

УДК 025.3:004
ББК 78.362.5+78.364.4

КЛЮЧЕВЫЕ СЛОВА В СТРУКТУРЕ ЗАПИСИ ЭЛЕКТРОННОГО КАТАЛОГА¹

© В. А. Белов, Л. В. Никанова, 2013

Федеральное государственное образовательное учреждение
высшего профессионального образования
«Череповецкий государственный университет»
162600, Вологодской обл., г. Череповец, пр. Луначарского, 5

Анализируются подходы к составлению набора ключевых слов в электронных каталогах и предлагаются правила выделения ключевых слов в тексте.

Ключевые слова: ключевое слово, набор ключевых слов, методика, запись, электронный каталог.

The target of the article is to analyze approaches to forming a set of keywords in electronic catalogs and to offer rules to determine keywords in a text.

Key words: a key word, a set of keywords, methods, a record, an electronic catalog.

В так называемую доинформационную эпоху основным средством систематизации информации в библиотеке выступали библиотечные классификации. Сейчас у них появился серьезный конкурент – *ключевые слова*. Этому способствует развитие информационных систем: поисковые системы Интернета позволили пользователям находить информацию с помощью ключевых слов.

В свою очередь, библиотеки, стараясь оперативно реагировать на потребности читателей, пытаются освоить работу с ключевыми словами. Однако отсутствие единого представления о том, как индексировать текст с помощью ключевых слов, приводит к затруднениям при поиске литературы. Каждый составитель библиографических описаний использует свою, индивидуальную стратегию подбора ключевых слов, что порождает разнообразие формулировок, понятий и терминов. Отсюда необходимость в разработке методики выделения ключевых слов, которая помогла бы библиотекарям в работе. Представленная в статье методика не ставит целью заменить ББК. Напротив, она стремится использовать достижения этой библиотечной классификации.

Набор ключевых слов

Ключевыми словами называются языковые единицы (слова и словосочетания), выделенные непосредственно из текста и предназначенные для свернутого представления его содержания. Такое пони-

мание ключевых слов отражено в работах отечественной психолингвистической школы [3, 5, 7]. Ключевые слова формируются путем компрессии (обобщения) содержания текста. С помощью набора ключевых слов (НКС) можно в обобщенном виде представить содержание текста: НКС замещает текст в поисковом массиве документов.

Со структурной точки зрения НКС – простое перечисление ключевых слов. Как правило, ключевые слова независимы друг от друга, хотя все же прослеживается тенденция располагать их от более важных к менее важным, т. е. первыми размещают слова, несущие концептуальный смысл. Ключевые слова, раскрывающие частные моменты содержания текста, располагают в конце НКС.

Например, составляя описание учебника по анатомии, ключевые слова логично размещать следующим образом (от общего к частному): «анатомия», «органы слуха», «органы зрения». Анатомия, безусловно, является более общим понятием по отношению к органам слуха и зрения.

Первым должно стоять наименование предметной рубрики – обобщенное название отрасли знания. Здесь пересекаются традиционные библиотечные классификации (ББК) и логика выделения ключевых слов: разумно использовать названия разделов классификации в качестве первого ключевого слова.

Первое ключевое слово выступает в качестве классификационного индекса текста: относит его к той или иной области знания. Так, раскрывая

¹ Работа проведена в рамках реализации ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009–2013 гг.

содержание учебника «Современный русский язык» под редакцией П. А. Леканта [6], важно указать, что книга посвящена языкознанию. Поэтому именно это слово должно стать первым ключевым словом.

Объем НКС не регламентируется: чем сложнее текст, тем больше нужно использовать ключевых слов. Говоря о количественном составе ключевых слов, нельзя сформулировать универсальное правило для всех типов текстов. Тексты отличаются не только объемом, но и содержанием.

Например, учебники имеют внушительный объем и затрагивают обширное количество вопросов. Цель учебника – дать общее представление о большом количестве тем. Как показывает библиографическая практика, именно у учебников самые большие НКС. Однако не следует в НКС воспроизводить все раскрываемые темы: важно перечислить основные темы книги.

Так, в учебнике по анатомии есть информация о большинстве органов человека, но материал дан кратко, реферативно. Представим ситуацию, когда студенту нужно написать реферат о строении указательного пальца руки. Он делает информационный запрос «указательный палец» в надежде найти литературу, в которой детально исследуется указательный палец. Информации в учебнике ему будет недостаточно, чтобы написать реферат.

Важной причиной ограничивать НКС учебников является то, что они встречаются в большом количестве в библиотеках и создадут «шумовые» помехи при поиске специальной, узкой литературы. В то же время нужно формировать НКС из достаточного количества слов, чтобы он дал представление о содержании текста.

Следует избегать употребления в НКС сложных словосочетаний. Их необходимо разбивать на ряд простых. При составлении НКС исходят из того, что читатель вводит простые, общепринятые понятия. Поэтому при аналитико-синтетической переработке содержания документа его уникальное содержание необходимо представить с помощью общепринятых понятий, придерживаясь правила: сложное передавать через простое.

Вот некоторые примеры того, как можно упростить сложные словосочетания: «история русской литературы» = «история литературы» + «русская литература»; «особенности русской средневековой культуры» = «культура» + «русская культура» + «средневековая культура»; «дети с нарушением речи» = «дети» + «нарушение речи»; «метод конструирования учебно-научного текста» = «метод конструирования» + «учебно-научный текст».

Для ключевых слов следует использовать научные термины, устоявшиеся понятия и словосочетания. Избегать словосочетаний типа: «световоздушная перспектива», «проклятие природных ресурсов», «глубинная композиция». Подобные язы-

ковые обороты обладают высокой степенью эмоциональности, что недопустимо для НКС. Таким образом, ключевые слова формируются с помощью устоявшихся языковых единиц с высокой языковой частотностью (т. е. встречаемостью в языке) и с нейтральной оценкой.

Принцип передавать сложное через простое нужно сохранять и при воспроизведении атрибутивных словосочетаний (построенных по формуле прилагательное + существительное): их лучше представлять как сочетание существительных. Например: «литературный портрет» = «литература» + «портрет»; «конкурентное преимущество» = «конкуренция» + «преимущество». Слова должны воспроизводиться в начальной форме – в именительном падеже и единственном числе.

Однако большая группа атрибутивных словосочетаний – это научные термины и устоявшиеся понятия. В таких случаях следует сохранять словосочетание: «русская литература», «корпусная лингвистика», «дискретная математика», «возрастная психология» и т. д.

При выделении НКС важно избегать слов с нулевой семантикой, таких, как «особенности», «уровень», «возможности» и др. Такие слова не несут информации о содержании текста, а выполняют восполняющую языковую функцию.

Приведем несколько примеров словосочетаний: «особенности русской культуры» = «русская культура», «уровень развития ребенка» = «развитие ребенка», «гендерная специфика» = «гендер». Изъятия подобных асемантических элементов позволит сделать НКС более концептуальным.

В некоторых лингвистических работах [9 и др.] подчеркивается особая роль фактора частотности в тексте, т. е. ключевым словом становится единица, которая часто встречается в самом тексте. Однако надо понимать, что не каждая частотная единица способна быть ключевым словом, поскольку кроме частотности, она должна нести важный для понимания текста смысл.

Покажем на примере, как частотность в тексте способна помочь найти ключевые слова. В небольшой статье «“Темный этап” в истории организации знаний в библиотеках» [8] речь идет о систематизации знаний человеком. Автор статьи выделяет следующие ключевые слова: «знания», «организация», «библиотеки», «карточные каталоги», «электронные каталоги», «Международное общество по организации знаний».

Однако в тексте часто встречается понятие «классификация» (9 употреблений), которое несет важный смысл для понимания содержания статьи (Э. Р. Сукиасян показывает историю различных классификаций знаний). Поэтому включение этого слова в НКС было бы вполне оправданным. Таким образом, ключевым словом может стать единица,

часто встречающаяся в этом тексте и при этом несут определенный смысл.

В НКС допускается включать имена собственные – писателей, ученых, общественных деятелей, – если текст посвящен повествованию об этих лицах. Однако следует понимать, что нестандартное использование имен собственных в НКС может привести к снижению адекватности поиска литературы. Одну и ту же фамилию можно внести по-разному: Пушкин А. С., А. С. Пушкин, Пушкин.

Соблюдение конвенциональных принципов позволит снизить «шумовые» помехи: представление фамилий должно быть единообразно. Лучшей формой записи является тип, когда в инициальной позиции пишется фамилия, а потом – инициалы (или имя для иностранного автора): Пушкин А. С., Брунер Джером. Инициалы автора полезны для различения однофамильцев: Толстой Л. Н., Толстой А. К. и пр.

Такого же принципа следует придерживаться при описании текстов по истории. Так, статьи, посвященные одному историческому событию, допустим, Отечественной войне 1812 года, могут иметь разные ключевые слова: «История войны 1812 года», «Бородинское сражение», «Отечественная война 1812 года».

Разногласие в подходах описания исторических событий порождает затруднения при поиске литературы.

Следует сконцентрироваться на обобщенном периоде (век, эпоха), избегая частных событий. Важно разработать конвенциональные правила описания крупных исторических событий: например, во всех описаниях использовать только наименование «Отечественная война 1812 года». При этом выбор обобщенного наименования должен опираться на научную традицию и речевую практику.

Особого внимания заслуживают случаи употребления сокращений и иностранных слов. Рекомендуется приводить аббревиатуры как в сокращенном виде – например, ФСБ, так и в полном – Федеральная служба безопасности. Иностранные слова так же можно употреблять в двух формах – на русском языке (если такое слово существует в русском языке) и на языке оригинала: например, «Тойота»/«Toyota». Такое требование объясняется тем, что читатель может искать литературу, используя разные стратегии.

Композиция текста и НКС

Определим, с помощью каких структурных компонентов текста можно эксплицировать ключевые слова. Основные формальные «помощники» в определении ключевых слов: заглавие, аннотация, предисловие и оглавление.

Заглавие представляет основную тему текста и, как правило, включает основные ключевые слова.

Они отражают максимально обобщенное содержание текста, поэтому НКС не следует ограничивать лишь ими. Так, заглавие учебного пособия «Сравнительная типология английского и русского языков» дает возможность выделить следующие ключевые слова: «языкознание», «лингвистика», «язык», «русский язык», «английский язык», «типология», «сравнительная типология». Однако приведенные ключевые слова не дадут полноценного представления о содержании текста. Ключевые слова не должны повторять жанр текста: не стоит использовать в НКС такие слова, как «рецензия», «книга», «словарь», «учебное пособие».

Аннотация позволяет выделить важные ключевые слова, поскольку представляет в сжатом виде основное содержание текста. При этом аннотация, в отличие от НКС, оформлена в виде связных предложений. НКС – простое соположение важных для смысла текста слов и словосочетаний. Поэтому ключевые слова являются более концентрированным содержанием текста.

Перевод аннотации в НКС осуществляется путем выделения самых важных, информативных слов и словосочетаний. Покажем это на примере аннотации книги «Современный русский язык»: «В учебнике представлены все разделы современного русского языка (лексика, фразеология, лексикография, фонетика, орфоэпия, графика, орфография, морфемика, словообразование, морфология, синтаксис). Книга дает целостное представление о современном освещении всех уровней литературного русского языка» [6, с. 1].

Из этой аннотации выделяем следующие ключевые слова: «языкознание», «лингвистика», «русский язык», «лексика», «фразеология», «лексикография», «фонетика», «орфоэпия», «графика», «орфография», «морфемика», «словообразование», «морфология», «синтаксис».

Здесь же коснемся вопроса, который часто тревожит библиотекарей: включать ли в электронные каталоги аннотации книг и статей? Как нам представляется, здесь следует руководствоваться правилом: если аннотацию можно полноценно заменить НКС, то аннотацию лучше не включать (это экономит время); заносить в электронные каталоги следует аннотации, содержащие информацию, которую нельзя представить в виде НКС.

Предисловие содержит общую тематику книги, основное содержание текста. Темы, затронутые в предисловии, могут выступать в роли ключевых слов. Однако при написании предисловия могут быть использованы различные стратегии, в том числе идти речь об авторе, ценности текста, истории создания.

Например, в предисловии редактора (известного психолога А. Р. Лурия) к книге Дж. Брунера «Психология познания: За пределами непосредственной

информации» читаем: «Настоящая книга принадлежит перу одного из выдающихся прогрессивных американских психологов Джерома Брунера, ранее в течение 30 лет бывшего профессором Гарвардского университета (США) и директором Центра по исследованию познавательных процессов, а ныне являющегося профессором Оксфордского университета (Англия)» [2, с. 5].

Задача такого предисловия, в первую очередь, – познакомить читателя с автором книги. Однако повествование о биографии сопровождается информацией о тематике научных исследований Брунера, что вполне может быть использовано для НКС: «Книга Дж. Брунера состоит из четырех разделов. Первый из них посвящен центральной проблеме психологии – анализу процессов *восприятия* [курсив. – А. Р. Лурия]» [2, с. 7]. Восприятие является одной из основных тем книги. Поэтому включение слова «восприятие» в НКС обязательно.

Оглавление (содержание) – также важный маркер НКС: названия основных разделов, частей, параграфов содержат ключевые слова.

Однако НКС не должен механически повторять оглавление. Наверное, не существует универсального правила отбора из него ключевых слов. Здесь надо помнить, что ключевые слова несут важную (сжатую) информацию о содержании текста. В НКС должны попадать те элементы, которые действительно существенны для всего произведения: не стоит включать в НКС частную, ценную лишь для отдельных фрагментов текста информацию. В больших текстах (например, в учебниках, монографиях, сборниках работ) для НКС можно использовать заглавия крупных фрагментов текста (раздел, часть). Кроме того, не следует выбирать в качестве ключевых слов «поэтические» (или модусные [1]) заглавия, которые нацелены на передачу не фактуальной информации, а на привлечение внимания читателя. К таким заглавиям можно отнести, например, «Скажи мне, кто твой партнер», «Дом, который построил... говорящий» (из книги Б. Ю. Нормана «Грамматика говорящего: От замысла к высказыванию») [4].

Заключительные замечания

Задача информационно-библиотечной системы – быстрый и качественный поиск нужной литературы. Поэтому, составляя НКС текста, нужно прогнозировать, какие стратегии будет использовать читатель, чтобы найти нужную ему литературу.

Главным принципом работы с ключевыми словами является конвенциональность, т. е. сотрудники библиотеки, создающие электронный каталог, должны принять согласованную методику формирования НКС. Это нужно для того, чтобы у читателей сформировалось единое представление о том, как осуществлять поиск по ключевым словам. При этом конвенциональные правила обязательно следуют за речевой практикой. Попытки «навязать» читателям не свойственные языку нормы и правила обязательно приведут к неудаче при поиске.

Ключевыми словами должны становиться языковые единицы, устоявшиеся в научном мире и обладающие высокой языковой частотностью.

Конечно, идеальный вариант, когда подобным образом договорятся все библиотеки страны и мира, но пока они применяют разные стратегии подбора НКС. Но думается, что со временем в информационной сфере сформируются единые правила выделения ключевых слов. Носители языка неосознанно будут выбирать наиболее продуктивные стратегии. Тем не менее задача библиотекарей сегодня – попытаться предугадать эти успешные модели, чтобы уже сейчас использовать все возможности автоматизированного поиска.

Литература

1. Белов В. А. Пропозициональная организация текста (на материале романа А. Белого «Петербург») : автореф. дис. ... канд. филол. наук. – СПб., 2011. – 23 с.
2. Брунер Дж. Психология познания: за пределами непосредственной информации. – М., 1977. – 413 с.
3. Новиков А. И. Текст и его смысловые доминанты. – М., 2007. – 224 с.
4. Норман Б. Ю. Грамматика говорящего: от замысла к высказыванию. – СПб., 1994. – 229 с.
5. Сахарный Л. В., Сибирский С. А., Штерн А. С. Набор ключевых слов как текст // Психолого-педагогические и лингвистические проблемы исследования текста. – Пермь, 1984. – С. 81–83.
6. Современный русский литературный язык / под ред. П. А. Леканта. – 6-е стер. изд. – М., 2004. – 462 с.
7. Соколов А. Н. Внутренняя речь и понимание // Ученые записки государственного научно-исследовательского института психологии. – М., 1941. – Т. 2. – С. 99–146.
8. Сукиасян Э. Р. «Темный» этап в истории организации знаний в библиотеке // Науч. и техн. б-ки. – 2011. – № 10. – С. 35–41.
9. Ягунова Е. В. Эксперимент и вычисления в анализе ключевых слов художественного текста // Философия языка. Лингвистика. Лингводидактика. – Пермь, 2010. – № 1. – С. 83–89.

Материал поступил в редакцию 13.12.2012 г.

Сведения об авторах: Белов Вадим Алексеевич – кандидат филологических наук, директор библиотеки, тел.: (8202) 55-91-66, e-mail: belovva@chsu.ru,

Никанова Любовь Владимировна – главный библиотекарь, тел.: (8202) 55-36-72