

Корпоративная каталогизация
Методические рекомендации
по формированию ключевых слов (КС)

Введение

"Ключевое слово — это информативное слово, приведенное к стандартной лексикографической форме и используемое для координатного индексирования"

"Координатное индексирование — это индексирование, предусматривающее многоаспектное выражение основного смыслового содержания документа или смыслового содержания информационного запроса множеством ключевых слов или дескрипторов". Совокупность ключевых слов составляет поисковый образ документа (ПОД).

Общие методические рекомендации

В качестве КС при свободном индексировании используются лексические единицы (ЛЕ) представляющие слова (унитермы), терминологические словосочетания, фразы (в исключительных случаях), аббревиатуры, численные характеристики, хронологические данные, имена собственные, символические обозначения.

В качестве КС не рекомендуется использовать:

- 1) служебные слова (предлоги, союзы, частицы, междометия);
- 2) часто встречающиеся, но весьма общие термины (например, «описание», «устройство», «конструкция», «метод», «система») следует либо исключить, либо употреблять в сочетании с другими словами, сужающими их значение. Например: Принцип свободного падения;
Уравнение турбулентной диффузии.
- 3) полисемичные термины включаются только в тех значениях, в которых они употребляются в данной тематической области, с соответствующими пояснениями. Например: лук (спорт)

При составлении ключевых слов целесообразно ориентироваться на использование единичных ключевых слов. Использование устойчивых словосочетаний допускается в тех случаях, когда они удовлетворяют приведенным ниже лингвистическим или прагматическим критериям распознавания устойчивых словосочетаний.

Правила распознавания устойчивых словосочетаний

Правило	Пример
1. Словосочетание является устойчивым, если над ним без потери смысла нельзя провести ни одной из следующих операций: а) заменить составляющие его слова синонимами; б) заменить прилагательное существительным с предлогом; в) заменить основное слово его производным; г) изменить порядок слов в словосочетании.	ни одной из этих операций нельзя произвести над словосочетанием «железная дорога»
2. Словосочетание является устойчивым, если при его образовании одно из слов изменяет свое значение, т.е. происходит переосмысление одного из компонентов словосочетания.	«легкая промышленность» (переосмысливается прилагательное «легкая»); «земная ось» (переосмысливается существительное «ось»)
3. Словосочетание можно считать устойчивым, если оно употребляется только в единственном числе, либо преимущественно во множественном числе.	«автоматический перевод», «термореактивная бумага», «библиотечное дело», «английский язык», «восточные языки», «строительные материалы», «продукты питания».
4. Словосочетание является устойчивым, если оно имеет один или несколько синонимов.	«стеклоотливные машины» – «линотипы»;

	«полупроводниковые триоды» - «транзисторы»; «суперпозиционные перфокарты» – «просветные карты» – «визуальные перфокарты»; «символическая логика» – «математическая логика» – «алгебра логики»
5. Устойчивыми являются словосочетания, представляющие собой имена собственные или включающие собственные имена в качестве компонентов.	«Новая Гвинея», «Латинская Америка», «таблица Менделеева», «эффект Доплера», «маятник Фуко»

Словосочетание является устойчивым, если оно удовлетворяет хотя бы одному из перечисленных в таблице критериев.

Следует разделять словосочетания,

- если ЛЕ - словосочетание объединяют следующие понятия
 - процесс и объект, который подвергается воздействию (производство азота- производство, азот)
 - часть и целое (катоды радиоламп- катоды, радиолампы)
 - объект и его назначение (станки для распиловки карандашей - станки, распиловка, карандаши);
- если ЛЕ - словосочетание объединяет два существительных (история России- история, Россия; искусство Франции - искусство, Франция);
- если ЛЕ - содержит несколько прилагательных - преобразуется в несколько двухсловных ключевых слов в которых одно и то же существительное сопровождается поочередно одним прилагательным (торированные вольфрамовые катоды - торированные катоды, вольфрамовые катоды)

Прагматические критерии распознавания устойчивых словосочетаний:

- 1) **Следует сохранять словосочетание**, если в используемом словаре ключевых слов очень мала вероятность использования его компонентов по отдельности;
- 2) **Следует сохранять** часто встречающиеся словосочетания, разделение которых в принципе возможно, но неэкономично с точки зрения ввода в автоматизированную информационную систему.

Правила написания ключевых слов

Существительные, прилагательные и причастия записываются в именительном падеже и множественном числе, кроме некоторых случаев.

При выборе между формами единственного и множественного числа рекомендуется придерживаться правил, представленных в таблице.

Использование единственного и множественного числа при нормализации ключевых слов

Тип термина	Критерий выбора числа	
	Единственного	множественного
1. Термины, обозначающие субстанции, вещества (химические соединения, смеси, материалы).	Если термин выражает видовое (единичное понятие). Например: хлоратан, целлофан.	Если термин выражает родовое (общее понятие). Например: амины, растворители.
2. Термины, обозначающие свойства, условия, характеристики.	Если термин выражает видовое (единичное понятие). Например: мощность, напряжение, частота.	Если термин выражает родовое (общее понятие). Например: условия залегания, физические свойства.
3. Термины, служащие для обозначения приборов, оборудования.	Не применяется	Применяется всегда. Например: аккумуляторы, распылители.
4. Термины, обозначающие процессы.	Применяется всегда. Например: поиск, хранение, чтение.	Не применяется
5. Термины, в состав которых входят имена собственные.	В зависимости от смысла	
	Теорема Бернулли	Законы Ньютона
6. Наименование научных дисциплин.	Применяется всегда. Например: гидравлика, физика, биология.	Не применяется

7. Термины, обозначающие исчисляемые события и явления.	Не применяется	Применяется всегда. Например: разряды, взрывы, войны.
---	----------------	--

Общее правило использования единственного и множественного числа заключается в том, что форма множественного числа употребляется для исчисляемых понятий, а форма единственного числа – для неисчисляемых понятий.

В тех случаях, когда существует смысловое различие в употреблении терминов в единственном или множественном числе, следует использовать обе формы:

Например: гражданское право
гражданские права
игра
игры

В одном случае речь идет об отрасли права, в другом — о правах граждан и т. д. Иногда один и тот же термин в одних отраслях знания употребляется только в единственном числе, в других — только во множественном:

Например: вода.
но:
подземные воды.
сточные воды.

Прилагательные и причастия записываются в форме именительного падежа множественного числа мужского рода (например, ручные, автоматизированные, информационные).

Прилагательные, наречия, числительные, местоимения, которые сами по себе могут выступать в роли КС, предлагается употреблять только в словосочетаниях.

Например:

При определении порядка слов в словосочетании для обеспечения экономичности и единообразия формулировок ключевых слов рекомендуется использовать прямую (неинвертированную) форму записи (например, информационные технологии; библиотечные системы автоматизированные). Исключение составляют только те словосочетания в которых одно из слов является личным именем.

Например: Ома закон
Благоева группа

Кроме перечисленных общих правил при образовании ключевых слов следует ориентироваться на следующие дополнительные правила:

Географические названия используются в том же виде, как и в географических справочниках или картах (например, Северная Америка). В некоторых КС, обозначающих названия географических объектов, указывается вид объекта: озеро, горы, река:

Например: Байкал, озеро.
Эверест, горы.
Томь, река.

Географическое название может быть записанным как КС и как предметная рубрика по правилам, обусловленным форматом, который поддерживается определенной АБИС и в зависимости от принятых правил работы библиотеки.

Официально принятые и устоявшиеся понятия (светские и религиозные праздники, термины народного календаря и лечебника) записываются полными словосочетаниями и без кавычек.

Например: День Победы,
Татьянин день,
День Благодарения.

При наличии нескольких вариантов названий предпочтение отдается более краткому. Например, Рождество Христово - записывается Рождество;

Названия технических средств записываются с конкретизацией названия или номера модели. При этом отдельным ключевым словом записывается тип изделия.

Например: автомобили,
ВАЗ 2106;
холодильники,
ЗИЛ 204;
самолеты,
Boeing 408;

Названия иностранных фирм и торговые марки записываются на языке оригинала либо по-русски, если в издании фигурирует только русская транскрипция.

Например: IBM,

Panasonic,
Коника;

Названия войн записываются как устойчивое словосочетание: названия войны и отдельно – слово «войны». Например, Великая Отечественная война, войны; Первая мировая война, войны. В случае неустоявшегося названия войны записывается слово «войны», название государства и годы. Например, Афганская война, войны, Афганистан, 1980-1992.

Ключевыми словами могут быть ГОСТированные и общепринятые **сокращения и аббревиатуры**. Предпочтение отдается, как правило, полной форме. **Аббревиатуры** используются только в том случае, когда краткая форма вытеснила полное наименование и когда сокращение чаще используется не только в терминосистемах, но и обычной речи.

Например: ЮНЕСКО, АСУ. СССР, ЛДПР, ГОСТ, ТУ и т.д.;

но

всеобщее обучение (а не всеобуч).

Всемирная организация интеллектуальной собственности (а не ВОИС).

Глаголы записываются в виде отглагольного существительного. Например, преобразуют – преобразование.

Наречия заменяются на соответствующие прилагательные. Например, автоматически – автоматические.

Деепричастия преобразуются в соответствующие существительные. Например, поставляя – поставка.

Хронологические данные. Все даты записываются арабскими цифрами. При необходимости используются другие слова, уточняющие дату.

Например: 19 век, 20 век.

Годы записываются арабскими цифрами и словом «год» для одного года или «годы» для интервала лет.

Например: 1954 год, 1941/1945 годы.

Десятилетия записываются арабскими цифрами без дефиса «е» со словом «годы» в сочетании с веком.

Например: 20 век 20 годы.

При невозможности четкого отображения охвата времени употребляются слова «первая половина», «вторая половина», «начало», «конец».

Например: 20 век начало.

Как дополнительная характеристика в качестве КС возможно включение устоявшихся названий веков.

Например: Серебряный век; Бронзовый век; Средние века

Символические обозначения, например Sr-90

Термины-предложения следует оставлять, не деля их на словосочетания, лишь в том случае, когда такое разделение может привести к потере смысла исходного предложения.

Например, такими КС могут быть различные команды.

Численные характеристики (диапазоны температур, давлений и т.д.) следует давать со словесной формулировкой.

Например: диапазон температур

Законы, правила, явления, теории, формулы и т.д., названные по именам авторов, открывателей и др. записываются с инверсией.

Например: Ньютона бином; Пифагора теорема; Доплера эффект.

Понятия, относящиеся к территориальному и политическому делению записываются как понятие и географическое название. Пример: английские колонии – колонии; Великобритания.

Исключения составляют названия государств, биологических видов и пород животных с географическими определениями.

Пример: Австрийская империя; гималайские медведи; английские бульдоги

Понятия связанные с культурой и народным творчеством- как существительное и прилагательное.

Пример: Английские сказки

Названия языков записываются как одно ключевое слово.

Пример: английский язык.

Народные промыслы описываются нормативной лексикой как устоявшиеся словосочетания.

Пример: Хохломская роспись; Каслинское литье, Федоскинская миниатюра, Дымковкая игрушка и т.д.

Персоналия. Имя лица имеет место быть записанным как КС только в том случае, если это принято правилами работы библиотеки. Как правило имя лица вносится поле 600 (форматы RUSMARC-USMARC) по правилам, обусловленным форматом, который поддерживается определенной АБИС.

Наименование организаций как предмет раскрытия содержания издания записывают по тем же правилам, что и Персоналия. Наименование организации записывается преимущественно в полной форме и/или в общепринятом сокращении. Пример: Министерство по чрезвычайным ситуациям, МЧС; МВД.

КС вносятся каждое в отдельную строку повторяемых полей. Порядок следования КС не имеет значения и носит произвольный характер. На поиск документов это не влияет.

Количество КС определяется необходимостью полного описания всех важных для читателя аспектов содержания индексируемого документа (10–12 КС). КС, включенные в запись не должны быть синонимами.

Качественные показатели индексирования определяются тремя категориями: глубиной (или полнотой), точностью и нормализованным языком понятий (или унифицированным, стандартизованным). Для улучшения качества поиска по ключевым словам необходимо:

- деления классификации, присвоенные обрабатываемой единице, выражать словами и словосочетаниями.
- вставлять понятия, называющие отрасль знания, отраженную в документе, или (что примерно то же самое) понятия, характеризующие верхний индекс УДК или ББК.
- должно быть выдержано отражение документа сверху вниз или снизу вверх, от общего к частному или наоборот, от основного дескриптора к нижестоящему дескриптору, затем к ассоциативному дескриптору и т.д. Для машинного поиска не имеет значения, в каком порядке отражены понятия. Важно во всей полноте и глубине отразить содержание документа.
- отражать вид издания по характеру информации — диссертация, автореферат диссертации, препринт, тезисы докладов, энциклопедия, учебное пособие и т.д.

Используемая литература:

1. Воройский Ф.С. Индексирование документов в АБИС// Библиотека. – 1996. - № 9. – С. 42-44.
2. Гендина Н.И. Лингвистическое обеспечение автоматизированных библиотечных систем. – Алма-Ата: Гылым, 1991. – 222 с.
3. Методика разработки и ведения тезауруса информационно-поискового (ИПТ) // Создание электронных каталогов: Из опыта работы Российской государственной библиотеки/ Рос. гос. б-ка; Сост. О.А. Лавренова. - М, 1996. – С. 65 – 77.
4. Скипор И. Л. Методика создания дескрипторного словаря

Составитель: Грехова Н. М.

Март 2004 г.