

**А. М. Федотов, В. Б. Баракнин, О. Л. Жижимов, О. А. Федотова**

Институт вычислительных технологий СО РАН  
пр. Акад. Лаврентьева, 6, Новосибирск, 630090, Россия

Новосибирский государственный университет  
ул. Пирогова, 2, Новосибирск, 630090, Россия

Государственная публичная научно-техническая библиотека СО РАН  
ул. Восход, 15, Новосибирск, 630090, Россия

E-mail: fedotov@nsu.ru, bar@ict.nsc.ru

## **ТЕХНОЛОГИЯ СОЗДАНИЯ КОРПОРАТИВНЫХ ИНФОРМАЦИОННЫХ СИСТЕМ УЧЕТА ТРУДОВ НАУЧНЫХ РАБОТНИКОВ\***

Работа посвящена описанию технологического подхода к созданию информационной системы для поддержки научных исследований, организованной в виде электронной библиотеки, предназначенной для интеграции разнородных информационных ресурсов (публикации, отчеты, фактографические данные, презентации, материалы конференций, информационные материалы и т. п.). Определен базовый профиль электронной библиотеки, предложена структура информационной системы на основе системы управления электронными библиотеками СО РАН.

*Ключевые слова:* поиск информации, электронные библиотеки, библиографические базы данных, распределенные информационные ресурсы.

### **Введение**

Разработка механизмов, обеспечивающих функционирование общей информационной среды, является приоритетным направлением для задач информационной поддержки научных исследований. Эти вопросы приобретают особую важность, когда различные группы исследователей, разделенные географически, должны осуществлять совместную работу, обмен данными и знаниями и координировать свои действия с целью оптимизации использования информационно-вычислительных ресурсов, сервисов и приложений, в частности для исследований экологических систем. Например, тесное кооперирование информационных технологий и наук об окружающей среде способствует пониманию как глобальных, так и региональных природных процессов, взаимодействия процессов, формирующих природную окружающую среду [1].

Накопленные в настоящее время массивы экспериментальных данных и данных наблюдений, представленных в различных публикациях, таблицах, презентациях и т. п., столь велики, а организация и динамика экосистем так сложна, что без современных информационных технологий, методов анализа информации их осмысление займет слишком много времени. Создание информационной среды для аккумуляции разнородных информационных источников и накопления в базах данных информации об экосистемах приобретает особую ак-

---

\* Работа выполнена при частичной поддержке РФФИ (проекты № 08-07-00229, 09-07-00277, 10-07-00302), президентской программы «Ведущие научные школы РФ» (грант НШ 6068.2010.9) и интеграционных проектов СО РАН.

туальность. Оценка темпов, масштабов и степени необратимости наблюдаемых изменений глобальных параметров биосферы представляет собой, без сомнения, одну из актуальнейших задач, стоящих перед современной наукой. Очевидно, что такую задачу нельзя решить без информационной системы, которая обеспечивает интеграцию всех сведений о биосфере с целью оценки ее состояния.

Такая система предусматривает объединение сведений о разнородных научных информационных ресурсах, обеспечение актуальности этих сведений и широких возможностей для достаточно точного поиска научных ресурсов на основе этих сведений, поддержку средств научной коммуникации, сервисов, связанных с возможностью оперативного информирования пользователей о необходимых им ресурсах и т. п. Система обеспечит пользователей актуальными данными о текущем состоянии и характеристиках научно-информационной базы. Облегченный доступ к информации (к публикациям и другим информационным ресурсам) изменит способы ведения научной деятельности, способы обучения.

В существующих информационно-поисковых системах, когда сведения о ресурсах представлены в виде слабоструктурированного текста и полнотекстовый поиск нужных данных осуществляется по запросам в свободной форме, пользователь получает огромное количество «шумовой» информации, среди которой очень трудно выбрать действительно полезные знания. Учитывая это обстоятельство, для представления сведений о ресурсах необходимо использовать каталогизацию ресурсов, структурное представление и метаданные, описывающие содержимое ресурса в виде набора именованных значений, в том числе связей с другими ресурсами [2; 3].

Метаданные используются для автоматизированного анализа содержимого ресурса, построения поисковых индексов и позволяют обеспечить достаточно высокую точность и эффективность поиска разнородной информации. Эти требования приводят к необходимости создания специализированных информационных систем, обличенных в форму электронных библиотек (ЭБ) [4–6], позволяющих решить основные проблемы интеграции разнородных распределенных информационных ресурсов на основе технологий и принципов построения открытых систем [7].

### **Электронные библиотеки**

В настоящее время научно-исследовательский процесс неотделим от использования различных электронных ресурсов доступных в сети Интернет. Значительную часть своего времени научные сотрудники проводят за компьютерами в поиске и анализе информации. Все большую роль в этом процессе начинает играть использование электронных библиотек или электронных каталогов обычных библиотек.

Электронная библиотека (ЭБ) – структурированная каталогизированная коллекция разнородных электронных документов (в отличие от печатных изданий, микрофильмов и других носителей), снабженная средствами навигации и поиска. Электронные библиотеки – явление достаточно новое, но популярное. Тем не менее электронные библиотеки сегодня следует рассматривать как множество слабосвязанных сущностей, объединяемых, на первый взгляд, только общим названием. Под термином «электронная библиотека» могут фигурировать совершенно различные объекты, такие как архивы цифрового контента и наборы программного обеспечения для управления этим контентом. Электронной библиотекой может называться система сетевых сервисов, предоставляющих доступ к цифровому контенту, объединенных единой системой управления. Кроме того, некоторые организации, которые берут на себя ответственность не только за исполнение функций управления цифровым контентом и предоставления к нему доступа всем заинтересованным лицам. Такое определение ЭБ полностью соответствует определению традиционной библиотеки как организации в системе, например, Министерства культуры [8].

Однако задача ЭБ не только обеспечить многосторонний поиск в каталоге, но и предоставить пользователю непосредственно найденный ресурс (публикацию, фотографию, описание научного факта и др.), а также дополнительные сведения о нем, например, об авторах редакторах, библиографии, организации и т. п. Важным фактором электронных библиотек является

ся определение метаданных для описания ресурсов и выделение ключевых видов субъектов и объектов.

В настоящее время нет какой-либо универсальной системы поддержки ЭБ, которая отвечала бы всем требованиям и ожиданиям пользователей. Анализ существующих систем ЭБ (см., например, [9]) показывает их разнородность на нескольких уровнях:

- на уровне информационной модели, которую они обеспечивают;
  - на уровне поддержки пользователей и групп пользователей;
  - на уровне функциональных возможностей.
- Из-за этой разнородности ЭБ и игнорирования нужд их пользователей возникает ряд проблем:

- интеграция информации из различных ЭБ;
- сравнение ЭБ по предоставляемой функциональности;
- оценка и сравнение производительности различных систем ЭБ;
- добавление новых типов хранимых объектов;
- добавление новых функциональных возможностей;
- резервное копирование.

Отметим, что основу разработки электронной библиотеки составляют, прежде всего, международные стандарты и рекомендации, формирующие профиль ЭБ, под которым понимается набор из одного или нескольких базовых нормативно-технических документов (стандартов и спецификаций), ориентированных на решение определенной задачи (реализацию заданной функции либо группы функций приложения или среды) с указанием, при необходимости, выбранных классов, подмножеств, опций базовых стандартов, которые являются необходимыми для выполнения конкретной функции [10]. Наиболее важным являются профили метаданных информации, циркулирующей в системе. Выбор профиля должен основываться на выполнении следующих требований:

- включать основные типы информации, требующейся для поддержки научной работы;
- быть открытыми, т. е. обеспечивать доступ к соответствующей информации по этим описаниям;

- быть расширяемыми, т. е. обеспечивать возможность детализации описаний;
- обеспечивать возможности интеграции информации;
- обеспечивать возможности уникальной идентификации информации;
- обеспечивать возможности размещения и поиска информации в распределенной среде;
- быть ориентированными на современные и перспективные технологии описания и использования информации;

- обеспечивать возможности интероперабельности с внешней средой.

С точки зрения потребностей научных сотрудников существенным недостатком многих схем метаданных электронных библиотек является то, что они работают лишь с так называемыми документоподобными объектами, определяют метаданные, описывающие только такие ресурсы, не выделяют другие виды важных объектов, например, персоналии, организации, коллекции и т. п. В итоге, например, встретив упоминание персоны в одном месте, невозможно точно установить соответствие с ее упоминанием в другом месте. Это обусловлено тем, что метаданные рассматриваются как нечто, связанное только с документом, их используют как средства идентификации ресурсов только для документов и только для целей их извлечения. Решение этой проблемы может быть достигнуто следующим способом. При формировании метаданных того или иного ресурса (при его каталогизации) необходимо использовать словари объектов и авторитетные базы данных (авторитетные файлы), с помощью которых устанавливать конкретные ссылки на объекты [11].

Набор элементов в специализированном профиле метаданных для электронных библиотек основан на предложениях наиболее влиятельных сообществ и организаций, выдвигающих или поддерживающих проекты стандартов (хотя значительное влияние на выбор решений оказал и анализ специфики работы научных сотрудников).

Отметим, что при работе с цифровыми объектами человечество уже выработало определенный набор стереотипов, отсутствие которых вызывает дискомфорт [8]. Одним из элементов этого набора являются требования наличия взаимных ссылок между цифровыми объектами, проявляющихся, например, в виде гиперсвязей в пользовательских графических

интерфейсах просмотра информации. Реализация взаимных ссылок в цифровых документах не представляет большой сложности, однако при этом проявляются специфические моменты. Во-первых, электронный объект с реализованными связями уже не совсем соответствует своему печатному оригиналу. Это уже другой объект. Во-вторых, внедренные в объект связи должны быть гарантированно актуальными. Никого, например, не интересуют гиперссылки, ссылающиеся на несуществующие документы. Так появляется требование обеспечения ссылочной целостности данных. Это очень жесткое требование, которое тяжело обеспечить даже в хорошо формализованных системах управления базами данных. Результат – новый цифровой объект как самосогласованное хранилище цифрового контента, или база данных цифровых объектов.

С другой стороны, в электронной библиотеке объекты хранения могут содержать информацию, которая не имеет к объектам хранения традиционных библиотек вообще никакого отношения. Речь может идти:

- об электронных копиях элементов хранения традиционных архивов;
- об изображениях элементов хранения традиционных музеев;
- о видео-, аудиоинформации, полученной разными способами, например, видеозапись доклада, сделанного на конференции;
- о научных или других фактах и т. д.

### **Профиль электронной библиотеки**

Придание конкретной информационной системе (ИС) перечисленных выше свойств открытых систем реализуется с помощью разработки ее профиля (функционального стандарта ЭБ). В соответствии с этим открытые системы по определению IEEE определены как системы, в которых реализован «исчерпывающий и согласованный набор базовых международных стандартов информационных технологий и профилей функциональных стандартов, которые специфицируют интерфейсы, службы и поддерживающие форматы данных, чтобы обеспечить интероперабельность и мобильность приложений, данных и персонала».

Каждую сложную интегрированную ИС, как уникальную, так и типовую, тиражируемую для определенной области применения, предлагается сопровождать ее профилем, включающим в себя совокупность базовых стандартов и спецификаций, которым должны отвечать как ИС в целом, так и ее составные части.

Необходимость стандартизации интерфейсов и протоколов для области телекоммуникаций была понята еще 20 лет назад. В отрасли связи сложились подходы и методология, без которых немисливо было бы построение сетей передачи данных, локальных и глобальных вычислительных сетей. Были разработаны эталонная модель взаимосвязи открытых систем – OSI/RM [12] и соответствующие ей функциональные стандарты. При этом то, как должны быть построены открытые системы, между которыми устанавливается взаимосвязь, модель OSI/RM не устанавливает. В зависимости от сферы распространения профилей ИС рассматриваются следующие их категории:

- профили конкретных ИС, определяющие стандартизованные проектные решения в пределах проекта данной ИС и имеющие статус документации проекта в части нормативных требований или статус стандарта предприятия, для которого создается эта ИС;
- профили группы типовых тиражируемых ИС, предназначенных для определенной области применения, имеющие статус отраслевого (ведомственного) стандарта для этой области или статус стандарта организации, разрабатывающей и поставляющей такие ИС (системного интегратора).
- стратегические профили для определенной области применения ИС, определяющие ориентацию информатизации этой области на долгосрочный период, например, профили переносимости приложений между разными ИС в этой области.

Для указанных целей предложена эталонная модель среды открытых систем – OSE/RM. Модель OSE/RM, принятая в качестве основы для предлагаемой методики, закреплена документами ISO/IEC [12].

В крупном плане концептуальная модель предусматривает разбиение ИС на приложения (прикладные программные комплексы), реализующие заданные функции ИС, и среду, обес-

печивающую подготовку и выполнение приложений. Между ними определяются стандартизованные интерфейсы прикладного программирования (API). Кроме того, определяются стандартизованные интерфейсы взаимодействия данной ИС с внешней для нее средой – другими ИС и сетью Интернет и (или) корпоративными сетями (ЕЕИ).

Спецификации функций компонентов ИС рассматриваются по четырем функциональным группам:

- функции, обслуживающие интерфейс ИС с пользователями;
- функции организации процессов обработки данных (системные функции среды);
- функции представления и хранения данных;
- коммуникационные функции.

Выбранная нами концептуальная модель профиля ЭБ основывается на модели предложенной в работе [13]. Объекты, подлежащие стандартизации в рамках ЭБ, представлены в таблице.

Технологическая часть, связанная с реализацией ЭБ, пристроена на основе Системы управления электронными библиотеками (СУЭБ), реализованной в ИВТ СО РАН [5]. Поэтому основными задачами стандартизации является информационная часть ЭБ. В настоящий момент определены следующие стандарты, связанные с информационными ресурсами электронных библиотек:

- построение электронных каталогов ЭБ и доступ к ним, а также доступ к библиографическим базам данных;
- электронные представления полных текстов изданий и аудиовизуальных материалов, хранящихся в ЭБ.

Для идентификации периодических изданий и отдельных публикаций (статей) в этих изданиях также применяется стандарт ANSI/NISO Z39.56-1991. Serial Issue and Contribution Identifier (SICI) [15; 16]. Стандарт SICI в первоначальной версии широко используется на уровне идентификации изданий во многих библиотечных системах мира как важный элемент сообщений электронного обмена данными.

Для обмена библиографическими данными применяется стандартный формат MARC (Machine Readable Cataloguing, ISO-2709), являющийся основой для формирования электронных каталогов библиотек и библиографических баз данных [17]. Правда, в последнее время физический формат записи ISO-2709 чаще заменяется на его XML-представление в соответствии с различными правилами, например, marcXML [21] или OAI marcXML. Традиционные схемы MARC (MARC21, RUSMARC) также вытесняются более гибкими METS и MODS.

Для формирования простых метаданных применяются несколько стандартов, являющиеся расширениями рекомендаций Dublin Core. Используемый профиль определяет список элементов данных (полей), необходимых для создания записи соответствующего типа (вида) и раскрывает содержание элементов данных [8].

### **Каталогизация и метаданные**

В традиционных библиотеках каталогизация реализует основную парадигму упорядочивания информации и обеспечения ее поиска по заранее определенным критериям. Здесь следует еще раз обратить внимание на то, что, с одной стороны, в электронных библиотеках могут существовать цифровые объекты, не имеющие аналогов в традиционных библиотеках и, как следствие, не попадающие под действующие правила каталогизации. С другой стороны, развитие пользовательских интерфейсов для доступа к информации требует возможности расширения списка атрибутивной информации, подлежащей вводу при каталогизации первичных объектов [8]. Например, уже сегодня прослеживается потребность привязки контента к географическим координатам, которая полностью игнорируется действующими правилами и сложившейся практикой каталогизации. Наконец, в-третьих, существует необходимость описания не только информационного контента объекта, но и общего контекста существования объекта с фиксацией всех событий в процессе его существования и непосредственной связи с другими объектами и персонами, имеющими к нему отношение [11].

Отдельно можно подчеркнуть, что при попытке описания контекста перестают работать все действующие библиотечные правила каталогизации. В качестве иллюстрации можно

## Объекты стандартизации профиля электронной библиотеки

	Пользователь	ПО	Хранилище	Коммуникации
Приложения ЭБ	Формулирование запросов пользователей ЭБ к поисковым услугам через посредника с тезаурусом и рубрикаторм (стандарт ISO 5964). Web-браузеры на клиентских рабочих местах (стандарты Интернет)	Услуги автоматического индексирования документов для поиска информации. Услуги поиска информации в распределенных хранилищах данных ЭБ. Протокол LDAP (X.500)	Хранилище электронных каталогов. Форматы MARC (MARC21 <sup>*</sup> , RUSMARC, МЕКОФ [14] и др.). Хранилище электронных копий полнотекстовых документов. Форматы PDF, TIFF, JPEG и др. База метаданных. Форматы Dublin Core <sup>**</sup> , MODS <sup>***</sup> , METSXML. Базы данных тезаурусов и классификаторов (БДТК), ISO 5974, ГОСТ 7.25	Доступ к онлайн-вым публичным каталогам (OPAC). Протокол Z39.50, SRW/SRU. Обмен полными текстами документов. Протокол электронной почты Интернет. Доступ удаленных пользователей к базе метаданных, протокол OAIPMH. Доступ к БДТК. Протоколы Z39.50, SRW/SRU.
ПО промежуточного слоя	Средства защиты информации ЭБ от несанкционированного доступа (руководящие документы Гостехкомиссии РФ)	Серверы приложений (спецификации COM/DCOM и EJB) Серверы обработки транзакций к базам данных.	Серверы реляционных баз данных. Серверы иерархических БД. Каталоги LDAP. Индексы полнотекстовых данных.	Услуги Web-сервера, в том числе шлюз Z39.50 – WWW. Услуги телекоммуникационной среды на прикладном уровне. Протоколы FTP, HTTP.
Операционные системы	Аутентификация пользователей	Стандартные функции операционных систем типа Unix (стандарты POSIX) и MS Windows	Файловые системы ОС типа Unix и MS Windows.	Стек протоколов телекоммуникационной среды на транспортном и сетевом уровнях TCP/IP

\* Форматы MARC21 (<http://marc21.rsl.ru>).

\*\* DCMI – Dublin Core Metadata Initiative (<http://www.dublincore.org/>).

\*\*\* Metadata Object Description Schema (MODS) / O-cal Web Site. (<http://www.loc.gov/standards/mods/mods-schemas.html>)

привести попытку создать разумное описание цифрового объекта, который является изображением глиняной таблички, найденной в точке с координатами  $(x_1; y_1)$  в момент времени  $t_1$ , помещенной в хранилище с координатами  $(x_2; y_2)$  в момент времени  $t_2$ , сфотографированной в момент времени  $t_3$  на выездной выставке в точке с координатами  $(x_3; y_3)$ . При этом на глиняной табличке описано событие, имеющее место быть в момент времени  $t_0$  в точке  $(x_0; y_0)$ . Обязательным требованием к структурированному описанию первичного объекта должно быть требование возможности поиска по всем временным и пространственным характеристикам как контекста, так и контента.

Современная основа каталогизации – это придание информационным объектам специализированной дополнительной информации, называемой метаданными. Метаданные играют в ЭБ двоякую роль. С одной стороны, они служат «обменными правилами (схемами)», с разными уровнями детализации, для обмена данными между системами, входящими в состав ЭБ. С другой стороны, они дают описание конкретных информационных систем для научных институтов. Современные технологии позволяют, во-первых, внедрять метаданные в информационные объекты и, во-вторых, организовывать поиск по ним. При этом метаданные образуют с объектом единое целое, а функциональность сервисов доступа к массивам информации не страдает. Заметим, что внедрение атрибутивной информации в цифровой объект могло существенно упростить технологии атрибутивного поиска.

Другой возможный тип поиска – поиск по заданным шаблонам. Наконец, поиск с привлечением онтологии является поиском более интеллектуальным, для его реализации требуется дополнительная информация о предметной области, включающая определения терминов, сущностей и связей. Следует отметить, что представление этой дополнительной информации должно соответствовать глобальным договоренностям – международным стандартам, иначе, поиск с привлечением онтологии всегда будет ограничен текущей системой, а интероперабельность не будет реализована [3].

Использование публикаций (информационных ресурсов) в научно-исследовательском процессе выдвигает необходимость быстрого ознакомления с содержанием публикации, и аннотации здесь может оказаться недостаточной. В связи с этим должны быть разработаны средства полуавтоматического выделения оглавления и фактов (научных результатов в соответствии с онтологией, понятиями) с обеспечением ссылок на соответствующие разделы документа, а также средства работы с библиографическими ссылками. Достаточно важными отношениями при составлении метаописания являются связи с предметной областью и с ее понятиями (концептами).

В целях обеспечения поддержки различных уровней детализации информации о публикациях, необходимых различным приложениям, библиографическая специализация разделена на базовую и расширенную подсхемы, а также выделяется академическая подсхема, отражающая специфику научных публикаций. Уже на базовом уровне требуется структурировать информацию обо всех вышестоящих библиографических уровнях для каждой публикации. Например, для описания ряда статей в журнале необходимо описать сам журнал как издание сводного уровня, далее описать интересные выпуски этого журнала как издания монографического уровня и, наконец, сами статьи как издания аналитического уровня<sup>1</sup>. И статья, и выпуск, и журнал как таковой являются полноценными структурированными ресурсами, которые описываются лишь единожды и связываются с помощью ссылок.

Такой структурированный подход требует некоторого усилия со стороны систем с «планарным» описанием публикаций. Однако структуризация информации обо всех библиографических уровнях необходима и крайне важна для схем данных ЭБ. Она позволяет избежать дублирования информации, эффектов наличия опечаток в названиях группирующих выпусков, серий и пр., позволяет представить пользователю информацию в целостном и непротиворечивом виде.

---

<sup>1</sup> При этом, конечно, весьма желательно описывать и факты, изложенные в данной статье.

## Информационные объекты

Информационный объект является корневым объектом в представляемой модели, он охватывает все объекты, информация о которых хранится в электронной библиотеке. В соответствии с рекомендациями Dublin Core информационный объект должен обладать базовым набором атрибутов. Набор атрибутов объекта расширяется в зависимости от его типа.

В базовый набор входят следующие атрибуты:

- идентификатор объекта;
- название;
- авторы (редакторы и другие персоны, имеющие отношение к объекту);
- тема;
- ключевые слова;
- версия;
- аннотация;
- издатель (или собственник) и характеристики издания;
- описание объекта (например, библиографическое описание публикации).

Эти атрибуты наследуются всеми другими объектами иерархии типов и подтипов объектов.

В системах ЭБ помимо информационных объектов необходимо предусмотреть хранение некоторых других объектов (сущностей), имеющих отношение к информационным объектам (в стандартах описаний именуемые словарями или авторитетными файлами):

- организации, отделы организаций и издательства, где создавались или публиковались рукописные объекты;
- люди (сущность, персона), работающие в этих организациях (отделах) – авторы объектов;
- проекты, в рамках которых создаются объекты;
- научные журналы (периодические издания);
- конференции их публикации;
- и т. п.

Объект Коллекция может быть применим к любой совокупности (группировке, агрегации) информационных объектов. Коллекциями могут быть как совокупности информационных объектов, так и совокупности персон, организаций, журналов и т. д. Критерии для таких совокупностей могут определяться, например, общностью местоположения, авторов, хронологией, тематикой, происхождением или принадлежностью и т. д. Коллекции могут содержать любое число объектов, причем критерии отбора этих объектов со временем могут изменяться.

В иерархии объектов нужно также перечислить классификаторы, используемые при задании некоторых их атрибутов. Так, например, атрибуты тема и ключевые слова, как правило, должны задаваться с помощью распространенных тематических или предметных классификаторов: ГРНТИ, УДК, ББК, тематического классификатора ВАК и некоторых других: DDC, LCC, LCSH, MESH. Атрибут язык желательно определять в соответствии со стандартами RFC 1766 (ISO 639-2, ISO 3166); географическое положение – в соответствии со спецификациями FGDC CSDGM или ISO-19115; форматы файлов задавать контролируемым словарем MIME; при описании сущности персона использовать набирающий все большую популярность FOAF.

Еще одна задача, которую должны решать ЭБ, – это идентификация документов. Эту задачу следует решать путем установления связей со словарями и авторитетными файлами. Словарь (авторитетный файл) является списком сущностей, которые определяют принадлежность документа: Персоны (авторы и редакторы), Организации, Журналы, Месторасположение и т. д. Для идентификации документов необходимо установление жестких связей с этими сущностями. Связи между объектами и записями в авторитетных файлах моделируются путем формирования списка ссылок. При таком способе установления связей они сохраняются даже при изменении значения записи (изменение фамилии), а идентификация может быть произведена и при совпадении значений (это свойство, к сожалению, не выполняется в большинстве существующих в настоящее время ЭБ) [11].

Отметим, что помимо общепринятых описательных метаданных основные сущности электронной библиотеки должны быть снабжены именованными отношениями, из которых можно выделить следующие:

- входит в состав (ссылка: объект) – данный ресурс является физически или логически частью указанного ресурса;
- включает (ссылка: объект) – данный ресурс физически или логически включает указанный ресурс;
- работал (ссылка: субъект; атрибут: время);
- преподавал (ссылка: субъект; атрибут: время);
- изображен (ссылка: объект; атрибут: время);
- ученик (ссылка: субъект);
- автор (ссылка: объект; атрибут: время);
- персонаж (ссылка: объект; атрибут: время).

В качестве примеров таких отношений, представленных в ЭБ, можно привести, в частности:

- документ, описывающий книгу, и совокупность документов, описывающих отдельные главы этой книги;
- описание конференции и список докладов, тезисов и презентаций на этой конференции.

Такого сорта отношения между документами моделируются путем задания связей типа «родитель-потомок». Последний вид связи, реализованный в модели ЭБ, определяет списки подчиненных документов в зависимости от условий истинности заданных администратором предикатов.

В соответствии с правилами классификации документов, описывающих публикации [15], с учетом расширений, необходимых для каталогизации таких документов, как электронные информационные ресурсы, конференции, мероприятия и т. п., была проведена систематизация информационных объектов ЭБ. Выделено шесть видов информационных объектов и 10 классов объектов, каждый из которых разбивается на подклассы (от 2 до 10). Кроме того, определено понятие тип объекта (например, препринт, фотография, диссертация, отчет и т. д. – всего 63 типа), каждый из которых имеет собственное расширение схемы метаданных.

## Заключение

В данной работе рассмотрены основные понятия теории информационных систем и принципы создания электронных библиотек. Предложена технология создания информационных систем, интегрирующих разнородные документы, такие как публикации, презентации, конференции и электронные ресурсы. Разработана структура каталогов, словарей, авторитетных файлов, словарей-классификаторов и связей между ними.

На основе изложенной технологии была создана «Электронная библиотека по моделям динамики изменения биосферы<sup>2</sup>», предназначенная для поддержки работ, проводимых по проекту СО РАН «Модели изменения биосферы на основе баланса углерода (по натурным и спутниковым данным и с учетом вклада бореальных экосистем)».

Для управления ресурсами использована Система управления электронными библиотеками (СУЭБ), разработанная в ИВТ СО РАН. В указанной СУЭБ определение всех интерфейсов системы дается через семейство шаблонов: шаблон коллекции, шаблон документа, шаблон связи и т. п. Для решения задачи представления информации пользователю (определение интерфейса пользователя) в СУЭБ предусмотрены два типа шаблонов: шаблон коллекции (ШК) и шаблон документа (ШД).

Шаблоны определяют правила формирования документов в коллекции. Содержательная часть документа состоит из трех правил:

- 1) правило представления содержания метаданных документа, записанных в базе данных;
- 2) правило формирования списков документов, связанных с данным документом (дочерние документы);

---

<sup>2</sup> [http://www.sbras.ru/win/elbib/data/show\\_page.dhtml?2+330](http://www.sbras.ru/win/elbib/data/show_page.dhtml?2+330)

3) правило задания области поиска документов.

Поддержка интероперабельности обеспечивается использованием открытых стандартов (OSI, DC, Z39.50, SRW/SRU, OAI PMH).

Все связи реализованы при помощи ссылочных механизмов, как с использованием механизма связей в реляционной модели (связи каталог – коллекция цифровых документов), так и с использованием механизмов направленных связей в иерархической модели (отображенной на реляционную СУБД). Ссылки реализованы с помощью встроенных процедур.

Разработаны интерфейсы для ввода информации и шаблоны для вывода и представления списков подчиненных документов, формируемых как виде отношений типа родитель-потомок, так и в виде заданий логических условий отбора документов.

### Список литературы

1. Ермаков Н. Б., Столяров С. В., Федотов А. М. Модели данных для формирования биологических коллекций // Вестн. Новосиб. гос. ун-та. Серия: Информационные технологии. 2007. Т. 5, вып. 2. С. 35–41.

2. Федотов А. М., Барахнин В. Б. Проблемы поиска информации: история и технологии // Вестн. Новосиб. гос. ун-та. Серия: Информационные технологии. 2009. Т. 7, вып. 2. С. 3–17.

3. Шокин Ю. И., Федотов А. М., Барахнин В. Б. Проблемы поиска информации. Новосибирск: Наука, 2010.

4. Федотов А. М. Концептуальные подходы к построению распределенных систем // Тр. Междунар. конф. по вычислительной математике МКВМ-2004. Новосибирск: Изд-во ИВМиМГ СО РАН, 2004. С. 132–143.

5. Федотов А. М. Методологии построения информационных систем // Вычислительные технологии. 2006. Т. 11. С. 3–17.

6. Шокин Ю. И., Федотов А. М., Жижимов О. Л., Гуськов А. Е., Столяров С. В. Электронные библиотеки – путь интеграции информационных ресурсов Сибирского отделения РАН // Вестн. Казах. нац. ун-та. Спец. выпуск. Алматы, 2005. № 2. С. 115–127.

7. ISO/IEC 7498-1:1994. Information technology – Open Systems Interconnection – Basic Reference Model: The Basic Model.

8. Жижимов О. Л., Мазов Н. А., Федотов А. М. Некоторые заметки об эволюции цифровых репозитариев традиционных библиотек к полнофункциональным электронным библиотекам // Вестн. Владивосток. гос. ун-та экономики и сервиса. Территория новых возможностей. 2010. № 3 (7). С. 55–63.

9. Candela L., Castelli D., Fuhr N., Ioannidis Y., Klas C.-P., Pagano P., Ross S., Saidis C., Schek H.-J., Schuldt H., Springmann M. Current Digital Library Systems: User Requirements vs Provided Functionality. IST-2002-2.3.1.12. Technology-enhanced Learning and Access to Cultural Heritage. March 2006.

10. ГОСТ Р ИСО / МЭК ТО 10000-2-99. Информационная технология. Основы и таксономия функциональных стандартов. Ч. 2: Принципы и таксономия профилей ВОС.

11. Федотов А. М., Жижимов О. Л., Князева А. А., Колобов О. С., Мазов Н. А., Турчановский И. Ю., Федотова О. А. Проблемы авторитетного контроля для распределенных электронных библиотек и библиографических баз данных // Вестн. Новосиб. гос. ун-та. Серия: Информационные технологии. Т. 9, вып. 1. С. 89–101.

12. ISO/IEC 14252-1996 (ANSI/IEEE Std1003.0-1995 ). Information technology – Guide to the POSIX Open Systems Environment (OSE).

13. Филинов Е. Н. Архитектура и структура среды распределенной обработки данных, методы и средства формального описания среды // Распределенная обработка информации: Тр. VI Междунар. семинара. Новосибирск, 1998. С. 101–105.

15. ANSI/NISO Z39.50-1995. Information Retrieval (Z39.50): Application Service Definition and Protocol Specification. Z39.50 Maintenance Agency O-cal Text for Z39.50-1995, July 1995.

16. ISO 23950:1998. Information and documentation – Information Retrieval (Z39.50) – Application Service Definition and Protocol Specification.

17. Функциональные требования к библиографическим записям: окончательный отчет / Рос. библ. ассоц. М.: РГБ, 2006.

*Материал поступил в редколлегию 03.03.2011*

**A. M. Fedotov, V. B. Barakhnin, O. L. Zhizhimov, O. A. Fedotova**

**THE TECHNOLOGY OF CREATION OF CORPORATE INFORMATION SYSTEMS  
FOR ACCOUNTING RESOURCES, CREATED BY RESEARCHERS**

This paper describes the technological approach to creation of information system to support scientific research, organized in the form of electronic library and intended for integration of heterogeneous information resources (publications, reports, factographic information, presentations, conference materials, information materials, etc.). The baseline profile of e-library is defined, the structure of information system based on digital library management system SB RAS is proposed.

*Keywords:* information retrieval, digital library, bibliographic databases, distributed information resources.