КОНЦЕПТУАЛЬНАЯ МОДЕЛЬ НАУЧНОЙ ЭЛЕКТРОННОЙ БИБЛИОТЕКИ

А. М. Федотов[1,3], *О. А. Федотова*[2,3]
[1]Институт вычислительных технологий СО РАН, 630090, Новосибирск
[2]Государственная публичная научно-техническая библиотека СО РАН, 630090, Новосибирск
[3]Новосибирский государственный университет, 630090, Новосибирск

Одной из основных современных тенденций социально-экономического развития общества является информатизация всех областей человеческой деятельности. Инструментом реализации этого всеобъемлющего процесса являются **информационные системы (ИС)** — системы сбора, хранения, обработки, преобразования, передачи и обновления информации. Работа посвящена вопросам описания и создания информационных моделей интенсивно развивающегося класса ИС — электронным библиотекам.
Ключевые слова: информационная система, электронная библиотека, модель, метаданные, данные, информация, знание.

REFERENCE MODEL OF THE SCIENTIFIC DIGITAL LIBRARY

A. M. Fedotov[1,3], *O. A. Fedotova*[2,3]
[1]Institute of Computational Technologies SB RAS, Novosibirsk, Russian Federation
[2]State Public Scientific and Technical Library SB RAS, Novosibirsk, Russian Federation
[3]Novosibirsk State University, Novosibirsk, Russian Federation

One of the main modern trends in the socio-economic development of society is the informatization of all areas of human activity. The tool for implementing this comprehensive process is **information systems (IS)** — systems for collecting, storing, processing, converting, transmitting and updating information. The work is devoted to the questions of description and creation of information models of intensively developing class of IS — to digital libraries.
Keywords: information system, digital library, model, metadata, data, information, knowledge

Введение

В зависимости от конкретной области применения информационные системы могут сильно различаться по своим функциям, архитектуре, реализации. Однако можно выделить одно свойство, которое являются общими для всех информационных систем (ИС). Любая ИС предназначена для сбора, хранения и обработки информации. Поэтому в основе любой ИС, в том числе электронной библиотеки (ЭБ), лежит среда хранения и доступа к данным [13]. Среда должна обеспечивать уровень надежности хранения и эффективность доступа, которые соответствуют области применения ИС.

Итак, ИС предназначены для:

- организации хранения информации (организация хранилищ, поддержка систем хранения данных);
- управления информацией (добавление, модернизация, изменение данных);
- управления доступом к информации (контроль исполнения правил регламентации доступа к данным), идентификация данных;
- поиска информации;

– извлечения информации и предоставление ее пользователю (компьютерному приложению) в необходимом ему виде (формате);
– визуализации (представления) информации в соответствии требованиями пользователя.

Цель работы ИС — обеспечение конечного пользователя необходимой информацией. Пользователя, как правило, не интересует, как устроена технологическая «кухня» информационной системы. Более того, чем меньше эта «кухня» пользователю заметна, тем лучше построена та или иная ИС. Следует думать, что использование проверенных временем и практикой типовых решений позволит обеспечить эту технологическую «прозрачность» [16].

«Проектирование действенной информационной системы требует глубокого понимания психологии пользователей и знания их социального контекста. Мы не можем исходить из предположения, что люди хотят получить от нас информацию, даже когда знаем, что они нуждаются в ней. В основе большинства проваливших веб-сайтов, невостребованных локальных сетей и никому не нужных интерактивных продуктов лежали ошибочные представления о пользователях и неверные модели их поведения при поиске информации. Пользователи — явление сложное. Пользователи — явление социальное. Информация — тоже».

«Таким образом, я предполагаю,» — писал К. Муэрс — *«что многие люди, возможно, не захотят получать информацию — и будут избегать пользоваться системой именно потому, что она снабжает их информацией. Обладание информацией причиняет беспокойство и создает неудобства. Всем нам знакомо это чувство. Получив информацию, вы должны прочитать ее, а это не всегда просто. Затем вам нужно постараться ее понять. Понимание может выявить, что вы действовали не в том направлении или что ваша работа была бесполезна. Получается, что отсутствие информации создает меньше проблем, чем ее наличие и использование.»*

Востребованность ИС подчиняется двум законам. *Закон первого лица*: «Информационная система любого предприятия работает только в том случае, если ее работа поддержана соответствующими нормативными актами предприятия и с ней работает первое лицо предприятия» и *закон Муэрса* [21]: «Степень использования информации прямо пропорциональна легкости ее получения». Система получения информации окажется невостребованной, если обладание информацией будет вызывать у клиента больше неудобств и беспокойства, чем ее отсутствие [21].

Очевидно, что автоматическая или автоматизированная переработка «информации» возможна лишь при наличии ее описания с помощью некоторого алгоритма, т. е. при наличии формальной модели данных и некоторой системы, которая ее воспринимает. Как отмечал А. А. Ляпунов, «информация всегда относительна, она зависит ... от того, какой информационной системой она воспринимается» [8].

Статья посвящена описанию информационной концептуальной модели ЭБ. Модель должна описывать, какие сущности могут быть представлены в ЭБ, должна фиксировать правила и отношения (связи) между сущностями, что в частности предполагает классификацию сущностей, абстрагирование, обобщение.

Большой интерес представляют «эталонные» модели цифровой библиотеки. Заметим, что русский термин «электронная библиотека» не совсем точно отражает содержа-

ние, которое в него вкладывается, точнее было бы этот класс информационных систем называть «цифровая библиотека» (как английский эквивалент «digital library»).

Концептуальные эталонные модели цифровой библиотеки (например, DLRM [17] или OAIS RM [20]) опираются на мировой опыт и включает такие понятия, как содержание, функциональность, пользователь, политика, качество и архитектура, что приводит к общему пониманию сущностей электронных библиотек [10]. Основная цель разработки модели ЭБ — описать фундаментальные понятия, существенные объекты и отношения, стандартные функциональные и структурные блоки и процессы, из которых состоит универсум ЭБ. Эталонная модель предназначена для разработки более узких моделей с конкретной архитектурой для последующей реализации в виде программной системы.

1 Определения ИС

В этом вопросе нет единства определений. В обиходе **информационными системами** обычно называют различные комплексы программно-аппаратных средств, позволяющие работать с данными, структурированными при помощи той или иной формальной модели.

При этом до сих пор присутствует путаница с понятиями «информационная система» (ИС) и «архитектура ИС»; она вовсе не безобидна и часто мешает на практике четко определить, что же является предметом разработки в конкретном проекте: ИС, только ее КСА (комплекс средств автоматизации) или автоматизированная система (АС) целиком. Для примера приведем два разных определения из ГОСТ'ов:

Информационная система: Комплекс, состоящий из процессов, технических и программных средств, устройств и персонала, обладающий возможностью удовлетворять установленным потребностям или целям [3];

Информационная система: Система, предназначенная для сбора, передачи, обработки, хранения и выдачи информации потребителям и состоящая из следующих основных компонентов: программное обеспечение, информационное обеспечение, технические средства, обслуживающий персонал [4].

Стоит напомнить, что в 1950-е и 1960-е годы в СССР науки *об информации* занимали весьма достойное место, хотя их развитие и было затруднено спецификой социального устройства общества. Произошедший в последующие годы разворот к работе с данными и размытие термина «информатика» привели к тому, что многое из накопленного оказалось если не потерянным, то невостребованным, а культура работы с информацией была утеряна. Сегодня для большинства пользователей важнее потребление информационных сервисов, а не обеспечивающие его технологии.

Но наиболее серьезной проблемой является кризис в сфере представления информации. Деструктивным моментом является отсутствие единых общепринятых определений в сфере информационных технологий, когда речь идет об обработке «информации». Прежде всего потому, что со времен А. Н. Колмогорова и Клода Шеннона на инженерном уровне произошло смешение понятий, объединение представлений об информации и данных или сигналах, кодирующих эту информацию, и под «информацией» стали понимать, по существу, наборы данных.

До последнего времени, пока ИС были относительно просты, отсутствие четкого раз-

деления на «данные» и «информацию» не имело практического значения. Но с появлением сложных ИС, где функции распределены между человеком и машиной, а также с развитием таких дисциплин, как поддержка принятия решений и управление знаниями, требуются более точные определения базисных понятий: «данные», «информация» и «знание».

На сегодняшний день имеется два определения ИС (технологическое и инженерное):

Информационная система — это набор технологий, направленных на поддержку жизненного цикла «информации» и включающих три основные составляющие процесса: обработку и управление данными, управление информацией и управление знаниями [11];

Информационная система — это программно-аппаратный комплекс, включающий вычислительное и коммуникационное оборудование, программное и лингвистическое обеспечение, информационные ресурсы, а также обслуживающий (системный) персонал.

Часть реального мира, которая моделируется ИС, называется ее *предметной областью.* Поскольку модель предметной области, поддерживаемая информационной системой, материализуется в форме организованных необходимым образом информационных объектов, она называется *информационной моделью* (см. рис. 1). Информационные объекты характеризуются метаданными, описывающим реальный объект, и могут быть снабжены аннотациями. Информационные объекты могут иметь информационное содержание (контент).

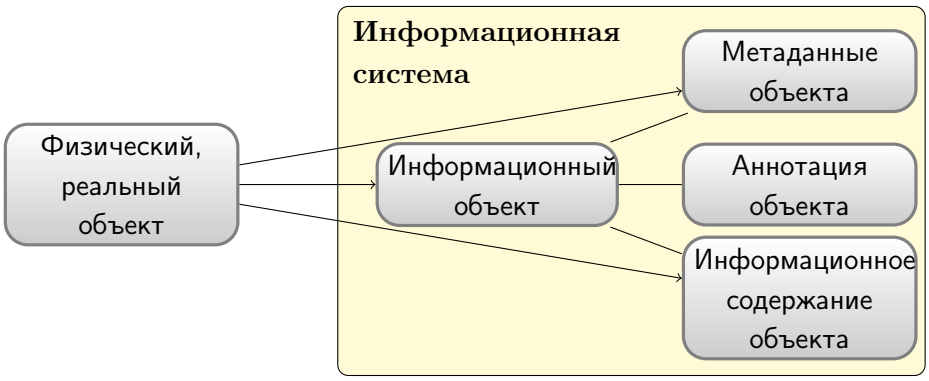

Рис. 1: Информационная модель ИС

Приведенное определение охватывает все классы ИС, в частности фактографические системы, которые основаны на технологиях баз данных и оперируют структурированными данными, документальные системы, оперирующие документами на естественных языках.

2 Информационный ресурс

Информационный ресурс — это понятие, включающее любую сущность ИС или ЭБ. В информационном пространстве все сущности (документы, публикации, персоны, события, факты, программы и любые другие сущности реального или виртуального мира) существуют только в форме некоторых информационных объектов. Информационный ресурс — это абстрактное понятие, выражаемое экземплярами одной из своих специализаций. В частности, экземплярами понятия информационного ресурса являются экземпляры информационного объекта любого типа.

Огрубляя сказанное, функционально любая ИС представляет собой систему управления информационными ресурсами с присущими ей функциями (методами), отношениями и связями [16].

Реализация информационного ресурса (информационный объект) — это единица информации, представляющая собой уникально именованный набор данных (см. рис. 2), структурированных в виде присущих ему именованных атрибутов и методов, характеризующих его свойства и связи (отношения) с другими ресурсами. Таким образом, каждый информационный ресурс должен:

– иметь идентификатор;
– быть организованным в соответствии с описанием ресурс (ресурсы могут быть сложными и структурированными, а с организационной точки зрения они могут группироваться в наборы ресурсов, которые рассматриваются как единая сущность);
– регулироваться функциями, управляющими его жизненным циклом, характеризуется набором присущих ему атрибутов и методов, характеризующих его свойства и связи с другими ресурсами;
– выражаться через информационный объект.

Рис. 2: Определение ресурса в ИС

Каждый информационный ресурс хранится в некотором репозитории как глобально и уникально именуемый набор структурированных данных (сведений о ресурсе, его свойств, атрибутов, связей) и, возможно, информационного содержания, например, одного или более форматов представления каталогизируемого ресурса. Эти структурированные данные, описывающие ресурс, а в связи с этим называемые метаданными, используются для получения представления о его свойствах, содержании, структуре, организации поиска, способах использования и т. п.

Дадим определения.

Репозиторий — это независимая система долговременного, надежного хранения и доступа к разнородным цифровым объектам, которая предназначена для предоставления электронных (цифровых) версий документов (книг, научных статей, репринтов, писем, изображений и других материалов, представленных в электронном виде), и предоставляющая некоторый четко специфицированный способ управления (схему данных, модель операций), включающий в той или иной мере способы доступа, выборки и манипулирования информационными ресурсами.

Коллекция — это совокупность информационных объектов (информационных ресурсов), объединенных общими свойствами (например, принадлежностью к одному классу объектов, одинаковой структурой, общей тематической направленностью и т. п.).

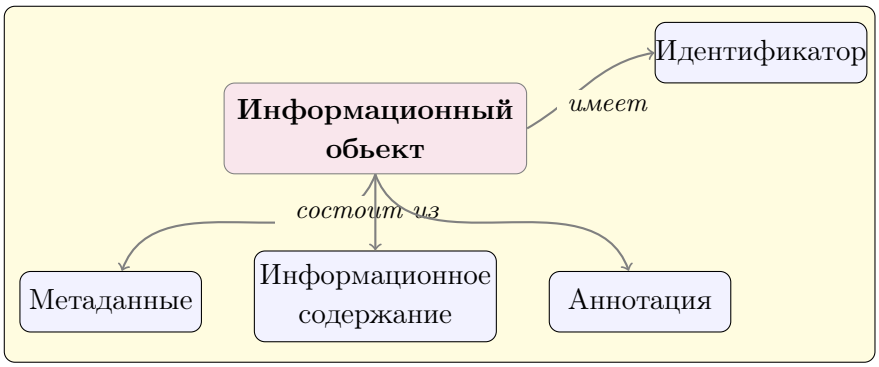

Рис. 3: Модель информационного объекта в ИС

В информационной системе каждому ресурсу соответствует информационный объект, который является традиционным вторичным информационным объектом, содержащим описание первичного ресурса, т. е. информационный объект — это объект, который хранит информацию о объектах ИС (физических объектах, ресурсах, информационных объектах) (см. рис. 3).

Каждый информационный объект в ИС состоит из следующих объектов:

- *метаданных* — объект, главная цель которого состоит в том, чтобы дать информацию о ресурсе;
- *аннотаций* — объект, главная цель которого состоит в том, чтобы аннотировать ресурс или его часть. Примеры таких аннотаций включают примечания, структурированные комментарии и связи. Объекты аннотации помогают интерпретировать ресурс, содержат детальные объяснения, либо информацию о том, как можно использовать ресурс.
- *информационного содержания* — объект, который может отсутствовать и может использоваться самостоятельно, как первичный информационный объект: например, изображение, полный текст и т. д. (первичный ресурс) [16].

Информационный объект — наиболее общее понятие в системе, представляющий произвольную единицу информации в ИС. Информационные объекты также могут быть сложными объектами и могут быть сгруппированы в коллекции информационных объектов, которые, в свою очередь, тоже являются информационными объектами. Коллекции наследуют все аспекты моделирования информационных объектов и средства их обслуживания, например, они могут аннотироваться.

Информационные объекты описывают все *классы сущностей* научного информационного пространства такие, как публикация, персона, ключевой термин или понятие, словарная статья, факт, функция, организация, пользователь и т. д., а также связи между ними [12, 15].

3 Определение «документа» и ЭБ

Одним из фундаментальных понятий в современных ИС является понятие «документ». Содержание понятия документ многозначно и зависит от того, для каких целей оно используется. Документы появились как дополнительное (к звуковой речи) средство общения людей. Они были вызваны к жизни прежде всего необходимостью подтвердить,

запечатлеть, закрепить и передать во времени и пространстве то или иное сообщение или свидетельство. Носители, материальные предметы, на которых фиксировалась информация, выполняли в древние времена главным образом функцию свидетельства [22]. С начала XX века вводится новое, более широкое понимание понятия «документ»: известный бельгийский ученый Поль Отле в своем Трактате о Документации (Traite de documentation [22]) определяет это понятие так: *Документ — «материальный объект, содержащий информацию, специально предназначенный для ее передачи в пространстве и времени»* [9].

Он писал, что не только графические изображения и текстовые записи могут считаться документами; документами можно считать и объекты, если они являются носителями некоторой информации. *Документом является любой источник информации, который может использоваться в качестве ссылки, для изучения или в качестве доказательства.* В качестве примера подобных документов Отле приводит природные объекты, артефакты, объекты, отмеченные следами человеческой деятельности (например, археологические находки), произведения искусства. Примеры документов: рукописи, печатные материалы, иллюстрации, диаграммы, музейные экспонаты.

Важно заметить, что термин «документ» более старый, нежели термин «информация».

В ИС понятие документа несколько сужается из-за того, что мы ограничиваемся цифровыми объектами: *Документ — это целостный информационный объект, представленный в цифровом виде, снабженный структурированными метаданными, имеющий некоторый стандартный набор атрибутов и функций и допускающий однозначную идентификацию.*

Соответственно определение Электронной библиотеки (ЭБ) можно записать так — *Набор структурированных каталогизированных коллекций разнородных электронных (цифровых) документов, снабженный средствами навигации и поиска.*

ЭБ способна не только обеспечить многосторонний поиск в каталоге, но и предоставить пользователю непосредственно найденный ресурс (публикацию, документ, фотографию, описание факта и др.), а также дополнительные сведения о нем, например, об авторах, библиографии, организации и т. п.

За популярностью словосочетания «электронная библиотека» стоит не только и не столько дань моде, сколько попытка охарактеризовать новый феномен — возникновение принципиально нового класса информационных систем, призванных аккумулировать и распространять информацию в цифровой форме [1, 2]. А большой интерес к самим системам данного класса объясняются актуальными потребностями общества и наличием развивающихся возможностей по их удовлетворению. В связи с этим можно сформулировать основные цели, стоящие перед электронными библиотеками:

– обеспечение доступа к информации;
– сохранение научного и культурного наследия;
– повышение эффективности научных исследований и обучения.

В большинстве случаев ЭБ представляет собой веб-сайт, где накапливаются различные тексты (литературные, научные и технические, в том числе и публикации, компьютерные программы, электронные карты и т.п.) и медиа-файлы. Отличие ЭБ от сайта журнала состоит в том, что ЭБ не делится на номера (выпуски) и обновляется по мере появления

новых материалов. Отличие ЭБ от сайта свободных публикаций заключается в том, что ЭБ, как правило, подбирается администратором проекта по определенным правилам и не всегда предусматривает создания вокруг публикуемых текстов коммуникативной среды.

В существующих разработках электронных библиотек, как правило, поиск и доступ к информации обеспечивается только посредством визуальных графических интерфейсов. Это хорошо для пользователя-человека, но очень плохо для пользователя-системы. Для обеспечения функций поиска вне графических интерфейсов требуется поддержка специальных сетевых сервисов и языков запросов. В идеальном случае все информационные системы должны поддерживать единый поисковый профиль и единый язык запросов.

Однако в общем случае 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 этим доступом [15]. Кроме того, некоторые организации, которые берут на себя ответственность не только за исполнение функций управления цифровым контентом, но и за предоставление к нему доступа всем заинтересованным лицам. Такое определение электронной библиотеки полностью соответствует определению традиционной библиотеки как организации в системе, например, министерства культуры [5].

Выделим три понятия для разграничения того, что обычно понимается под термином «электронная библиотека» (см. рис. 4):

- ЭБ — конкретная ЭБ с ее пользователями, политикой, содержимым и ответственной организацией, которая может быть виртуальной (DL).
- Система ЭБ — система программного обеспечения (например: DSpace [19]), которая основана на определенной (возможно, распределенной) архитектуре и обеспечивает все функциональные возможности, необходимые конкретной цифровой библиотеки. Пользователи взаимодействуют с цифровой библиотеки через соответствующие интерфейсы цифровой библиотечной системы (DLS).
- Система Управления Электронными Библиотеками (СУЭБ) — обобщенное системное программное обеспечение (например: СУЭБ ИВТ СО РАН [16]), которое обеспечивает соответствующую программную инфраструктуру (I) для администрирования системы цифровой библиотеки, включающей в себя набор функциональных возможностей, которые считаются основополагающими для цифровых библиотек и (II) позволяющее интегрировать дополнительное программное обеспечение предлагающее специализированные или расширенные функциональные возможности для создания и управления ЭБ (DLMS).

В ЭБ каждый ресурс определяется следующим образом (см. рис. 4):

- имеет идентификатор;
- организован в соответствии с форматом ресурса, формат здесь описан структурными метаданными, являющимися онтологией (ресурс может быть сложным и структурированным, поскольку он, в свою очередь, может состоять из меньших ресурсов и иметь связи с другими ресурсами) ;
- описан структурированными метаданными и аннотациями;
- может характеризоваться параметрами качества;

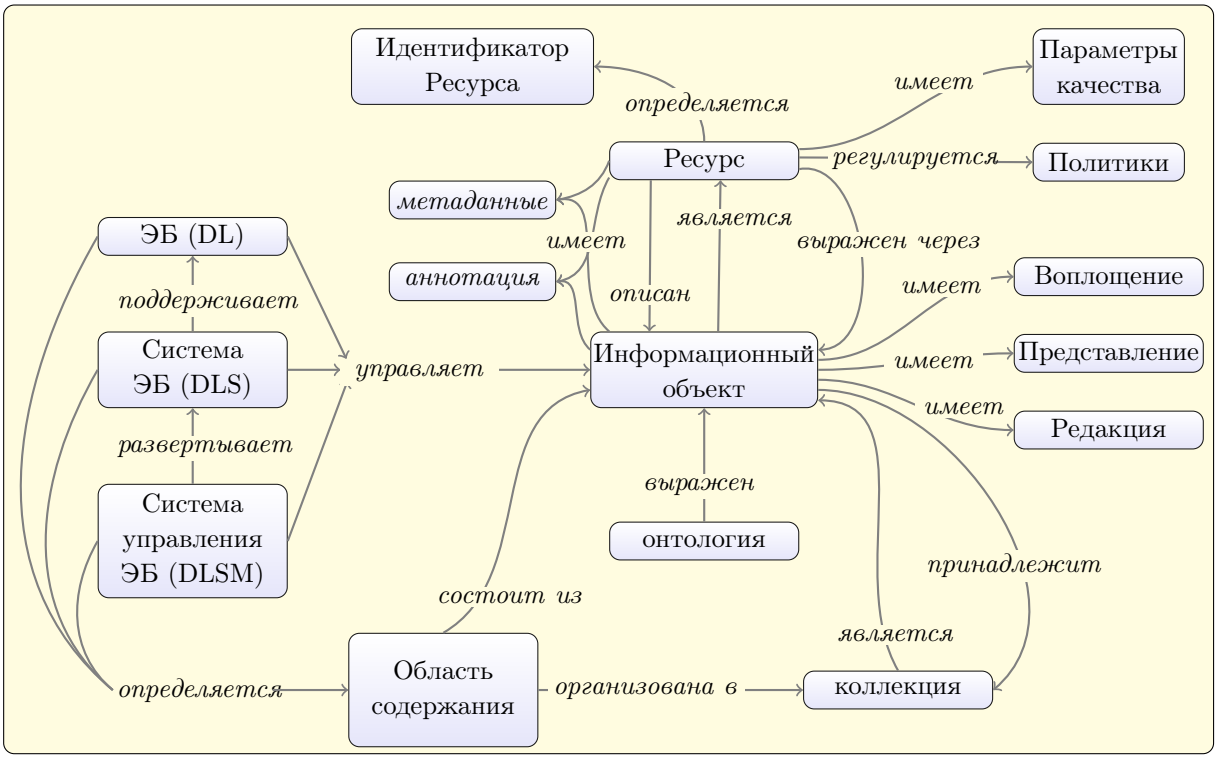

Рис. 4: Область содержания ЭБ в эталонной модели

– может регулироваться политиками, политикой управления его жизненным циклом;
– выражается через информационный объект;
– может быть описан или дополнен информационным объектом.

С организационной точки зрения, ресурсы могут группироваться в наборы ресурсов — коллекции, которые рассматриваются как единая сущность.

4 Информационная модель ЭБ

Исходя из целей ЭБ и анализа существующих ИС, направленных на поддержку научных исследований, можно сформулировать следующие функциональные требования к модели научной электронной библиотеки [14]:

– Надежное долговременное и неисчезающее хранение информации.
– Актуальность, полнота, достоверность происхождения документов.
– Историчность информации.
– Географическая привязка информации.
– Наличие большого числа словарей-классификаторов (справочников), для обеспечения идентификации и классификации ресурсов.
– Поддержка неоднородных и слабо структурированных информационных ресурсов.
– Поддержка взаимосвязей информационных ресурсов. Идентификация информационных ресурсов.
– Предоставление информации пользователю в виде, выбранном пользователем.
– Наличие интеллектуальных служб обслуживания запросов пользователя.
– Наличие программных интерфейсов для поддержки аналитической работы пользователя с помощью программных приложений.

– Поддержка требований интероперабельности как на программном, так и на семантическом уровне.
– Поддержка работы с внешними источниками (например, каталогами библиотек и журналов, цифровыми депозитариями информационных ресурсов и т. п.).

Наиболее важным выводом из вышесказанного является то, что (см. рис. 5) информационная модель ЭБ должна быть многоуровневой и состоять как минимум из следующих компонент [12]:

– хранилища данных — репозиторий,
– сервера метаданных,
– сервера приложений (диспетчера),
– словарей-справочников.

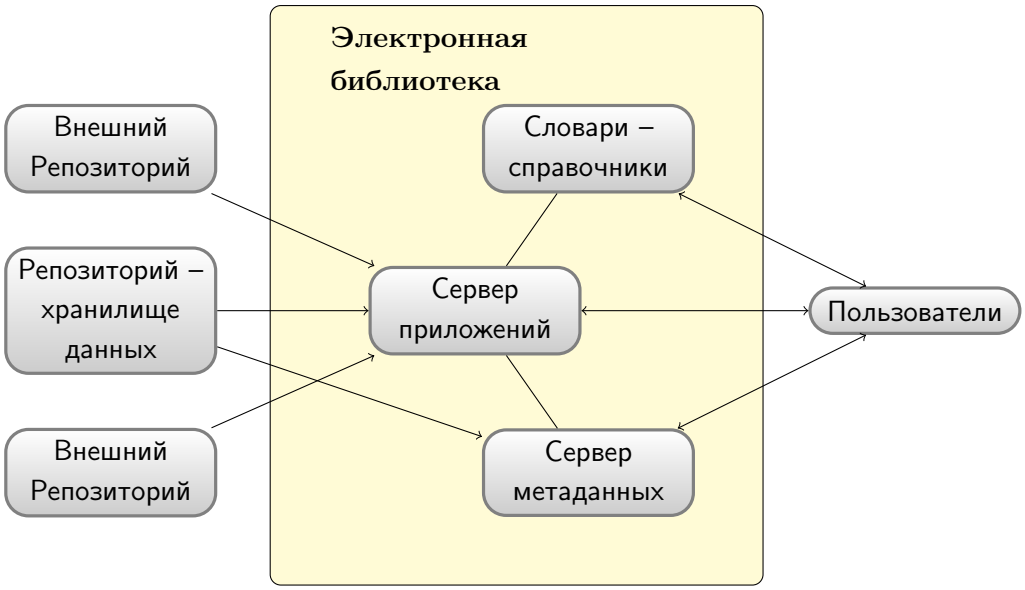

Рис. 5: Архитектура ЭБ

Репозиторий — это независимая система долговременного хранения и доступа к разнородным цифровым объектам, см. определение на стр. 397.

Сервер метаданных должен обеспечить работу с метаданными — каталогизацию всех информационных ресурсов в соответствии с общепринятыми международными стандартами.

Сервер приложений должен обеспечить работу основных сервисов ЭБ. Обеспечивает сервисы необходимые для формирования информационных ресурсов с использованием и без использования диалоговых пользовательских интерфейсов. Сервисы позволяют использовать метаданные других информационных систем в диалоговом режиме и пакетных режимах. Их функциональность должна обеспечивать поиск и извлечение метаданных из других систем, конвертирование полученных метаданных в схемы и структуры локальной системы.

Справочники — управляемые словари (ключевые признаки, ключевые термины) — это особый вид метаданных, отражающих наиболее существенные свойства *информационного объекта* и имеющие наиболее важное значение с точки зрения ЭБ. Специфика словарей определятся терминологией конкретной предметной области, которой посвящена ЭБ. Необходимо рассматривать различные типы ключевых терминов (ключевые термины

в стандартном понимании; ключевые термины, описывающие персону; ключевые термины, описывающие организацию; ключевые термины, описывающие временные периоды; ключевые термины, описывающие географические понятия). Это набор баз данных (нормативных словарей), содержащих информацию об авторах и других персонах (авторитетные записи), географических пунктах, городах, издательствах, имеющих отношение к конкретной теме или разделу ЭБ (например, к научной школе), тематические словари-классификаторы, тезаурусы, рубрикаторы, описания предметной области и классификаторы документов.

Основу содержания ЭБ составляют *информационные объекты*, представляющие основные типы сущностей:

– субъекты: актор, персона, организация, действующие лица, приложение и т. п.;
– объекты: публикация, журнал, документ, факт, научный результат, мероприятие, проект, фотография и др.;
– отношения: понятие, ключевой термин, событие, время, место и т. п.

Ниже изображена иерархия сущностей или объектов, представленных в электронной научной библиотеке (см. рис. 6). Информационный объект является корневым объектом в представляемой модели, он охватывает все объекты, информация о которых хранится в электронной библиотеке.

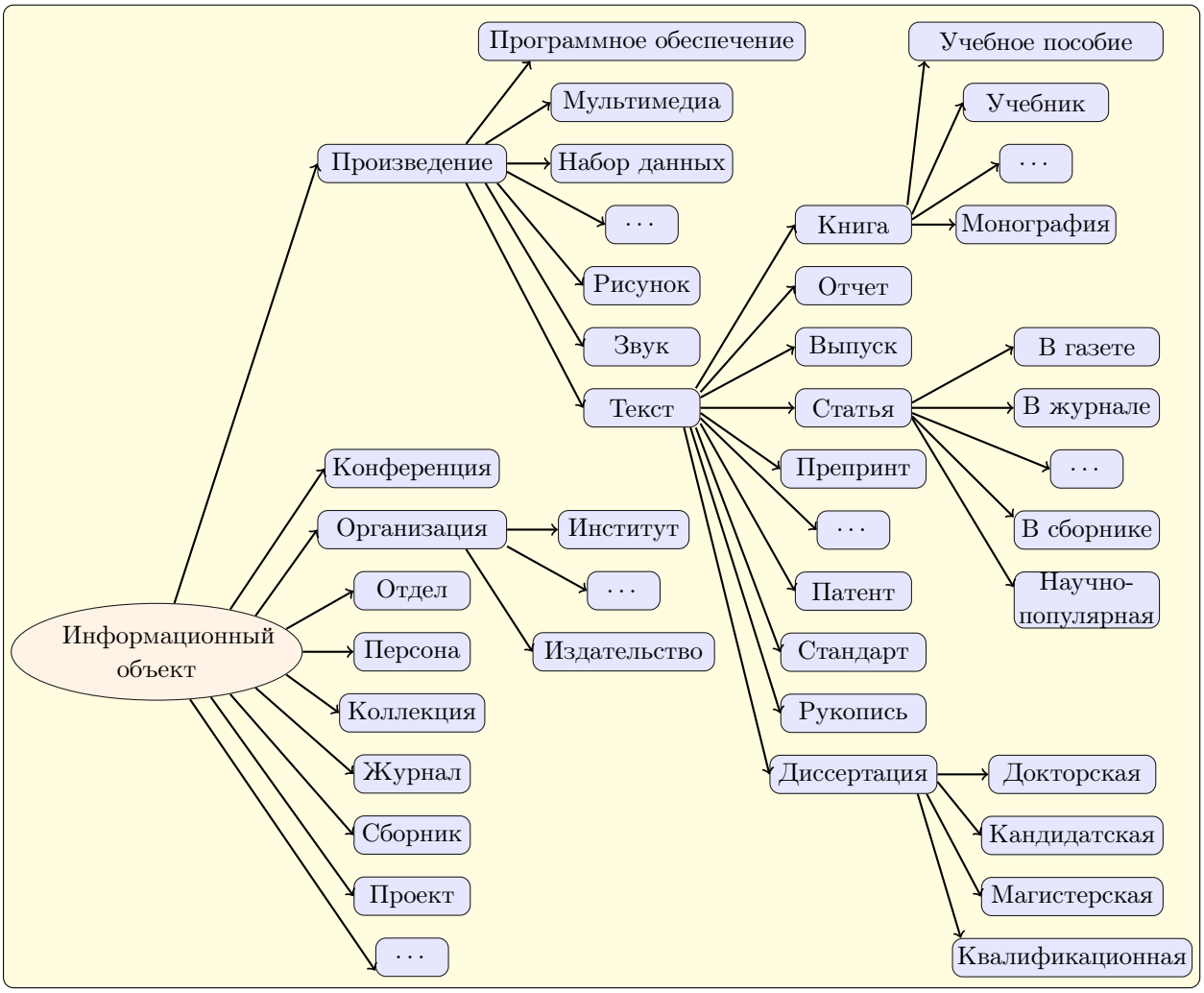

Рис. 6: Иерархия классов в информационной модели научной ЭБ

Используемый профиль определяет список элементов данных (полей), необходимых для создания записи соответствующего типа, и раскрывает содержание элементов данных. Для эффективной работы сервера приложений используется набор словарей-классификаторов, содержащих как классификационные признаки, так и наборы ключевых терминов (с отношениями порядка), по которым производится систематизация и классификация материала.

Как уже было упомянуто, каждый *информационный объект* имеет следующее атрибуты:

- идентификатор объекта;
- название;
- метаданные, включая ключевые слова;
- версия;
- аннотация.

Эти атрибуты наследуются всеми другими сущностями и объектами представленной иерархии. На рис. 6 представлена часть иерархии классов в информационной модели научной ЭБ.

Как правило, в системах ЭБ предусматривается хранение документов, являющиеся объектами-произведений — основного типа объектов информационного содержания (контента), а также некоторых других объектов, имеющих к ним отношение:

- описания организаций, отделов организаций и издательств, где создавались или публиковались объекты-произведения;
- описание людей (на схеме это сущность персона), работающие в этих организациях (отделах) — авторы объектов-произведений;
- описание проектов, в рамках которых создаются объекты-произведения;
- научные журналы (периодические издания) и конференции, их публикующие.

Особым типом объекта является объект *коллекция* может быть применим к любой совокупности (группировке, агрегации) информационных объектов. Информационные объекты здесь могут быть любого типа, т.е. коллекциями могут быть как совокупности субъектов, так и объектов, например, совокупности организаций, журналов и т.д. Критерии для таких совокупностей могут определяться, например, общностью местоположения, общностью авторов, хронологией, тематикой, происхождением или принадлежностью и т. д. Коллекции могут содержать любое число объектов и критерии отбора этих объектов со временем могут изменяться.

5 Метаданные

Ключевым моментом в работе с документами (информационными объектами) является использование метаданных. Метаданные — структурированная информация, которая описывает, поясняет и указывает местоположение информационного объекта [6]. Метаданные необходимы для решения следующих задач:

- предоставление сведений о документах, для получения представления о их свойствах, содержании, структуре, способах использования и т. д.;
- систематизация информации о документах и ведения каталога системы;
- выбор из множества документов определенного подмножества по формальным при-

знакам и сопоставление документов по формальным признакам;
– внутрисистемные технологические задачи, связанные с обеспечением подготовки документов, размещением документов в информационной среде и т. д.;
– внешние технологические задачи, связанные, прежде всего, с обменом данными с внешними информационными системами.

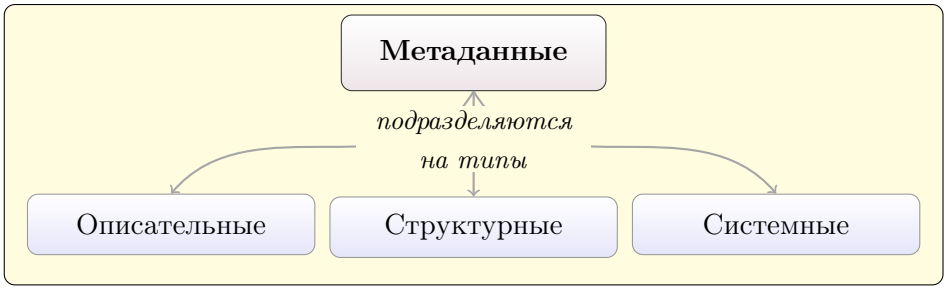

Рис. 7: Основные типы метаданных

Метаданные отражают наиболее существенные свойства объекта, имеющие наибольшие значение с точки зрения информационной системы. Метаданные подразделяются на три типа: описательные, структурные и системные (см. рис. 7).

Описательные метаданные — метаданные, которые описывают содержание и свойства информационного ресурса, например, библиографические данные или аннотация, основная задача которых однозначное представление цифрового объекта для внешнего мира и в различных приложениях.

Структурные метаданные — метаданные, которые характеризуют общую структуру информационного ресурса и его компоненты, объем и другие свойства информационного ресурса;

Системные или административные метаданные служат для обеспечения системы управления информационными ресурсами и администрирования информационных ресурсов, например, даты создания или модификации ресурса, идентификатор владельца и т. п.

Состав имен атрибутов, ограничения, накладываемые на их значения, набор правил, определяющих структурирование атрибутов, их семантику задаются *схемой метаданных*. Правила структурирования метаданных в определенном смысле аналогичны правилам, предлагаемым онтологией для построения отношений между понятиями. Правила, определяющие представление метаданных в информационной системе, а также правила их интерпретации, образуют *формат метаданных*.

Схема метаданных — это набор элементов метаданных, каждый из которых обладает некоторым именем и семантикой, принимает значения с установленной семантикой или значения из *управляемого словаря*, называемым *схемой кодировки*. В соответствии с рекомендациями Dublin Core (DCMI) [18], информационный объект должен обладать базовым набором атрибутов. Набор атрибутов объекта расширяется в зависимости от его класса.

Схема кодировки — система записи или правила анализа значений элементов метаданных. Значение, определенное с помощью схемы кодировки, представляет собой код (символ), выбранный из управляемого (контролируемого) словаря (например, индекс си-

стемы классификации или значение из набора предметных рубрик), либо строку определенной структуры (например, `"2000-01-01"` как стандартное обозначение даты).

Управляемый словарь (Controlled Vocabulary – контролируемый словарь) — это список предварительно определенных кодов, терминов, слов, фраз или нотаций, предназначенный для обозначения предметных рубрик или состава допустимых значений атрибутов элементов метаданных. Все коды (термины) в словаре должны иметь однозначное определение.

Особым видом метаданных являются метаданные однозначно характеризующие (идентифицирующие) документы, которые необходимы для систематизации документов и для эффективного поиска, получившие название *авторитетных.*

Авторитетный контроль (Authority controls, Нормативный контроль) — предоставление доступа к документам посредством специального класса элементов метаданных (имен собственных, предметных рубрик, классификационных индексов, видов деятельности, географических названий, имен/наименований создателей документов и т. п.). Значения элементов этих метаданных (являющиеся ключевыми словами) выбираются согласно схемам кодировки из управляемого словаря.

Существует проблемы в использовании авторитетных данных: во-первых, авторитетный контроль должен обеспечивать использования повторяющихся географических названий и имен/наименований создателей документов, которые пишутся одинаково, но обозначают разные места или разных людей, а во-вторых, имена могут меняются во времени и пространстве: например, германизированные или англизированные имена и названия не являются подлинными (аутентичными), поэтому для изучения истории необходимо также знать подлинные имена.

Особым классом метаданных является метаданные, описывающие отношения и связи между информационными ресурсами — документами.

Отношение — связь между экземпляром некоторой сущности и тем, что с ней соотнесено. По Аристотелю «есть то, что оно есть», лишь «в связи с другим или находясь в каком-то ином отношении к другому». Количество типов отношений в информационной системе определяется, исходя из конкретных целей. В реальном мире их число стремится к бесконечности. С позиций удовлетворения информационных потребностей пользователей нас будут интересовать отношения только между документами, например, «Публикация – Публикация», «Публикация – Персона», «Публикация – Словарная статья», «Публикация – Ключевой термин», «Персона – Словарная статья» и так далее. Связи существуют между всеми классами документов.

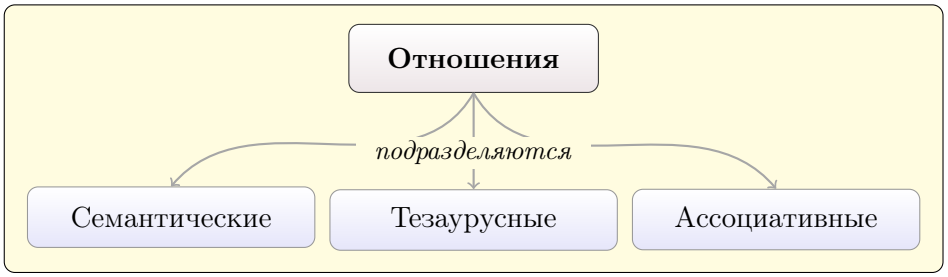

Рис. 8: Типы отношений

В зависимости от условий использования отношения между документами подразде-

ляются на следующие типы: тезаурусные отношения, семантические отношения и ассоциативные отношения (см. рис. 8):

- *тезаурусные отношения:* отношения, применяемые в описании информационно-поисковых тезаурусов — это иерархические отношения и отношение ассоциации. Основным иерархическим отношением — является родовидовое отношение (родитель-потомок, шире-уже, выше-ниже, часть-целое). Основное назначение установления ассоциативных отношений между документами — указание на дополнительные связи [7]. Тезаурусные отношения специфичны для отношений между ключевыми терминами, значительно реже используются при задании отношений между публикациями и словарными статьями.
- *семантические отношения:* именованные отношения между документами, например, «Персона является автором Публикации»; «Публикация посвящена Персоне»; «Публикация посвящена Факту, описанному в Словарной статье».
- *отношения ассоциации:* отношения между двумя документами которые близки по содержанию, например, ключевые слова в описании Публикации, Персоны, Словарной статьи.

В информационной системе возможно два способа реализации связей (отношений) между документами: жесткие и мягкие. Жесткие связи реализуются средствами СУБД путем ссылок на первичные ключи записи. К сожалению, такой тип связи не защищен от нарушения целостности (в случае неправильного изменения или удаления записи). Мягкие связи реализуются через процедуру поиска соответствий. Такой способ установления связей защищен от любых нарушений целостности БД и достаточно удобен пользователям, поскольку для указания на необходимость связи используются наглядные мнемонические определения.

Работа выполнена при частичной финансовой поддержке гранта Президента РФ для государственной поддержки ведущей научной школы РФ № НШ-7214.2016.9.

ЛИТЕРАТУРА

[1] *Акимов С. И., Елизаров А. М., Ершова Т. В., Когаловский М. Р., Федоров А. О., Хохлов Ю. Е.* Научно-методическая поддержка разработки научных электронных библиотек [Электронный ресурс] // Электронные библиотеки: рос. науч. электронный журн. 2005. Т. 8, вып. 1. `http://www.elbib.ru/index.phtml?page=elbib/rus/journal/ 2005/part1/AEEKFH` (дата обращения: 04.05.2010).

[2] *Вигурский К. В.* Что такое электронная библиотека? // Доклад на конференции «Информационные технологии в образовании — 2005». [Электронный ресурс] `http://rd.feb-web.ru/library.htm` (дата обращения: 28.01.2013)

[3] ГОСТ Р ИСО/МЭК 12207-99. Информационная технология. Процессы жизненного цикла программных средств. Госстандарт России. Москва, 1999.

[4] ГОСТ 34.003-90. Информационная технология. Комплекс стандартов и руководящих документов на автоматизированные системы. Термины и определения.

[5] *Жижимов О. Л., Мазов Н. А., Федотов А. М.* Некоторые заметки об эволюции цифровых репозитариев традиционных библиотек к полнофункциональным электронным библиотекам // Вестник Владивостокского государственного университета экономики и сервиса. Территория новых возможностей. 2010. Т.7, № 3. С. 55–63.

[6] *Когаловский М. Р.* Метаданные, их свойства, функции, классификация и средства представления // Труды 14-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» — RCDL2012, Переславль-Залесский, Россия, 15-18 октября 2012 г.

[7] *Лукашевич Н. В.* Тезаурусы в задачах информационного поиска. Москва: Издательство МГУ. 2011. 512 с.

[8] *Ляпунов А. А.* О соотношении понятий материя, энергия и информация // В кн.: Ляпунов А. А. Проблемы теоретической и прикладной кибернетики. Новосибирск: Наука, 1980. С. 320–323.

[9] *Отле П.* Библиотека, библиография, документация: Избранные труды пионера информатики. Пер. с англ. и фр. Москва: ФАИР-ПРЕСС: Пашков Дом, 2004. 348 с.

[10] *Резниченко В. А, Проскудина Г. Ю., Кудим К. А.* Концептуальная модель электронной библиотеки [Электронный текст] // Труды XI Всероссийской научной конференции RCDL'2009, Россия, г. Петрозаводск, 17–21 сентября 2009 г. С. 23–31.

[11] *Советов Б.Я.* Информационные технологии : учебник для вузов М.: Высшая школа, 2005.

[12] *Федотов А. М.* Методологии построения распределенных систем // Вычислительные технологии. 2006. Т. 11, Избранные доклады X Российской конференции «Распределенные информационно-вычислительные ресурсы» (DICR-2005), Новосибирск 6-8 октября 2005 г. С. 3–16.

[13] *Федотов А. М., Шокин Ю. И.* Электронная библиотека Сибирского отделения РАН // Информационное общество. 2000. № 2. С.22–31.

[14] *Федотов А. М., Барахнин В. Б., Жижимов О. Л., Федотова О. А.* Модель информационной системы для поддержки научно-педагогической деятельности // Вестник НГУ. Серия: Информационные технологии. 2014. Т.12, вып. 1. С.89–101.

[15] *Шокин Ю. И., Федотов А. М. Жижимов О. Л., Гуськов А. Е., Столяров С. В.* Электронные библиотеки — путь интеграции информационных ресурсов Сибирского отделения РАН // Вестник КазНУ, спец. вып. г. Алматы, Казахстан, Казахский национальный университет им. аль-Фараби. 2005. № 2. С. 115–127.

[16] *Шокин Ю. И., Федотов А. М., Жижимов О. Л., Федотова О. А.* Эволюция информационных систем: от Web-сайтов до систем управления информационными ресурсами // Вестник НГУ. Серия: Информационные технологии. 2015. Т. 13, вып. 1. С. 117–134.

[17] D3.2b The Digital Library Reference Model // Funded under the Seventh Framework Programme, ICT Programme — "Cultural Heritage and Technology Enhanced Learning" Project Number: 231551. April 2011. (http://www.dlorg.eu) — Full Text: http://db4.sbras.ru/elbib/data/show_page.phtml?22+269.

[18] DCMI — Dublin Core Metadata Initiative // (http://www.dublincore.org/).

[19] DSpace [Электронный ресурс]: an open source solution for accessing, managing and preserving scholarly works // dspace.org [web-сайт] / MIT Libraries; HP Labs. 2007. (http://www.dspace.org/).

[20] ISO 14721:2012 Reference model for an Open archival information system (RM OAIS) / Recommended Practicle: CCSDS 650.0-M-2 (Magenta Book). June 2012. Available at: http://public.ccsds.org/publications/archive/650x0m2.pdf

[21] *Mooers C.N.* «Mooers» law, or why some retrieval systems are used and other are not // American Documentation. 1960. Vol.11, No.3.

[22] *Otlet P.* Traite de documentation. Bruxelles: Ed. Mundaneum, 1934.