

# Подход к созданию биографо-библиографической электронной библиотеки

Федотова О.А. (ГПНТБ СО РАН)

В докладе предлагается подход к созданию биографо-библиографической электронной библиотеки (ЭБ), основанный на известных моделях описания электронных библиотек. Модель разрабатывается для описания научных школ СО РАН.

## 1. Научные школы СО РАН

Изучение научного наследия основателей сибирских научных школ является важной задачей. Период их деятельности совпал со временем стремительного развития науки. Сегодня, спустя десятилетия после смены поколения, отчетливо вырисовывается истинное значение идей и событий, масштабность личностей. Среди сибирских ученых есть всемирно признанные, сделавшие весомый вклад в мировую науку.

Нашему поколению оставлено богатейшее научное наследие, которое следует беречь. Промедление в этой работе может привести к невозможным утратам, связанным с временным фактором: потерей документов, уходом из жизни очевидцев событий.

Одним из первоочередных способов решения задачи сохранения такого наследия может стать комплекс мероприятий, направленный на быструю и качественную публикацию материалов наследия в виде электронных библиотек в сети Интернет.

## 2. Электронные библиотеки

Электронной, или цифровой, библиотекой (ЭБ, или DL) называется структурированная каталогизированная коллекция разнородных электронных документов, снабженная средствами навигации и поиска.

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

Важным фактором при создании ЭБ является определение метаданных для описания ресурсов и выделение ключевых видов субъектов и объектов.

В существующих реализациях ЭБ как правило не решается проблема идентификации документов и субъектов, поскольку метаданные рассматриваются только для целей извлечения документов. *(Это может привести, например, к тому, что, встретив упоминание персоны в одном месте, невозможно точно установить соответствие с ее упоминанием в другом месте).*

На сегодняшний день не существует всеохватывающей модели ЭБ, которую можно было бы назвать эталонной. Главной задачей при проектировании ЭБ является построение хорошей модели, учитывающей мировой опыт подобных разработок.

В настоящей работе принято решение построить модель электронной библиотеки по научным школам СО РАН на основе международных рекомендаций к концептуальным моделям электронных библиотек: CIDOC CRM, DELOS DLRM.

На данном слайде представлена модель отношений, предложенная организацией DELOS. В нашей работе в соответствии с этой моделью информационный объект также описывается информационным ресурсом, который

- имеет уникальный идентификатор;
- организован в соответствии с *форматом ресурса* (схемой данных). Формат здесь выражен *онтологией*. Ресурс может быть сложным и структурированным, состоять из меньших ресурсов и иметь связи с другими ресурсами;
- может выражаться через *информационный объект*;
- должен быть снабжен *метаданными*, в том числе *аннотацией* и *биографическим описанием*, которые также являются информационными объектами.

Для формирования метаданных применяется стандарт Dublin Core, расширенный рекомендациями CIDOC CRM и DELOS DLRM. А для библиографического описания документов используется ГОСТ.

### 3. Типы сущностей ЭБ

Мы выделяем следующие основные типы сущностей, составляющие основу содержания электронной библиотеки:

1. Субъекты: *люди (персоны) и организации*.
2. Объекты – единицы хранения: *произведение, выражение, воплощение, экземпляр, факт, научный результат, мероприятие, фотография и др.*
3. Отношения: *понятие, событие, время и место*.

Субъекты и Объекты-хранения описываются схемами данных в соответствии с профилем ЭБ. Каждая единица хранения снабжается максимально подробным описанием в соответствии с ГОСТом, включая коды известных классификаторов (например, ББК, УДК, Current Contents, MSC2000), а также ссылками на *понятие* из словарей-классификаторов ЭБ.

При возможности дается ссылка на репозиторий (полнотекстовую БД). В отличие от общепринятых библиографических БД указание на субъекты: *персоны* (автор, редактор, персонаж и т.д.) и *организации* (институт, университет, издательство, журнал и т.д.) дается ссылкой на экземпляр сущности *субъект*, что позволяет корректно решать задачу идентификации *объектов*.

Помимо общепринятых описательных метаданных основные сущности ЭБ снабжены именованными отношениями.

Структура отношений в настоящий момент находится в стадии определения, но из главных отношений мы выделяем следующие:

- *Входит в состав* (ссылка: объект) - Данный ресурс является физически или логически

частью указанного ресурса.

- *Включает* (ссылка: объект) - Данный ресурс физически или логически включает указанный ресурс.
- *Работал* (ссылка: субъект; атрибут: время)
- *Преподавал* (ссылка: субъект; атрибут: время)
- *Изображен* (ссылка: объект; атрибут: время)
- *Ученик* (ссылка: субъект)
- *Автор* (ссылка: объект; атрибут: время)
- *Персонаж* (ссылка: объект; атрибут: время)

Наиболее важным для описания научной школы (НШ) являются отношения, связанные с предметной областью *понятиями* (концептами).

Для связи с предметной областью используется тезаурус. Понятия (словарь – онтология), которые имеют отношение к конкретной научной школе, строятся как подмножество тезауруса предметной области, дополненное словарями из предметных указателей монографий. Важным дополнением стандартного тезауруса являются списки (словари) фактов: научных достижений НШ с соответствующими отношениями.

Использование публикаций в научно-исследовательском процессе выдвигает необходимость быстрого ознакомления с содержанием публикации, и аннотация здесь может оказаться недостаточной. В связи с этим должны быть разработаны средства полуавтоматического выделения оглавления и выделения фактов (научных результатов в соответствии с онтологией, понятиями) с обеспечением ссылок на соответствующие разделы документа, а также средства работы с библиографическими ссылками.

#### **4. Сущность Персона**

Планируемая система управления ЭБ является *персоноцентричной*: все объекты и отношения, понятия, факты, мероприятия, публикации и др. привязываются к персонам.

Стандартная схема данных персоны расширяется биографическим описанием (или ссылками на биографические описания) и свойствами связывающие данную персону с НШ и основными понятиями из тезауруса НШ.

Основные свойства (атрибуты) структуры отношения персона-объект следующие:

- *Принадлежит* (школе) (атрибуты: время, место)
- *Участствует* (мероприятие, событие и т.п.) (атрибуты: время, место)
- *Имеет* (награды, звания и т.п.) (атрибуты: время, место)
- *Создатель* (публикация, результат, школа и т.п.) (атрибуты: время, место)
- *Действующее лицо* (персонаж) (атрибуты: время, место)

Основные свойства (атрибуты) структуры отношения персона-персона следующие:

- *Учитель*

- *Ученик*

Принадлежность персоны к научной школе определяется либо отношением учитель-ученик («генеалогическим деревом»), либо отношением совместной работы с основателем НШ или с другими членами НШ, определяемым *местом работы и научными результатами*.

Построение «генеалогического дерева» предполагается вести двумя способами:

- автоматически по базе данных авторефератов (учеником считается тот, у кого данная персона была руководителем кандидатской или консультантом по докторской диссертациями – естественно ученики ищутся не только у основателя школы, но и у его учеников);
- вручную, используя информацию «очевидцев».

## **5. Электронная библиотека по моделям динамики изменения биосферы**

Основные типы отношений между сущностями субъект-объект, представленные выше, были реализованы в «Электронной библиотеке по моделям динамики изменения биосферы».

Разработанная ЭБ содержит коллекцию «полнотекстовых» документов (публикации, презентации, файлы данных и др.) и следующие коллекции метаданных (таблицы):

1. Основной каталог документов, каждая запись в которой содержит метаописания документов, представленных в коллекции «полнотекстовых» документов.
2. Авторитетные файлы:
  - 2.1.Список авторов.
  - 2.2.Список организаций.
  - 2.3.Список журналов.
3. Словари-классификаторы:
  - 3.1.Вид документа (жанровый тип ресурса в соответствии с спецификацией DC).