять и модифицировать компьютерные программы на условиях, что все эти модификации также будут свободно распространяться
Greenstone	Название программы по созданию электронных библиотек, которая описана в данной книге (greenstone.org)
HTML	Язык разметки гипертекста (сокращ. от HyperText Markup Language), т.е. язык, на котором написаны веб-документы
HTML Tidy	Программная утилита, которая позволяет конвертировать устаревшие форматы HTML в XHTML
HTTP	Протокол передачи гипертекста (сокращ. от Hypertext transfer protocol)
Importing	Импортирование; процесс переноса коллекций документов в систему Greenstone
Index	Индекс, каталог; информационная структура, используемая для поиска объектов в коллекции или для ее просмотра
ISBN	Международный стандартный номер книги (сокращ. от International Standard Book Number)
JPEG	Стандарт на сжатие с потерями полноцветных неподвижных изображений, называемый по названию компании “Joint Photographic Experts Group”
JPEG-2000	Последняя версия стандарта сжатия изображений JPEG
LCSH	Сокращ. от Library of Congress Subject Headings, предметные рубрики Библиотеки Конгресса США; утвержденный словарь для назначения предметных дескрипторов
LOM	Сокращ. от Learning Object Metadata; стандарт метаданных для учебных объектов
MARC	Сокращ. от Machine-readable cataloging; формат для машиночитаемых каталогов; схема метаданных, разработанная в конце 60-х гг. для использования профессиональными составителями библиотечных каталогов

Metadata	Метаданные; структурированная информация об авторе, заголовке, дате, ключевых словах и т.п., ассоциированная с конкретным документом (или коллекцией документов)
METS	Схема METS, являющаяся стандартом кодирования описательных, структурных и служебных метаданных для объектов, хранящихся в электронной библиотеке
MIDI	Сокращ. от Musical Instrument Digital Interface; проигрывание музыки с помощью музыкальных синтезаторов
Mirror	Зеркало; процесс копирования веб-сайта или его части, и его переноса на другой, доступный для пользователей, компьютер
MIME	Многоцелевые расширения почтовой службы в Интернет (сокращ. от Multipurpose Internet Mail Extensions); стандарт для включения различных типов файлов — текстовых, графических, аудио, видео или данных прикладных программ, например, в электронных сообщениях
MODS	Сокращ. от Metadata Object Description Schema; схема описания метаданных объекта
MPEG	Стандарт представления мультимедийного материала, названный по имени компании “Motion Picture Experts Group”
New Zealand Digital Library Project	Исследовательский проект Департамента компьютерных наук Университета Вайкато (Computer Science Department at the University of Waikato), Новая Зеландия, создавшего программу Greenstone (<i>nzdl.org</i>)
OAI	Сокращ. от Open Archives Initiative; название протокола, созданного для эффективного распространения библиотечного контента
OCR	Оптическое распознавание символов (сокращ. от Optical Character Recognition); процесс создания оцифрованной версии текста в виде его изображения
PDF	Формат переносимого документа (сокращ. от Portable Document Format); язык описания страниц, предназначенный для интерактивного использования и являющийся преемником языка PostScript
Perl	Язык программирования, используемый при выполнении многих действий по обработке текста при использовании Greenstone
Plug-in	Подключаемая программа, встроенное приложение; ключевой модуль для обработки документов, представленных в различных форматах, использующийся в процессе импортирования и создания коллекций

PNG	Переносимая сетевая графика (сокращ. от Portable Network Graphics); открытый стандарт для графики, переносимой без потерь
PostScript Protocol	Первый язык описания страниц, появившийся в 1985 году
RDF	Протокол; набор правил, в соответствии с которым две системы могут обмениваться друг с другом информацией (например, в Greenstone это сервер коллекции и реципиент)
RDF	Сокращ. от Resource Description Framework; схема, разработанная для управления интероперабельностью метаданных
RTF	Расширенный текстовый формат (сокращ. от Rich Text Format); стандартный формат обмена текстовыми документами
SCORM	Модель совместного использования связанных учебных объектов (сокращ. от Sharable Content Object Reference Model); стандарт, состоящий из множества компонентов, описывающих использование целого ряда существующих стандартов, включая LOM.
SGML	Язык обобщенной разметки документов (сокращ. от Standard Generalized Markup Language); метаязык для описания форматов разметки, который был стандартизован в 1986 году и положен в основу XML
SQL	Язык структурированных запросов (сокращ. от Structured Query Language); язык, использующийся как промышленный стандарт при формировании запросов в базах данных
SVG	Сокращ. от Scalable Vector Graphics; основанный на XML, формат отображения векторной графики на веб-страницах
TEI	Инициатива по кодированию текста (сокращ. от Text Encoding Initiative); проект, начатый в 1987, по разработке языка SGML DTD для представления учебных текстовых материалов по гуманитарным и общественно-политическим дисциплинам
TIFF	Файловый формат тега для изображений (сокращ. от Tagged Image File Format); широко использующийся файловый формат для растровых изображений со встроенными средствами формирования описательных метаданных
UCS	Набор символов Unicode (сокращ. от Unicode Character Set); набор символов, поддерживающих формат Unicode
Unicode	Стандартная схема представления символов, использующихся в национальных алфавитах

URI	Унифицированный идентификатор ресурса (сокращ. от Uniform Resource Identifier); общее название для всех URL и URN
URL	Унифицированный указатель местонахождения ресурса (сокращ. от Uniform Resource Locator); стандартный способ адресации объектов, размещенных в Вебе (но этот термин уже считается устаревшим, поскольку ему на смену появился URI)
URN	Унифицированное имя ресурса (сокращ. от Uniform Resource Name); способ наименования ресурса без включения информации о его местонахождении
UTF	Формат преобразования (сокращ. от UCS Transformation Format); схема представления символов в формате Unicode в одном из трех вариантах: UTF-32, UTF-19 и UTF-8
XHTML	Современная версия языка HTML, которая основывается на строгих синтаксических правилах языка XML
XML	Расширяемый язык разметки (сокращ. от Extensible Markup Language); метаязык для описания форматов структурированных документов и данных, размещаемых в Вебе
XML Schema	Способ определения структуры документов особого вида, для создания которых использовались более сложные средства формирования структуры и типов данных, нежели в DTDs
Z39.50	Международный стандарт коммуникационного протокола, разработанный для использования в системах библиотечных каталогов

Приложение Б

Библиография, журналы и веб-сайты

Библиография включает статьи, на которые есть ссылки в материалах курса, включая материалы для чтения по данному курсу, а также те библиографические ссылки, которые авторы рекомендуют для дальнейшего самостоятельного чтения по данной теме.

Список литературы

- Apperley, M., Keegan, T. Cunningham, SJ and Witten, I.H. (2002) Delivering the Maori-Language Newspapers on the Internet, In *Rere Atu, TakuManu! Discovering History, Language & Politics in the Maori-Language Newspapers*, edited by Jennifer Curnow, Ngapare Hopa & Jane McRae. Auckland University Press, 211-232.
- Brown, D. (2003) *Lost in Cyberspace: The BBC Domesday Project and the Challenge of Digital Preservation*, Cambridge Scientific Abstracts; June. <http://www.csa.com/hottopics/cyber/oview.html>
- Drake, M.A. (2004) Institutional Repositories: hidden treasures, *Journal Searcher*, 12(5); May. <http://www.infotoday.com/searcher/may04/drake.shtml>
- Duval, E. (2004) Learning technology standardization: making sense of it all, *International Journal on Computer Science and Information Systems*, No. 1, pp. 33–43. <http://www.comsis.fon.bg.ac.yu/ComSISpdf/Volume01/InvitedPapers/ErikDuval.pdf>
- Duval, E. and Hodgins, W. (2003) A LOM research agenda, *Proc International Conference on World Wide Web*, edited by G. Hencsey, B. White, Y. Chen, L. Kovacs and S. Lawrence, pp. 1–9. <http://www2003.org/cdrom/papers/alternate/P659/p659-duval.html>
- Friesen, N. Mason, J. and Ward, N. (2003) Building educational metadata application profiles. *Proc International Conference on Dublin Core and Metadata for e-Communities*, pp. 63–69. <http://www.bncf.net/dc2002/program/ft/paper7.pdf>
- Gartner, R.(2002) *METS: Metadata Encoding and Transmission Standard*, Oxford University Library Services, JISC. http://www.jisc.ac.uk/uploaded_documents/tsw_02-05.pdf

- Gartner, R.(2003) MODS: *Metadata Object Description Schema*, Oxford University Services, JISC.
http://www.jisc.ac.uk/uploaded_documents/tsw_03-06.pdf
- Marchionini, G. and Geisler, G. (2002) The Open Video Digital Library, *D-Lib Magazine*, 8(12); December.
<http://www.dlib.org/dlib/december02/marchionini/12marchionini.html>
- Marchionini, G. and Maurer, H. (1995) The roles of digital libraries in teaching and learning. *Communications of the ACM* (1995), 38(4), 67–75.
- Masullo, M. and R. Mack. (1996) Roles for digital libraries in K-12 education. *D-Lib Magazine*. 2(8); September.
<http://www.dlib.org/dlib/september96/eduport/09masullo.html>
- Mendel, J.M. (1999) Education using digital libraries, in *WTEC Panel Report on Digital Information Organization in Japan*. 13-22. World Technology Division, International Technology Research Institute, Loyola College, Baltimore, MD, USA.
<http://www.wtec.org/loyola/pdf/dio.pdf>
- Najjar, J. Ternier, S. and Duval, E. (2004) User behaviour in learning objects repositories: an empirical analysis, Proc *ED-MEDIA World Conference on Educational Multimedia, Hypermedia and Telecommunications*, edited by L. Cantoni and C. McLoughlin, pp. 4373-4379.
<http://www.cs.kuleuven.ac.be/~najjar/papers/edmedia2004.pdf>
- Nixon, W.J. (2003) DAEDALUS: Initial experiences with EPrints and DSpace at the University of Glasgow, *Ariadne*, No. 37; October.
<http://www.ariadne.ac.uk/issue37/nixon/>
- Roes, H. (2001) Digital libraries and education: trends and opportunities. *D-Lib Magazine*, 7(7/8); July/August.
<http://www.dlib.org/dlib/july01/roes/07roes.html>
- Wallace, R., Krajcik, J. & Soloway, E. (1996) Digital libraries in the Science Classroom: Opportunities for Inquiry, *D-Lib Magazine*. 2(8); September.
<http://www.dlib.org/dlib/september96/umdl/09wallace.html>
- Witten, I.H. and Bainbridge (2003) *How to Build a Digital Library*, Morgan Kaufmann, CA.
- Witten, I.H., Bainbridge, D., Tansley, R., Huang, C-Y. and Don, K.J. (2005) StoneD: A bridge between Greenstone and DSpace. Working Paper 2005/02, Department of Computer Science, University of Waikato, New Zealand.

Библиография

- ABBYY Software (2000) *FineReader User's Guide*. ABBYY Software, 123015 Moscow, P.O. 72, Russia.
- Adobe Systems Incorporated (1985) *PostScript Language Tutorial and Cookbook*. Addison Wesley, Boston, MA.
- Adobe Systems Incorporated (1999) *PostScript Language Reference*. Addison Wesley, Boston, MA, third edition.
- Adobe Systems Incorporated (2000) *PDF Reference*. Addison Wesley, Boston, MA, second edition (version 1.3).
- American National Standards Institute (1968) *American Standard Code for Information Interchange (ASCII)* Standard No. X3.4–1968; updated as X3.4–1986.
- Andrews, N. (1987) Rich Text format standard makes transferring text easier. *Microsoft Systems Journal*, Vol. 2, No. 1, pp. 63–67; March.
- Arms, W.Y. (2000) *Digital Libraries*. MIT Press, Cambridge, Massachusetts.
- Atkinson, R. (1986) Selection for preservation: a materialistic approach. *Library Resources and Technical Services* 30, pp. 344–348; October/December.
- Baeza-Yates, R. and Ribeiro-Neto, B. (1999) *Modern information retrieval*. ACM Press, New York.
- Bainbridge D. and Cunningham S.J. (1998) Making oral history accessible over the World Wide Web. *History and Computing*, Vol. 10, No. 1/3, pp. 73–81.
- Berners-Lee, T., Hendler, J., and Lassila, O. (2001) The Semantic Web. *Scientific American*, Vol. 284, No. 5, pp. 34–43; May.
- Borgman, C.L. (2000) *From Gutenberg to the Global Information Infrastructure: access to information in the networked world*. MIT Press, Cambridge, Massachusetts.
- Bryan, M. (1988) *SGML: An author's guide to the Standard Generalized Markup Language*. Addison Wesley, Boston, MA.
- Chang, S.J. and Rice, R.E. (1993) Browsing: a multidimensional framework. *Annual Review of Information Science and Technology*, Vol. 28, pp. 231–276.
- Chapman, N. and Chapman, J. (2000) *Digital Multimedia*. Wiley, New York.
- Chen, S.S. (1998) *Digital Libraries: The life cycle of information*. BE (Better Earth) Publisher, Columbia, MO.
- Committee on Intellectual Property Rights, Computer Science and Telecommunications Board (2000) *The Digital Dilemma: Intellectual Property in the Information Age*. National Academy Press, Washington DC.

- Cooper, M.D. (1996) *Design of Library Automation Systems*. Wiley, New York.
- Cox, I., Miller, M. and Bloom, J. (2001) *Digital Watermarking*. Morgan Kaufmann, San Francisco, CA.
- Crane, G. (1998) The Perseus project and beyond: how building a digital library challenges the humanities and technology. *D-Lib Magazine*, Vol. 4, No. 1; January.
- Crawford, W. and Gorman, M. (1995) *Future Libraries: Dreams, Madness, and Reality*. American Library Association, Chicago, IL.
- Dartois, M., Maeda, A., Sakaguchi, T., Fujita, T., Sugimoto, S. and Tabata, K. (1997) A multilingual electronic text collection of folk tales for casual users using jff-the-shelf browsers. *D-lib Magazine*, Vol. 3, No. 10; October.
- de Stefano, P. (2000) Selection for digital conversion. In *Moving Theory into Practice: digital imaging for libraries and archives*, edited by A.R. Kenney and O.Y. Rieger. Research Libraries Group, Mountain View, CA, pp. 11–23.
- Frakes, W.B. and Baeza-Yates, R. (Editors) (1992) *Information Retrieval: Data Structures and Algorithms*. Prentice Hall, Englewood Cliffs, NJ.
- Friesen, N. (2003) Three Objections to Learning Objects, Learning Objects and Metadata, Kogan, London. http://phenom.educ.ualberta.ca/not_vert_similarnfriesen/.
- Gaines, B.R. (1993) An agenda for digital journals: the socio-technical infrastructure of knowledge dissemination. *Journal of Organizational Computing*, Vol. 3, No. 2, pp. 135–193.
- Gapen, D.K. (1993) The virtual library: Knowledge, society, and the librarian. In *The Virtual Library: Visions and Realities*, edited by L.M. Saunders. Information Today, Medford, NJ, pp. 1–14.
- Giles, C.L., Bollacker, K.D. and Lawrence, S. (1998) CiteSeer: An automatic citation indexing system. *Proc ACM Digital libraries*, Pittsburgh, PA, pp. 89–98; June.
- Goldfarb, C.F. (1990) *The SGML Handbook*. Oxford University Press, New York.
- Gore, D. (Editor) (1976) *Farewell to Alexandria*. Greenwood Press, Westport, CT.
- Gorman, M. and Winkler, P.W. (Editors) (1988) *Anglo-American Cataloguing Rules*, American Library Association, Chicago, IL, second edition.
- Gorn, S., Bemmer, R.W. and Green, J. (1963) American standard code for information interchange. *Communications of the ACM*, Vol. 6, No. 8, pp. 422–426; August.

- Harold, E.R. (2001) *XML Bible*. IDG Books, Boston, MA, Gold edition.
- Hyman, R.J. (1972) *Access to Library Collections: An inquiry into the validity of the direct shelf approach, with special reference to browsing*. Scarecrow Press, Metuchen, NJ.
- Jones, S., McInnes, S. and Staveley, M.S. (1999) A graphical user interface for boolean query specification. *International J Digital Libraries*, Vol. 2, No. 2/3, pp 207–223.
- Kahle, B. (1997) Preserving the Internet. *Scientific American*, Vol. 276, No. 3, pp. 82–83; March.
- Kientzle, T. (1995) *Internet File Formats: your complete resource for sending, receiving, and using internet files*. Coriolis Group, Scottsdale, AZ.
- Kientzle, T. (1997) *A Programmer's Guide to Sound*. Addison Wesley, Boston, MA.
- Korfhage, R.R. (1997) *Information Storage and Retrieval*. Wiley, New York.
- Kuny, T. (1998) A digital dark ages? Challenges in the preservation of electronic information. *International Preservation News*, No. 17; May.
- Lagoze, C. and Fielding, D. (1998) Defining collections in distributed digital libraries. *D-Lib Magazine*, Vol. 4, No. 11; November.
- Lagoze, C. and Payette, S. (2000) Metadata: principles, practices and challenges. In *Moving Theory into Practice: Digital Imaging for Libraries and Archives*, edited by A.R. Kenney and O.Y. Rieger. Research Libraries Group, Mountain View, CA, pp. 84–100.
- Lagoze, C. and Van de Sompel, H. (2001) The open archives initiative: building a low-barrier interoperability framework. *Proc Joint Conference on Digital Libraries*, Roanoke, Virginia, pp. 54–62; June.
- Lesk, M. (2005) *Understanding Digital Libraries*. Second Edition, Morgan Kaufmann, San Francisco.
- Library of Congress (1998) *Library of Congress Subject Headings*. Library of Congress Cataloging Policy and Support Office, Washington DC, 21st edition.
- Lovins, J.B. (1968) Development of a stemming algorithm. *Mechanical Translation and Computation*, Vol. 11, No. 1–2, pp. 22–31.
- Lynch, C. (1999) Canonicalization: A fundamental tool to facilitate preservation and management of digital information. *D-Lib Magazine*, Vol. 5, No. 9; September.
- Mason, J., Mitchell, S., Mooney, M., Reasoner, L. and Rodriguez, C. (2000) INFOMINE: Promising directions in virtual library development. *First Monday*, Vol. 5, No. 6; June.
- Miller, E. (1998) An introduction to the resource description framework. *D-Lib Magazine*, Vol. 4, No. 5; May.

- Miller, P. (Editor) (2000) *D-Lib Magazine Special Issue on Collection-Level Description*, Vol. 6, No. 9; September.
- Murray, J.D. and van Ryper, W. (1996) *Encyclopedia of Graphics File Formats*. O'Reilly and Associates, Sebastopol CA, second edition.
- Nack, F. and Lindsay, A. (1999) Everything you wanted to know about MPEG-7: Part I. *IEEE Multimedia*, Vol. 6, No. 3, pp. 65–77; July–September.
- Nack, F. and Lindsay, A. (1999) Everything you wanted to know about MPEG-7: Part II. *IEEE Multimedia*, Vol. 6, No. 4, pp. 64–73; October–December.
- Paepcke, A., Baldonado, M., Chang, C.-C. K., Cousins, S. and Garcia-Molina, H. (1999) Using distributed objects to build the Stanford digital library Infobus, *IEEE Computer*, Vol. 32, No. 2, pp. 80–87; February.
- Pennebaker, W.B. and Mitchell, J.L. (1993) *JPEG: Still image data compression standard*. Van Nostrand Reinhold, New York.
- Pohlmann, K.C. (2000) *Principles of Digital Audio*. McGraw-Hill, New York, fourth edition.
- Polsani, P.R. (2003) Use and abuse of reusable learning objects, *Journal of Digital Information* Vol. 3 (4); February. <http://jodi.ecs.soton.ac.uk/Articles/v03/i04/Polsani/>.
- Price-Wilkin, J. (2000) Access to digital image collections: system building and image processing. In *Moving Theory into Practice: Digital Imaging for Libraries and Archives*, edited by A.R. Kenney and O.Y. Rieger. Research Libraries Group, Mountain View, CA, pp. 101–118.
- Ranganathan, S.R. (1931) *The Five Laws of Library Science*. Madras Library Association, Madras.
- Rothenberg, J. (1995) Ensuring the longevity of digital documents. *Scientific American*, Vol. 272, No. 1, pp. 42–47; January.
- Rothenberg, J. (1997) Digital information lasts forever – or five years, whichever comes first. Rand Corporation Video V-079.
- Salton, G. and McGill, M.J. (1983) *Introduction to Modern Information Retrieval*. McGraw Hill, New York.
- Salton, G. (1989) *Automatic Text Processing: The transformation, analysis, and retrieval of information by computer*. Prentice Hall, Englewood Cliffs, NJ.
- Samuelson, P. (1998) Encoding the law into digital libraries. *Communications of the ACM*, Vol. 41, No. 4, pp. 13–18; April.
- Samuelson, P. and Davis, R. (2000) The digital dilemma: a perspective on intellectual property in the information age. Presented at the Telecommunications Policy Research Conference, Alexandria, Virginia; September.

- Sanders, L.M. (Editor) (1999) *The Evolving Virtual Library II: Practical and Philosophical Perspectives*. Information Today, Medford, NJ.
- Shank, J.D. (2005, to appear) The emergence of learning objects: the reference librarian's role. *Research Strategies*.
- Sperberg-McQueen, C.M. and Burnard, L. (Editors) (1999) *Guidelines for Electronic Text Encoding and Interchange*. Chicago and Oxford.
- Sun Microsystems (2000) *The Digital Library Toolkit*. Sun Microsystems, Palo Alto, CA. Available at <http://www.sun.com/edu>.
- Svenonius, E. (2000) *The intellectual foundation of information organization*. MIT Press, Cambridge, MA.
- Thiele, H. (1998) The Dublin Core and Warwick Framework: A Review of the Literature, March 1995–September 1997. *D-Lib Magazine*, Vol. 4, No. 1; January.
- Unicode Consortium (2000) *The Unicode Standard, Version 3.0*. Addison Wesley, Materials for reading, MA.
- U.S. Congress (1990) *Taking a Byte out of History: The Archival Preservation of Federal Computer Records*. House Committee on Government Operations Report 101-987, Washington DC.
- van Rijsbergen, C.J. (1979) *Information Retrieval*. Butterworths, London, second edition.
- Weibel, S. (1999) The state of the Dublin Core metadata initiative. *D-Lib Magazine*, Vol. 5, No. 4; April.
- Weller, M.J., Pegler, C.A. and Mason, R.D. (2003) Putting the pieces together: What working with learning objects means for the educator, *Elearn International*. Scotland: Edinburgh; February. <http://iet.open.ac.uk/pp/m.j.weller/pub/>
- White, J. (Editor) (1999) *Intellectual Property in the Age of Universal Access*. ACM Press, New York.
- Witten, I.H., Moffat, A. and Bell, T.C. (1999) *Managing Gigabytes: Compressing and Indexing Documents and Images*. Morgan Kaufmann, San Francisco, CA, second edition.
- Witten, I.H., McNab, R., Jones, S., Cunningham, S.J., Bainbridge, D. and Apperley, M. (1999) Managing complexity in a distributed Digital Library, *IEEE Computer*, Vol. 32, No. 2, pp. 74–79; February.
- Witten, I.H., Loots, M., Trujillo, M.F. and Bainbridge, D. (2001) The promise of digital libraries in developing countries. *Communications of the ACM*, Vol. 55, No. 5, pp. 82–85; May.
- World Bank (2000) *World Development Indicators 2000*. World Bank, Washington DC.

Yeates, S., Bainbridge, D. and Witten, I.H. (2000) Using compression to identify acronyms in text. *Proc Data Compression Conference*, IEEE Press Los Alamitos, CA, p. 582.

Журналы

Computers and Education – Компьютеры и образование

D-Lib Magazine <http://www.dlib.org/> – Журнал D-Lib

The Electronic Library – Электронная библиотека

International Journal of Digital Libraries – Международный журнал по электронным библиотекам

Journal of the American Society for Information Science and Technology – Журнал Американского общества информатики и технологии

Library Hi-Tech – Библиотека высоких технологий

Конференции

American Society for Information Science and Technology (ASIST) Annual Meeting – Ежегодная конференция Американского общества информатики и технологии

Digital Libraries for Knowledge Communities (DLKC) – Электронные библиотеки для сообществ знания

European Conference on Digital Libraries (ECDL) – Европейская конференция по электронным библиотекам

International Conference of Asian Digital Libraries (ICADL) – Международная конференция по азиатским электронным библиотекам

International Conference of Digital Libraries (ICDL) – Международная конференция по электронным библиотекам

Joint Conference on Digital Libraries (JC DL) – Объединенная конференция по электронным библиотекам

Librarian Information Technology Association (LITA) Forum – Форум Ассоциации библиотечно-информационных технологий

Веб-сайты, упомянутые в данном курсе

<http://www.google.com/>

<http://www.greenstone.org/>

<http://ltsc.ieee.org/>

<http://matti.usu.edu/nlvm>

<http://nzdl.org/>

<http://www.w3c.org/>

<http://www.cancore.ca/>

<http://www.careo.org/>
<http://www.dlconsulting.co.nz/>
<http://www.dlese.org/>
<http://www.loc.gov/standards/mets/>
<http://www.merlot.org/>
<http://www.moodle.org/>
<http://www.nzdl.org/niupepa>
<http://www.opensource.org/>
<http://www.reload.ac.uk/>
<http://www.si.umich.edu/CAMILEON/>

Другие рекомендуемые веб-сайты

<http://acm.org/dl/>

Электронная библиотека Ассоциации по компьютерному оборудованию (Association for Computing Machinery, ACM), представляющая собой большую коллекцию материалов научных исследований в области компьютерных наук.

<http://memory.loc.gov/ammem/>

Сайт по американскому наследию (The American Memory) Библиотеки Конгресса США, дающий свободный доступ через Интернет к различным артефактам, отражающим американские традиции: текстам и аудиозаписям, статичным и динамичным изображениям, печатным изданиям, картам и музыкальным альбомам. По сути, этот сайт можно назвать цифровой записью американской истории и творчества.

<http://scholar.google.com/>

Ученый Google (Google Scholar) – это коллекция ресурсов, специально подобранных для академической аудитории. Она включает письменные работы с рецензиями, диссертации, книги, препринты, аннотации и технические доклады, собранные по различным областям знаний.

<http://www.adlnet.org/>

Инициатива по передовому распределенному обучению (The Advanced Distributed Learning (ADL) Initiative), спонсируемая офисом государственного секретаря по обороне США (Office of the Secretary of Defence, OSD). На сайте отражены результаты совместной деятельности правительства, индустрии и научного сообщества по созданию новой распределенной среды обучения, которая допускает интероперабельность средств обучения и содержания учебных курсов в глобальном масштабе.

<http://www.archive.org/>

Архив Интернет (The Internet Archive) – проект создания электронной библиотеки Интернет-сайтов и других культурных артефактов, представленных в цифровом формате. Это один из способов получения (в отдельных случаях) контента с тех веб-сайтов, которые были удалены из Веба.

<http://www.diglib.org/>

Электронная библиотека «Федерация: международная ассоциация библиотек и связанных с ними институтов» (Federation: an International Association of Libraries and Allied Institutions). Ее миссией является поддержка новых исследований и стажировок для ее членов, студентов, аспирантов, взрослой категории обучаемых и других целевых групп граждан путем развития международной сети электронных библиотек.

<http://www.perseus.tufts.edu/>

Персей (The Perseus) – электронная библиотека, которая нацелена на совершенствование доступа к источникам информации по гуманитарным дисциплинам. В самом начале создания этой коллекции она специализировалась на классическом древнегреческом и древнеримском материале, но позже распространилась и на другие области.

ЭЛЕКТРОННЫЕ БИБЛИОТЕКИ В ОБРАЗОВАНИИ

Специализированный учебный курс

ПРАКТИЧЕСКОЕ РУКОВОДСТВО

Корректор Людмила Ермолаева
Дизайн и верстка Екатерина Орехова

Изд. лиц. № 05683 от 24 августа 2001 г.

Уч.-изд. л. 22,7. Печ. л. 15,5. Формат издания 70x100 ¹/₁₆

Печать офсетная

Бумага офсетная № 1. Тираж 500 экз. Гарнитура Kudrashov

Цена договорная. Заказ

ЗАО «Издательский дом «Обучение-Сервис».

113046, Москва, а/я 46