

ВАРИАНТ ИНТЕГРАЦИИ ЦИФРОВОЙ БИБЛИОТЕКИ И БИБЛИОГРАФИЧЕСКОГО КАТАЛОГА.

Палей Д.Э. (paley@yars.free.net), Курчинский Д.Н. (reno@econom.uniyar.ac.ru),

Смирнов В.Н. (smirnov@uniyar.ac.ru)

Ярославский государственный университет.

В докладе авторы рассматривают задачу интеграции электронной библиотеки (ЭБ) и библиографических каталогов.

В настоящее время существует несколько подходов к созданию таких систем. К ним можно отнести концепции связанные с обоснованием и разработкой возможности трансляции элементов описания структуры данных отдельных информационных систем (в том числе MARC записей) в некоторый общий формат метаданных. Одним из наиболее разработанных библиотечных решений является создание распределенных систем на основе сервисов протокола Z39.50.

В рассматриваемом авторами варианте предполагается, что библиографические записи хранятся в MARC формате и используются реально действующей информационной системой. Также предполагается, что электронная библиотека реализует объектный подход к описанию и хранению данных. Т.е. она состоит из

- репозитория, содержащего определения классов, их атрибутов, методов и (возможно) связей между классами и их атрибутами;
- хранилища объектов (экземпляров классов), которые связаны между собой;
- интерфейсов доступа к репозиторию и объектам.

В докладе не рассматривается задача репликации данных библиографического каталога в ЭБ. Предлагаемый вариант интеграции в своей основе содержит установление связей объектов ЭБ и соответствующих MARC записей.

В статье выделены возможные уровни интеграции.

1) Интеграция на основе ссылки 1:1 (запись - объект). При этом на одну MARC запись ссылается только один объект ЭБ (каждая MARC запись ссылается на один объект ЭБ)

2) Интеграция на основе ссылки 1:М (запись - объекты). Одна MARC запись может ссылаться на один или несколько различных объектов ЭБ (например на объекты классов “автор”, “организация” и т.д.). При этом возникает две основные задачи: экспорт метаданных MARC формата в репозиторий ЭБ, семантическая идентификация объектов ЭБ, связанных с конкретной библиографической записью. В докладе предложены различные варианты решения этих задач, в частности экспорт/импорт метаданных на базе SGML и RDF. Также приводятся варианты алгоритма идентификации конкретного экземпляра класса по значению поля MARC записи.

3) Интеграция на основе семантического анализа содержимого MARC записи. В основе этого уровня интеграции лежит достаточно очевидное положение о возможности семантической связи одного атрибута библиографической записи с несколькими объектами ЭБ. Причем класс этих объектов заранее неизвестен. В общей постановке решение этой задачи затруднено. В докладе рассмотрен частный вариант, основанный на идентификации объектов из заранее определенного ограниченного множества классов.

В заключение доклада рассмотрен пример реальной системы, функционирующей на описанных принципах.

INTEGRATION VARIANT OF DIGITAL LIBRARY AND BIBLIOGRAPHIC CATALOGUE.

Paley D. (paley@yars.free.net), Kurchinsky D. (reno@econom.uniyar.ac.ru),

Smirnov V. (smirnov@uniyar.ac.ru)

Yaroslavl State University

In the report authors examine the task of integration of digital library (DL) and bibliographic catalogues.

Nowadays there are a few approaches for creating such systems. It is possible to refer to them the concepts describing the proving and development of facilities for translation of elements data structure description of single information system (including MARC records) to some common format of metadata. One of the most elaborated library approaches is the creation of the distributed systems on the base of Z39.50 services.

In the presented by authors variant it is supposed, that bibliographic data are kept in MARC format and are used by really working information system. It is also supposed, that a digital library realises the object approach in description and storing data, i.e. it consists of:

- repositorium, containing definitions of classes, their attributes, methods and (possibly) relations between classes and their attributes;
- objects' repository (classes' copies), which are concerned with each other;
- access interfaces to repositorium and objects.

The report does not touch the task of bibliographic catalogue data replication in the DL. The supposed integration variant has in its base the establishment relations between DL objects and the corresponding MARC records.

The possible levels of integration are marked out in the article:

1) Integration on the base of reference 1:1 (record - object). This means that each MARC record refers on only one object of DL.

2) Integration on the base of reference 1:M (record - objects). One MARC record can refer to one or a few different objects of DL (for example, to classes' objects "author", "organisation" etc.). Here appear two main tasks: export of MARC format metadata to repositorium of DL, semantic identification of DL objects, related to the specific bibliographic record. Authors suggest different variants how to solve these tasks, particularly metadata export/import on the base of SGML and RDF. Variants of algorithm for identification of concrete class specimen according the value of MARC record field are also given in the article.

3) Integration on the base of semantic analysis of MARC record contents. The base of this integration level is in the evident thesis about possibility of semantic relation between one attribute of bibliographic record and a few DL objects. And these objects class is beforehand unknown. In general the solving of this task is difficult. Authors show a particular variant, based on identification of objects got from in advance defined limited classes' multitude.

In the end of the report the example of the real system, working on the mentioned principals, is presented.