

Информационная система учреждения культуры - как система обработки объектов электронного каталога.*

© Курчинский Д. Н., Палей Д. Э., Смирнов В. Н.

Ярославский государственный университет им. П. Г. Демидова

Аннотация

В данном докладе рассматривается один из вариантов построения системы автоматизации учета в учреждениях культуры (библиотек, музеев, архивов). В качестве технологической основы такой системы авторы предлагают использовать универсальную электронную библиотеку.

1 Введение в задачу

В настоящее время на рынке присутствует достаточно большое количество информационных систем (ИС) предназначенных для автоматизации учета фондов библиотек, музеев, архивов. К наиболее распространенным библиотечным ИС можно отнести такие: как “Библиотека 2000”, “Ирбис”, “Марк-SQL” и т.д., к ИС автоматизации музеев – “КАМИС”, “АС-Музей”

С точки зрения методологии построения, все эти системы являются специализированными приложениями, предназначенными для решения конкретных задач на четко определенных предметных областях. Для всех этих систем можно выделить следующие общие черты:

- Основу хранения и представления данных составляет стандартный для данной предметной области формат данных, например для библиотек MARC (Machine Readable Cataloguing) формат в различных вариациях [1,2].

- При моделировании предметных данных реализован ставший ‘классическим’ подход, при котором выделяют несколько стандартных сущностей, основными из которых являются книги, экземпляры коллекции, читатели, авторы и т.д. Собственно коллекции экземпляров этих сущностей и обрабатываются приложениями.

- Взаимодействие между сущностями осуществляется и обрабатывается на уровне бизнес

логики. Подчеркнем, логические и семантические связи между объектами систем жестко заданы и поддерживаются на уровне бизнес логики (за исключением некоторых типов описания литературы, которые связываются на уровне MARC формата). Эти связи обычно определяются при проектировании системы на этапе постановки технического задания. Соответственно, создание новых типов связей, или их изменение, внесение дополнительных функций в бизнес логику связано с изменением системы (т.е. собственно с ‘программированием’) и доступно только непосредственно разработчикам.

Описанные подходы, являются достаточно эффективным для выполнения традиционных задач, таких как создание электронных каталогов литературы и документов, создание коллекций описаний музейных и архивных документов, выполнения бизнес процессов, связанных с обработкой массивов этих данных. Вместе с тем, при попытках расширить функциональность информационных систем для решения задач, которые возникают в настоящее время, появляются ощутимые проблемы. К таким задачам, в первую очередь можно отнести:

- хранение и обработка полных текстов документов, изображений;

- хранение альтернативных описаний или версии документов или экземпляров коллекций;

- обработка многоверсионности документов и значений их атрибутов;

- изменение в процессе работы стандартных описаний сущностей, данные о которых хранятся в системах;

- моделирование средствами системы новых сущностей с неизвестными заранее наборами атрибутов;

- изменение существующих, создание нестандартных связей между сущностями системы и поиск, извлечение данных с их учетом.

Труды 7^{ой} Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» - RCDL'2005, Ярославль, Россия, 2005.

В данный момент некоторые из перечисленных проблем частично решаются путем создания надстроек или дополнительных модулей.

Наиболее логичным путем 'эволюционного' развития в этом случае выглядит максимально возможное сохранение существующей функциональности и модели данных – ядра упростить переход пользователей на новые версии, позволяя использовать без лишних накладных расходов данные старых версий и т.д.

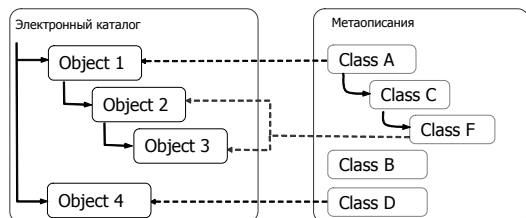
Вместе с тем, после доработок, с расширением не заложенной изначально функциональности 'под крышей' одной системы с течением времени начинают существовать несколько различных подсистем. Многие из этих подсистем во многом не зависят, а иногда даже дублируют друг друга. Являясь 'системной' эта проблема должна решаться на этапе проектирования информационной системы.

2 Электронная библиотека - основа системы

Существует несколько вариантов комплексного решения описанной проблемы. Один из них и предлагается авторами данной статьи. Решить перечисленные выше задачи в полной мере представляется возможным, если сделать технологической основой учетной системы электронную библиотеку (ЭБ).

Проекты по созданию специализированных ЭБ существовали и существуют ныне [3,4,5,6]. Работы по созданию различных типов цифровых каталогов уже давно ведутся в техническом центре Ярославского государственного университета им. П. Г. Демидова (ЯрГУ). Результаты работы по одному из таких проектов подробно описаны в [7]. Его и предлагается положить в основу системы.

В данном варианте – основу системы автоматизации учета учреждения культуры предполагается сделать в виде хранилища электронных объектов, каждый из которых будет описывать ту или иную сущность (книгу библиотечного фонда, предмет коллекции музея, читателя библиотеки и т.д.). Предполагается, что в информационной системе будет описано достаточно большое количество артефактов (физических объектов) реального мира.



Все объекты связаны друг с другом различными отношениями. При этом необходимо учитывать, что полное формальное описание структуры части объектов достаточно хорошо стандартизировано (например, библиографическое описание книги), а часть неизвестно или может меняться с течением времени, и более того, иметь несколько вариантов (например, архивный документ с комментариями

системы. А вся дополнительная логика хранения и обработки информации выносится в дополнительные модули. Это позволяет обеспечить максимальную совместимость версий, исследователей). Это же можно сказать и про связи между объектами.

Моделирование данных и связей между ними будет осуществляться в метаданных электронного каталога (ЭК). Каждая сущность описывается соответствующим ей некоторым определением класса. Там же определяются методы обработки объекта (большинство из которых будут, несомненно, методами представления объекта для других объектов и для пользователей системы). Все объекты (экземпляры классов) объединены связями "главный-подчиненный" и образуют некоторый системный каталог объектов. Подчеркнем, что связи между сущностями определяются не на уровне метаданных, а на уровне каталога объектов (экземпляров классов). Существенно упростить моделирование, использование и доступ к информации помогут следующие возможности ЭК:

2.1 Иерархии классов

Возможность создания иерархий классов позволяет сделать универсальным процесс моделирования и описания схожих объектов, имеющих много общих атрибутов.

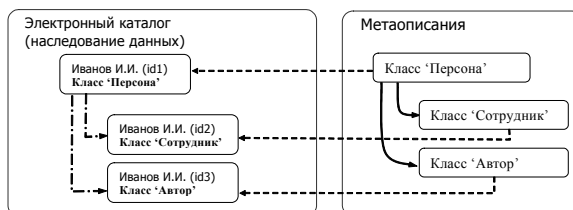
'Классическим' примером могут служить описания сущностей:

- 'персона' -> 'специалист'
- 'персона' -> 'специалист' -> 'сотрудник'
- 'персона' -> 'автор'

Класс 'персона' содержит основное описание произвольного человека, а его подклассы 'автор', 'специалист', 'сотрудник' детализируют его для конкретных предметных областей. При необходимости описание базового класса может быть изменено. Соответственно описание производных классов так же расширится.

2.2 Наследование данных

- Наследование данных [7] позволяет использовать одни и те же данные в описаниях различных объектов, при условии, что они принадлежат классам одной и той же иерархии. Поясним на простом примере.



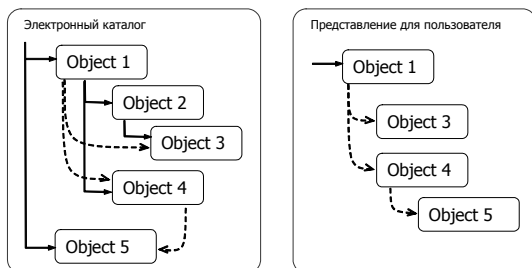
Положим, мы включили в ЭК объект класса 'персона', описывающий Иванова И.И. с

идентификатором id1. Далее мы включаем в электронный каталог объект класса 'автор', описывающий того же самого Иванова И.И. с идентификатором id2. Если класс 'автор' является подклассом 'персона', то наследование данных позволяет связать объекты (экземпляры этих классов), в данном случае id2 и id1, так что данные в объект id2 будут автоматически переноситься из объекта id1. Все это, конечно, касается только атрибутов, определенных в классе 'персона'.

2.3 Создание дополнительных 'виртуальных' каталогов

- Возможность включения одного и того же объекта в различные каталожные (иерархические структуры). В ЭК предлагается ввести некоторую "основную" иерархию объектов. "Основанная" иерархия будет определяться и модифицироваться администратором системы. В нее будет встраиваться каждый вновь создаваемый объект. Это позволит отделить этап ввода данных от этапа моделирования связей между объектами. Также это позволит избежать потерь информации при ошибках моделирования связей.

Вместе с тем, должна быть предусмотрена возможность создавать "виртуальные" каталоги объектов для конкретного пользователя (групп пользователей). Т.е. каждый объект может иметь некоторый набор связей с другими объектами, которые не отражаются в "основной" иерархии и описываются при его создании или модификации. Эти связи в общем случае могут быть как произвольны, так, и типизированы и организуют "сеть" объектов. Представление данных для конечного пользователя определяется выбранным им из существующего набора некоторым правилом (типом связей, набором и типом исходных объектов для просмотра и т.д.), по которому из исходного набора связанных сущностей формируется иерархический каталог объектов. При этом пользователь получает дерево объектов, оптимальное для решения его конкретных задач.



Отношения, не укладывающиеся в полученную иерархию, выступают в качестве дополнительных ссылок между объектами. Таким образом, появляется возможность совместить максимальную гибкость и полноту описания достаточно произвольных отношений между сущностями со

всеми преимуществами хранения и обработки данных в виде иерархических структур.

3 Проблемы и пути их решения

Таким образом, предлагается сделать электронную библиотеку ядром учетной системы (системы автоматизации). Отметим, что фактически ЭБ представляет собой независимую информационную систему, данные которой могут быть опубликованы для конечного пользователя тем или иным способом. Прикладные модули конкретной учетной системы обращаются к объектам ЭК через его интерфейсы. При этом, например, для системы учета библиотеки модуль каталогизации использует объекты книг, документов и т.д., Модуль абонемент еще и объекты 'персона'.

Вместе с тем, хорошо известно, что чем большую универсальность имеет система, тем менее она эффективна для решения конкретных задач (по сравнению с узкоспециализированными системами). Это в полной мере относится и к ЭБ. Для разработчиков основная проблема состоит в совмещении универсальных походов хранения и обработки данных в ЭК и узкоспециальных практических задач, решаемых в конкретных модулях учетных систем.

По итогам практического опыта разработки и эксплуатации АБИС [8,9,10] и различных электронных библиотек [7], разработчики считают, что при реализации проекта возникнут основные следующие проблемы:

3.1 Сложность описания сущностей в ЭК для их пользователей

Например, правила библиографического описания книги (руководство по применению одного из MARC форматов) представляют собой солидную книгу. Понятно, что работник библиотеки не сможет без специальной подготовки самостоятельно сделать адекватную модель книги в виде описания класса ЭК.

Выход из этой ситуации возможен в использовании имеющихся на сегодняшний день представлений в виде xml-schema или профилей протокола Z39.50, которые уже прошли практическую апробацию и достаточно широко используются [11]. Задача решается путем создания механизма автоматического импорта описаний в том или ином виде в ЭБ и генерации на их основе определений классов.

Конечные пользователи просто используют уже готовые наборы определений классов ЭК или автоматически их создают с использованием сервисов ЭБ.

3.2 Скорость доступа и поиска данных в ЭК

Специализированные библиотечные системы и системы учета музеев обеспечивают сейчас очень хорошие показатели по скорости поиска и обработки информации. Чтобы достичь этого уровня для ЭК (как системы обработки объектов) авторы предлагают сделать технической основой хранилища высокопроизводительную RDBMS. Проблемы моделирования данных и представления объектной модели в реляционном виде обсуждалась уже неоднократно (например [12,13,14,15,16,17]).

Авторы предлагают совместное использование вариантов поиска, основанных на построении справочных индексных таблиц (по значению атрибута) и поиска с учетом связей между объектами и их атрибутами.

Механизм построения индексных таблиц является широко распространенным в настоящее время. Его эффективность для поиска и извлечения данных подтверждена практикой. Применительно к ЭБ особенностью (и во многих случаях преимуществом) является то, что таблицы могут строиться автоматически по значениям атрибутов выделенных множеств объектов [13,14]. Идентификация объектов на принадлежность множеству может выполняться по нескольким критериям, таким как класс объекта, положение объекта в иерархии каталога, наличие связей с заданными объектами. Подчеркнем, что все эти критерии могут быть автоматически обработаны на уровне ЭК. Например, при создании объекта класса 'персона' в индексной таблице 'ФИО', настроенной на все объекты этого класса, возникнет новая учетная запись.

Поиск так же можно осуществлять с учетом связей объектов. Фактически, в этом случае, индексной таблицей выступает набор всевозможных путей на заданном множестве объектов и/или их атрибутов. Простейшим примером такого поиска является извлечение всех объектов, зависящих от указанного.

3.3 Возможности интеграции с существующими системами учета.

Т.к. важнейшей частью любой информационной системы являются содержащиеся в ней данные, то для успешного функционирования ЭБ необходима возможность непосредственного доступа к данным ИС музея или библиотеки с возможностью автоматического конвертирования информации в обе стороны.

При решении задач такого рода всегда возникают трудности (описанные выше) совмещения универсальных и специализированных подходов. Так для систем автоматизации библиотек это требование выливается в необходимость обработки MARC форматов обмена

библиографическими данными на уровне ядра электронной библиотеки.

Предлагается решить эту задачу, поместив описание соответствующих форматов непосредственно в объекты электронного каталога. Т.е. создать классы метаописаний, например, класс описания поля MARC формата. Таким образом, создав необходимый набор объектов этого класса и поместив его в соответствующий каталог ЭБ, мы получим возможность осуществлять импорт - экспорт данных в ЭК на основе метаописаний, хранящихся в этом же ЭК.

В общем случае можно выделить несколько уровней 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:M (запись - объекты). Одна MARC запись может ссылаться на фиксированный набор из одного или нескольких объектов ЭБ заранее известных типов (например на объекты классов "автор", "организация" и т.д.). При этом основной задачей является семантическая идентификация объектов ЭБ, связанных с конкретной библиографической записью. Т.к. структура MARC формата четко стандартизирована, то построение алгоритма идентификации возможно по значениям ключевых атрибутов.

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. Вместе с тем возможно ограниченное решение, основанное на идентификации объектов из заранее определенного ограниченного множества классов.

Заключение

Представленные в докладе подходы воплощаются в практические решения группой разработчиков технического центра ЯрГУ в рамках выполнения проектов поддержанных РФФИ и РГНФ.

Литература

- [1] Российский коммуникативный формат представления библиографических записей в машиночитаемой форме (Рос. вариант UNIMARC), СПб.: Изд-во РБН, - 1998.
- [2] Форматы USMARC. Краткое описание. В 3-х ч. / Пер. с англ.; ГПНТБ России. - М., 1996.
- [3] Кулагин М.В., Лопатенко А.С. Интеграция электронных библиотек и систем управления документами, как задача, решаемая системой обмена информации между объектно-

- ориентированными информационными системами. // Сборник трудов второй Всероссийской научной конференции “Электронные библиотеки: Перспективные методы и технологии, электронные коллекции”, Протвино, 26-28 сент., 2000, С. 107-116.
- [4] Kalinichenko L.A., Integration of Heterogeneous Semistructured Data Models in the Canonical one, First Russian National Conference on Digital Libraries: Advanced Methods and Technologies, Digital Collection.
- [5] Бездушный А.Н., Меденников А.М., Серебряков В.А., Власова С.А., Каленов Н.Е. Подход к интеграции информационных коллекций в ИСИР РАН. // Сборник трудов второй Всероссийской научной конференции “Электронные библиотеки: Перспективные методы и технологии, электронные коллекции”, Протвино, 26-28 сент., 2000, С. 129-135.
- [6] Рябев В.С., Соколова Н.В. Реализация электронной библиотеки на базе типовых модулей распределенной библиотечной системы. // Сборник трудов первой Всероссийской научной конференции “Электронные библиотеки: Перспективные методы и технологии, электронные коллекции”, Санкт-Петербург, 1999, С. 128-138.
- [7] Палей Д. Э., Курчинский Д. Н., Смирнов В.Н. Цифровая библиотека Ярославского региона. Итоги работы, перспективы развития. // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Труды пятой всероссийской научной конференции RCDL'2003, (Санкт-Петербург, 29 – 31 октября 2003 г.). - Санкт-Петербург: НИИ химии СПбГУ, 2003. - С. 315 – 319.
- [8] Палей Д. Э., Курчинский Д. Н., Смирнов В. Н. Принципы программно-технической организации Ярославской корпоративной библиотечной сети // Труды девятой Международной конференции “Крым 2002” “Библиотеки и ассоциации в меняющемся мире: новые технологии и новые формы сотрудничества”. г. Судак, 08 - 16 июня 2002 г. Т. 2, С. 582-588.
- [9] Палей Д. Э., Курчинский Д. Н., Смирнов В. Н. АБИС “Буки”. Первые итоги развития, новые возможности, перспективы на будущее // Информационные технологии, компьютерные системы и издательская продукция для библиотек: Доклады и тезисы докладов. Седьмая Международная конференция и Выставка “LIBCOM-2003”, пансионат “Ершово”, Звенигород, Московская область, 17 – 21 ноября 2003 г. – Москва: ГПНТБ России, 2003. – С. 192 – 200.
- [10] <http://www.buki.yar.ru>
- [11] DTD (Document Type Definition)
<http://lcweb.loc.gov/marc/marcsgml.html>

- [12] Представления идентифицируемых сложных объектов в реляционной базе данных. Е. Григорьев. “Открытые системы”, 2000, №1-2.
- [13] М. Гринев. Системы управления полуструктурированными данными. // Открытые системы, 1999, №5-6
- [14] Палей Д. Э. Моделирование квазиструктурированных данных // Открытые системы. - М.: Открытые системы, 2002. - № 9, С. 57 – 64
- [15] J. McHugh, S. Abteboul, R. Goldman, D. Quass, J. Widom. Lore: A Database Management System for Semistructured Data. www.ds.stanford.edu/lore
- [16] М.Р. Когаловский. Абстракции и модели в системах баз данных. // СУБД, 1998, № 4-5
- [17] Джим Грей. Управление данными. Прошлое настоящее и будущее. // СУБД, 1998, № 3

Culture Institution Information System As Processing System Of Digital Objects Catalogue

Kurchinsky D. (reno@econom.uniyar.ac.ru),
Paley D. (paley@yars.free.net),
Smirnov V. (smirnov@uniyar.ac.ru)
P.G. Demidov Yaroslavl State University

One of the variant culture institutions information system design is considered in this paper. For example culture institutions are libraries, archives, museums. Authors suggest using the universal digital library as the foundation of this system. The positive and negative aspects of this solution are discussed in the article.

* Работа выполняется при финансовой поддержке РФФИ (гранты № 03-07-90178, № 04-07-90154) и РГНФ (грант № 04-06-12016)