

Новосибирский государственный
университет экономики и управления



Координатное индексирование документов в библиотеке НГУЭУ

Докладчик: Чумова Н. Ю.

СЕКЦИЯ БИБЛИОГРАФИИ

НБ НГТУ

Новосибирск
13 марта 2007 г.

1. Общие положения

Для обеспечения эффективного информационного поиска в электронном каталоге нашей библиотеки используется индексирование при помощи ключевых слов.

Индексирование – выражение содержания документа и/или смысла информационного запроса на информационно-поисковом языке.

Координатное индексирование – индексирование, цель которого состоит во всестороннем отражении содержания документа или запроса путем включения в поисковый образ всех необходимых для этого терминов индексирования.

Ключевое слово – информативное слово, приведенное к стандартной лексикографической форме и используемое для координатного индексирования.

Ключевыми словами называются также полнзначные слова, устойчивые сложные и сложносокращенные слова и терминологические словосочетания, несущие существенную смысловую нагрузку в текстах документов соответствующих отраслей знания.

При индексировании представляется с необходимой и достаточной полнотой и точностью в поисковом образе документа (ПОД) в виде терминов индексирования основное содержание документа, а при необходимости — его форма и назначение для обеспечения эффективного информационного поиска.

Поисковый образ документа (ПОД) – это поисковый образ, выражающий основное смысловое содержание документа.

Информационно-поисковый язык (ИПЯ) – формализованный искусственный язык, предназначенный для индексирования документов, информационных запросов и описания фактов с целью последующего хранения и поиска.

Процесс индексирования включает в себя следующие взаимосвязанные процессы:

1. Анализ содержания документа;
2. Выявление и отбор смысловых компонентов в содержании документа (выбор понятий, характеризующих содержание документа);
3. Формирование перечня ключевых слов, используемого для создания поискового образа документа (принятие решения о составе ПОДа);
4. Нормализация ключевых слов по форме и содержанию (оформление отобранных смысловых компонентов как понятий в терминах индексирования в соответствии с системой грамматических средств данного ИПЯ);
5. Избыточное индексирование;
6. Заполнение поля ПОДа «Ключевые слова».

2. Формирование перечня ключевых слов

В результате анализа содержания документа определяется набор информативных слов, которые используются для формирования перечня ключевых слов, необходимого для создания поискового образа документа.

В качестве ключевых слов при координатном индексировании используются лексические единицы (Обозначение отдельного понятия, принятое в системе естественного или специально разработанного искусственного языка и далее в этой функции неделимое).

Лексические единицы могут представлять собой слова, устойчивые словосочетания, аббревиатуры, символы, даты, общепринятые сокращения, а также лексически значимые компоненты сложных слов, принятые в естественном языке, и эквивалентные им кодовые или символические обозначения искусственного языка, например коды классов классификационной системы.

В нашей библиотеке в качестве ключевых слов используются общепринятые термины, отраженные в словарях и энциклопедиях, классификационных таблицах ББК, УДК и др., а также в словарях ключевых слов, предметных рубрик, имеющих в электронном каталоге библиотеки (в ПО «Библиотека-4.0» был внутренний словарь (из которого можно было копировать ключевые слова и вносить их в библиографическую запись), и в Virtua тоже есть – это список предметов, который выходит при поиске по полю «Предмет»).

Методические решения по сложным случаям индексирования фиксируются в картотеке методических решений НБО.

Координатное индексирование документов в нашей библиотеке производится в соответствии с утвержденной Инструкцией.

При формировании перечня ключевых слов не допускается использование аббревиатур и сокращений слов, не предусмотренных ГОСТом.

Выбранные из текста информативные слова, применяемые в качестве ключевых, при необходимости, дополняются, уточняются, изменяются. Существуют определенные правила приведения ключевых слов к стандартной лексикографической форме:

Прилагательные не считаются самостоятельными ключевыми словами, они включаются в словосочетания:

а) когда характеризуют существительное, например,

биржевые операции;

б) когда прилагательное является неотъемлемой частью устойчивого словосочетания, например,

магнитные бури, производственные отношения

Сочетания существительного с существительным даются, если это широко распространенные и устойчивые, часто употребляемые словосочетания, например,

источники света, средства связи.

Словосочетание из одного существительного и нескольких прилагательных преобразуется в несколько двухсловных ключевых слов, в которых одно и то же существительное сопровождается поочередно одним прилагательным, например,

мировые финансовые кризисы = мировые кризисы, финансовые кризисы.

В сочетаниях существительных с прилагательными инверсия не применяется.

В словосочетаниях числительного с существительным применяется инверсия. Порядковые числительные инверсируются тогда, когда они обозначают очередность явления или события при последовательном их повторении, например,

Олимпийские игры, 18-е

3. Нормализация ключевых слов по форме и содержанию

Каждое ключевое слово необходимо правильно сформулировать.

Язык ключевых слов является условным формализованным языком с использованием слов естественного языка и арабских цифр.

При формулировке ключевых слов используются как русские, так и иностранные термины.

Ключевое слово может быть выражено одним словом или словосочетанием. Для этого используют существительные, прилагательные, причастия.

Словосочетания, как правило, используют только те, которые представляют собой термины или другие нерасторжимые сочетания слов, например,

деловые игры, красная строка

Словосочетания выступают в качестве ключевых слов в следующих случаях:

- а) если они обозначают названия различных законов, методов, теорий:

единство и борьба противоположностей

б) названия научной дисциплины или раздела науки:

экономика промышленности

в) названия различных типов организаций, партий, документов:

Организация Объединенных Наций

г) если один из компонентов словосочетания имеет слишком широкое терминологическое значение и потому целесообразно использовать его в качестве самостоятельного слова ввиду его неинформативности:

политические партии, математическое обеспечение

д) если в состав словосочетаний входит имя собственное:

Ортеги гипотеза

Ключевыми словами могут быть:

- Имя собственное:

имена ученых, писателей, общественных деятелей, например,

Лермонтов М. Ю., Шаляпин Ф. И.

административно-территориальные и географические наименования, например,

Сибирь, Западная Сибирь, Томская область, Россия

названия исторических событий, которые приводятся в полной форме в соответствии с научной традицией, например,

Великая Отечественная война, Бородинское сражение

- Хронологические данные. Все даты записываются арабскими цифрами, годы пишутся без «г», века пишутся с «в», например,

XX век=20 в.,

1812 год=1812

Слова, используемые для написания ключевых слов, как правило, формулируются в именительном падеже и во множественном числе, например,

Автомобили, Антибиотики, Сеялки

Исключение представляют термины, которые не употребляются во множественном числе, например,

Дыхание, Транспорт, Эффективность

В единственном числе в формулировке ключевых слов употребляются названия отдельных учреждений, организаций, индивидуальных предметов, собственных имен и т. д., например,

Ломоносов М. В., Никитинский ботанический сад, Солнце

При формулировке ключевых слов большое значение имеет употребление полной или краткой формы наименования предмета. Как правило, отдается предпочтение полной форме, например,

Министерство внутренних дел, а не МВД

Но если краткая форма вытеснила полное наименование предмета, то используется краткая, например,

ЮНЕСКО

Чаще же всего словосочетания употребляются в полной и сокращенной форме, например,

Байкало-Амурская магистраль (БАМ)

Универсальная десятичная классификация (УДК)

4. Избыточное индексирование

В состав ИПЯ входят множество ЛЕ и грамматические (парадигматические и синтагматические) отношения между ними. Грамматические отношения позволяют организовать ЛЕ в систему.

Парадигматические отношения – это родовидовые (иерархические) отношения.

На этапе избыточного индексирования происходит дополнение ПОДа лексическими единицами ИПЯ, связанными парадигматическими отношениями с лексическими единицами исходного ПОДа.

Избыточное индексирование осуществляется путем включения в ПОД близких по смыслу лексических единиц ИПЯ для повышения полноты и качества поиска.

Для удобства поиска документа в ЭК допускается дублирование ключевых слов, например,

Анализ хозяйственной деятельности

АХД

Внешиэкономическая деятельность

ВЭД

Планирование

Налоговое планирование

5. Порядок заполнения поля «Ключевые слова»

Когда в нашей библиотеке была внедрена АБИС «Библиотека-4.0», при заполнении шаблона библиографической записи использовалось поле 650 – предметная рубрика, подрубрика. Потом было принято решение отказаться от заполнения данного поля, оставив поле 653 – ключевые слова, - но вносить в него те же данные – рубрику, подрубрику, ключевые слова. Т. е. ключевые слова приводятся в ЭК нашей библиотеки в порядке от общего к частному.

По Инструкции ключевые слова вносятся в поле «Ключевые слова» в любом порядке, но стремимся выстраивать иерархические отношения, чтобы прослеживалась логика.

Количество ключевых слов определяется не их числом, а необходимостью возможно более полного описания всех важных аспектов содержания индексируемых документов. Опытом установлено, что для достижения этой цели достаточно 10-12 слов, но порой вводится намного больше.

653 поле MARC-записи в ПО Virtua – повторяющееся поле. Копируется метка поля (в столбик), и ключевые слова записываются в поле строчными буквами.

Не употребляются знаки препинания: запятые, точки, скобки, кавычки и др. Исключение составляют запятые при заполнении хронологических данных, например,

Олимпийские игры, 18-е

и скобки, если необходимо устранить многозначность (полисемию) одного и того же ключевого слова, например,

Нефрит (геол.)

Нефрит (мед.)

Нью-Йорк (город)

Нью-Йорк (штат)

Все ключевые слова и словосочетания даются полностью, без сокращений.

В библиографических записях на статьи мы даем ключевое слово – две последние цифры года выхода статьи (06, 07 и т. д.) для поиска статей по году издания.

Список литературы

1. ИНСТРУКЦИЯ ПО КООРДИНАТНОМУ ИНДЕКСИРОВАНИЮ В ЭЛЕКТРОННОМ КАТАЛОГЕ БИБЛИОТЕКИ НГУЭУ.
2. ГОСТ 7.59-2003 ИНДЕКСИРОВАНИЕ ДОКУМЕНТОВ. ОБЩИЕ ТРЕБОВАНИЯ К СИСТЕМАТИЗАЦИИ И ПРЕДМЕТИЗАЦИИ.
3. ГОСТ 7.66-92 (ИСО 5963-85) ИНДЕКСИРОВАНИЕ ДОКУМЕНТОВ. ОБЩИЕ ТРЕБОВАНИЯ К КООРДИНАТНОМУ ИНДЕКСИРОВАНИЮ.
4. ГОСТ 7.74-96 ИНФОРМАЦИОННО-ПОИСКОВЫЕ ЯЗЫКИ. ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ.