

ВЫБОР МЕТОДА ПРЕДСТАВЛЕНИЯ БИБЛИО- ГРАФИЧЕСКОЙ ИНФОРМАЦИИ

А. Н. Толстикова*

Е. П. Павленко

Кандидат технических наук, доцент*

Контактный тел.: (057) 702-14-51

E-mail: evg-pavl@mail.ru

*Кафедра информационных управляющих систем
Харьковский национальный университет
радиоэлектроники
просп. Ленина, 14, г. Харьков, 61166

Н. Г. Толстикова

Кандидат технических наук, заведующий кафедрой
Кафедра информатики
Харьковский гуманитарно-технический институт
ул. Кандаурова, 2, г. Харьков
Контактный тел.: (057) 335-24-29
E-mail: tols—alex@yandex.ru

Досліджені переваги і недоліки методів представлення бібліографічної інформації. Розглянута задача обліку вказаної інформації і вибраний метод для її ефективного вирішення

Ключові слова: електронна бібліотека, каталог, база даних

Исследованы преимущества и недостатки методов представления библиографической информации. Рассмотрена задача учета указанной информации и выбран метод для ее эффективного решения

Ключевые слова: электронная библиотека, каталог, база данных

Advantages and disadvantages of methods of bibliographic information presentation were analyzed. The problem of account of this information was considered and also was chosen a method for its effective decision

Keywords: electronic library, catalogue, database

1. Введение

По мере информатизации общества, развития информационных технологий, доступ к информации и обмен ею упрощается, но в связи с увеличением объема информации и ее многократным дублированием в разных формах затрудняется поиск и выборка необходимой информации.

В области библиографии [1] наблюдаются такие же тенденции – постоянно выходят новые издания, старые издания переиздаются, библиографическая информация рассредоточена по большому количеству электронных каталогов и библиотек, причем информация хранится в различных форматах. Задача поиска информации решается путем сложной, длительной и затратной процедуры многочисленных локальных поисковых запросов ко всем имеющимся ресурсам, что неприемлемо в современных быстроменяющихся условиях жизни. Поэтому становится актуальной задача создания универсального каталога, объединяющего данные разных форматов и обеспечивающего возможности редактирования информации.

2. Постановка задачи

Для обеспечения требований, предъявляемых к электронным библиотекам [1], таких как: компактное

и длительное хранение больших объемов информации; оперативный многоаспектный поиск информации; доступ к удаленным ресурсам и электронная доставка документов; совмещение библиографических и полнотекстовых баз данных, - необходимы совместные усилия библиографов и разработчиков информационных систем. А также база данных с гибкой структурой и максимально приспособленная к внутренним изменениям, т.к. реляционные модели данных не удовлетворяют в полной мере всем изложенным требованиям. Поэтому разрабатываются библиографические модели хранения данных.

Рассмотрим существующие электронные модели представления библиографической информации MARC, Dublin Core и FRBR.

3. Сравнение библиографических моделей

Наиболее распространенная библиографическая модель – модель MARC (Machine–Readable Cataloging Record). MARC для библиографических данных предназначен для представления в машиночитаемом виде библиографических сведений о печатных и рукописных текстах, компьютерных файлах, картографиче-

ских материалах, нотах, продолжающихся ресурсах, видеоматериалах и собраниях разнотипных материалов.

Эта модель является основным стандартом каталогизации в США (поддерживается библиотекой конгресса США), также используется в Великобритании. Библиографическая запись в формате MARC имеет такой вид:

```
Маркер *****nam#a22*****#i#4500001 00031272
003 RuMoRGB
005 20000915102711.0
008 000915s2000###ru#a###frm####000#0#rus#d
017 ##ад530-8-01#bPГБ
072 #7#a13.00.01#2nsnr
040 ##aRuMoRGB
041 0#agus
084 ##aЧ426.24,6(2)-25,0#2rubbk
100 1#aЯрковая, Ирина Петровна.
245 00#aПедагогические условия применения ди-
дактических стихов в процессе обучения предметам
естественно-математического цикла: #bДис. : канд.
пед. наук : 23.00.02.
260 ##aХабаровск, #c1000.
300 ##a217 с. : #бил. 650 #7#aПедагогика#2<код
системы предметизации>
852 4#aPГБ bФБ#i61:01-13/56-1
```

В процессе развития формата MARC в 1970-х гг. появились более 20 его различных версий, ориентированных на национальные правила каталогизации. Для преодоления несовместимости этих форматов в 1977 г. Международной федерацией библиотечных ассоциаций (ИФЛА) было выпущено издание «Универсальный формат MARC» (Universal MARC Format, UNIMARC). Сравнение различных разновидностей формата MARC описано в [2].

Dublin Core (DC) [3] – широко распространенный стандарт метаданных, простой и эффективный набор для описания разнообразных сетевых ресурсов. Стандарт разделён на два уровня:

- простой (неквалифицированный, simple), состоящий из 15 элементов;
- компетентный (квалифицированный, qualified), состоящий из 18 элементов и группы т. н. тонкостей (или квалификаторов), которые уточняют семантику элементов для повышения полезности поиска ресурсов. Схема DC более приспособлена к использованию поисковыми инструментами, однако записи DC не обеспечивают такую полноту и детальность, как MARC – записи.

Простой набор элементов метаданных Дублинского ядра (Dublin Core Metadata Element Set; DCMES) состоит из 15 элементов метаданных: Title – название; Creator – создатель; Subject – тема; Description – описание; Publisher – издатель; Contributor – участник создания ресурса; Date – дата; Type – тип; Format – формат документа; Identifier – идентификатор; Source – источник; Language – язык; Relation – отношения; Coverage – покрытие; Rights – авторские права. Квалифицированный (компетентный) набор элементов метаданных Дублинского ядра, помимо 15 вышеперечисленных, может включать: Audience – аудитория (зрители); Provenance – происхождение; RightsHolder – правообладатель. Каждый элемент опционален и может повторяться. Инициатива метаданных Дублинского ядра

(Dublin Core Metadata Initiative; DCMI) описала стандартные пути определения элементов и поощряет использование схем кодирования и словарей.

В 1997 году DC был расширен с помощью двух типов квалификаторов, позволяющих уточнять утверждения о ресурсах. Первый из них касался уточнения самого элемента, а второй – уточнения допустимых значений элемента. Квалификаторы первого типа делают значение элемента более узким и специфичным. Уточненный («квалифицированный») элемент имеет тот же смысл, что и неуточненный элемент, но с более узкой и строгой областью действия. Пользователь, который не понимает смысла данного уточнения, должен иметь возможность игнорировать квалификатор и работать с элементом, как если бы он был неквалифицированным. Квалификаторы второго типа определяют схему интерпретации значений элементов. Эти схемы включают в себя контролируемые словари, условные обозначения или правила разбора значения. Если схема кодирования непонятна обработчику, значение элемента все равно должно остаться полезным для читателя-человека. Строгое описание схемы кодирования для квалификаторов должно находиться в свободном доступе.

Модель FRBR (Functional Requirements for Bibliographic Records) – функциональные требования к библиографическим записям, разработанные для реструктуризации баз данных каталогов [4]. Такая реструктуризация имеет целью отражение концептуальной структуры информационных ресурсов в библиографических базах данных. FRBR использует модель метаданных «сущность-связь» для информационных объектов. Модель FRBR, которая была разработана в 1992-1995 годах группой ИФЛА по исследованию функциональных требований к библиографическим записям, является моделью связей сущностей как обобщенного представления библиографической среды.

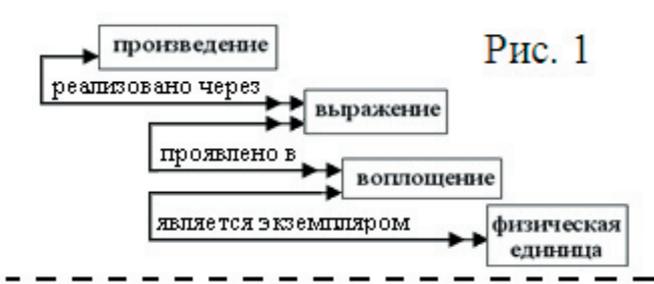
Под сущностью понимается идентифицированный «предмет», реальный или абстрактный. По сути, это обобщенное название множества однотипных предметов. Модель FRBR предназначена для того, чтобы поставить данные, хранящиеся в библиотечных записях, в соответствие с нуждами пользователей этих записей, и для повышения способности каталогов к взаимодействию.

Объекты первой группы (рис. 1) представляют аспекты интересов пользователя к продуктам художественного или интеллектуального труда. Объекты, определённые как произведение (творение) и выражение (реализация произведения), отражают интеллектуальное или художественное содержание. Объекты, определённые как воплощение (физическое проявление выражения произведения) и физическая единица (один экземпляр воплощения), с другой стороны, отражают физическую форму.

Объекты второй группы (рис. 2) (лицо и организация) представляют тех, кто несёт ответственность за интеллектуальное или художественное содержание, физическое производство и распространение или хранение объектов первой группы.

Объекты третьей группы (рис. 3) представляют дополнителный набор объектов: концепция (абстрактное понятие или идея), предмет (материальная вещь), событие (действие или явление), место, которые слу-

жат темами произведений. Атрибуты объекта – средство, с помощью которого пользователи формулируют запросы и интерпретируют ответы при поиске информации о конкретном объекте. Атрибуты выражены с точки зрения характеристик объекта так, как их могут воспринимать пользователи.



Связи, отраженные в библиографической записи, дают дополнительную информацию, которая помогает пользователю установить связи между найденным объектом и другими объектами, которые относятся к этому объекту.

Концепция FRBR лучше всего описывает, какой должна быть система работы с библиографической информацией. Однако модель FRBR находится только в процессе развития, поэтому библиографических данных, совместимых с FRBR, мало.

4. Выводы

Существующие модели хранения библиографических данных ограничены по своей структуре и их сложно конвертировать друг в друга.

Тогда необходимо создание такой структуры хранения

библиографической информации, которая представляла бы возможность редактирования понятий предметной области, то есть в которую можно было бы интегрировать данные других форматов.

Поэтому была создана универсальная контекстно-независимая структура базы данных для библиографической информации,

в которой хранятся данные и понятия предметной области, описываемые данными.

Универсальность структуры позволяет изменять состав предметной области базы данных – появляется возможность представления данных другого источника с дополнением понятий предметной области в разработанной базе. Для работы с базой данных разработана программная система с определенной архитектурой.

Литература

1. Каталогизация и организация электронных ресурсов: практическое руководство для библиотекарей. – Энн М. Митчелл, Брайан Э. Сэррэтт. – М.: Омега-Л, 2008. – 231с.
2. О. Н. Жлобинская. MARC- форматы в современной информационной среде// Российская национальная библиотека. [Электронный ресурс] http://www.rba.ru/rusmarc/publish/MARC_now.pdf.
3. А.В. Манцивода. Система метаописаний Dublin Core [Электронный ресурс] <http://teacode.com/concept/eor/dc.html>.
4. The FRBR Blog [Электронный ресурс] <http://www.frbr.org/>.