

УДК 02 : 004 + 002.5 : 004  
ББК 78.30 + 32.973.202 - 018.2

## СОЗДАНИЕ КОМПЛЕКСА ПРОГРАММ ДЛЯ ЦИФРОВОЙ ПОЛНОТЕКСТОВОЙ БИБЛИОТЕКИ НАУЧНО-ТЕХНИЧЕСКОЙ ЛИТЕРАТУРЫ

© М. С. Угаров, А. В. Шабанов, 2009

*Государственная публичная научно-техническая библиотека  
Сибирского отделения Российской академии наук  
630200, г. Новосибирск, ул. Восход, 15*

Рассмотрены ключевые функции системы и основные компоненты приложения для цифровой полнотекстовой библиотеки, структура данных, а также создание и тестирование пробного варианта базы данных полных текстов. Показана программная реализация связи с другими цифровыми библиотеками.

*Ключевые слова:* цифровая библиотека, научно-техническая литература, программный комплекс, полнотекстовая база данных.

Key functions of the system and main components of appendices for digital full-text library, data structure, as well as generation and testing of the experimental full-text data base are considered. Software realization of link with other digital libraries is shown.

*Key words:* digital library, scientific and technical literature, software complex, full-text database.

**В** настоящее время библиотеки, становясь частью среды Интернета, должны обеспечить общество полной и оперативной информацией в электронном виде. Таким образом, актуальной становится задача создания цифровой библиотеки как средства хранения электронной информации, информационного обслуживания читателей библиотеки и расширения доступа к документированным знаниям.

В рамках разработки комплекса программ для цифровой полнотекстовой библиотеки был создан прототип системы, который предоставляет доступ к книгам научно-технической тематики в электронном виде посредством глобальной сети.

В данной работе рассматриваются основные функции системы, структура данных и основные компоненты приложения, создание и тестирование (на объем и быстродействие) пробного варианта базы данных (БД) полных текстов, а также программная реализация связи с другими цифровыми библиотеками [1].

Основные *функции* системы следующие:

- обеспечение поиска нужной литературы по наименованию, автору, рубрикам;
- обеспечение комфортного просмотра найденной литературы с помощью специального просмотрщика;
- обеспечение одновременного доступа множества пользователей (до тысячи) к полным текстам через локальную сеть;

- обеспечение масштабируемости системы без существенного увеличения времени доступа к книгам (т. е. система работает быстро как с 10 тыс. книг, так и со 100 тыс. или миллионами);

- возможность сравнения разных цифровых полнотекстовых библиотек и пополнение собственной конечной библиотеки из нескольких внешних источников;

- возможность оперативного обновления данных о книгах, авторах и рубриках;

- возможность автономного функционирования системы с DVD.

Основными *компонентами* разрабатываемого комплекса являются [2]:

1. СУБД MySQL.
2. Веб-сервер Apache Tomcat.
3. Веб-приложение библиотеки.

Остановимся на каждом компоненте подробно. MySQL – компактный многопоточный сервер баз данных.

Tomcat – это контейнер сервлетов и JSP-страниц, готовых для выполнения при запуске их из браузера.

MySQL и Tomcat характеризуются большой скоростью, устойчивостью и легкостью в использовании.

*Данные* представляют собой архивы полных текстов в форматах pdf, djvu, html, txt и doc, которые разбиты на несколько основных разделов, таких как математика, физика, химия, биология и техника.

**Структура данных следующая:**

tbl_rubrics «Рубрики»	
ID int NOT NULL AUTO_INCREMENT	Ключевое поле

RUBNAME varchar(100)	Название рубрики
tbl_subrubrics «Подразделы»	
ID int NOT NULL AUTO_INCREMENT	Ключевое поле
IDRUB int NOT NULL	Номер рубрики
SUBRUBNAME varchar(100)	Название подраздела

tbl_authors «Авторы»	
ID int NOT NULL AUTO_INCREMENT	Ключевое поле
AUTHNAME varchar(100)	Имя автора (или группы авторов)

tbl_rubtab «Кросс-таблица – книги-подразделы»	
IDBOOK int NOT NULL	Номер книги

IDSUBRUB int NOT NULL	Номер подраздела
tbl_authtab «Кросс-таблица – книги-авторы»	
IDBOOK int NOT NULL	Номер книги

IDAUTH int NOT NULL	Номер автора
tbl_books «Книги»	
ID int NOT NULL AUTO_INCREMENT	Ключевое поле
TITLE varchar(100)	Название книги
FILENAME varchar(100)	Относительный путь к файлу
DATE varchar(100)	Год издания
ISBN varchar(100)	ISBN книги
PUBLISHER varchar(100)	Название издательства
NOTICE varchar(255)	Примечания к книге

В результате была получена следующая схема данных [2] (рисунок).

Начальное формирование БД происходит с помощью скриптов MySQL [3]. В свою очередь, сами скрипты формируются комплексом программ, которые анализируют списки источников в html виде. Анализ списков происходит следующим образом:

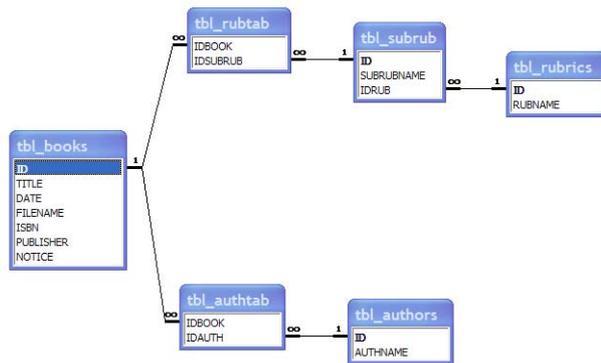


Схема данных

1. Выделение ссылки на источник.
  2. Разбиение строки с описанием источника на части для определения автора, названия и года издания.
  3. Найденные автор, название, год издания и ссылка на источник формируют скрипт MySQL.
- Поиск в базе данных будет производиться по автору, наименованию источника и году издания. Также предусмотрено четыре варианта сортировок найденных данных:

1. Автор-название.
2. Название-автор.
3. Дата-автор-название.
4. Дата-название-автор.

Веб-приложение библиотеки состоит из двух частей: пользовательской и администраторской [2]. Пользовательская часть содержит следующие компоненты:

1. Рубрикатор. Выводит подразделы, относящиеся к указанной рубрике. Иерархия рубрик и подразделов выполнена в соответствии с двумя уровнями ГРНТИ.

2. Поиск. Осуществляет поиск по автору или по наименованию книги.

Опишем более подробно каждый из разделов:

1. Рубрикатор. Для каждой рубрики выводится список подразделов, которые принадлежат к данной рубрике. Затем при выборе подраздела выдается список авторов, книги которых принадлежат к данному подразделу. При выборе автора выдается список книг, принадлежащих к данному подразделу. Как говорилось ранее, рубрики и их подразделы соответствуют двум уровням ГРНТИ.

2. Поиск. Здесь можно выбрать критерий поиска либо по имени автора, либо по названию книги. Далее надо ввести в поле ввода строку поиска и нажать на кнопку «Искать». В зависимости от критерия появится либо список авторов, имена которых содержат строку поиска, либо список книг.

В каждом из разделов при появлении списка книг можно «прочитать» полный текст книги. Справа в верхнем или нижнем углу окна вывода будет находиться навигатор. Навигатор позволяет

перемещаться между страницами, а также показывает, на какой странице вы находитесь и сколько страниц содержит книга.

Также можно посмотреть наличие выбранной книги в других библиотечных проектах, например в Google Book Search. Для этого используются стандартные и динамические ссылки.

Стандартный формат ссылок или статические ссылки позволяют разработчикам ссылаться на книги, используя номера ISBN и др. Ссылаться можно не только на саму книгу, но и на описание, титульный лист, страницу об авторском праве, оглавление, алфавитный указатель.

Динамические ссылки реализованы при помощи интерфейса программирования Book Viewability API. Данный интерфейс позволяет разработчикам:

- ссылаться на книги в Google Book Search, используя номера ISBN, LCCN и OCLC;
- знать о том, имеется ли заголовок книги в Google Book Search и какой возможностью просмотра этот заголовок обладает;
- создавать ссылки на страницу с информацией о книге.

Для ввода и редактирования данных используется администраторский модуль. С помощью него

можно добавлять, удалять и редактировать авторов, рубрики, подразделы и книги.

Полученные результаты позволяют сказать, что разработан и отлажен тестовый вариант цифровой библиотеки. Реализована связь с другими цифровыми библиотеками.

### Список литературы

1. Угаров М. С., Шабанов А. В. Создание комплекса программ для цифровой полнотекстовой библиотеки. Ч. 1. Основные функции системы // Итоги и перспективы научной работы ГПНТБ СО РАН : материалы науч. сес. (г. Новосибирск, 14 февр. 2007 г.). – Новосибирск, 2007. – С. 262–264.
2. Угаров М. С., Шабанов А. В. Создание комплекса программ для цифровой полнотекстовой библиотеки: структура данных. Основные компоненты приложения // Электронные ресурсы библиотек региона : материалы регион. науч.-практ. конф. (г. Новосибирск, 24–28 сент. 2007 г.). – Новосибирск, 2008. – С. 237–243.
3. Угаров М. С. Создание комплекса программ для цифровой полнотекстовой библиотеки. Ч. 3. Создание тестового варианта базы данных полных текстов // Сборник материалов научной сессии молодых ученых информационно-библиотечной сети СО РАН. – Новосибирск, 2008. – С. 97–100.

Материал поступил в редакцию 19.09.2008 г.

Сведения об авторах: Угаров Михаил Станиславович – аспирант ГПНТБ СО РАН,  
тел.: (383) 266-75-79, e-mail: ugarovms@spsl.nsc.ru,

Шабанов Андрей Васильевич – кандидат технических наук, старший научный сотрудник  
отдела редких книг и рукописей,  
тел.: (383) 266-10-91, e-mail: shabanov@spsl.nsc.ru