

Интеграция библиотечных и архивных информационных систем

Антопольский А.Б. (НП ЭЛБИ)

Маркарова Т.С. (ГНПБ им К.Д.
Ушинского)

Исходные данные для разработки

- Единая тематическая область
- Организационное единство
- Общность пользователей
- Опыт автоматизации каталога библиотеки
- Отсутствие опыта автоматизации в архиве
- Различия в организации фондов, моделях данных, типологии документов, лингвистических средствах,

Постановка задачи

Разработка информационной системы,
интегрирующей электронный каталог
(ЭК) ГНПБ им К.Д. Ушинского и
Справочно-поисковый аппарат
Научного архива РАО

Цели интегрированной системы

- Ввести в активный научный оборот ценные материалы Научного архива РАО;
- Повысить эффективность информационного обслуживания пользователей архива и библиотеки;
- Автоматизировать процессы справочного информационного обслуживания в Научном архиве РАО;
- Производить поиск по одному запросу одновременно в СПА и ЭК;
- Заложить основы для создания единой автоматизированной технологии комплектования и информационного обслуживания в библиотеке и архиве;
- Автоматизировать подготовку статистической отчетности по Научному архиву РАО

Задачи разработки

- Создать единое лингвистическое обеспечение интегрированной системы на основе тезауруса ГНПБ
- Разработать структуру БД для СПА, совместимой с БД ЭК на основе АБИС «Ирбис»
- Разработать технологию ввода существующих данных СПА в БД (ретроспективная оцифровка описей)
- Разработать технологию дальнейшего ведения БД СПА при пополнении Научного архива и автоматизированного составления новых описей в БД СПА
- Разработать технологию дополнительного индексирования СПА
- Разработать технологию поиска в СПА
- Разработать технологию сквозного поиска в БД ЭК и БД СПА
- Разработать технологию ведения объединенного лингвистического обеспечения ЭК и СПА

Содержание фондов Научного архива РАО

- Научный архив РАО является уникальным собранием по истории отечественной педагогики и народного образования, начиная с 18-го века и до наших дней. Он содержит разнообразные материалы, характеризующие деятельность образовательных и научных учреждений, личные фонды крупнейших ученых в сфере педагогики и психологии, материалы съездов и конференций, а также фонд диссертаций, защищенных в научных учреждениях АПН и РАО

Структура архива

- Научный архив РАО включает около 140 фондов. Фонды состоят из дел, всего в совокупности во всех фондах содержится св. 100 тыс. дел
- Фонды имеют 3 разновидности, отличающиеся по источнику комплектования, составу документов и по структуре СПА:
 - Фонды организаций, в которых выделяются два подтипа – документы организаций и персональные данные.
 - Личные фонды
 - Архивные коллекции
- Для каждого фонда составлена одна или несколько описей, в зависимости от однородности дел в фонде. Некоторые новые фонды не имеют описей.
- Особая структура у диссертационного фонда (около 6 тыс. диссертаций). Опись отсутствует, но имеется ручная картотека, включающая минимум библиографических сведений (автор, наименование, место защиты, дата).

Справочно-поисковый аппарат

СПА реализован в виде описей. Для каждого фонда (кроме фонда диссертаций) составлена опись, иногда описей для одного фонда несколько. Всего СПА включает до 200 описей. Каждая опись включает;

- Титульный лист, содержащий номер и наименование фонда, номер описи и дату составления.
- Пояснительную записку (предисловие), в которой в свободной форме описывается содержание фонда, история его формирования, для личных фондов – приводятся биографические сведения о фондообразователе и другая информация. Может быть несколько отдельных предисловий.
- Оглавление, в котором перечисляются разделы описи, иногда снабженные цифровыми индексами с указанием первой страницы раздела описи или диапазона страниц.

Структура описи

- **№ дела**
- **Индекс дела**
- **Заголовок дела**
- **Дата**
- **Количество листов**
- **Примечание**

Классификации в СПА

Классификации дел могут иметь вид:

- Систематической классификации
- Хронологической
- Алфавитной (только для персональных данных)

Систематическая классификация используется для каждого фонда своя. Для фондов организаций характерными являются рубрики, именующие структурные подразделения организаций, для коллекций - содержание классификаций уникальное.

Общие принципы интегрированной системы

1. БД ЭК остается в неизменном виде, в соответствии с АБИС «ИРБИС»
2. Для СПА создается отдельная БД в «ИРБИС», состоящая из 2 связанных файлов:
 - Фонды и описи
 - Дела
3. Библиографические описания диссертаций вводятся в общий ЭК ГНПБ по стандартным правилам каталогизации.
4. Поиск в БД СПА отдельный, исходя из структуры данной БД.
5. При просмотре описания фонда выдаются идентификаторы всех дел фонда, при просмотре дела – описание фонда.

Сквозной поиск в БД ЭК и СПА

Сквозной поиск для БД СПА организуется

- по полю «Наименование дела» и «Рубрика» - лексический поиск по ненормированной лексике ;
- по наименованию организации - по нормированному словарю и (фасету тезауруса), но на первом этапе - просто показывая пользователю значения, имеющиеся в БД;
- по виду документов - по нормированному словарю
- по ФИО - показывая пользователю значения, имеющиеся в БД
- по тематической классификации (ББК) – поиск только фонда целиком
- по дескрипторам тезауруса – основной рекомендуемый режим предметного и тематического поиска

Лингвистическое обеспечение интегрированной системы

Основной инструмент сквозного поиска – тезаурус по педагогике и психологии

Рубрики систематических классификаций включаются в тезаурус

- 1) в виде дескрипторов, если рубрики дескриптороспособны
- 2) в виде аскрипторов, заиндексированных дескрипторами – существующими или новыми

Дополнительно формируются фасеты тезауруса

- 1) Виды архивных документов, объединенный со словарем видов документов из ЭК
- 2) Хронологический
- 3) наименования организаций со специальной парадигматикой, указывающей переименования, слияния и проч..
- 4) Словарь личных имен, объединенный с словарем значений полей «Автор» из ЭК.

Основные технологические операции создания БД СПА

- Оцифровка описей и ввод их в БД,
- Обработка систематических классификаций описей, и пополнение тезауруса ГНПБ
- Индексирование рубрик и ввод ПОДа в БД
- Создание формализованных описаний фондов и описей,
- Интеллектуальная обработка наименований дел, включая выделение корпоративных и личных имен, и определение вида документов дела,

Перспективы

- Представление тезауруса и нормированных словарей средствами OWL, т.е. создание единой онтологии по педагогике и психологии
- Представление форматов СПА и ЭК средствами DC XML
- Создание единой системы метаданных по тематической области в семантическом вебе