



Седьмая Международная Конференция «Крым 2000»

Конференция проводится в рамках мероприятий ИФЛА 2000 г.

***Библиотеки и ассоциации в меняющемся мире:
новые технологии и новые формы сотрудничества***

***Тема 2000 года:
Библиотеки, издательства,
книгораспространение и образование
в едином информационном и
социокультурном пространстве***

Труды конференции

Том 1

СУДАК

(основная программа)

Ялта, Алушта, Феодосия, Симферополь,

Коктебель, Старый Крым

(выездные заседания)

Автономная Республика Крым, Украина

3-11 июня 2000

Отображение электронного каталога ГПНТБ России в Z39.50

Electronic Catalog of Russian National Public Library for Science and Technology in Z39.50

Відображення електронного каталогу ДПНТБ Росії в Z39.50

Мазов Н.А., Жижимов О.Л.

*Объединенный институт Геологии, Геофизики и Минералогии
Сибирского Отделения РАН, г. Новосибирск
Бродовский А.И.*

Государственная публичная научно-техническая библиотека России, г. Москва

Nikolai A. Mazov, Oleg L. Zhizhimov

*Joint Institute of Geology, Geophysics and Mineralogy of the Russian Academy of Sciences' Siberian Branch,
Novosibirsk, Russia
Alexander I. Brodovsky*

Russian National Public Library for Science and Technology, Moscow, Russia

Мазов Н.А., Жижимов О.Л.

*Об'єднаний інститут геології, геофізики і мінералогії СВ РАН, Новосибірськ, Росія
Бродовський О.Й.*

Державна публічна науково-технічна бібліотека Росії, Москва, Росія

В докладе описуються особливості отображення каталога ГПНТБ России для работы по протоколу Z39.50. Рассмотрены особенности отображения данных в формат типа MARC как для стандартного CDS/ISIS, так и для провайдера данных Z-ISIS, применяемого для работы с базами данных системы CDS/ISIS, в среде которой ведется электронный каталог ГПНТБ России.

Described are the features of mapping NPLS&T catalog for Z39.50 protocol. Considered are mapping of data in MARC-type format both for standard CDS/ISIS software package and for Z-ISIS data provider, used for databases of the CDS/ISIS environment, under which the electronic catalog of Russian National Public Library for Science and Technology is maintained.

У доповіді висвітлюються особливості відображення каталогу ДПНТБ Росії для роботи по протоколу Z39.50. Розглядаються особливості відображення даних у формат типу MARC як для стандартного CDS/ISIS, так і для провайдера даних системи Z-ISIS, що застосовується для роботи з базами даних системи CDS/ISIS, у середовищі якої ведеться електронний каталог ДПНТБ Росії.

Система CDS/ISIS [1], разработанная в ЮНЕСКО в отделе разработки и применения программного обеспечения, получила широкое применение и распространение как у нас в стране, так и за рубежом. Более восьми тысяч крупных учреждений разных стран являются зарегистрированными пользователями системы, среди которых национальные библиотеки, международные, общественные, культурные, религиозные, государственные организации, библиотеки и архивы в России, в странах ближнего и дальнего зарубежья.

В настоящее время существует несколько центров, развивающих и поддерживающих программное обеспечение CDS/ISIS на различных программно-аппаратных платформах и в различных операционных средах.

Одной из таких организаций России, использующей в своей работе CDS/ISIS является ГПНТБ России. Опыт практического использования системы CDS/ISIS в ней насчитывает более пятнадцати лет. Неудивительно, что основное детище ГПНТБ России – электронный каталог – ведется под управлением CDS/ISIS.

Одним из основных достоинств CDS/ISIS является использование системой записей в структуре стандарта ISO-2709 [2] как для записей баз данных, так и для записей, экспортируемых (импортируемых) системой. ISO-2709 - известный стандарт, определяющий структуру записей, на которой базируются форматы семейства MARC. Это обстоятельство упрощает пользователям, использующих при работе с базами данных CDS/ISIS общение с форматами семейства MARC (RUSMARC, UNIMARC, USMARC и др.). Следует отметить, что форматы типа MARC в настоящее время широко используются при построении автоматизированных библиотечно-информационных систем, как средство обмена между различными библиотечными системами. Так, например, формат UNIMARC (RUSMARC) является как внутренним, так и форматом для обмена интегрированной библиотечно-информационной системы ИРБИС, разработанной и распространяемой ГПНТБ России.

MARC это аббревиатура словосочетания MACHine READable Cataloguing (машиночитаемая каталогизация). Фактически, это формат представления библиографической информации, разделенной на записи. При этом обычно каждая запись соответствует описанию библиографического документа. Формат MARC

разработанный в конце шестидесятых годов в США (до недавнего времени известный как USMARC, теперь MARC21), в настоящее время реализован во многих национальных вариантах. Существует также его международная версия - UNIMARC, разработанная в конце семидесятых годов и рекомендованная для использования ИФЛА, с целью обеспечения унификации конвертирования информации, представленной в различных национальных форматах. В некоторых странах UNIMARC (или его диалекты) принят в качестве национального формата. Так в России он известен как RUSMARC [3].

Пионером использования в практике обмена библиографической информацией в России по праву можно считать ГПНТБ России. Именно ГПНТБ России первой предприняла попытку перевода описания таких форматов как UNIMARC[4] и USMARC[5]. Здесь сложился коллектив высококвалифицированных специалистов в области трактовки и использования вышеуказанных форматов. Поэтому неудивительно и то, что при установке сервера Z39.50 авторам удалось в сжатые сроки произвести отображение электронного каталога ГПНТБ России для возможности обмена по протоколу Z39.50 в этих форматах.

В июне 1999 года в ГПНТБ России был установлен сервер Z39.50 (ОИГТиМ СО РАН) разработки авторов на платформе Linux. Архитектура этого сервера позволяет динамически настраивать структуру конкретных СУБД к работе с Z39.50 через специальные модули – провайдеры данных. В частности, сервер ГПНТБ России кроме прочих использует провайдер данных CDS/ISIS (Z-ISIS) – независимая динамическая библиотека, реализующая основные функции этой системы. В настоящее время электронный каталог ГПНТБ России доступен по протоколу Z39.50 на сервере star.gpntb.ru:210 база данных katb. Сервер включен в реестр тестируемых ресурсов Z39.50 компании IndexData как один из немногочисленных российских информационных ресурсов.

Сервер Z39.50 ГПНТБ России поддерживает протокол Z39.50-95(v3)[6], обрабатывает запросы RPN (type 1), выдает записи в форматах: SUTRS, GRS1, USMARC, UNIMARC, RUSMARC, XML, HTML, RTF. Дополнительно поддерживается Explain в категориях: CategoryList, TargetInfo, DatabaseInfo, SchemaInfo, TagSetInfo, AttributeSetInfo. Сервер имеет встроенные возможности работы с распределенными данными. Детальная информация о возможностях сервера приведена в [7].

Ниже рассматриваются некоторые особенности отображения электронного каталога ГПНТБ России для работы по протоколу Z39.50.

Поскольку идеология Z39.50 предполагает отображение конкретной структуры данных в абстрактную в соответствии с используемой схемой данных [6], основная задача адаптации данных – формулирование правил такого отображения. В основе провайдера данных Z-ISIS [8], разработанного авторами, лежит модуль форматирования данных CDS/ISIS, который позволяет не только сформулировать эти правила, но и снять ряд ограничений, накладываемых стандартным CDS/ISIS на форматирование данных. Так, при формировании записи в MARC формате используется привычный для пользователя CDS/ISIS язык форматирования данных, допускающий некоторые отклонения от стандартного языка форматирования CDS/ISIS для более корректного представления данных в форматы типа MARC, которые являются одними из основных при работе по протоколу Z39.50. Все правила настройки соответствия данных вынесены в отдельные текстовые серверные таблицы и допускают изменения при необходимости.

Ниже рассмотрены особенности отображения данных в формат типа MARC как для стандартного CDS/ISIS, так и для провайдера данных Z-ISIS.

| CDS/ISIS | Z-ISIS |
|---|---|
| Маркер ISO 2709 | |
| <p>Маркер записи недоступен для изменений, поэтому ввести коды статуса записи или библиографического уровня в соответствии с требованиями формата MARC невозможно. Здесь следует работать с выходным файлом CDS/ISIS с помощью специальной программы, либо использовать дополнительный модуль обмена, который бы позволил создавать маркер записи, исходя из содержимого библиографических полей записи, зарезервированных для этих целей</p> | <p>Для маркера записи в провайдере данных Z-ISIS определена специальная команда в настроечной таблице с именем tag. Поэтому коды статуса записи или библиографического уровня в соответствии с требованиями формата MARC формируются как обычное поле фиксированной длины, которое подставляется в запись в момент окончания ее формирования.</p> |

| Метки ISO 2709 | |
|--|---|
| <p>Формат MARC предполагает использование только трехсимвольных меток, поэтому проблем при хранении и экспорте данных в CDS/ISIS не возникает.</p> | <p>Как и для стандартного CDS/ISIS проблем с выводом меток не существует.</p> |
| Подполя ISO 2709 | |
| <p>CDS/ISIS поддерживает подполя с алфавитными идентификаторами, следовательно, можно использовать подполя MARC. Первый символ идентификатора подполя, а именно ASCII код 31 в формате MARC, представлен в CDS/ISIS при помощи символа "^". Замену символа "^" на ASCII код 31 можно осуществить при помощи таблицы конвертирования на выходе, либо глобальным редактированием, либо специальной программой. К сожалению, повторяющиеся подполя в пределах одного поля, широко используемые в большинстве форматов MARC, обработать корректно не представляется возможным. При выводе и обработке посредством таблицы выбора полей (ТВП) учитывается только первое повторение. Повторяющиеся подполя можно обработать корректно экспорте, если не используется таблица ТВП. Для решения этой проблемы пишутся форматные выходы, которые обеспечивают корректный вывод повторяющихся подполей; их можно использовать при экспорте. Существуют и другие способы, а именно, определять как повторяющиеся те поля, которые обычно таковыми не являются, например поле выходных данных. Такое представление корректно обрабатывается системой и при экспорте данных конвертируется в одно поле. Такое решение является специфическим, которое можно принять для отдельного поля, но применять его для других полей следует с осторожностью.</p> | <p>Z-ISIS также как и стандартный CDS/ISIS поддерживает подполя с алфавитными идентификаторами. Первый символ идентификатора подполя, а именно ASCII код 31 в формате MARC, представляется специальным образом в настроечной таблице, однако, если необходимо включить в конкретное повторение поля еще одно подполе, в этом случае ASCII код 31 кодируется напрямую как литерал. Возможен вывод повторяющихся подполей в пределах одного поля, широко используемый в большинстве форматов MARC. Для корректности такого вывода, в провайдере данных Z-ISIS имеется возможность обработки указанного повторения поля.</p> |
| Индикаторы | |
| <p>Индикаторы можно выводить в начале каждого поля так, как они появляются в записи ISO 2709. Вывод их осуществляется как обычных литералов, в зависимости от содержимого полей конкретной записи.</p> | <p>Для вывода индикаторов в настроечной таблице существует специальный синтаксис. В случае, когда индикатор не указывается, принимается значение по умолчанию.</p> |
| Поля фиксированной структуры формата MARC | |
| <p>Поля фиксированной структуры формируются как строки литералов в соответствии с языком форматирования.</p> | |

Из вышеизложенного видно, что в своем первоначальном виде CDS/ISIS не может обеспечить вывод данных в полном соответствии с форматами MARC. Однако наличие дополнительных настроечных таблиц Z-ISIS позволяет частично скрыть некоторые проблемы в провайдере данных. На наш взгляд без радикального изменения структуры данных CDS/ISIS и языка форматирования невозможно достичь полного функционального соответствия схем данных CDS/ISIS и MARC.

В заключение хотелось бы отметить, что MARC – это формат внешнего представления данных, цель его – служить средством обмена данными. Нет необходимости хранить данные во внутреннем формате такой же структуры. И все же, чем ближе формат ввода и хранения данных к формату обмена, тем легче

осуществлять однозначные конвертации данных. Поэтому желательно, чтобы при проектировании систем, разработчики структур данных учитывали этот момент.

Литература

1. Методические материалы и документация по пакетам прикладных программ. Вып.70. ППП CDS/ISIS Версия 2.3. М.: МЦНТИ.- 1991.- 256 с.
2. International Organization for Standardization. Documentation: format for bibliographic information interchange on magnetic tape. [2nd ed.] Geneva, ISO, 1981 (ISO 2709-1981). The first edition was published in 1973.
3. Российский коммуникативный формат представления библиографических записей в машиночитаемой форме: (Рос. вариант UNUMARC). СПб.: Изд-во РНБ. – 1998.
4. Руководство по UNIMARC: Руководство по применению международного коммуникативного формата UNIMARC. – М.: ГПНТБ России. - 1992. - 320 с.
5. Форматы USMARC. Краткое описание: В 3-х ч. М.: ГПНТБ России. - 1996.
6. ANSI/NISO Z39.50-1995. Information Retrieval (Z39.50): Application Service Definition and Protocol Specification. Z39.50 Maintenance Agency Official Text for Z39.50-1995, July 1995.
7. Жижимов О.Л., Мазов Н.А. Модель распределенной информационной системы Сибирского Отделения РАН на базе протокола Z39.50. Электронные библиотеки, 1999, т.2, вып.2, ISSN 1234-5678.
8. Мазов Н.А., Жижимов О.Л. Интеграция Z39.50 и CDS/ISIS: состояние и перспективы развития. 6 Международная Конференция "Крым-99", Материалы конференции, 1999, т.2, стр. 249-251.