

**ИНДЕКСИРОВАНИЕ ДОКУМЕНТОВ:
ОПРЕДЕЛЕНИЕ, ОСНОВНЫЕ
ПРОЦЕССЫ**

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, предметной рубрики или дескриптора, ключевого слова

Информационно-поисковый язык (ИПЯ) –это специально созданный искусственный язык, предназначенный для выражения содержания документов и (или) запросов с целью их последующего поиска. ИПЯ – это основной элемент логико-семантического аппарата информационно-поисковой системы (ИПС)

Основные требования, предъявляемые к ИПЯ:

- Однозначность;
- Достаточная семантическая сила
- Открытость (возможность корректировки языка)

Каждый ИПЯ имеет определенный словарный состав, представляющий **совокупность лексических единиц (ЛЕ)** – обозначения отдельного понятия, минимального и неделимого в этой функции. В качестве ЛЕ в ИПЯ используется лексика естественных языков – слова, словосочетания предметной рубрики, цифровые или буквенно-цифровые коды и т.д.

Важным моментом при индексировании является создание поискового образа документов (ПОД). **Поисковый образ документа** – основное смысловое содержание документа (а не вся информация, содержащаяся в нем), выраженное в терминах формализованного ИПЯ. ПОД ставится в однозначное соответствие этому документу, по нему производится отыскание документов в массиве документов.

Индексирование запроса также осуществляется путем перевода его содержания на ИПЯ. Таким образом, совокупность терминов индексирования, выражающих смысловое содержание запроса, называется **ПОИСКОВЫМ образом запроса (ПОЗ)**.

Поиск ведется по совокупности терминов индексирования. С целью повышения эффективности поиска информации по запросу ПОЗ может быть дополнен специальными указаниями о последовательности выполнения логических операций в процессе информационного поиска, которые называются **поисковым предписанием**.

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 определенного классификационного ИПЯ (КС). Такой принцип индексирования принято считать классификационным. Он обеспечивает возможность организации информационного поиска по иерархическому признаку. Специалисты, осуществляющие процесс систематизации в библиотеках и информационных центрах, называются систематизаторами

Предметизация— вид индексирования, при котором содержание документа и (или) запроса выражено предметной рубрикой (ПР) в соответствии с правилами определенного предметизационного ИПЯ. Предметизационный принцип индексирования основан на использовании ЛЕ, которые входят в состав предметных рубрик, естественного языка (т.е. предметные рубрики, выражающие ПОД и ПОЗ составляются на основе естественного языка). Предметизационный принцип индексирования обеспечивает возможность организации информационного поиска по алфавитному признаку. Процессы предметизации осуществляются в библиотеках и информационных центрах предметизаторами.

Координатное индексирование —вид индексирования, при котором смысловое содержание документа и (или) запроса многоаспектно выражаются множеством ключевых слов или дескрипторов. ИПЯ, предназначенный для координатного индексирования, называется дескрипторным языком. Специалисты, осуществляющие координатное индексирование, называются индексаторами.

Итак, в зависимости от вида индексирования, выделяются предметизационные, классификационные и дескрипторные ИПЯ. В их состав входит множество ЛЕ и грамматические (парадигматические и синтагматические) отношения между ними. ЛЕ – обозначение отдельного понятия, принятое в ИПЯ и неделимое в этой функции. ЛЕ 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.

Требования к ИПЯ:

- Полно и точно передавать содержание документа, отражаемого в данной ИПС;
- Обеспечить однозначное толкования 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.

Необходимость создания искусственного языка (для выражения смыслового – семантического содержания документов с целью их поиска) обусловлена тем, что естественный язык обладает рядом свойств, препятствующих его использованию для записи и поиска информации: это неоднозначность и многозначность слов естественного языка. Точное значение многих слов можно определить только из контекста, в котором они употреблены (ударный инструмент, ударная доза и т.д.). Эти факторы не позволяют добиться точного соответствия между содержанием документа и средствами выражения этого содержания.

Трудности использования естественного языка в качестве ИПЯ усиливаются еще и тем, что в ИПС в качестве входных документов могут быть использованы не полные тексты, а рефераты, аннотации, библиографические описания, которые являются результатом свертывания содержания документов.