



- ▶ Журнал ЭБ
- ▶ О журнале
- ▶ Редакционная коллегия и редакция
- ▶ Что нового?
- ▶ 2014 год
- ▶ 2013 год
- ▶ 2012 год
- ▶ 2011 год
- ▶ 2010 год
- ▶ 2009 год
- ▶ 2008 год
- ▶ 2007 год
- ▶ 2006 год
- ▶ 2005 год
- ▶ 2004 год
- ▶ 2003 год
- ▶ 2002 год
- ▶ 2001 год
- ▶ 2000 год
- ▶ 1999 год
- ▶ 1998 год

▶ ENGLISH

Электронные библиотеки - 2000 - Том 3 - Выпуск 2

Dublin Core в коробке

М.Е. Шварцман, А.С. Ильин
Российская государственная библиотека

Введение

В связи со стремительным ростом ресурсов Интернет перед библиотекарями встала задача сориентировать своих читателей в новом информационном пространстве. Для этого, прежде всего, нужно самим понять, что есть в Интернете и что из этого нужно читателям конкретной библиотеки. Несмотря на большое количество поисковых машин и справочников, библиотекари все равно продолжают вести свои собственные каталоги избранных ресурсов.

Это происходит по следующим причинам:

- В большом каталоге труднее ориентироваться
- В большом каталоге уже есть какая-то система классификации, и ею приходится пользоваться, хотя она может и не устраивать библиотекаря.
- В большом каталоге приводятся совсем не те сведения о ресурсе, которые нужны библиотекарю
- В большом каталоге зачастую нет именно того, что нужно Вашим читателям.

Хотя во многих системах предусмотрена возможность дополнительного ввода описаний ресурсов Интернет, этим пользуются только создатели ресурсов для увеличения посещаемости, а не библиотекари. С психологической точки зрения это понятно, трудно решиться на значительные трудозатраты для дополнения "АУ" или "RAMBLER". Никто не знает, что будет с ними завтра, и не пропадут ли напрасно все усилия.

Поэтому библиотекари создают своими силами небольшие каталоги. Кто в виде HTML файла на своей домашней странице, кто в виде большого количества закладок в браузере. При небольшом количестве ресурсов этот метод работал, однако при увеличении количества найденных ресурсов появляется задача оптимальной организации найденной информации. Раньше справочно-библиографические службы библиотек вели картотеку выполненных справок, сейчас пришло время организовать аналогичную систему на новом уровне.

В Российской государственной библиотеке (РГБ) при поддержке Российского фонда фундаментальных исследований (РФФИ) с 1998 проводится работа по созданию пакета прикладных программ (ППП) для ведения систематического каталога ресурсов Интернет.

В начале работы мы определили основные требования к ППП

- ППП должен быть прост в установке, использовании и не требовать мощного компьютера,
- ППП должен использовать для описания ресурсов простой набор полей, соответствующий мировым стандартам.
- ППП должен позволять проводить поиск по всем элементам библиографического описания
- ППП должен позволять проводить поиск по любой иерархической классификационной системе, выводя на экран все дерево или его часть.
- ППП должен распространяться бесплатно.
- ППП должен обеспечить поиск в базе данных (БД) и ввод новых описаний, как на локальном компьютере, так и через WEB.
- ППП должен содержать инструменты администрирования БД (поиск дублетов, редактирование, проверка существования ресурсов и т.п.).

Установка ППП

Для установки ППП требуется компьютер с операционной системой Windows 95/98 или Windows NT4.0. Оперативная память должна быть не менее 32МВ, на жестком диске ППП занимает 30 мб. В комплект ППП входит программа инсталляции, которая последовательно проводит пользователя через все этапы процесса инсталляции. Нужно только выбрать вариант установки (полный, частичный выборочный),

директорию для размещения и URL или IP адрес вашего компьютера. Для реализации возможности ввода и поиска через WEB компьютер должен быть подключен к Интернет и на компьютере должна быть установлена программа WEB сервера. В комплект включена свободно распространяема программа Personal Web Server для Windows 95/98. Также в комплект включен ряд свободно распространяемых утилит для администрирования БД.

Формат описания ресурсов

Вопрос, о формате описания был решен в пользу формата Dublin Core Metadata Set. Выбор, надо сказать, был невелик. Можно было использовать какую – либо разновидность формата MARC, придумать свой формат или взять Dublin Core Metadata Set.

В настоящее время о Dublin Core Metadata Set существуют противоречивые мнения. Одни считают его универсальным форматом и предлагают использовать его для каталогизации всех электронных изданий. Другие говорят о бедности набора полей этого формата и невозможности использования.

Истина на наш взгляд, как всегда, посередине. Для каждого формата необходимо определить область применения. Как не существует универсального лекарства, так и нет универсального формата. Ограниченность набора полей оборачивается простотой описания. Мы считаем, что формат описания ресурса должен быть настолько прост, чтобы его мог понять любой человек без специальной подготовки. С этой точки зрения Dublin Core Metadata Set является оптимальным выбором. Внедрение Dublin Core в России только начинается и подробнее о нем можно прочитать на специальном сайте www.rsl.ru/dc.

В Российской государственной библиотеке наш ППП используется для каталогизации ресурсов самими создателями ресурсов. Для этого нами была разработана интерактивная форма с набором полей в соответствии с форматом Dublin Core Metadata Set . Любой создатель ресурса может заполнить поля описания своего ресурса и отослать его в нашу базу данных. Взамен он получает описание своего ресурса в тегах Dublin Core Metadata Set. Ему остается только вырезать кусок текста и описанием и вставить его в свой ресурс между тегами <head> и </head>.

Рис 1. Краткая форма описания ресурса Интернет

Первоначально наш проект задумывался не столько как создание ППП для каталогизации ресурсов, а как инструмент создания метаданных в формате Dublin Core Metadata Set. Поэтому введенные описания ресурсов можно вывести на экран в виде таблицы с описанием:

Имя поля	Схема	Значение
Заглавие		О решениях сингулярно возмущенных задач, имеющих пограничный слой типа всплеска. "Фундаментальная и прикладная математика". Выпуск 1, 1995 г.
Имя индивид. Автора		А.Б. Васильева
Предмет, ключевые слова		22.151.3

Описание		Рассматривается сингулярно возмущенное уравнение второго порядка. Задается краевое условие, при котором решение имеет пограничный слой типа всплеска. Исследуется асимптотика такого решения. (51Kb)
Имя коллект.издателя		МГУ. Механико-математический факультет
Адрес коллект. Издателя		www.math.msu.ru/koi8/welcome.htm
Имя индивид. Соавтора		Пентус Анна
Адрес индивид. Соавтора		apentus@mech.math.msu.ru
Адрес коллект. Соавтора		fpm@mech.math.msu.ru
Дата созд./ модиф.		21.06.97
Тип ресурса		Text, Serial, Journal
Идентификатор	URL	http://www.math.msu.ru/fpm/koi/95/951/95105.htm
Первоисточник		ФУНДАМЕНТАЛЬНАЯ И ПРИКЛАДНАЯ МАТЕМАТИКА 1995, ТОМ 1, ВЫПУСК 1, СТР. 109-122
Составная часть		www.math.msu.ru/fpm/koi/95/95_1.htm

или в виде метаданных в формате Dublin Core Metadata Set

```
<head>
<!-- DC RSL, © 1998,2000 Russian State Library -->
<META NAME="DC.Title" CONTENT="О решениях сингулярно возмущенных задач, имеющих пограничный
слой типа всплеска. "Фундаментальная и прикладная математика". Выпуск 1, 1995 г.">
<META NAME="DC.Creator.PersonalName" CONTENT="А.Б. Васильева">
<META NAME="DC.Subject" CONTENT="22.151.3">
<META NAME="DC.Description" CONTENT="Рассматривается сингулярно возмущенное уравнение второго
порядка. Задается краевое условие, при котором решение имеет пограничный слой типа всплеска.
Исследуется асимптотика такого решения по параметру возмущения и его устойчивость.
Постскрипт статьи (без рисунков) (51Kb)">
<META NAME="DC.Publisher.CorporateName" CONTENT="МГУ. Механико-математический факультет">
<META NAME="DC.Publisher.CorporateName.Address" CONTENT="www.math.msu.ru/koi8/welcome.htm">
<META NAME="DC.Contributor.PersonalName" CONTENT="Пентус Анна">
<META NAME="DC.Contributor.PersonalName.Address" CONTENT="apentus@mech.math.msu.ru">
<META NAME="DC.Contributor.CorporateName.Address" CONTENT="fpm@mech.math.msu.ru">
<META NAME="DC.Date.Creation/Modification_of_present_form" CONTENT="21.06.97">
<META NAME="DC.Type" CONTENT="Text, Serial, Journal">
<META NAME="DC.Identifier" SCHEME="URL"
CONTENT="http://www.math.msu.ru/fpm/koi/95/951/95105.htm">
<META NAME="DC.Relation.Creative" CONTENT="ФУНДАМЕНТАЛЬНАЯ И ПРИКЛАДНАЯ МАТЕМАТИКА 1995,
ТОМ 1, ВЫПУСК 1, СТР. 109-122">
```

</head>

Поиск по иерархической классификации

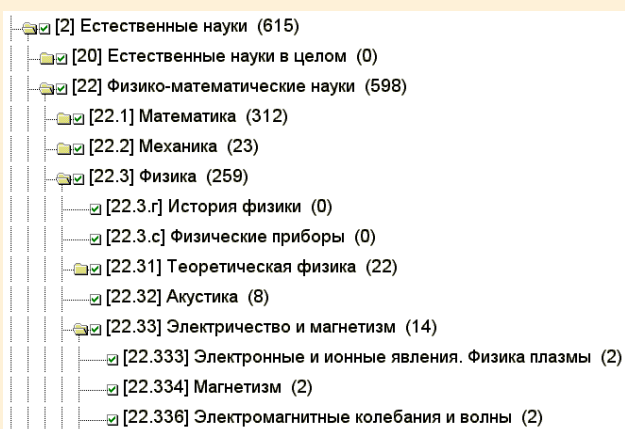
Для организации поиска по любой иерархической классификационной системе в ППП создано специальное средство загрузки рубрик из текстового файла. Нами разработана очень простая структура данных, которая применима к любой иерархической классификации.

Пользователь составляет в виде текстового файла список рубрик, указывая для каждой рубрики ее код, словесное наименование и код вышестоящей рубрики. Затем этот файл обрабатывается специальной программой, которая в соответствии с заданными соотношениями выстраивает дерево. По этому дереву происходит навигация аналогично движению по дереву директорий в программе "Проводник" Windows 95/98.

Загрузив данные для системы классификации, пользователь может вводить коды рубрик в поля библиографического описания ресурса. После ввода в дереве классификации будет указано, сколько записей имеют рубрики, соответствующие каждой ветви дерева.

Как пример реализации такой возможности, приведен текстовый файл, необходимой структуры с рубриками Библиотечно-библиографической классификации для массовых библиотек.

Рис. 2. Фрагмент Библиотечно-библиографической классификации для массовых библиотек, где указано количество записей в базе данных по каждой отрасли.

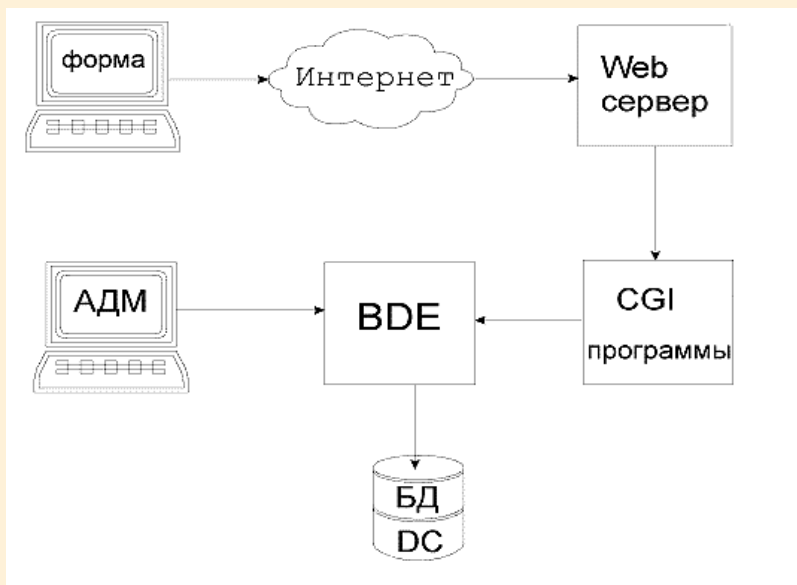


Структура программного обеспечения

ППП создан в двух вариантах:

- СУБД, разработанная на Microsoft SQL Server 7.0
- СУБД, разработанная на DELPHI, на основе Borland Database Engine и таблиц PARADOX .

Рис 3. Схема работы ппп



Программное обеспечение распространяется свободно. Все последние доработки будут публиковаться на сервере www.rsl.ru/dc. Необходимо учитывать, что для использования варианта с Microsoft SQL

Server 7.0 необходимо его иметь. В варианте с СУБД, разработанной на DELPHI, на основе Borland Database Engine и таблиц PARADOX все необходимое для работы уже включено в ППП.

Пакет DC RSL реализует подмножество спецификации Dublin Core Metadata Element Set и позволяет вести каталогизацию ресурсов Internet с автоматической генерацией метаданных в соответствии с этой спецификацией.

Пакет включает следующие компоненты:

- набор CGI-программ, работающих на стороне сервера и обеспечивающих ввод, поиск и отображение данных, накапливаемых в БД
- служебные утилиты для просмотра, проверки и поддержания целостности БД

Программы разработаны в Российской Государственной библиотеке с использованием Inprise Delphi 5.0.

Программу PdxRbld.exe создал Roman Crejci, Czech Republic <http://www.rksolution.cz>

Программу DbExpl32.exe разработали Федор Кожевников, Игорь Павлюк и Сергей Королев, АКБ "Мастер-Банк", Москва <http://rx.demo.ru>

В остальных программах использовались freeware-компоненты сторонних разработчиков, доступные на Torry Delphi Pages <http://www.torry.ru>

Назначение программ пакета:

CgiDcmes.cgi

CGI-программа, реализующая поиск по ключевым словам в базе данных (с ключом /f в URL). Вызывается из форм поиска f_srch*.htm

DcForm.cgi

CGI-программа, реализующая ввод информации в базу данных. Форма ввода генерируется "на лету" с использованием шаблонов f_*.htt, содержащих макрокоманды подстановки значений полей БД

DcBbk.cgi

CGI-программа, реализующая поиск информации по индексу ББК. Страницы генерируются "на лету" с использованием шаблонов bbk*.htt

PdxRbld.exe

Утилита для проверки и исправления физической структуры поврежденных таблиц Paradox

DataChk.exe

Утилита для проверки и исправления логической целостности БД

Dcrslmgr.exe

- Диалоговый редактор БД, позволяет просматривать и удалять ненужные данные, отображает статистику числа записей, введенных операторами
- Диалоговый редактор индекса ББК с возможностью выгрузки данных в текстовый файл и загрузки из файла
- Программа поиска дублетных записей в базе данных

Дальнейшие планы развития ППП

Наша разработка постоянно совершенствуется и все новые изменения будут оперативно предоставляться пользователям. Многие библиотеки в мире уже ведут каталоги ресурсов Интернет в виде баз данных с форматом Dublin Core в качестве внутреннего формата представления данных. Ряд библиотек объединяет свои усилия и создает распределенные каталоги, распределив свои усилия по тематике или территории. Для проведения поиска одновременно во всех каталогах участниках распределенной системы используется протокол Z39.50. В связи с вышеизложенным, было принято решение в первую очередь создать свободно распространяемый сервер Z39.50 для организации распределенных баз данных, способных взаимодействовать с зарубежными каталогами.

Для решения этой проблемы мы планируем к концу года завершить все работы по созданию программного обеспечения. Затем оно будет передано всем желающим. Любая организация, которая захочет вести свой каталог ресурсов Интернет может получить наши программы. Установив их на любом компьютере, подключенном к Интернет, вы получите базу данных для каталогизации, на вашем сервере в Интернете будут опубликованы интерактивные формы для каталогизации и поиска в базе. Также вы получите список адресов всех баз данных, работающих с аналогичным форматом и программу - клиент для одновременного поиска сразу во всех базах. Адрес вашей базы (при вашем желании) будет сообщен всем пользователям аналогичных баз данных, для того чтобы ваши усилия по

каталогизации стали доступны всем.

Вышеописанный подход уже реализован в некоторых странах. Наиболее успешны проекты электронных библиотек в Биллефельде и Геттингене. (Германия), европейский проект MATH NET и МЕТАРНУС.

Если наша работа будет успешно завершена, и, главное, поддержана российским Интернетовским сообществом, мы сможем стать частью мировой системы распределенных каталогов и это, несомненно, будет способствовать развитию науки.

© Шварцман М.Е., Ильин А.С., 2000

Последнее обновление страницы было произведено: 2003-12-09

Все предложения и пожелания по содержанию и структуре портала направляйте по адресу rdlp@iis.ru

