ДИСЦИПЛИНА: Аналитико-синтетическая переработка информации

Лекция №11

1. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ИНДЕКСИРОВАНИЯ ДОКУМЕНТОВ

План лекции:

1. Общие положения индексирования.
2. Требования к информационно-поисковым языкам.
3. Общие правила индексирования.

1. ОБЩИЕ ПОЛОЖЕНИЯ ИНДЕКСИРОВАНИЯ

Одной из важнейших операций аналитико-синтетической обработки документов как средства информационного поиска является индексирование документов и информационных запросов.

Индексирование – процесс выражения содержания документа и (или) запроса на информационно-поисковом языке (ИПЯ) с помощью терминов индексирования (классификационных индексов, предметных рубрик (ПР), ключевых слов, дескрипторов, кодов). Иными словами, индексирование – процесс перевода содержания документов и запросов с естественного языка на ИПЯ, в результате чего создается поисковый образ документа (ПОД) и поисковый образ запроса (ПОЗ). Таким образом, происходит «свертывание» информации, содержащейся в документе и изложение ее на ИПЯ в виде индекса, рубрики, кода (на классификационном языке) или дескриптора, ключевого слова (на дескрипторном языке).

Поисковый образ документа – это выраженное в терминах формализованного информационно-поискового языка основное смысловое содержание документа (а не вся информация, содержащаяся в документе), которое поставлено в однозначное соответствие этому документу, по которому производится отыскание данного документа в массиве других документов.

Индексирование запроса осуществляется путем перевода его содержания на информационно-поисковый язык. Совокупность терминов индексирования, выражающих смысловое содержание запроса, называется **поисковым образом запроса (ПОЗ). В** этом случае поиск документов ведется по совокупности терминов индексирования. Чтобы повысить эффективность поиска информации по запросу, поисковый образ запроса можно дополнить специальными указаниями о последовательности выполнения логических операций в процессе информационного поиска, которые являются **поисковым предписанием.**

Методика индексирования запросов за счет включение в поисковое предписание дополнительных терминов, понятий или дополнительных указаний позволяет расширить область поиска информации (например, за счет включения видовых и ассоциативных терминов) и в конечном итоге повысить функциональную возможность информационно-поисковой системы.

Индексирование реализуется в следующих процессах: *предметизации, систематизации и координатном индексировании.*

Систематизация – вид индексирования, при котором содержание документа и (или) запроса выражено классификационными индексами в соответствии с правилами какого-либо классификационного ИПЯ (классификационной системы). Такой принцип индексирования принято считать классификационным. Классификационный принцип индексирования обеспечивает возможность организации информационного поиска по иерархическому признаку.

Предметизация – вид индексирования, при котором содержание документа и (или) запроса выражено предметной рубрикой (ПР) в соответствии с правилами соответствующего предметизационного ИПЯ. Предметизационный принцип индексирования базируется на представлении содержания документа в понятийной системе определенного естественного языка и на использовании в качестве терминов индексирования лексических единиц (ЛЕ) естественного языка.

Предметизационный принцип индексирования обеспечивает возможность организации информационного поиска по алфавитному признаку.

Координатное индексирование – вид индексирования, при котором смысловое содержание документа и (или) запроса многоаспектно выражается множеством ключевых слов или дескрипторов. Информационно-поисковый язык, предназначенный для координатного индексирования документов (или частей документов) и запросов посредством ключевых слов и (или) дескрипторов, называется дескрипторным языком.

Термины «индексирование» и «индексатор» распространены в англоязычной литературе и широко применяются в международной и отечественной библиотечной практике.

В зависимости от того, какой принцип индексирования является ведущим (или единственным) различают **предметизационные, классификационные** и **дескрипторные** информационно-поисковые языки. В соответствии с характером ИПЯ индексирование осуществляется как предметизация, систематизация и координатное индексирование.

А в качестве терминов этих видов индексирования используют предметные рубрики, классификационные индексы, дескрипторы, ключевые слова.

В зависимости от типа используемого ИПЯ, различают два режима индексирования: **предкоординатное** (классификационное) и **посткоординатное.**

В предкоординатном индексировании связи между терминами индексирования устанавливаются заранее – при разработке соответствующего ИПЯ. К предкоординатным ИПЯ можно отнести следующие:

— перечислительные классификации;
— иерархические классификации;
— предметные классификации;
— фасетные классификации.

В посткоординатном индексировании связи (координация) между терминами индексирования устанавливаются во время индексирования документов или в процессе их поиска. К посткоординируемым ИПЯ относят:

— дескрипторные языки;
— посткоординируемые фасетные классификации.

Теоретические основы и методика процессов индексирования (предметизации, систематизации, координатного индексирования) подробно изложены в последующих главах методического пособия.

2. ТРЕБОВАНИЯ К ИНФОРМАЦИОННО-ПОИСКОВЫМ ЯЗЫКАМ

В состав ИПЯ входят множество **лексических единиц** (ЛЕ) и грамматические (парадигматические и синтагматические) отношения между ними.

ЛЕ – обозначение отдельного понятия, принятое в информационно-поисковом языке и неделимое в этой функции. Лексические единицы могут представлять собой принятые в естественном языке слова, устойчивые словосочетания, аббревиатуры, символы, даты, общепринятые сокращения, лексически значимые компоненты сложных слов, а также эквивалентные им кодовые или символические обозначения искусственного языка. Грамматические отношения позволяют организовать ЛЕ в систему.

Элементы ИПЯ могут быть реализованы в любой форме: машиночитаемой, карточной, на микроносителях, в форме книжного издания.

Информационно-поисковый язык должен:

— полно и точно передавать содержание (при необходимости – форму и назначение) документа, отражаемого в данной ИПС;
— обеспечить однозначное толкование терминов индексирования;
— допускать многоаспектное индексирование;
— допускать внесение изменений (дополнений, исправлений);
— обеспечивать простоту и удобство индексирования, информационного поиска и ведение данной ИПС;

— отражать современное состояние терминосистемы в данной области знания.

Правила использования ИПЯ фиксируют в справочно-методическом аппарате соответствующей ИПС.

3. ОБЩИЕ ПРАВИЛА ИНДЕКСИРОВАНИЯ

Сущность индексирования заключается в том, что в процессе индексирования осуществляется интеллектуальный анализ документов, в процессе которого из них извлекаются понятия и преобразуются в термины индексирования. Анализ и преобразование этих понятий производится с помощью таких средств индексирования, как системы классификации и тезаурусы.

Индексирование состоит из следующих взаимосвязанных этапов:

— анализа содержания документа как объекта индексирования;
— выявления и отбора смысловых компонентов в содержании документа;
— принятия решения о составе ПОД;
— оформления отобранных смысловых компонентов как понятий в терминах индексирования в соответствии с системой грамматических средств данного ИПЯ;
— редактирования терминов индексирования в составе ПОД.

Анализ содержания документа должен проводиться по всем элементам, которые позволяют выявить его содержание, идеологическую направленность, форму, назначение. Для этого тщательно изучается и оценивается информативность всех компонентов документов: заглавия, реферата, содержания, введения, начальных предложений глав и параграфов, заключения и т. д.

При изучении самого документа анализ его содержания проводится по заранее разработанному алгоритму (это зависит от вида индексирования). Одновременно выявляются и отбираются смысловые компоненты в содержании документа, выделяются компоненты, отражающие предмет (предметы) и аспект (аспекты) его рассмотрения. Алгоритм анализа и выявления и отбора смысловых компонентов содержания документов подробно описан в соответствующих главах. В процессе анализа можно использовать справочный аппарат, а также консультации отраслевых специалистов или специалистов-переводчиков. Например, по узкому вопросу, связанному с определенной предметной областью или по языку текста документа и др.

Индексирование – основное средство раскрытия содержания документа и соответственно документальных потоков. От качества индексирования зависят и эффективность поиска в информационных ресурсах библиотек и эффективность использования их фондов.

Принятие решения о составе ПОД и оформление терминов индексирования проводятся в соответствии с правилами ИПЯ, а также с необходимой и достаточной для данной ИПС глубиной, точностью и многоаспектностью.

Глубина индексирования характеризует полноту раскрытия содержания документа (запроса) в присвоенном ему поисковом образе и выражается степенью детализации классификационных индексов, предметных рубрик (в классификационном ИПЯ) или количеством дескрипторов (в дескрипторном ИПЯ). Чем полнее и глубже раскрыто содержание документа, тем точнее можно осуществить информационный поиск, позволяющий удовлетворить самые разнообразные запросы пользователя.

Точность (адекватность) индексирования характеризует степень отражения содержания документа (запроса) в поисковом образе и степень соответствия классификационного индекса, предметной рубрики или дескриптора, ключевого слова правилам применяемого ИПЯ, т. е. классификационной системе.

Многоаспектность индексирования характеризует степень отражения в классификационном индексе, предметной рубрике, дескрипторе, ключевом слове аспектов (точек зрения), признанных с позиций поиска существенными. Многоаспектность показывает, насколько всесторонне раскрыто содержание документа.

Принятое при индексировании решение фиксируется в БЗ документа (на каталожных карточках или в машиночитаемой форме). Термины централизованного индексирования (классификационные индексы или предметные рубрики) можно дополнять, уточнять или изменять.

Процесс редактирования терминов индексирования заключается в проверке правильности проведенного анализа документа, соответствия отобранных смысловых компонентов его содержанию. Кроме того, оцениваются глубина, точность и многоаспектность решения, его соответствие принятой в библиотеке (организации, учреждении) методике индексирования, техника оформления решения, то есть правильность всех элементов терминов индексирования.

Методические решения типового характера, а также решения по сложным случаям индексирования должны быть зафиксированы по правилам, принятым в конкретной ИПС.

В процессе индексирования прогнозируется будущий запрос потребителя, решение принимается с учетом возможных путей поиска документа.

Под методикой индексирования подразумевается совокупность приемов и правил образования поисковых образов документов или поисковых образцов запросов, то есть приемов и правил образования

индексов, предметных рубрик, дескрипторов, ключевых слов для понятий, отражающих основное содержание документа или запроса.

Методика индексирования делится на ***общую*** и ***частную. Общая методика*** включает принципы и правила, имеющие общий характер, то есть касающиеся документов любого содержания, независимо от отрасли знания и формы документа. ***Частная методика*** индексирования содержит принципы и правила, которые применяются в определенной отрасли знания или для определенных видов документов. Учет специальных потребностей определенной организации, системы библиотек и т. д. можно осуществлять с помощью ***специализированной методики*** индексирования.

Методика отдельных видов индексирования (предметизации, систематизации, координатного индексирования) имеет общие и специфические черты. Общие выражены общими правилами общей методики, а специфические – с привлечением правил частной методики, детально изложенных в конкретных главах данного учебника.

ИСПОЛЬЗОВАННЫЕ ИСТОЧНИКИ:

Зупарова Л. Б.
Аналитико-синтетическая переработка информации: Учебник / Л. Б. Зупарова, Т. А. Зайцева; под ред. Ю. Н. Столярова. - М.: «Издательство ФАИР», 2007. – 400 с. – (Специальный издательский проект для библиотек).