Лекция 5. Индексирование документов

План

1.Теоретические основы и методика индексирования и предметизации
2.Правила и основы систематизации
3. Виды классификационных систем – УДК и ББК
4. Определения информационно-поискового тезауруса

1.Теоретические основы и методика индексирования и предметизации

Индексирование — это процесс выражения содержания документа или информационного запроса на информационно-поисковом языке (ИПЯ) для целей систематизации и поиска. Оно включает присвоение документам индексов или предметных рубрик, которые отражают их тематическую принадлежность и содержание.

Предметизация — вид индексирования, при котором содержание документа выражается средствами языка предметных рубрик (ЯПР).

Предметная рубрика (ПР) — краткая словесная формулировка, отражающая основной предмет или аспект, рассматриваемый в документе.

Теоретические основы предметизации:

- В основе лежат две операции: выявление предмета документа и преобразование его в словесную формулировку.
- Предмет — всё, о чём в документе сообщается: материальные объекты, события, явления, факты, личности, организации и т. д…
- Цель предметизации — формирование предметных рубрик для раскрытия содержания фонда по предметному признаку и облегчения поиска документов.
- Язык предметных рубрик — ИПЯ, предназначенный для индексирования документов и запросов. Он должен обеспечивать однозначность толкования рубрик, устраняя синонимию, полисемию и омонимию.

Общая методика предметизации включает совокупность процессов, приёмов и правил, которые применяются при предметизации документов всех типов и видов, независимо от их тематики. Процесс включает:

1. Анализ содержания документа с использованием текста, титульного листа, аннотации, оглавления, предисловия, заключения, списка литературы, алфавитно-предметного указателя, глоссария.
2. Выбор основных предметов документа — смысловых компонентов, обозначающих его основной предмет.

3. Формулирование предметных рубрик на основе отобранных предметов и их аспектов.

При формулировке ПР обычно учитывают, что тема должна занимать не менее 20% текста документа. Чаще достаточно одной ПР, но может быть две-три.

2.Правила и основы систематизации

Систематизация — процесс присвоения документу классификационных индексов, позволяющих распределить документы или библиографическую запись по отделам или подразделениям документного фонда, систематическому каталогу, библиографическому указателю и т. д…

Теоретические основы систематизации:

- Основной принцип — приоритет содержания документа над другими его признаками (конструктивной формой, языком, читательским назначением и т. п.).
- Принцип объективности требует выявления основной сути документа и точного отражения его содержания с помощью правильного классификационного решения.
- Принцип последовательности предполагает единообразие классификационных решений для устранения разночтений.

Некоторые правила систематизации:

- Предпочтение частного вопроса общему: отделы строятся по принципу «от общего к частному».
- Систематизация по аспекту рассмотрения предмета, если документ посвящён предмету, который относится к одной области знания, но рассматривается в аспекте другой.
- Систематизация по области применения объекта, если в документе рассматриваются вопросы использования положений одной науки в других.
- Систематизация по объекту, испытывающему влияние, если документ содержит два объекта — принципы какой-либо науки и другие науки или отрасли, на которые они влияют.
- Систематизация документов широкого содержания, в которых рассматривается три и более предмета или предмет освещается в трёх и более аспектах.

Результатом систематизации является классификационный индекс, который определяет место документа в системе отраслей знания. Индексы могут быть полными, полочными (указывают место хранения при

систематической расстановке фонда) и каталожными (позволяют разместить библиографическую запись в соответствующем делении).

3. Виды классификационных систем – УДК и ББК

Иерархическая. Построена по принципу деления от общего к частному с использованием цифрового или буквенного кода. Каждый класс занимает фиксированное место в иерархии.

Фасетная. Объекты распределяются по параллельным независимым признакам (фасетам), например по характеру, регулярности поступления, стадии обработки.

Дескрипторная. Основана на использовании дескрипторов — ключевых слов, выбранных из группы условно эквивалентных терминов.

Универсальная десятичная классификация (УДК) — международная система классификации информации, используемая для систематизации произведений науки, литературы, периодики и других документов. Создана в 1895 году бельгийскими библиографами Полем Отле и Анри Лафонтеном на основе десятичной классификации Дьюи.

Основы и построение:

- Построена по иерархическому принципу с использованием цифрового десятичного кода.
- Основные таблицы охватывают всю совокупность знаний и делятся на классы, разделы, подразделы и т. д…
- Один и тот же предмет может встречаться в разных разделах в зависимости от отрасли знания и аспекта рассмотрения (многоаспектность).
- Используются общие и специальные определители (например, для обозначения языка, формы документа, места, времени).

Развитие:

На 2007 год УДК содержала 126 441 код.

Ведётся международным Консорциумом УДК. Эталонная база данных таблиц УДК ведётся на английском языке.

В России исключительным правом распоряжения таблицами УДК обладает Всероссийский институт научной и технической информации (ВИНИТИ).

Библиотечно-библиографическая классификация (ББК) — национальная классификационная система России в области библиотечного дела. Разработана в 1936–1968 годах специалистами Российской Государственной библиотеки, Российской Национальной библиотеки, Библиотеки Академии наук и Российской Книжной палаты.

Особенности:

1. Охватывает все отрасли знания (науки, дисциплины и соответствующие им сферы практической деятельности).
2. Предусмотрены возможности систематизации не только книг и брошюр, но и других видов документов и электронных ресурсов.

3. Обозначается латинскими буквами и цифрами, что упрощает визуальное восприятие разделов. Например, раздел «Искусство» обозначается буквой «Щ», а подкатегория «Живопись» — кодом «Щ1».

Структура

Таблицы ББК представлены совокупностью трёх составных частей:

1. Основные таблицы — раскрывают отрасли знания и их проблематику, помогают проанализировать содержание документа при его систематизации. Представлены в виде иерархического перечня классификационных делений.
2. Вспомогательные таблицы — отражают повторяющиеся классификационные признаки. Система вспомогательных (типовых) таблиц состоит из таблиц, используемых во всех или многих отделах классификации (таблицы общих типовых делений), и таблиц, обслуживающих отдельные отрасли (специальные типовые деления).
3. Алфавитно-предметный указатель — обязательная составная часть любого издания таблиц классификации, обеспечивает словесный вход в таблицы.

Варианты таблиц: полные, средние и сокращённые. Полные, средние и сокращённые таблицы согласованы между собой по структуре и индексации и различаются только по степени детализации.

Принцип

Многоступенчатая структура системы отражает принципы перехода от простого к сложному, от общего к частному, от низшего к высшему. Алфавитное расположение понятий применяется на наиболее глубоких ступенях деления.

Некоторые классификационные признаки: по объекту или предмету изучения, методу или аспекту рассмотрения объекта, территориальному признаку, по историческим периодам, хронологическим и другим временным понятиям, этническим и языковым признакам, целевому, читательскому назначению и форме издания документа.

ББК применяется в библиотеках для:

- организации систематической (по отраслям знания) расстановки книжных фондов;
- построения структуры карточных систематических каталогов и картотек;
- в качестве поискового инструмента в электронных каталогах;
- в целях библиотечной статистики.

Кроме того, ББК используется в библиографической и издательской практике.

Издательские стандарты включают индекс ББК в число основных элементов выходных сведений, определяя ему место на обороте титульного листа издания вверху слева отдельной строкой под индексом УДК.

4. Определения информационно-поискового тезауруса

Информационно-поисковый тезаурус (ИПТ) — контролируемый словарь терминов на естественном языке, явно указывающий отношения между терминами и предназначенный для информационного поиска.

Основные цели разработки ИПТ:

- обеспечение перевода естественного языка документов и запросов на контролируемый словарь;
- обеспечение последовательного использования единиц индексирования;
- описание отношений между терминами;
- использование как поискового средства при поиске документов.

Единицы ИПТ:

Дескриптор — авторизованный термин, выбранный из группы условно эквивалентных ключевых слов для представления данной группы при индексировании и поиске.

Аскриптор (недескриптор) — вспомогательные термины из синонимического ряда, помогающие найти подходящие дескрипторы.

Отношения в ИПТ могут быть иерархическими (род — вид, часть — целое) и ассоциативными (причина — следствие, функциональное сходство и т. д.).
ИПТ используются в автоматизированных системах поиска, для унификации терминологии в отраслевых базах данных и при обучении систем машинного поиска информации.

Список использованных источников:

1.Меркулова, А. Ш. Аналитико-синтетическая переработка информации: Систематизация документов [Текст]: учебно-методическое пособие для обучающихся по направлению подготовки 51.03.06 «Библиотечноинформационная деятельность», квалификация (степень) выпускника «бакалавр» / А. Ш. Меркулова. – Кемерово: КемГИК, 2018. – 107 с. ISBN 978-5-8154-0439-7

2. Учебные и научные издания. Библиографический список : метод. рекомендации / сост. Н. П. Лиханская, Е. А. Хвостова, А. А.

3. Панкова, Е. В. Электронное библиографическое пособие : практическое руководство для библиотечных работников / Е. В. Панкова, Л. С. Беркутова. – Москва : ФОРУМ, 2025 г.. -128 с.