

СОДЕРЖАНИЕ

| | |
|---|-------|
| ВВЕДЕНИЕ..... | 3 |
| ГЛАВА 1. ПОДОДЫ К ПОСТРОЕНИЮ ЭЛЕКТРОННЫХ БИБЛИОТЕК..... | 5 |
| ГЛАВА 2. ОБЗОР ИЗВЕСТНЫХ ПРОЕКТОВ..... | 9 |
| 2.1. Существующие эталонные модели. CIDOC CRM..... | 9 |
| 2.2. FRBR FRBR00..... | 10 |
| 2.3. Delos DLRM..... | 12 |
| ГЛАВА 3. АРХИТЕКТУРА ЭЛЕКТРОННОЙ БИБЛИОТЕКИ (ЭБ)..... | 17 |
| 3.1. Элементы Dublin Core..... | 17 |
| 3.2. Открытые архивы DSpace и EPrint..... | 18 |
| 3.3. Многоуровневая архитектура ЭБ..... | 19 |
| ГЛАВА 4. СИСТЕМАТИЗАЦИЯ ИНФОРМАЦИОННЫХ РЕСУРСОВ..... | 21 |
| 4.1. Систематизация электронных публикаций..... | 27 |
| ГЛАВА 5. ИНФОРМАЦИОННО-ПОИСКОВЫЙ ТЕЗАУРУС..... | 28 |
| 5.1. Построение информационно-поискового тезауруса..... | 29 |
| 5.2. Структура словарных статей..... | 29 |
| 5.3. Связи и типы данных..... | 31 |
| ЗАКЛЮЧЕНИЕ | 35 |
| ЛИТЕРАТУРА..... | 36 |
| ПРИЛОЖЕНИЕ..... | 38-43 |

ВВЕДЕНИЕ

В настоящее время научно-исследовательский процесс неотделим от использования сети Интернет. Значительную часть своего времени научные сотрудники проводят за компьютерами в поиске и анализе информации, в электронной переписке с коллегами во всем мире. В частности, все большую роль в этом процессе начинает играть использование электронных библиотек.

В настоящий момент значительная часть информационных ресурсов хотя и переведена в цифровую форму, но недоступна широкому кругу научной общественности, а ресурсы, представленные в Интернет, разрознены, недостаточно систематизированы и структурированы. Поэтому, в процессе научно-педагогической деятельности очень часто необходима систематизация и классификация имеющихся информационных ресурсов.

Целью работы является создание системы классификации информационных ресурсов посвященных созданию и развитию электронной библиотеки на основе информационно-поискового тезауруса.

Из поставленной цели вытекают следующие **подцели**, требующие решения:

- Исследовать различные подходы к построению информационной поисковой системы;
- Исследовать правила и системы для построения информационно-поискового тезауруса;
- Реализовать информационно-поисковый тезаурус на примере информационно-поисковой системы посвященной электронной библиотеке.

Достижение поставленной цели осуществляется путем решения следующих задач:

- 1) Изучить основные современные концептуальные модели электронной библиотеки:
 - Сделать аналитический обзор международных стандартов и рекомендаций по организации баз данных публикаций, информационных ресурсов и электронных библиотек;
- 2) Исследовать технологические средства для построения электронной библиотеки в области информационных систем;
- 3) Решить следующие вопросы для построения информационно-поискового тезауруса:
 - выяснить классификационные признаки документов(публикаций) и систематизировать их;
 - определить и установить связи между документами;
 - расклассифицировать документы;
 - построить поисковый образ документа;
 - составить информационно-поисковый тезаурус и создать базу данных;

- перевод языка авторов на контролируемый язык, используемый для индексации и поиска;
- обеспечение последовательности в присваивании индексных терминов;
- обозначение отношения между терминами.

Общая проблема исследования. Систематизация ресурсов электронной библиотеки на основе информационно-поискового тезауруса.

Научная новизна состоит в использовании базы данных ключевых терминов с автоматически генерируемыми связями для классификации ресурсов электронной библиотеки.

Разработан прототип информационно-поискового тезауруса для выбранной предметной области – классификации информационных ресурсов в области информатики и электронных библиотек в виде базы данных в информационной системе.

Результатами проекта является дальнейшее совершенствования методов и средств организации данных в электронных библиотеках и информационных фондах.

Результаты данного проекта важны для наукоемких исследовательских работ, способствуют совершенствованию организации и представления данных в адаптируемых сервисах электронных библиотек и других информационных фондов на основе информационно-поискового тезауруса.

ГЛАВА 1. ПОДХОДЫ К ПОСТРОЕНИЮ ЭЛЕКТРОННЫХ БИБЛИОТЕК

1. Определение электронной библиотеки

Проблема поиска информации – одна из вечных проблем человечества. Чтобы решить проблему доступа к информации, человечество создало библиотеки – универсальную систему хранения, систематизации и каталогизации «информации и знаний» [2].

Электронная библиотека (ЭБ) – это структурированная каталогизированная коллекция разнородных электронных документов, снабженная средствами навигации и поиска (в отличие от печатных изданий, микрофильмов и других носителей). 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, информацию об авторах, библиографию, организации и т. п.

За высокой популярностью слов «электронная библиотека» стоит не только и не столько дань моде, сколько попытка охарактеризовать новый феномен – возникновение принципиально нового класса систем, призванных аккумулировать и распространять информацию в электронной форме. А большой интерес к самим системам данного класса объясняется потребностями общества и наличием развивающихся возможностей по их удовлетворению.

В существующих разработках ЭБ, как правило, поиск и доступ к информации обеспечиваются только посредством визуальных графических интерфейсов. Это хорошо для пользователя – человека, но не годится для пользователя – системы. Для обеспечения функций поиска вне графических интерфейсов требуется поддержка специальных сетевых сервисов и языков запросов. В идеальном случае все ИС должны поддерживать единый поисковый профиль и единый язык запросов.

Однако в общем случае под словосочетанием «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. Такое определение ЭБ полностью соответствует определению традиционной библиотеки как организации в системе, например, Министерства культуры.

В настоящее время нет какой-либо универсальной системы поддержки ЭБ, которая отвечала бы всем требованиям и ожиданиям пользователей. Анализ существующих систем ЭБ 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 и игнорирования нужд 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.

В настоящее время существуют достаточно мощные информационные системы, удовлетворяющие в той или иной степени потребности научных работников в информации, однако основной недостаток большинства систем – ограниченность возможностей обеспечения интеграции ресурсов как внутри каждой из систем, так и с внешними системами. Основу разработки ЭБ составляют стандарты и международные 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. Наиболее важным является профиль метаданных информации, циркулирующей в системе. Выбор профиля должен основываться на выполнении следующих требований:

- включать в себя основные типы информации, требующейся для поддержки научной работы;
- быть открытым, т. е. обеспечивать доступ к соответствующей информации по этим описаниям;
- быть расширяемым, 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 ориентированным на современные и перспективные технологии описания и использования информации;
- обеспечивать возможность интероперабельности с внешней средой.

При работе с цифровыми объектами человечество уже выработало определенный набор стереотипов, отсутствие которых вызывает дискомфорт. Одним из элементов этого набора является требование 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. Во-вторых, внедренные в объект связи должны быть гарантировано актуальными. Так появляется требование обеспечения ссылочной целостности данных. Это очень жесткое требование, которое трудно 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 ЭБ объекты хранения могут содержать информацию, которая не имеет к объектам хранения традиционных библиотек вообще никакого отношения. Речь идет об электронных копиях элементов хранения традиционных архивов, о видео или аудиоинформации, полученной разными способами, о научных или других фактах и т. п.

Существует достаточно много технологических разработок информационных систем для электронных библиотек, так или иначе ориентированных на поддержку научных исследований. Среди них следует отметить информационные системы, близкие к фактографическим, например ИСИР (ЕНИП) РАН, ИРИС СО РАН, euroCRIS, и документальные, например eLibrary, Информика, MathNET. Названные системы в той или иной степени удовлетворяют потребности исследователей в информации, однако каждая из них имеет определенные недостатки (см., например, [2]).

1.2. Функциональные требования к модели электронной библиотеки

Исходя из целей ЭБ по научному наследию и анализа существующих систем, направленных на поддержку научных исследований, можно сформулировать следующие функциональные требования к модели ЭБ:

- актуальность, полнота, достоверность происхождения документов;
- наличие большого числа словарей-классификаторов (справочников), для обеспечения идентификации и классификации ресурсов;
- поддержка неоднородных и слабо структурированных информационных ресурсов;
- поддержка взаимосвязей 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.

Наиболее важным выводом из вышесказанного является то, что информационная модель ЭБ должна быть многоуровневой и состоять как минимум из следующих компонент: хранилище данных – репозиторий, сервер метаданных, сервер приложений, словари-справочники.

ГЛАВА 2. ОБЗОР ИЗВЕСТНЫХ ПРОЕКТОВ

Нацелившись на описание информационного пространства библиотечной системы, мы естественно не могли не рассмотреть что-то подобное, что уже сделано или делается сегодня в мире.

2.1 Существующие эталонные модели

CIDOC CRM

Концептуальная эталонная модель (Conceptual Reference Model, CRM) CIDOC, разработанная Международным комитетом по документации Международного совета музеев (The International Committee for Documentation of the International Council of Museums, ICOMCIDOC), предназначена для интеграции, посредничества и обмена информацией в области мирового культурного наследия и связанных областей. Четко обозначенные цели определили основные конструкции модели и уровень ее детализации. CIDOC CRM в терминах формальной онтологии определяет семантику схем баз данных и структур документов, используемых в культурном наследии и музейной документации. Модель не определяет терминологию, появляющуюся в конкретных структурах данных, но имеет характерные отношения для ее использования. Она не стремится предлагать то, что должны документировать учреждения культуры. Скорее она объясняет логику того, что они фактически документируют и таким образом предоставляет семантическую интероперабельность между музеями, библиотеками, архивами. Представляя определения и формальную структуру описания неявных (implicit) и явных (explicit) сущностей и отношений, модель CIDOC CRM претендует на общий язык для экспертов предметной области и специалистов по информационным технологиям. Она предназначена для покрытия контекстной информации исторического, географического и теоретического характера об отдельных экспонатах и музейных коллекциях в целом [8, 9].

Версия 5.0.1 модели CIDOC CRM [4] состоит из 90 классов и 148 свойств (бинарных отношений), связывающих классы между собой и описывающих предметы, понятия, людей, события, место, время и их отношения. Все концепты (классы и свойства) модели можно разделить на три группы. Первая группа включает классы и отношения, охватывающие наиболее общие понятия окружающего мира: постоянные и временные сущности, отношения участия, зависимости, совпадения во времени. Вторая группа содержит понятия, частично поддерживающие функции управления: приобретение и учет единиц хранения, передача прав собственности на объекты культуры. К третьей группе относятся классы и свойства, используемые для внутренней организации самой онтологии: средства, необходимые для подключения внешних источников терминов,

например, тезаурусов по отраслям культуры. На рис. 1 представлена часть иерархии классов CIDOC CRM. Все классы, за исключением класса простое значение (Primitive Value) и его подклассов (эти классы являются вспомогательными), прямо или опосредовано являются подклассами класса сущность CRM, охватывающего все сущности, которые могут быть описаны в CIDOC CRM. Далее иерархия классов делится на 2 ветви: постоянные сущности и временные сущности.

На самых нижних уровнях иерархии классов появляются понятия, характерные для сферы культуры: хранение, перемещение (ценностей), проект или процедура (в том числе техника производства), период (в том числе художественный стиль). Иерархия классов может быть гибко расширена с применением встроенного класса тип. Наибольший интерес представляют свойства. Классы на нижних уровнях иерархии имеют около 10-15 свойств, причем большая часть свойств наследуется от классов-предков. Названия свойств представляют собой глагольные фразы, выбранные так, что при последовательном связывании двух классов свойством получается осмысленная фраза с субъектом (первый, если считать слева направо, класс), предикатом (свойством) и объектом (второй класс).

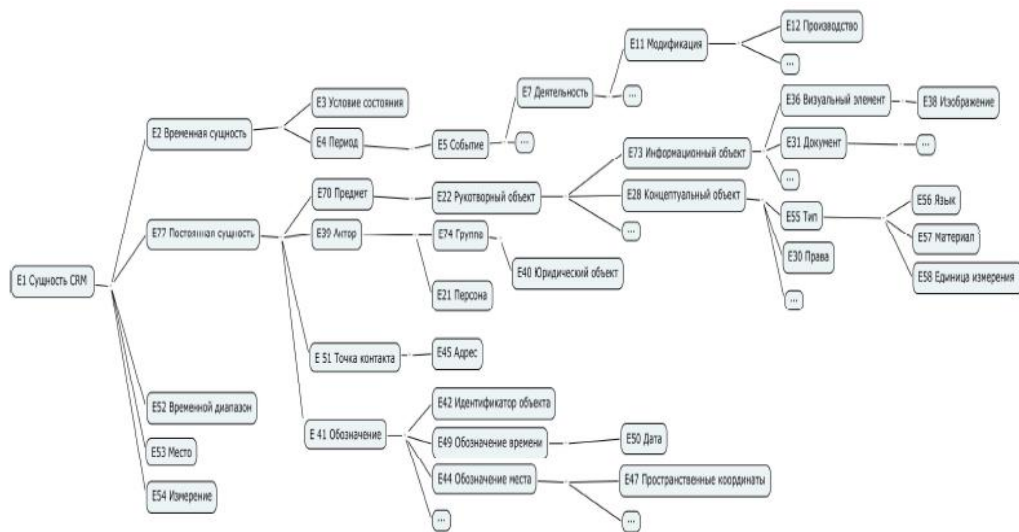


Рис.1. Часть иерархии классов в модели CIDOC CRM

2.2 FRBR и FRBRoo

Независимо от CIDOC CRM в 1991–1997 годах Международной федерацией библиотечных ассоциаций и учреждений (International Federation of Library Associations and Institutions, IFLA) была разработана ER-модель "Функциональные требования к библиографическим записям" (Functional Requirements for Bibliographic Records, FRBR) как обобщенное представление библиографического универсума, независимого от какого-либо кода каталогизации или реализации.

FRBR включает описание концептуальной модели (сущности, их отношения и атрибуты), предлагает универсальные библиографические записи для всех типов материалов и пользовательских задач, связанных с библиографическими ресурсами, описанными в каталогах, библиографиях и других библиографических инструментах [10, 11].

Модель FRBR различает три группы сущностей (рис. 2):

- для описываемых объектов: произведение (work), выражение (expression), воплощение (manifestation), экземпляр (item);
- для описателей-субъектов: человек (person) и организация (corporate body);
- для описателей-объектов: концепт, объект, событие и место (concept, object, event, place).

Большое внимание в модели уделено отношениям между сущностями. Отношения могут быть отражены в библиографических записях многими способами. Те, что изображены на ER-диаграмме FRBR (рис. 2), описывают логические связи между сущностями и часто реализуются простой конкатенацией одной сущности с атрибутами связанной сущности в одной записи. Помимо логических связей в модели выделена группа так называемых контентных связей (для первой группы сущностей). Они идентифицируют основные типы отношений, которые существуют между экземплярами сущности одного типа (например, сущности произведения) или между экземплярами разных типов сущностей (например, сущностей произведение и воплощение). Например, в группе отношений произведение-произведение выделены такие типы отношений: имеет адаптацию (свободный перевод); имеет приложение (сходство, соответствие), имеет продолжение; имеет резюме (обзор, аннотацию); имеет преобразование (стихотворную форму); имеет имитацию (пародию). В группе отношений выражение-выражение перечислены следующие типы отношений: имеет сокращение (корректировку, уплотнение); имеет пересмотр (исправленную редакцию, расширенную редакцию); имеет перевод (буквальный перевод) и некоторые другие типы отношений, касающиеся музыкальных произведений. И наконец, отношения часть/целое и часть в части также представлены в модели FRBR.

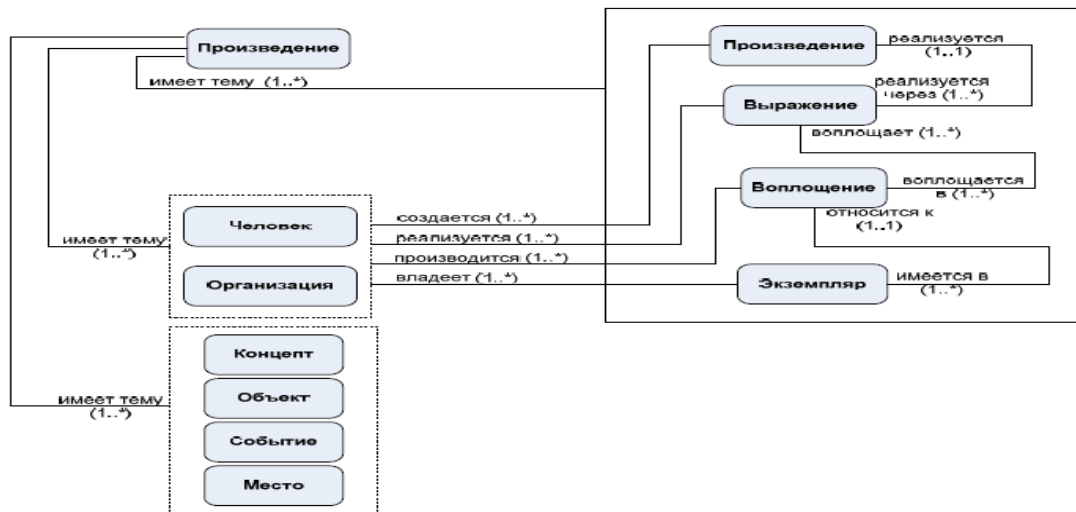


Рис. 2. Модель FRBR

Следует отметить, что в FRBR границы между различными типами основных сущностей (произведение и выражение) размыты и окончательное решение по тому, к какому типу отнести тот или иной объект, отдается на откуп каталогизатору. Кроме того, сущностей этих совсем немного и явно недостаточно для большинства конкретных библиотечных приложений. Для нас основной интерес представляет очень богатый набор атрибутов и отношений в этой модели. Как и в случае CIDOC CRM, в соответствии с решаемыми задачами, FRBR применима только для описания информационной составляющей концептуальной модели.

2.3 DELOS DLRM

Группа специалистов ассоциации в сфере ЭБ DELOS в 2006–2007 гг., основываясь на анализе имеющихся библиотечных систем [3], где большое внимание было уделено функциональным возможностям современных ЭБ, начали разработку эталонной модели ЭБ (Digital Library Reference Model, DLRM) [6]. Цель проекта – разобраться с фундаментальными понятиями, существенными объектами и их отношениями, стандартными функциональными и структурными блоками и процессами, из которых состоит универсум ЭБ.

Эталонная модель предназначена для разработки более узких моделей с конкретной архитектурой для последующей реализации программных систем. Прежде всего, в модели было выделено три понятия для разграничения того, что обычно называется ЭБ:

- ЭБ – конкретная ЭБ с ее пользователями, правилами, содержимым, интернет-сайтом и ведущей организацией. Например, библиотека Института программных систем <http://eprints.isofts.kiev.ua>;

- система ЭБ – программное обеспечение, на основе которого создаются ЭБ. Например, EPrints 3.0.
- система управления ЭБ – программное обеспечение для создания и управления системами ЭБ. Например, система OpenDLib.

Далее модель DELOS DLRM рассматривается в ролевом аспекте, т. е. с точки зрения разных категорий пользователей:

- конечный пользователь ЭБ;
- разработчик ЭБ;
- системный администратор ЭБ;
- разработчик приложений для ЭБ.

Соответственно DELOS DLRM имеет четыре уровня пользовательских представлений.

Весь универсум ЭБ разбит на шесть высокоуровневых ключевых областей (рис. 3):

- контент;
- пользователь;
- функциональные возможности;
- качество;
- политики;
- архитектура и несколько дополнительных.

Эти шесть областей объединены в одну область ресурса. В каждой из них вводятся и определяются свои сущности и их свойства. Теперь вкратце рассмотрим наиболее важные области ЭБ и их структуру.

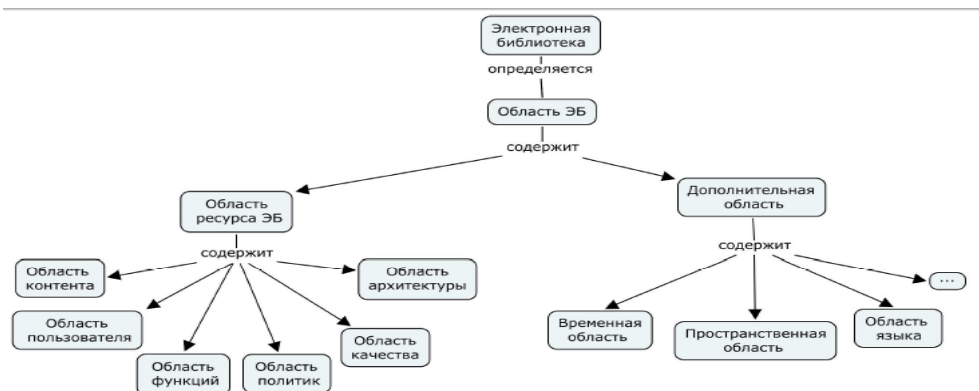


Рис. 3. Иерархия областей ЭБ в модели DELOS DLRM

С организационной точки зрения, ресурсы могут группироваться в наборы ресурсов, которые рассматриваются как единая сущность. Например, коллекции в области контента или группы в области пользователя.

Область контента (рис. 4) представляет все объекты, связанные с информацией, которой управляет ЭБ. Информационный объект – наиболее общее понятие в этой области, представляет произвольную единицу информации, управляемую в универсуме ЭБ. В DELOS DLRM различают информационный объект по уровню абстракции, где заимствуются типы объектов из модели FRBR (произведение, выражение, воплощение) и информационный объект по связи – "абстрактный концептуальный контейнер для классов, которые порождают эти объекты", а именно:

- первичный информационный объект – информационный объект, который используется самостоятельно, например, текстовые документы, изображения;
- метаданные – информационный объект, главная цель которого состоит в том, чтобы дать информацию о целевом ресурсе (как правило о первичном информационном объекте);
- аннотация – информационный объект, главная цель которого состоит в том, чтобы аннотировать целевой ресурс или его часть (на рис. 4 части ресурса соответствует сущность регион).

Примеры таких аннотаций включают примечания, структурированные комментарии и связи. Объекты аннотации помогают интерпретировать целевой ресурс, содержат либо поддержку, либо возражения, либо более детальные объяснения. Поскольку информационный объект является ресурсом, то он наследует все вышеперечисленные свойства ресурса.

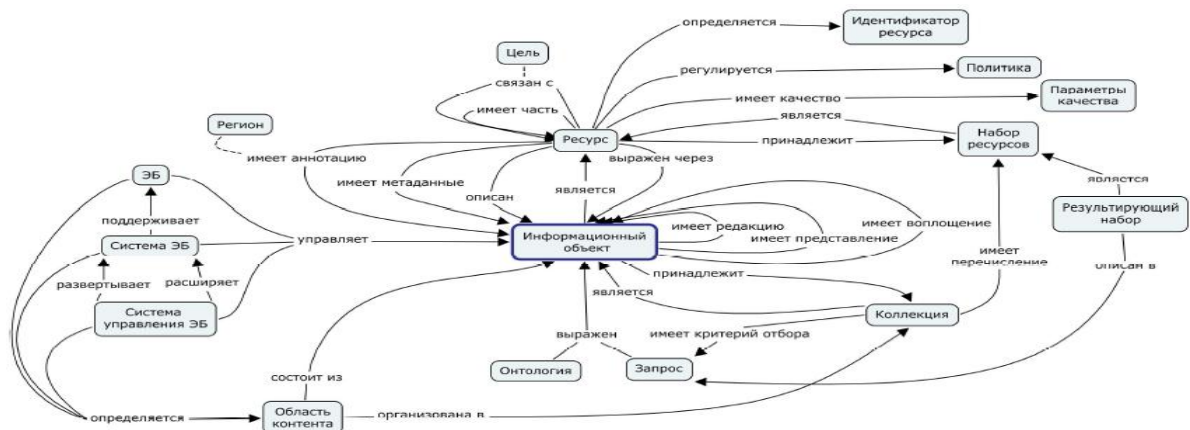


Рис. 4. Область контента ЭБ в модели DELOS DLRM

Информационные объекты также могут быть сложными объектами и могут быть сгруппированы в коллекции информационных объектов. Коллекции, в свою очередь, тоже являются информационными объектами, они наследуют все аспекты моделирования информационных объектов и средства их обслуживания, например они могут аннотироваться. Кроме того, коллекция – специализация понятия набора ресурсов.

Коллекции определяются критерием отбора (`hasIntension`) либо перечислением элементов (`hasExtension`).

Другая специализация понятия набор ресурсов в данной модели – результирующий набор. В традиционных ЭБ он представляет собой набор документов, которые извлекаются в ответ на запрос. Область пользователя в DELOS DLRM содержит все объекты, которые являются "внешними по отношению к системе ЭБ и с ней взаимодействуют: люди и неодушевленные объекты, например, программы или физические инструменты... или даже другая ЭБ может быть среди пользователей ЭБ". Поскольку главная сущность в этой области – персона является ресурсом и следовательно, наследует все его свойства, а именно:

- имеет уникальный идентификатор (идентификатор пользователя);
- организован в соответствии с форматом (модель пользователя);
- благодаря свойствам ресурса композиции и соединения может быть составлен в различные сложные и структурированные группы например, сотрудничество пользователей или соавторов;
- описан или дополнен метаданными и аннотациями.

Область функций представляет наиболее объемную и наиболее открытую часть модели DELOS DLRM, поскольку охватывает всю обработку ресурсов, а также действия пользователей в ЭБ. Здесь наиболее общим понятием является сущность функция. Функция – специфическая задача обработки, которая может быть реализована на наборе ресурсов или одном ресурсе в результате действий отдельного пользователя. Описание функций основано на пользовательском аспекте и ресурсе, представляющем все объекты, вовлеченные в ЭБ. Хотя функции в традиционных моделях ЭБ обычно связываются с контентом в ЭБ и выполняются людьми, здесь, в данной модели, функции могут выполняться неодушевленными пользователями на любом типе ресурсов.

В данной модели ЭБ каждая функция также является ресурсом и потому наследует все его характеристики. Функции разделены на пять классов:

- доступ к ресурсам;
- управление ресурсами;
- совместная работа;
- управление ЭБ;
- настройка ЭБ.

Внимательное изучение данной модели помогло не только обозреть всю сферу ЭБ, но и найти некоторые пробелы в самой модели. Вот некоторые из них:

- недостаточно формализованные определения, оставляющие размытыми границы многих сущностей (например, сущности, заимствованные из FRBR, или граница между метаданными и аннотацией);
- в некоторых местах остаются не ясными критерии выделения сущностей (в частности, область качества наименее убедительна в этом отношении);
- неоднородность описания различных областей ЭБ, скрытая за внешне однообразным описанием (достаточно сравнить простую иерархию области функций со сложной структурой области контента).

К преимуществам DELOS DLRM следует отнести наибольшую полноту охвата среди существующих концептуальных моделей ЭБ.

ГЛАВА 3. АРХИТЕКТУРА ЭЛЕКТРОННОЙ БИБЛИОТЕКИ (ЭБ)

3.1 Элементы Dublin Core

Для электронных библиотек в настоящее время получил распространение формат Dublin Core [19]. Стандарт Dublin Core состоит из двух частей:

- 1) минимальной Dublin Core Metadata Element Set, имеющей статус международного стандарта (ISO15836:2009), которая содержит только 15 атрибутов, применяемых к любым сущностям, в которых можно задать название объекта, его описание, автора, административные метаданные.
- 2) компетентной Dublin Core (квалифицированной, Qualified Dublin Core) состоящей из 18 элементов и группы т. н. тонкостей (или квалификаторов), которые уточняют семантику элементов для повышения полезности поиска ресурсов.

Обе версии стандарта не привязаны ни к какой конкретной предметной области и могут описывать ресурсы любых видов.

Семантика Дублинского ядра была создана международной междисциплинарной группой профессионалов библиотечного дела, компьютерных наук, кодирования текстов, музейного дела и других смежных групп [9].

В России с 2011-07-01 действует ГОСТ Р 7.0.10-2010 (ИСО 15836:2003) «НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ. Система стандартов по информации, библиотечному и издательскому делу. НАБОР ЭЛЕМЕНТОВ МЕТАДААННЫХ ДУБЛИНСКОЕ ЯДРО».

Простой набор элементов метаданных Дублинского ядра (Dublin Core Metadata Element Set; DCMES) состоит из 15 элементов метаданных:

- Title — название;
- Creator — создатель;
- Subject — тема;
- Description — описание;
- Publisher — издатель;
- Contributor — внёсший вклад;
- Date — дата;
- Type — тип;
- Format — формат документа;
- Identifier — идентификатор;
- Source — источник;
- Language — язык;

Relation — отношения;

Coverage — покрытие;

Rights — авторские права.

Квалифицированный (компетентный) набор элементов метаданных Дублинского ядра, помимо 15 вышеперечисленных, может включать:

Audience — аудитория (зрители);

Provenance — происхождение;

RightsHolder — правообладатель.

Каждый элемент опционален и может повторяться. Инициатива метаданных Дублинского ядра (Dublin Core Metadata Initiative; DCMI) описала стандартные пути определения элементов и поощряет использование схем кодирования и словарей. Не существует заранее заданного порядка перечисления этих элементов. DCMI также поддерживает небольшой общий словарь, который рекомендуется использовать с элементом Type (Тип) и который состоит из 12 слов.

3.2 Открытые архивы DSpace и EPrints

Открытые архивы строятся по принципу распределенной системы с общим сетевым протоколом и программным обеспечением, что позволяет говорить о единой информационной системе.

Платформа электронных библиотек DSpace разрабатывалась совместно компанией Hewlett-Packard и библиотеками MIT (Massachusetts Institute of Technology) [10].

DSpace формировалась под влиянием научных исследований в области систем электронных библиотек. Платформа EPrints произошла из движения Scholarly Communication, как следствием конфигурация по умолчанию направлена на поддержку научных статей. EPrints имеет много сходства с DSpace, но оптимизирована так, чтобы обеспечить доступ к самостоятельному депонированию (внесению) автором материалов (статей, отчетов, книг и пр.), в то время как DSpace обеспечивает платформу, для долгосрочного хранения цифровых материалов, используемых в академических исследованиях.

С помощью EPrints можно так же обеспечить подобную функциональность репозитория. Основные отличия EPrints и DSpace лежат в их подходе к организации структуры данных. В DSpace данные организуются таким образом, чтобы отобразить структуру организации, использующую данную систему.

В отличии от DSpace, EPrints не имеет строгого структурного деления на разделы и коллекции. В основе EPrints лежит равенство всех элементов репозитория и отсутствие

иерархии. Однако иерархия необходима для организации навигации по репозиторию. Для этого в EPrints существует представления при помощи информации из метаданных. Тем самым в EPrints можно получить более сложную иерархию. Сходство в моделях начинается на уровне элементов, это соответствует модели OAI-PMH [19]. Выигрышем DSpace, благодаря жесткой иерархии, будет физическая близость данных, находящихся в одном наборе, на накопителе. Как следствие уменьшается время на выдачу информации пользователям. Выигрышем EPrints будет возможность более сложной организации иерархии данных.

DSpace позволяет создавать, распространять и хранить электронные ресурсы. Здесь дается короткое описание функциональных возможностей DSpace (использование электронных библиотек, созданных на основе DSpace, обзор их содержимого и введение новых документов в коллекции). Информационное наполнение DSpace организовано вокруг понятия разделов, которые, как правило, отвечают некоторым организационно-административным структурам, таким как учебные заведения, департаменты, лаборатории, научные учреждения и исследовательские центры. В каждом разделе может быть создано любое количество подразделов и любое количество коллекций. Любая коллекция может содержать любое количество электронных документов [16].

3.3 Многоуровневая архитектура ЭБ

Для систематизации ресурсов электронной библиотеки используется многоуровневая архитектура ЭБ, состоящая из хранилища данных – репозитория, сервер метаданных, сервер приложений, словарь справочники (рис. 5).

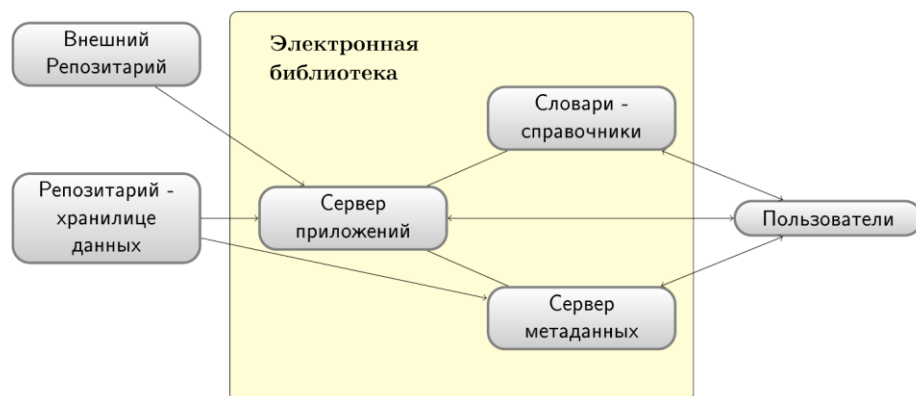


Рисунок 5. Архитектура ЭБ

Хранилище данных предназначено для хранения коллекций и дополнительных метаданных об их структуре. Они обеспечивают хранение электронных версий публикаций, статей, книг и т.д., и предоставляют авторизованный доступ к ним внешних систем и пользователей.

Для сопровождения размещаемых в библиотеке коллекций необходимыми метаданными (информация об авторах, библиографические данные и т.д.) существуют унифицированные форматы метаданных о персонах и их научных трудах. Для хранения и сопровождения этих метаданных в электронной библиотеке служат сервера метаданных, которые обеспечивают хранение и сопровождение метаданных о персонах и их трудах в унифицированном формате, а так же обеспечивают эффективный поиск и предоставление хранимых метаданных внешним системам и пользователям.

Для эффективной работы сервера приложений необходимо использовать набор словарей-классификаторов, содержащих как классификационные признаки, так и наборы ключевых терминов (с отношениями порядка), по которым производится систематизация и классификация материала.

Словарь справочник – это набор термов, которые образуют словарь описания содержимого документа. Его поддерживают органы по стандартизации для создания стандартного способа категоризации материалов, которые попадают в архив.

Точная категоризация материала с помощью словаря справочника увеличивает вероятность того, что в результате поиска будут найдены релевантные поисковому выражению документы при организации поиска по одной или нескольким электронным библиотекам [19].

ГЛАВА 4. СИСТЕМАТИЗАЦИЯ ИНФОРМАЦИОННЫХ РЕСУРСОВ

В системе образования информационные ресурсы выполняют разные функции. Во-первых, они могут содержать учебный материал и служить предметным педагогическим инструментарием. К данному классу относятся компьютерные учебники, задачки, тренажеры, системы тестирования и другие виды компьютерных средств обучения, а также традиционные учебно-методические пособия и их электронные представления. Во-вторых, большое число ресурсов имеет информационно-справочное назначение. Третий класс информационных ресурсов охватывает средства поддержки образовательных бизнес-процессов. В него входят как программные инструменты (системы управления учебным процессом, корпоративные информационные системы, средства разработки информационных ресурсов, информационно-поисковые системы, автоматизированные информационно-библиотечные функции и др.), так и информационные компоненты информационно-образовательных сред (базы данных, файловые массивы). Наконец, в-четвертых, ряд программных и информационных продуктов выступает в качестве предметов изучения.

Для обеспечения потребителей нужными информационными ресурсами используют механизмы включения средств формирования и публикации метаданных информационных ресурсов, каталогизации информационных ресурсов, ведения хранилищ (репозиториев) метаданных и информационных ресурсов, поиск информационных ресурсов по метаданным, управление доступом к информационным ресурсам и др. Использование данных механизмов необходимо для решения проблемы систематизации информационных ресурсов.

Под систематизацией понимается деятельность, направленная на обеспечение условий для эффективного управления информационными ресурсами. В рассматриваемом плане управление информационными ресурсами интерпретируется не в узком технологическом смысле как управление контентом, а охватывает процессы учета информационных ресурсов, публикации информационных ресурсов и их метаданных, поиска, выбора, агрегации, адаптации и распространения информационных ресурсов, оценивания качества информационных ресурсов, формирования заказов на разработку информационных ресурсов и др.

Цели систематизации информационных ресурсов можно считать достигнутыми, если:

- существует согласованный набор общепринятых классификаций информационных ресурсов, требования которых соблюдаются при создании, хранении и учете информационных ресурсов;

- классификации информационных ресурсов отделимы от использующих их систем;
- для каждого информационного ресурса сформированы метаданные, основанные на стандартной открытой модели, учитывающей особенности области применения и совместимой с моделями метаданных для смежных областей;
- метаданные описаны на языках, соответствующих аудитории, на которую ориентированы информационные ресурсы;
- в метаданных указаны глобальные уникальные идентификаторы и классификационные признаки информационных ресурсов;
- метаданные представлены в открытом стандартном формате и опубликованы (т.е. доступны для всех заинтересованных лиц);
- классификации и способы идентификации информационных ресурсов, схема и формат метаданных, а также средства хранения метаданных и информационных ресурсов обеспечивают эффективные возможности для поиска и выбора информационных ресурсов с учетом ограничений и предпочтений пользователей по доступности;
- средства учета, хранения и поиска метаданных и информационных ресурсов способны взаимодействовать друг с другом, образуя распределенную систему;
- определены показатели качества информационных ресурсов, учитываемые при создании, выборе и использовании информационных ресурсов;
- информационных ресурсов реализуются в виде, допускающем оценивание их качества в соответствии с установленными требованиями.

Систематизация предполагает создание многоуровневой инфраструктуры управления информационными ресурсами, формирующей в информационной образовательной среде пространство для взаимодействия всех участников процессов создания, распространения и применения информационных ресурсов: авторов, разработчиков и издателей информационных ресурсов, дистрибьюторов и провайдеров контента, учащихся и преподавателей, менеджеров образовательных заведений и др. Создатели и распространители информационных ресурсов в рамках такой инфраструктуры получают доступ к обобщенным сведениям о потребностях системы образования в информационных ресурсах и состоянии рынка соответствующих продуктов, средствам публикации информационных ресурсов и метаданных, службам регистрации, депонирования и сертификации информационных ресурсов. Потенциальным пользователям она предоставляет информацию о существующих информационных ресурсах, отвечающих их интересам, и дает возможность обоснованного выбора среди них наилучших по соотношению цена/качество ресурсов.

Выделяются следующие базовые задачи систематизации информационных ресурсов:

- классификация информационных ресурсов;
- идентификация информационных ресурсов;
- формирование метаданных информационных ресурсов, отражающих их характеристики, важные с точки зрения управления информационных ресурсов;
- учет (в том числе каталогизация) информационных ресурсов;
- регистрация и депонирование информационных ресурсов;
- публикация информационных ресурсов и метаданных (с обеспечением возможностей поиска);
- оценивание качества и сертификация информационных ресурсов.

Решение большинства задач систематизации информационных ресурсов связано с использованием метаданных. В самом широком смысле метаданные - это данные, характеризующие какие-либо другие данные. Метаданные информационного ресурса выступают в качестве описания этого информационного ресурса. Такое описание можно сравнить с записью в библиотечном каталоге. Оно представляет заглавие информационных ресурсов, его назначение, тип, объем, предметное содержание, технические параметры, сведения об авторах и разработчиках, целевую аудиторию, способы доступа к информационным ресурсам и другую информацию, которая может быть полезна при выборе ресурса.

Консорциум Meta Data Coalition определяет метаданные как описательную информацию о структуре и смысле данных, а также приложений и процессов, которые манипулируют данными. Для обеспечения возможностей автоматизированной обработки метаданных они должны быть структурированы.

Метаданные могут иметь разное назначение, связанное с отражением тех или иных характеристик информационных ресурсов и решением соответствующих задач их систематизации. С этой точки зрения принято выделять различные виды метаданных:

- описательные (библиографические описания информационных ресурсов и характеристики их содержания в виде рефератов и аннотаций);
- структурные (формат, объем и структура информационных ресурсов);
- административные (правообладатели, права на доступ и корректировку информационных ресурсов, сведения о пользователях, платежах и т.п.);
- идентифицирующие, служащие для однозначного представления описываемых объектов для внешнего мира и приложений, и др.

Метаданные могут характеризовать сущности, относящиеся не только к информационному пространству (документы, базы данных, информационные системы,

сервисы и т.п.), но и к реальному миру (персоны, организации, события и др.), а также абстрактные понятия.

Столь широкие описательные возможности метаданных связаны с обобщением понятия информационных ресурсов. Фундаментальная спецификация, определяющая формат URI (Uniform Resource Identifier), не ограничивает класс сущностей, называемых ресурсами. В ней термин «ресурс» обозначает то, что может быть идентифицировано (в частности, с помощью URI). Указанная трактовка зафиксирована в стандарте информационной модели системы метаданных «Дублинское ядро». При этом отмечается, что в контексте сферы использования метаданных Дублинского ядра информационных ресурсов обычно является электронным документом (т.е. Дублинское ядро создавалось в первую очередь для описания электронных документов).

В контексте образовательных информационных технологий метаданные - это структурированные данные, предназначенные для описания характеристик образовательного ресурса, объекта данных или компонента LTS. Таким образом, в информационно-образовательной среде метаданные могут представлять трактуемый в широком смысле информационный ресурс (документ на электронном, печатном или ином носителе, программу, сервис, источник информации, базу данных, информационную систему), информационный объект (экземпляр или элемент информационной модели, запись в базе данных и т.п.), компонент информационно-образовательной системы (техническое устройство, организационную структуру, персону, группу персон), а также событие (образовательное мероприятие - лекцию, семинар, урок, конференцию и т.д.). Система метаданных выступает в качестве центрального звена любой информационной системы. Метаданные могут быть как частью информационного ресурса, так и храниться отдельно от него. Например, выходные сведения издания, являющиеся по отношению к нему метаданными, приводятся в самом издании и, кроме того, включаются в библиотечный каталог.

Как и модель предметной области для базы данных, система метаданных специфицируется на нескольких уровнях:

- инфологическом, фиксируемым схемой метаданных, которая отражает состав и структуру элементов данных (полей) в экземпляре метаданных, их семантику, типы значений (включая словари и классификаторы) и ограничения целостности;
- даталогическом, фиксируемым форматом метаданных, который отражает способ представления (кодирования) информации в экземпляре метаданных.

Главная идея формирования и использования метаданных заключается в том, чтобы извлечь из информационных ресурсов значимые характеристики и представить их в виде,

доступном для интерпретации широкого круга систем, что позволит этим системам опосредованно обрабатывать массивы информационных ресурсов, не затрагивая содержимое самих информационных ресурсов. Разнообразие форм, видов и форматов информационных ресурсов является объективным фактором, затрудняющим обеспечение интероперабельности компонентов информационно-образовательных сред. В то же время для систематизации и управления информационными ресурсами требуется ограниченное множество их свойств, которые целесообразно отразить в метаданных. Определение состава и взаимосвязей этих свойств (системы метаданных) и способа представления (модели) метаданных проще унификации форматов информационных ресурсов. Таким образом, наличие метаданных, базирующихся на стандартных открытых моделях, создает новые возможности для взаимодействия систем - интероперабельность на уровне метаданных. Обеспечение подобной совместимости требует унификации структуры метаданных, интерпретации ее компонентов и способа их представления. Результаты анализа и обобщения характеристик множества информационных ресурсов служат основой для построения системы метаданных - семантической модели, определяющей структуру описания класса информационных сущностей, ориентированную на некоторую область применения (типы информационных ресурсов, задачи систематизации и управления ими). Система метаданных привязывается к формальному способу представления ее элементов (модели метаданных). На сегодняшний день основными моделями метаданных являются базовая иерархическая модель языка XML и модель RDF (Resource Description Framework). К настоящему времени в мире создано множество систем метаданных, обладающих разным статусом: международные, национальные и корпоративные стандарты, спецификации профессиональных консорциумов и др. Ряд разработок, не имеющих статуса официально утвержденных стандартов, широко применяются на практике и рассматриваются в качестве стандартов «де-факто». Ниже перечислены некоторые *наиболее распространенные системы метаданных ИР*:

- Дублинское ядро (Dublin Core) - инвариантный к предметной области набор наиболее общих элементов описания ИР, введенный для обеспечения глобальной интероперабельности приложений, работающих с метаданными;
- MARC (MACHine-Readable Cataloging Record) - предназначена для описания библиотечных ресурсов (как на бумажных, так и на электронных носителях).
- GILS (Global Information Locator Service, Government Information Locator Service) - предназначена для описания любых видов ИР, расширяет MARC и базируется на протоколе Z39.50;

- ONIX (ONline Information eXchange) - предназначена для описания продукции в системах книгоиздания и книготорговли;
 - LOM - предназначена для описания ИР в сфере образования;
 - METS (Metadata Encoding & Transmission Standard) - предназначена для организации обмена пакетами цифровых ИР в ЭБ; определяет формат пакета, включающего цифровой контент и его метаданные, описывающие общие характеристики, логическую структуру, состав физических компонентов (файлов), административные и поведенческие свойства;
 - MODS (Metadata Object Description Schema) - упрощенный вариант MARC, предназначенный для описания цифровых ИР в ЭБ и определенный на уровне формата как XML-схема;
 - UDDI (Universal Description, Discovery and Integration) - предназначена для описания web-сервисов;
 - INDECS (Interoperability of Data in E-Commerce Systems) - ориентирована на системы электронной коммерции и содержит элементы для управления правами интеллектуальной собственности;
 - EAD (Encoded Archival Description) и ISAD (International Standard Archival Description) - предназначены для описания архивных материалов;
 - GEM (Gateway to Education Materials) - расширение Дублинского ядра для описания электронных учебных материалов;
 - RSS (RDF Site Summary) - предназначена для описания новостных сообщений и оповещений об изменениях на сайтах; в качестве модели метаданных использует RDF;
 - CDWA (Categories for the Description of Works of Art) - предназначена для описания произведений искусства, экспонируемых в музеях;
 - CERIF (Common European Research Information Format) - предназначена для описания научных проектов;
 - МЕКОФ - международный коммуникативный формат, выступающий в качестве альтернативы MARC;
 - формат описания БД и машиночитаемых информационных массивов.
- С точки зрения ориентации на виды ИР и области применения системы метаданных принято делить на *универсальные* и *специализированные*. К универсальным системам относятся Дублинское ядро и GILS.

4.1 Систематизация электронных публикаций

Одним из самых распространенных средств систематизации публикаций являются различные рубрикаторы, в частности, Универсальный десятичный классификатор (УДК) и Государственный рубрикатор научно-технической информации (ГРНТИ).

УДК это хороший разветвленный классификатор, но он давно не обновлялся. УДК мог бы использоваться для построения контролируемого словаря ключевых терминов по теме «Электронная библиотека», но данная тема является узконаправленной, поэтому затруднительно будет достраивать его для удобного использования.

Однако ГРНТИ представляет собой достаточно грубую классификацию (3 иерархических уровня), которая позволяет производить систематизацию и поиск только по достаточно широким тематическим классам, что не всегда удовлетворяет пользователей, если их информационная потребность имеет конкретный узконаправленный характер.

В целом эти словари содержат только общенаучную информацию и не годятся для систематизации материалов по конкретной теме.

В рамках диссертационной работы для решения задач систематизации и организации поиска научной электронной информации нами разрабатывается подход, основанный на использовании ИПТ.

Систематизация публикаций предназначена для их разделения по темам с целью сокращения времени поиска по запросам и выполняется с использованием ИПТ. Предполагается, что массивы публикаций сопровождаются метаданными, в состав которых обязательно включаются заголовки публикаций и списки ключевых слов. Процесс систематизации с целью сокращения затрат времени на его выполнение разделяется на два этапа: предметизацию и индексацию.

В нашем случае предметизация публикаций осуществляется на основе анализа их названий и ключевых слов. В результате выполняется отнесение публикаций к соответствующим рубрикам.

Технология индексации состоит в «поэлементном» просмотре программой индексации текстов публикаций, выявления терминов, используемых в качестве индекса публикации. Результатом индексации массива публикаций является база их индексов, которая впоследствии используется для поиска по запросам.

ГЛАВА 5. ИНФОРМАЦИОННО-ПОИСКОВЫЙ ТЕЗАУРУС

При автоматизации поиска документальной информации важнейшей является задача формализации содержания документа и запроса. При этом поиск происходит по всему тексту документа или по его поисковому образу, а в качестве запроса чаще всего выступают отдельные ключевые слова или их логические комбинации. Именно на этой технологии основано действие информационно-поискового тезауруса (ИПТ).

Можно выделить два определения информационно-поискового тезауруса:

1. Информационно-поисковый тезаурус представляет собой словарь, отображающий семантические отношения между лексическими единицами дескрипторного информационно-поискового языка и предназначенный для поиска слов по их смысловому содержанию.
2. Информационно-поисковый тезаурус (ИПТ) – контролируемый словарь терминов предметной области, создаваемый для улучшения качества информационного поиска в этой предметной области.

И в том, и в другом случае речь идет о словаре, который призван облегчить поиск необходимой информации.

Возможно два способа расположения слов в словарях: по близости их буквенного состава и по смысловой близости.

По первому способу создаются алфавитные словари. По второму способу – тезаурусы. Алфавитные словари служат для раскрытия значения данного слова. Тезаурусы служат для поиска слов для выражения данного понятия. То есть, если в обычном словаре по слову ищется его смысл, то в тезаурусе по заданному смыслу ищутся слова, которые этот смысл выражают [15].

Основными целями разработки традиционных ИПТ являются следующие:

- обеспечение перевода естественного языка документов и пользователей на контролируемый словарь, применяемый для индексирования и поиска;
- обеспечение последовательного использования единиц индексирования;
- описание отношений между терминами;
- использование как поискового средства при поиске документов;
- единицы традиционных информационно-поисковых тезаурусов.

5.1 Построение информационно-поискового тезауруса

Тезаурусы разрабатываются для отдельных отраслей. Например, существует «Тезаурус по документации», «Дескрипторный словарь по информатике» и т.п. Тезаурус может быть разработан и для обслуживания конкретной информационно-поисковой системы.

Построение тезауруса состоит из нескольких взаимосвязанных этапов.

Первый этап - формирование словаря. Словарь – первоначальные множества ключевых слов. При этом рассматривается представительный массив наиболее информативных для данной предметной области документов. Например, реферативные журналы, словари, учебники, справочники, нормативные документы, и т.д. Выбираются слова, употребляемые в этих источниках, при всем этом устанавливается частота употребления слов и учитываются все формы, которые могут иметь слова.

Второй этап – формирование множества ключевых слов. Из словаря формируется множество ключевых слов. При отборе ключевых слов учитывается информативность слова, которая определяется исходя из частоты встречаемости слова, роли слова в этой предметной области. Процесс выбора ключевых слов достаточно сложно формализовать. Например, такой критерий, как частота встречаемости не может быть абсолютным. Если слово встречается в текстах очень часто, это может означать, что оно выражает чрезмерно широкое понятие, либо недостаточно четко определено, т.е. неинформативно. Если ключевое слово встречается очень редко, это может означать, что оно выражает новое понятие и таким образом является информативным.

Третий этап – формирование классов эквивалентности. Выделение дескрипторов.

Автоматические информационно-поисковые тезаурусы являются составным элементом автоматического индексирования документов и запросов. В словарной статье автоматического тезауруса, как правило, зафиксированы отношения условной эквивалентности, отношения подчинения и ассоциативные отношения.

5.2 Структура словарных статей

В данной коллекции отображаются словарные статьи и публикации. В свою очередь публикации делятся на статьи (опубликованы) и документы (не опубликованные статьи).

На рисунке 6 схематично представлена структура данных в словарных статьях.

При вводе новых данных в коллекцию необходимо заполнить обязательные и не обязательные поля. Если не все основные поля были заполнены, то при сохранении документа всплывает ошибка. К «обязательным» полям относятся:

Поле (author) - фамилии авторов публикации (тип STRING). При создании новой статьи или публикации указывается фамилия автора, а при вводе новой словарной статьи в этом поле указывается название термина.

Поле (title) - заголовок публикации или документа, который будет отображаться в списке (тип STRING).

Поле (bib_description) - библиографическое описание (тип TEXT). Запись этого поля будет отображаться на странице.

Поле (year) – год издания (тип WORD). При вводе словарной статьи указывается текущий год.

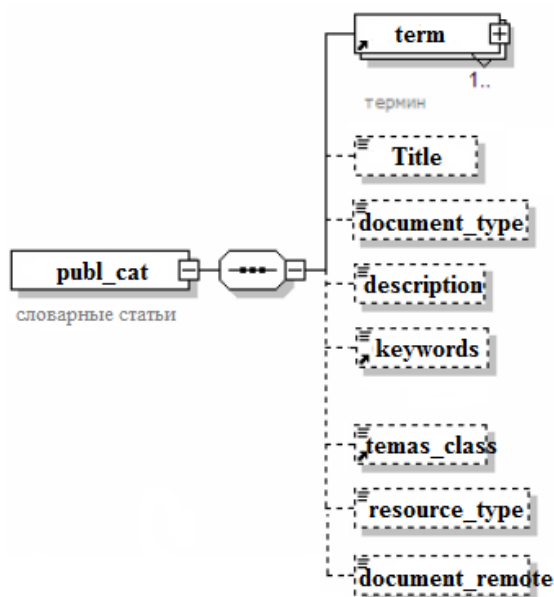


Рисунок 6. Структура данных «Словарных статей»

Поле (document_type) - жанровый тип ресурса, статья или документ выбирается из списка (тип SELECT): web сайт; база данных; библиотеки; документы; журналы; информационные материалы; информационные сообщения; коллекции; конференции и совещания; математическое обеспечение; научные общества и советы; общие материалы; отчеты; презентации; программы; проекты; словари; события статьи.

Поле (keywords) - ключевые слова, вводятся через двоеточие и нужны для связи с другими коллекциями (тип WORD_LIST).

Поле (temas_class) - связь со списком тем (тип TLINK).

Поле (resource_type) - вид ресурса (тип SELECT): main – корневой; add – вспомогательный; info – информационный.

Поле (date) - дата записи (тип DATE).

Поле (last_mod) - дата модификации (тип DATE).

Поле (document_owner) - владелец записи (тип WORD).

Поле (document_correct) - правильность заполнения (тип SELECT).

Поле (document_remote) - тип ресурса (тип SELECT).

К «не обязательные» поля относятся:

Поле (organization) - фамилии персон, кому посвящена публикация (тип STRING).

Поле (description) - описание ресурса (тип TEXT).

Поле (abstract) - подробное содержание материала (аннотация) ресурса (тип TEXT).

Поле (document_largeobject) - HTML-файл с дополнительным описанием (тип LARGE_OBJECT).

Поле (document_appl) - PDF файл с дополнительным содержанием (тип LARGE_OBJECT).

Поле (url) - основной URL адрес в сети (тип URL).

Поле (source_url) - URL-адрес источника (тип URL).

Поле (DSpace_handel) - Указатель на DSpace (тип URL).

Поле (document_modifier) - кто исправил запись (тип WORD).

Поле (document_template) - шаблон ресурса отличного от публикации (тип TEXT).

Поле (document_code) - коды рубрикаторов (тип WORD).

Поле (link_to_dict) - связь со страницами (словарем) (тип LINK).

Интерфейс словарной статьи удобен для пользователя и отображает основные поля и связи с другими коллекциями:

- название термина (статьи);
- описание термина (статьи);
- ключевые термины, связанные с данным термином;
- ссылки на публикации;
- ссылка на персон;
- ключевые слова;
- контекстный поиск.

5.3 Связи и типы данных

Главной задачей дипломной работы было построение контролируемого словаря для систематизации ресурсов электронной библиотеки. На данный момент в мире нет универсальных контролируемых словарей. Построение такого словаря необходимо для организации информационно-поискового тезауруса.

Под информационно-поисковым тезаурусом (ИПТ) понимается словарь лексических единиц информационно-поискового языка, в котором заданы парадигматические (базисные) смысловые отношения между этими единицами.

Контролируемый словарь описывает отношения между терминами, которые относятся к теме «Электронные библиотеки». Такой словарь предназначен для помощи при информационном поиске и представляет собой список тем и ключевые термины (основные и дополнительные) (рис. 7). Ключевые термины привязываются к словарным статьям через id.

Связи между темами и словарем задается жестко. А связи между терминами задаются легко, т.е. зависит от того присутствует этот термин в описании или нет.

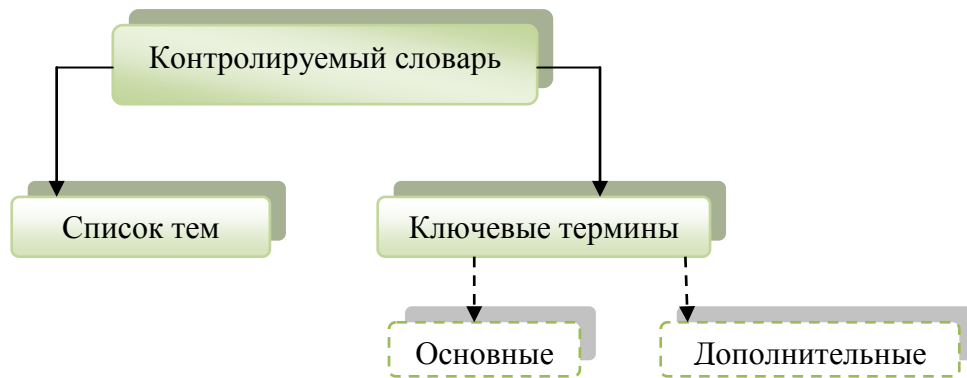


Рисунок 7. Контролируемый словарь

Контролируемый словарь состоит из следующих элементов:

- Персоны (persons_cat);
- Ресурсы в коллекции (publ_cat);
- Список разделов (тем) (temas_class).

Объекты, которые хранятся в системе, являются статьями. К типам этих объектов относятся: персоны, публикации, которые имеют выходные данные, прочие документы (автобиография, письма), и словарные статьи. Между ними устанавливаются соотношения связи по ключевым терминам id.

Словари (ключевые признаки, ключевые термины) – это особый вид метаданных, которые отражают наиболее существенные свойства объекта и их специфика определяются терминологией конкретной предметной области. Необходимо рассматривать различные типы ключевых терминов, а именно:

- ключевые термины в стандартном понимании;
- ключевые термины, описывающие персону;

- ключевые термины, описывающие организацию;
- ключевые термины, описывающие временные периоды;
- ключевые термины, описывающие географические понятия.

Классификация тем и терминов представлена в виде десятичного рубрикатора по правилам классификации Дьюи.

Порядковый номер тем для курса «Современные проблемы информатики» нумеруются следующим образом темы «Электронные библиотеки» нумеруются:

001

002

003 и т.д.

Такая классификация предназначена для удобства и систематизации тем. Для каждой темы иерархически привязываются статьи по ключевым словам.

Задачи при разработке ИПТ в рамках диссертационной работы:

Установление связей между:

- ключевой термин → общие термины;
- ключевой термин → публикация;
- ключевой термин → персона;
- ключевой термин → организация;
- ключевой термин → время;
- ключевой термин → географическая привязка.

Пример. Установленная связь между корневым термином «Информационно-поисковый тезаурус» и общими терминами (Приложение А).

Пример: 3 установленных связей: 1) между корневым термином «Гипертекст» и общими терминами; 2) между корневым термином «Гипертекст» и связанными с этим термином публикациями; 3) между корневым термином «Гипертекст» и персонами (Приложение Б).

Наш подход к описанию терминологии с помощью тезауруса во многом опирается на работы [Никитина, 1978, 1987, 1996]. Свойства терминов предметной области – системность, устойчивость и регулярность связей, установка на объективность описания – делают возможным адекватное описание терминологии с помощью тезаурусов. Ключевой момент такого подхода – учет системных свойств терминов предметной области (понятийной структуры терминологии по [Шелов, 2001]).

Формат должен выполнять также *коммуникативную функцию*, т.е. способствовать повторному использованию, обмену и интеграции терминологических данных в виде тезаурусов.

Кроме того, при разработке перед нами стояла традиционная задача сохранения баланса между развитостью структуры, выразительными возможностями формата – с одной стороны – и простотой, прозрачностью описания – с другой.

Исходя из этих требований, в качестве формата представления тезауруса мы выбрали язык XML. Формат тезауруса описывается в виде XML Schema [XML].

В качестве аналогов при работе мы рассматривали формат словаря Virtual HyperGlossary и документы проекта SALT [Virtual, SALT]. При разработке формата описания тезауруса мы старались ориентироваться на отечественный и международный стандарты [ГОСТ, ISO]. В частности, мы стремились привести словарь описания тезауруса к словарю стандарта ISO 12620: 1999 Computer applications in terminology – Data categories.

ЗАКЛЮЧЕНИЕ

В результате проделанной работы были исследованы элементы Dublin Core и все его составляющие для работы с репозиторием, дан обзор используемых технологий и работ, которые в настоящий момент существуют в ЭБ ИВТ СО РАН.

В процессе работы были решены следующие задачи:

- сделан аналитический обзор международных стандартов и рекомендаций по организации баз данных публикаций, информационных ресурсов и электронных библиотек;
- Исследованы технологические средства для построения ЭБ в области ИС;

Решены следующие вопросы для построения ИПТ:

- выяснены классификационные признаки документов(публикаций) и систематизированы;
- определены и установлены связи между документами;
- проклассифицированы документы;
- построен поисковый образ документа;
- составлен ИПТ и реализована БД;
- обозначены отношения между терминами.

Изучены более 200 статей и монографий, посвященных организации ИС связанных с ЭБ, которые были занесены в хранилище данных DSpace, для части было сделано мета-описание, т.е. информация была занесена в сервер метаданных и на основе обработки информации построен информационно-поисковый тезаурус посвященный тематике «Электронные библиотеки» и «Информатика». Исходя из анализа литературы посвященной электронной библиотеке были отобраны соответствующие ключевые термины и дана их классификация.

Практическим результатом исследований является систематизация ресурсов электронной библиотеки на основе информационно-поискового тезауруса, позволяющей повысить эффективность работы информационной системы. Создан контролируемый словарь ключевых терминов, определены связи и классифицированы документы, а также построен автоматизированный отбор публикаций в электронной библиотеке.

ЛИТЕРАТУРА

1. Когаловский М.Р., Паринов С.И. Информационные ресурсы, наукометрические показатели и показатели качества метаданных системы Соционет // Тр. IX Всероссийской конф. "Электронные библиотеки: перспективные методы и технологии, электронные коллекции". - RCDL'2007, г. Переславль-Залесский, Россия. - 2007. - С. 45-54.
2. Шокин Ю.И., Федотов А.М., Барахнин В.Б. Проблемы поиска информации. Новосибирск: Наука, 2010.
3. 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.
4. Crofts N., Doerr M., Gill T., Stead S., Stiff M.(editors), Definition of the CIDOC Conceptual Reference Model, January 2008. Version 4.2.4.
5. *Functional Requirements for Bibliographic Records, Final Report / IFLA Study Group on the Functional Requirements for Bibliographic Records.* – München: K.G. Saur, 1998. (UBCIM Publications, New Series; v.19)
<http://archive.ifla.org/VII/s13/frbr/frbr.html>
6. Candela L., Castelli D., Dobрева M., Ferro N., Ioannidis Y., Katifori H., Koutrika G., Meghini C., Pagano P., Ross S., Agosti M., Schuldt H., Soergel D. The DELOS Digital Library Reference Model Foundations for Digital Libraries. IST-2002-2.3.1.12. Technology- enhanced Learning and Access to Cultural Heritage. Version 0.98, December 2007.
http://www.delos.info/files/pdf/ReferenceModel/DELOS_DLReferenceModel_0.98.pdf
7. Goncalves M.A., Fox E.A., Watson L.T. and Kipp N.A. Streams, structures, spaces, scenarios, societies (5s): A formal model for digital libraries // ACM Trans. Inf. Syst. 22(2). - 2004. - P. 270–312.
8. The CIDOC Conceptual Reference Model.
<http://cidoc.ics.forth.gr>
9. Doerr M., Iorizzo D. The Dream of a Global Knowledge Network - A New Approach //ACM J. on Computing and Cultural Heritage, June 2008. - Vol. 1, N. 1, Article 5.
10. Tillett, Barbara B. "Bibliographic Relationships." In: Relationships in the Organization of Knowledge, edited by Carol A. Bean and Rebecca Green. - Dordrecht: Kluwer Academic Publishers, 2001. - P. 19-35.
11. Barbara Tillett What is FRBR? A Conceptual Model for the Bibliographic Universe.
<http://www.loc.gov/cds/downloads/FRBR.pdf>

12. Функциональные требования к библиографическим записям : окончат. отчет / Рос. библ. ассоц., Рос. гос. б-ка ; пер. с англ.
[В.В. Арефьев ; науч. ред. пер.: Т.А. Бахтурина, Н.Н. Каспарова, Н.Ю. Кулыгина]. -М.: РГБ, 2006. - 150 с.
13. Лапо Петр Михайлович, Соколов Андрей Владимирович : ВВЕДЕНИЕ В ЭЛЕКТРОННЫЕ БИБЛИОТЕКИ. – 2005, с. 27-32.
14. Филозова Ирина Анатольевна: ОТКРЫТЫЕ АРХИВЫ НАУЧНОЙ ИНФОРМАЦИИ. -2010. с. 4-6.
15. Ивасенко А.Г. Информационные технологии и управления: учебное пособие / А.Г.Ивасенко, А.Ю.Гридасов, В.А.Павленко – 2 изд., стер. – М.:КНОРУС,2007, -106с.
16. Организация работы с документами: Учебник / Под.ред.проф. В.А.Кудрякева, - 2-изд.,перераб. и доп.- М.: -ИНФРА-М,2001,-592с.
17. Браславский П.И. Тезаурус для расширения запросов к машинам поиска Интернета: структура и функции.
18. Овчинников Д.В. «Разработка системы автоматизации функций формирования SCORM совместимых метаданных для информационных ресурсов», 2011г.
19. Федотов А.М., Байдилдаева А.Б., Ручка Е.В, Федотова О.А. Модель информационной системы для поддержки научно-педагогической деятельности. // Вестник Новосибирского государственного университета. Серия: Информационные технологии. 2013. Т. 11. Вып. 3. С. 7-9.

ПРИЛОЖЕНИЕ А

Ключевые термины, связанные с термином "*Информационно-поисковый тезаурус*":

1. [Вспомогательный дескриптор](#)
2. [Графический указатель информационно-поискового тезауруса](#)
3. [Дескриптор](#)
4. [Дескрипторный информационно-поисковый язык](#)
5. [Дескрипторный словарь](#)
6. [Иерархический указатель информационно-поискового тезауруса](#)
7. [Ключевое слово](#)
8. [Ключевые слова в контексте](#)
9. [Код дескриптора](#)
10. [Лексико-семантический указатель](#)
11. [Макротезаурус](#)
12. [Микротезаурус](#)
13. [Многоязычный информационно-поисковый тезаурус](#)
14. [Нондескриптор](#)
15. [Одноязычный информационно-поисковый тезаурус](#)
16. [Пермутационный указатель информационно-поискового тезауруса](#)
17. [Политематический информационно-поисковый тезаурус](#)
18. [Систематический указатель информационно-поискового тезауруса](#)
19. [Специализированный информационно-поисковый тезаурус](#)
20. [Язык ключевых слов](#)
21. [Контролируемое индексирование](#)

Ключевые слова: [тезаурус](#); [информационный поиск](#); [Автоматизированный информационный поиск](#);

● *Контекстный поиск*: Задайте образец для поиска:

Найти

ПРИЛОЖЕНИЕ Б

Словарная статья по термину «Гипертекст».

Словарь терминов в коллекции "Современные проблемы информатики"

Гипертекст [hypertext]

Гипертекст - принцип организации информационных массивов, при котором отдельные информационные элементы связаны между собой ассоциативными отношениями, обеспечивая взаимосвязанных данных, взаимосвязанных указанными отношениями.

Термин гипертекст предложен в 1967 году [Теодором Нельсоном](#), представившим гипертекстовую структуру в виде ориентированного графа, в вершинах которого находятся записи. Средством организации связей в гипертексте и ориентации в нем является навигатор.

Идея использования ассоциативных отношений между документами впервые была высказана [Полем Отле](#) и частично реализована [Эмануэлем Гольдбергом](#) в статистической машине. В 1939 году, в которой описывает работу гипотетической машины MEMEX:

[Описание MEMEX из статьи В.Буша "As We May Think"](#)

Необходимость реализации прозрачного поиска информации в 1941 году формализует [Хорхе Борхес](#) в своем эссе "[Вавилонская библиотека](#)", описывая Человека Книги и алгоритм.

Ключевые термины, связанные с термином "гипертекст":

1. [Навигатор \[navigator\]](#)
2. [Гиперссылка \[hyperlink\]](#)
3. [Ассоциативная связь](#)
4. [Гипертекстовая база данных](#)
5. [Гипертекстовый документ](#)
6. [Протокол передачи гипертекста](#)
7. [Язык гипертекстовой разметки HTML](#)
8. [Микрофиша](#)
9. [MEMEX](#)
10. [Гипермедиа](#)

Ссылки на публикации:

1. [Черняк Л. Статистическая машина Эмануэля Гольдберга // Открытые системы №03, 2004](#)
2. [Bush Vannevar. As We May Think // The Atlantic Monthly, July 1945.](#)
3. [Memex and Beyond \(Memex и его окружение\) Web-сайт \[http://cs.brown.edu/memex\]](#)
4. [Иванова И. А. Электронные гипертекстовые тезаурусы в дистанционном обучении/ Открытое образования. № 1. 2002. С.15- 19.](#)
5. [Борхес Х. Вавилонская библиотека // Хорхе Луис Борхес «Письмена Бога»: Республика, Москва, 1994.](#)

Ссылки на персон:

1. [Буш Ваннивер](#)
2. [Гольдберг Эмануэль](#)
3. [Нельсон Тея](#)
4. [Отле Поль](#)
5. [Энгельбаот Дуглас Карл](#)

Ключевые слова: поиск; [гипертекст](#); [навигатор](#); [Статистическая машина](#);

● **Контекстный поиск:** Задайте образец для поиска:

ПРИЛОЖЕНИЕ В

Персоны в коллекции "Современные проблемы информатики"

Руджеро Сергеевич Гиляревский

Родился в Москве 31 августа 1929 г.



Основные достижения:

Гиляревский, Руджеро Сергеевич, доктор филологических наук, профессор, заслуженный деятель науки России, ведущий специалист в области социальной информатики, жена - Этторе Макки и балерина Калаянское оперного театра Екатерина Владимировна Крайнова. После отъезда отца из страны в 1929 г. его судьба неизвестна, мать в 1937 - 1954 гг. - у него женой, школьной учительницей Зинаидой Владимировной Гиляревской.

Краткая биография:

В 1947 - 1948 гг. Руджеро Сергеевич учился в Московском энергетическом институте на факультете электрорадиотехники промышленности, затем в МГУ им. М. В. Ломоносова на заочном кандинатурском курсе "Описание книг на иностранных языках для каталогов советских библиотек". 1959 г. - докторская диссертация "Общие закономерности в ра-

боте научных интересов Р.С. Гиляревского очень широк. Начатые еще в ВГБИЛ вместе с покойным Б.А. Старостиним работы по транскрипции иностранных имен привели к созданию справочника «Иностр. «Практическая транскрипция фамильно-именных групп» (2004, 2006). Книга «Основы научной информатики», написанная в соавторстве с А.И. Михайловым и А.И. Черным, вышла на семи языках: в Берлине

В сферу научных интересов Р.С. Гиляревского входят проблемы научной информации и коммуникации, компьютерной технологии и, в частности, гипертекста и электронных книг, информационного менеджера

В настоящее время Руджеро Сергеевич Гиляревский заведует Отделом теоретических и прикладных проблем информатики, Отделением научных исследований по проблемам информатики ВИННИТИ, прог "Высшая школа экономики", является главным редактором сборников «Научно-техническая информация», заместителем главного редактора журнала «Международный форум по информатике», входит в Гиляревский награжден несколькими медалями, в 1999 г. ему присвоено звание "Заслуженный деятель науки Российской Федерации"

Научные труды:

1. Гиляревский Р.С. Информационный менеджмент : управление информацией, знаниями, технологиями. - СПб. : Профессия, 2009.
2. Гиляревский Р.С. Развитие принципов книгописания : кратк. очерк. - 2-е изд., доп. - СПб. : Профессия, 2008. - 240 с.
3. Гиляревский Р.С. Рубрикатор как инструмент информационной навигации / Р.С.Гиляревский, А.В.Шалкин, В.Н.Воловоерок. - СПб. : Профессия, 2008. - 352 с.
4. Справочник информационного работника : науч. ред. Р.С. Гиляревский, В.А. Мижкина. - 2-е изд., перераб. и доп. - СПб. : Профессия, 2007. - 584 с.
5. Информатика как наука об информации : информ., техн., экон., социал. и орг. аспекты / под ред. Р.С. Гиляревского. - М. : ФАИР-ПРЕСС, 2006. - 592 с.
6. Гиляревский Р.С. Основы информатики : курс лекций / Р.С. Гиляревский. - М. : Экзамен, 2003. - 319 с.
7. Рынок информационных услуг и продуктов / И.И. Родионов, Р.С. Гиляревский, В.А. Цветкова, Г.В. Залев. - М. : МК-Периодика, 2002. - 549 с.
8. Информационное пространство новых независимых государств / Ю.М. Арский, Р.С. Гиляревский, И.И. Родионов, В.А. Цветкова и др. - М. : ВИННИТИ, 2000. - 200 с.
9. Информатика : информационные структуры, системы и процессы в науке и обществе / Ю.М. Арский, Р.С. Гиляревский, И.С. Туруев, А.И. Черный. - М. : ВИННИТИ, 1996. - 489 с.
10. Михайлов А.И. Научные коммуникации и информатика / А.И. Михайлов, А.И. Черный, Р.С. Гиляревский. - М. : Наука, 1976. - 433 с.
11. Михайлов А.И. Основы научной информатики / А.И. Михайлов, А.И. Черный, Р.С. Гиляревский. - М. : Наука, 1965. - 653 с.
12. Михайлов А.И. Информатика - новое название теории научной информации / А.И. Михайлов, А.И. Черный, Р.С. Гиляревский // Научно-техническая информация. - 1966. - № 12.

Публикации о жизни и деятельности персоны:

1. [Человек и Неограниченная информатика, или кто и где read // Открытые системы №03, 2004](#)

Публикации персоны:

1. [Гиляревский Р. С. Основы информатики: курс лекций. - М.: Экзамен, 2003. - 319 с.](#)

Ключевые слова: [Информатика](#), [Библиотечное дело](#).

Документальный номер: Задайте образец для поиска:

[найти](#)

ПРИЛОЖЕНИЕ Г

Словарная статья по термину «Метаданные»

Метаданные

Метаданные (Metadata) - данные о данных - данные, описывающие сущности, представленные в информационных системах, представляют собой характеристики описываемого информационного ресурса.

Метаданные - это содержание каталогов, справочников, реестры, и т.п., содержащие сведения о составе данных, содержании, статусе, происхождении, местонахождении

В настоящее время существует большое количество систем метаданных, предназначенных для описания различных классов информационных объектов. Использование оригинальных систем метаданных и форматы метаописаний, а также разные подходы к решению прикладных задач.

Различаются описательные, структурные и административные метаданные.

Описательные метаданные описывают содержание информационного ресурса (например, это — набор значений элементов метаданных Дублинского ядра - Dublin Core).

Структурные метаданные характеризуют общую структуру ресурса и ее компоненты, объем и другие подобные свойства описываемого ресурса.

Административные метаданные описывают даты создания и обновления ресурса, кем он создан или модифицирован, кто владелец прав на этот ресурс, полномочия пост

Стандарты метаданных:

Для обеспечения interoperability и повторного использования метаданных важное значение имеет стандартизация средств их представления. Деятельность в этой сфере стандартов метаданных как независимых от сферы применения, так и предназначенных для специфических применений. К первой группе относятся, например, дескриптивные консорциума W3C: XML Schema, RDF, RDFS, OWL, OWL2 с его профилями и язык описания интерфейсов веб-сервисов WSDL; набор элементов Дублинского ядра (Dublin

Ключевые термины, связанные с термином "Метаданные":

1. [Дублинское ядро](#)
2. [EDIFACT](#)
3. [MARC](#)
4. [MARC](#)
5. [Стандарт для цифровых геопространственных данных \(SDGM\)](#)
6. [Формат обмена для справочников геопространственных данных \(DIF\)](#)
7. [Глобальная \(правительственная\) служба поиска информации \(GIS\)](#)
8. [Инициатива по копированию текстов \(TEI\)](#)
9. [Шаблонно-ориентированные метаданные \(IAFA/WHOIS++\)](#)
10. [Копировка военных списаний \(EAD\)](#)

Ссылки на публикации:

1. [Understanding Metadata // NISO Press - 2004. 20 с.](#)
2. [Коваленский М.Р. Метаданные, их свойства, функции, классификация и средства представления // Труды 14-й Всероссийской научной конференции «Электронные библиотеки» - 2005. - 92 с.](#)
3. [Лето П.М., Соколов А.В. Введение в электронные библиотеки // \[Электронный ресурс\] Москва, - 2005. - 92 с.](#)
4. [Иванов И.Г. Подходы к созданию системы поддержки функционирования репозитория порталов // Интернет-порталы: содержание и технологии: Сб науч. ст. Вып.1](#)
5. [Бичковский В.И., Корсаков А.Г., Гурский Л.М. Технологии использования метаданных при создании электронных библиотек. Менеджмент библиотек учреждений образования // Библиосфера. - 2006. - № 1. - с. 43-54.](#)
6. [Журавлев О.В., Лисовский К.А., Тимченко Г.С., Тимченко Э.С. Функциональные методы обработки слабоструктурированных данных и их применение для построения репозитория // Библиосфера. - 2006. - № 1. - с. 43-54.](#)
7. [Васильев А., Козлов Д., Саломеев С., Шамина О. Извлечение метаданных и библиографических ссылок из текстов русскоязычных научных статей // Труды конференции «Информационные ресурсы России» - 2006. - № 1. - с. 43-54.](#)
8. [Воробей Х. Вавилонская библиотека // Хорхе Луис Борхес «Письмена Бога». Республика. Москва. 1994.](#)

Ключевые слова: Метаданные.

Документальный поиск: Задайте образец для поиска:

[найти](#)

ПРИЛОЖЕНИЕ Д

В качестве примера приведено описание одной из словарных статей в виде xml кода.

```
<publ_catEntry publID="title">
<publEntry lang="ru"
<publUID="112">
<title>Электронная библиотека</title>
<years>2013</years>
<document_type>Словари</document_type>
<description><LI><B>ЭЛЕКТРОННАЯ БИБЛИОТЕКА</B> – это структурированная
каталогизированная коллекция разнородных электронных документов,
снабженная средствами навигации и поиска.</description>
<subject>Репозиторий</subject>
<subject>Архив</subject>
<temas_class>Электронные библиотеки</temas_class>
<resource_type>main</resource_type>
<category>process</category>
<tables_links="relatedConcept" publ_catEntry="temas_class"/>
</publ_catEntry>
```

ПРИЛОЖЕНИЕ Ж

Словарная статья по термину «Электронная библиотека»

Электронная библиотека

- 1) Электронная библиотека - распределенная информационная система, позволяющая надежно сохранять и эффективно использовать разнородные коллекции электронных документов;
- 2) Электронная библиотека - это информационная система, объединяющая свои ресурсы единой идеологией структуризации.

Электронная библиотека - это вид информационных систем, в котором документы хранятся и могут использоваться в машиночитаемой ("электронной") форме, причем при использовании электронных коллекций документов. Электронные издания на оптических компакт-дисках включаются в ЭБ только при условии, если Библиотека выставляет их в сети (и

Ключевые термины, связанные с термином "электронная библиотека":

1. [Документ в электронной форме](#)

Ссылки на публикации:

1. [Дале П.М., Соколов А.В. Введение в электронные библиотеки // \[Электронный ресурс\] Москва, - 2005. - 92 с.](#)
2. [Бричковский В.И., Корсаков А.Г., Гурский Д.М. Технологии использования метаданных при создании электронных библиотек. Менеджмент библиотек учреждений образования. - 2003. - Том 6 - Выпуск 2.](#)
3. [Ковалевский М.Р. Стандарты XML и электронные библиотеки // Электронные библиотеки. ИРИО - 2003. - Том 6 - Выпуск 2.](#)
4. [Голышев Е., Вышинский К. Развитие электронных библиотек: мировой и российский опыт, проблемы, перспективы // Интернет и российское общество. Под ред. И. Семигуса. - М.: ИРИО, 2009.](#)
5. [Павлов Д.Э., Смирнов В.Н., Курочкин Д.Н. Электронная библиотека ВУЗа - как инструмент автоматического формирования учебных мультимедийных коллекций // Петрозаводск, 2009.](#)
6. [Добряков М. Методы изучения пользователей электронных библиотек: современное состояние и перспективы // Вестник Пермского университета. Выпуск 2 \(16\), 2011.](#)
7. [Антопольский А.В., Майборода Г.В., Чуваев А.В. Формирование электронного документного пространства и перспективы создания Российской ассоциации электронных библиотек // Электронные библиотеки. ИРИО - 2003. - Том 6 - Выпуск 2.](#)
8. [Воробей Ф. С. Развитие электронных библиотек как подсистем АБИС - перспективное направление автоматизации библиотек \[Электронный ресурс\]. - Ф. С. Воробей, 2009.](#)
9. [Ковалевский М.Р. Тенденции развития технологий управления информационными ресурсами в электронных библиотеках // Тр. 8-ой Всерос. научн. конф. «Электронные библиотеки». - 2009. - № 2. - С. 2.](#)
10. [Васильев А., Козлов Д., Савицкая С., Шамина О. Создание электронной библиотеки русскоязычных научных статей // Сборник работ стипендиатов гранта "Интернет - наука". - 2010. - № 2. - С. 2.](#)
11. [Ковалевский М.Р. Перспективы развития технологий управления информационными ресурсами в электронных библиотеках // Электронные библиотеки. ИРИО - 2003. - Том 6 - Выпуск 2.](#)
12. [Журавлева О.В., Лысовский К.Ю., Тажичев Г.С., Тажичев Э.С. Функциональные методы обработки слабоструктурированных данных и их применение для построения систем управления контентом // Электронные библиотеки. ИРИО - 2003. - Том 6 - Выпуск 2.](#)
13. [Париков С.И., Ковалевский М.Р. Технологии семантического структурирования контента научных электронных библиотек // Электронные библиотеки. ИРИО - 2003. - Том 6 - Выпуск 2.](#)
14. [Колычева Э.А., Гинзбург Л.Ю., Палицына Н.А. Результаты мониторинга и основные рекомендации по вопросам развития электронных библиотек ВУЗов. // Труды докладной комиссии по развитию науки, 2008. - С. 380 - 386.](#)
15. [Резниченко В.А., Прокшудина Г.Ю., Кудин К.А. Концептуальная модель электронной библиотеки \[Электронный текст\]. - В. А. Резниченко, Г. Ю. Прокшудина, К. А. Кудин. - 2009.](#)
16. [Бричковский В.И. Создание программного обеспечения управления информационными ресурсами Web портала виртуальной библиотеки В.И. Бричковский, А.М.Андреев // Электронные библиотеки. ИРИО - 2003. - Том 6 - Выпуск 2.](#)
17. [Антопольский А.В. Научно-образовательная электронная библиотека вуза / А.Г. Абросимов, Д.С. Зубов // Тр. 10-й Всерос. науч. конф. «Электронные библиотеки: перспективы». - 2010. - № 2. - С. 2.](#)
18. [Антопольский А.В. О социально-экономических моделях электронных библиотек: вопросы стратегии // Интернет и современное общество. Труды X Всероссийской научной конференции «Электронные библиотеки». - 2010. - № 2. - С. 2.](#)
19. [Кудин К.А., Прокшудина Г.Ю., Резниченко В.А. Сравнение систем электронных библиотек EPrint 3.0 и DSpace 1.4.1 // Деятая всероссийская научная конференция «Электронные библиотеки». - 2010. - № 2. - С. 2.](#)
20. [Ле Лави, Гилберт А. Ф. Разработка семантической электронной библиотеки // Доклады ТУСУРа, № 2 \(24\), часть 2, декабрь 2011.](#)
21. [Семиниченко О.В., Воробей Ф. Электронные библиотеки и проблемы интеграции информационных ресурсов в гетерогенном сетевом пространстве \[Текст\] // Электронные библиотеки. - 2009. - № 2. - С. 2.](#)
22. [Ковалевский М.Р. Систематика коллекций информационных ресурсов в электронных библиотеках. // Программирование. № 3, 2000 г., с. 31-51.](#)
23. [Ковалевский М.Р., Новиков Е.А. Электронные библиотеки - новый класс информационных систем. \(От сост. вып.\) \[Текст\] // Программирование. - 2000. - № 3. - Вып. Тр. 10-й Всерос. науч. конф. «Электронные библиотеки: перспективы». - 2010. - № 2. - С. 2.](#)
24. [Вдовиченко В.Г., Лебедев В.А., Лысовский К.Ю., Сорокин А.Д., Старченко В.Г. Разработка и развитие технологии публикации и поиска документов в электронных коллекциях // Электронные библиотеки. ИРИО - 2003. - Том 6 - Выпуск 2.](#)
25. [Ковалевский М.Р., Палицына С.И. Семантическое структурирование контента научных электронных библиотек на основе онтологий // В сб. «Современные технологии в библиотечном деле». - 2009. - № 2. - С. 2.](#)
26. [Васильев А.В. Электронные научные информационные ресурсы для поддержки инвестиционной деятельности в регионе // Вдовиченко В.Т., Лысовский К.Ю., Сорокин А.Д., Резниченко В.А. - 2009. - № 2. - С. 2.](#)
27. [Ковалевский М.Р. Технологии XML и XML-данные. Базы данных и информационные технологии XXI в. // мат.-лы Междунар. науч. конф. 29-30 сентября 2003 г.](#)
28. [Иванов Е.Е. Построение системы защиты электронных библиотек от несанкционированного копирования документов // Труды Докладной Всероссийской научной конференции «Электронные библиотеки», 2007. С. 300-306.](#)
29. [Васильев А.В., Лысовский К.Ю., Старченко В.Г. Изменение концептуальных графов в системах поддержки электронных библиотек // Тр. 9-й Всерос. науч. конф. «Электронные библиотеки». - 2009. - № 2. - С. 2.](#)
30. [Васильев А.В., Гинзбург Л.Ю., Елизаров А.М. О доступе к электронным коллекциям в виде реляционных баз данных на основе онтологий // Труды 9-й Всерос. науч. конф. «Электронные библиотеки». - 2009. - № 2. - С. 2.](#)
31. [Лысовский К.Ю. Морализация семантики научно-технических текстов для АИС и его теоретические основы // Электронные библиотеки. ИРИО - 2003. - Том 6 - Выпуск 2.](#)
32. [Лысовский К.Ю. Тезаурусный анализ мировой культуры // сб. науч. трудов. Вып. 20 / под общ. ред. Вл. А. Лукова. - М.: Изд-во Моск. гуманит. ун-та, 2010. - 80 с.](#)
33. [Гладких А.Я. Применение тезауруса предметной области для повышения релевантности поиска в Интернете // А.Я. Гладких, Ю.В. Рогункина // Искусственный интеллект. - 2009. - № 2. - С. 2.](#)
34. [Найдионов Л.В. Технологии создания методов автоматического построения онтологий с применением семантического и автоматного программирования // Л.В. Найдионов. - 2009. - № 2. - С. 2.](#)
35. [Антопольский А.В., Дикунин Е.А., Манжарова Г.С. Проблемы и технологические проблемы создания и функционирования электронных библиотек. - М.: Патент, 2009.](#)
36. [Иванов Е.Е. Электронные гипертекстовые тезаурусы в дистанционном обучении // Открытое образование. № 1, 2002. С. 13-19.](#)
37. [Васильев А., Козлов Д., Савицкая С., Шамина О. Извлечение метainформации и библиографических ссылок из текстов русскоязычных научных статей // Труды конференции «Электронные библиотеки». - 2009. - № 2. - С. 2.](#)
38. [Воропеев А.Н. Электронная книга и электронно-библиотечные системы России \[Текст\] // \[Электронный ресурс\] : отраслевой доклад. / А.Н. Воропеев, К.Б. Леонтьев. - 2009.](#)

Ключевые слова: репозиторий, архив.

Документальный поиск: