

Опыт написания программы для создания распределенного каталога ресурсов Интернет

М. Е. Шварцман

Российская Государственная библиотека

Адрес: 1101000 Москва, Воздвиженка, 3

тел. (095) 202-66-3

E-mail: shvar@rsl.ru

В связи с большим интересом в мире к ресурсам Интернет, многие библиотеки 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 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 формате Dublin Core как локально, так и для организации распределенных баз данных способных взаимодействовать и с зарубежными каталогами. Для решения этой задачи мы приступили к созданию свободно распространяемого программного обеспечения (ПО), удовлетворяющего следующим требованиям:

- 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.).

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.

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 .

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.

В 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 METAPHYS.

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.

В настоящее время ПО сделано, за исключением сервера Z39.50, и может быть получено в РГБ. Планируется к концу года завершить все работы. Любая организация, которая захочет вести свой каталог ресурсов Интернет может получить наши программы. Они устанавливаются на любом компьютере, работающем на Windows 95 или Windows NT и подключенном к Интернет, Это - готовый инструмент для публикации в Интернете интерактивных форм для ввода новых библиографических описаний и поиска в базе. Ресурсов Internet с автоматической генерацией метаданных в соответствии с форматом DC. Пакет включает набор CGI -программ, работающих на стороне сервера и служебные утилиты для просмотра, проверки и поддержания целостности БД

Для организации распределенного каталога необходимо будет вести работу по координации деятельности всех организаций, использующих это ПО, для определения сфер интересов каждого участника этой системы.

Шварцман Михаил Ефремович - зав. отделом использования компьютерных

сетей Российской государственной библиотеки

С 1978 года работает в Российской государственной библиотеке в отделе автоматизации. За это время занимался автоматизацией каталогизации, комплектования, межбиблиотечного абонементов.

Участвует в проектах:

Память России

Центр текущей национальной библиографии в Интернет

Встреча на границах

Проект TACIS по информатизации Российской государственной библиотеки

Создание систематического каталога российских ресурсов Интернет с использованием Dublin Core Metadata Set

Занимается созданием локальной сети в РГБ и разработкой Web сервера РГБ

Последние труды:

Kasparova N. Shvartsman M. Creation of the electronic resources Meta-database; problems and prospects // 66th IFLA General Conference, Jerusalem, Israel, 13-18 August 2000. - <http://www.ifla.org/IV/ifla66/papers/139-168e.htm>

Шварцман М.Е. Опыт написания программы для создания распределенного каталога ресурсов Интернет// Труды международной конференции "Крым 2000" июнь 2000. - М., 2000. - Том 1. - С.224-226

Международный библиотечный проект "Встреча на границах" Г. Василенко, М.Шварцман И. Куприянов, В.Чебанов // Труды международной конференции "Крым 2000" июнь 2000. - М., 2000. -Том 1.

Шварцман М.Е. Ильин А.С. Dublin Core в коробке: доклад на Международной конференции "Управление электронными ресурсами библиотек" 17-18 апр. 2000 г. Москва // Электронные библиотеки - 2000 г.- Том 3 - Выпуск 2. - <http://www.elbib//2000/200002/SI/si.ru.html>

Давыдов В.М. Шварцман М. Е. Давыдов Е. В. Состояние работ по внедрению системы универсальных имен ресурсов Интернет для электронных информационных служб национальных библиотек // Электронные библиотеки. - N2, вып. 2. - 1999. - <http://www.elbib//1999/199902/davydov/davydov.ru.html>

Опыт создания цифровой библиотеки первопечатных славянских книг XV-XVI вв /Л.М. Еремина, И.В. Морозова, М.Е. Шварцман, Н.М. Корценштейн, В.Ю. Левашов, И.М. Мазилін библиотек // Электронные библиотеки. - N2, вып. 2. - 1999. <http://www.elbib//1999/199902/eremina/eremina.ru.html>

Центр информационных технологий (ЦИТ) Российской Государственной Библиотеки.

Основной задачей Центра является внедрение интегрированной автоматизированной системы РГБ, базирующейся на современной информационно-вычислительной среде. ЦИТ поддерживает программное обеспечение для обработки текущих поступлений документов в библиотеку и электронного каталога, локальных баз данных для пользователей; осуществляет техническое перевооружение библиотеки, объединяя существующие и вновь приобретаемые средства вычислительной техники локальными вычислительными сетями;

разрабатывает компьютерные технологии и информационно-лингвистическое обеспечение АИБС РГБ.

В развитии информационных технологий ЦИТ ведет тесное международное сотрудничество с Европейским сообществом. По соглашению с Европейской комиссией, в рамках проекта ТАСИС "Интегрированная автоматизированная система для Российской государственной библиотеки" приобретено новое программное обеспечение "Алеф-500" (модуль ОПАС) для перевода на него существующего электронного каталога библиотеки. В ближайшей перспективе – приобретение и внедрение модуля каталогизации.

В ИНТЕРНЕТ представлен электронный каталог отечественных и иностранных книг, карт, авторефератов и диссертаций, поступающих в РГБ и в совокупности насчитывающий около 600000 библиографических записей. ЦИТ поддерживает Web-site библиотеки (www.rsl.ru). Сегодня в рамках Центра начинается формирование будущей электронной. Центр оказывает научно-методическую помощь библиотекам по внедрению информационных технологий.

Шварцман М.Е. ©