

ВЕДЕНИЕ СЛОВАРЯ КЛЮЧЕВЫХ СЛОВ ЭЛЕКТРОННОГО КАТАЛОГА ЦЕНТРАЛЬНОЙ НАУЧНОЙ БИБЛИОТЕКИ НАН БЕЛАРУСИ

Л. Л. Астапович

Центральная научная библиотека им. Я. Коласа НАН Беларуси, Минск

Представлена практика ведения словаря «Неконтролируемые тематические термины» электронного каталога (ЭК) Центральной научной библиотеки им. Я. Коласа НАН Беларуси (ЦНБ НАН Беларуси). Рассмотрен вопрос создания словаря ключевых слов (КС) как одного из средств лингвистического обеспечения ЭК. Отражены особенности процесса редактирования КС и их унификации с целью представления в Сводном электронном каталоге (СЭК) библиотек Беларуси.

В ЦНБ НАН Беларуси ведется постоянная работа по сопровождению и актуализации справочно-поискового аппарата (СПА) ЭК, обеспечивающего проведение информационного поиска и выполнения запросов пользователей. В СПА каталога входят словари авторитетных (нормативных) записей (АЗ) на различные наименования и словарь «Неконтролируемые тематические термины», содержащий КС.

Словарем, используемым для формирования поискового образа документов (ПОД) и обеспечения тематического поиска, первоначально стал словарь КС. Вопрос о выборе средств лингвистического обеспечения ЭК и составлении словаря КС решался в процессе создания ЭК библиотеки. При составлении словаря термины выбирались из заглавия, оглавления, аннотации, введения и заключения индексируемых документов. Контроль над формированием КС первое время отсутствовал. Индексаторы, которые и занимались составлением тематического словаря, не обладали опытом определения КС и выбирали термины без учета того, какие КС уже использовались ранее для индексирования близких по тематическому содержанию документов. В этой ситуации не устранялась синонимия и омонимия КС, не проводилась лексико-семантическая обработка терминов. Недостатки такой работы очень быстро проявились при составлении запросов для поиска в ЭК [1].

Изучая методы разных библиотек по созданию КС и используя терминологию из словарей, тезаурусов, а также рубрики из классификационных таблиц, были установлены требования к КС и разработана методика их составления для ЭК библиотеки. При определении КС на документы проводилась обязательная сверка со словарем, выполнялась нормализация новых КС – приведение их к принятой форме записи.

С 2005 г. ЦНБ НАН Беларуси полностью перешла на автоматизированную библиотечную информационную систему (АБИС) «БИТ-2000и», разработанную в Объединенном институте проблем информатики НАН Беларуси (ОИПИ НАН Беларуси). ЭК библиотеки с помощью разработчиков АБИС «БИТ-2000и» был конвертирован в формат BELMARC [2]. Из КС был сформирован словарь «Неконтролируемые тематические термины». Для представления КС в библиографических записях (БЗ) каталога использовалось 610-е поле формата BELMARC. Структура словаря представляла собой алфавитный перечень КС, в котором применялась система отсылок, использовались реляторы, обозначающие принадлежность к определенной отрасли (рис. 1).

Словарь «Неконтролируемые тематические термины» состоял уже из двух частей: терминов на русском и белорусском языках (кириллице) и терминов на иностранных языках (латинице). Значительная часть КС формировалась на русском языке, а для документов, изданных на белорусском языке, они составлялись параллельно на двух язы-

ках – белорусском и русском. Для поискового образа на иностранные документы, поступающие в фонд библиотеки, также определялись КС. Чтобы обеспечить многоаспектное выражение смыслового содержания научных документов, опубликованных на английском, французском, немецком и польском языках, КС приводились на русском языке и языке оригинала. На документы, изданные на других иностранных языках, КС определялись только на русском языке.

первобытнообщинная формация
см. Первобытное общество
первобытнообщинный строй
см. Первобытное общество
первоочередные права
см. Преимущественные права

а)

Морфология (биология)
Морфология (мед.)
Морфология (языкознание)
Тромб (мед.)
Тромб (метеорология)

б)

Рис. 1. Примеры отсылок в словаре КС (а) и КС с реляторами (б)

К участию в системе корпоративной каталогизации (СКК) и ведению СЭК библиотек Беларуси ЦНБ НАН Беларуси приступила в 2007 г. В технологию корпоративного взаимодействия вовлекались локальные информационные ресурсы библиотек – участниц СКК, в том числе и ЦНБ НАН Беларуси. При объединении информационных ресурсов значительная часть КС была заменена на АЗ СЭК, которые были отнесены к сформированным в ЭК библиотеки словарям на отдельные наименования. Однако в результате объединения ресурсов в каталоге появились синонимы, выраженные как КС, так и АЗ, в словарях присутствовали различные варианты названий организаций, географических объектов, родовых имен. Поэтому возникла необходимость в редакции словаря КС, которая была начата в 2008 г.

Помимо постоянно проводимой редакции КС, составленных на тематический предмет, в течение 2010–2016 гг. выполнялась работа по выводу из словаря персоналий, родовых имен, заглавий произведений, заголовков организаций и предприятий, а также географических наименований. Данная работа была проведена, чтобы избежать использования таких КС другими библиотеками. КС выявлялись в словаре «Неконтролируемые тематические термины» и заменялись АЗ. При отсутствии необходимых АЗ составлялись списки КС, на основании которых сотрудники ЦНБ НАН Беларуси и Национальной библиотеки Беларуси (НББ) создавали новые АЗ. КС на перечисленные наименования были полностью выведены из словаря «Неконтролируемые тематические термины» и заменены на АЗ, которые в БЗ соотносились с соответствующими полями 6 Блока определения тематики формата BELMARC [3].

При ведении словаря КС уже учитывались не только потребность в постоянном улучшении качества словаря и возможность проведения тематического поиска в ЭК, но и унификация БЗ для СЭК. Так, для единообразия поискового образа на иностранные документы от специалистов НББ поступило предложение отказаться от составления КС на иностранных языках. Полагая, что отказ от КС, сформулированных на языке оригинала на иностранные научные документы, мог сказаться на качестве лингвистического обеспечения ЭК и ограничить поисковые возможности пользователей, предложение не было принято, и КС на иностранных языках продолжали формировать на все наименования предметов. Однако необходимость в унификации ПОД требовала решения этого вопроса.

С 2011 г. словарь «Неконтролируемые тематические термины» перестал пополняться иноязычными КС на следующие наименования: имя лица, родовое имя, организации, торговая марка, форма/жанр и географические названия. В поисковом образе

иностраннных документов эти понятия были представлены только АЗ. КС продолжали составлять только на тематический предмет.

В рамках СКК осуществлялась работа по редакции АЗ и актуализации словарей СЭК. На заседаниях Технологического комитета СКК продолжалась проработка вариантов решения вопросов по обеспечению единообразия отражения ПОД в БЗ, в том числе и вопрос о целесообразности составления иноязычных КС. В результате в 2014 г. в ЦНБ НАН Беларуси было принято решение удалить КС на иностранных языках из словаря «Неконтролируемые тематические термины» СЭК.

Иноязычные КС, ранее поступившие в СЭК, были «отсечены» путем программной корректировки сотрудниками ОИПИ НАН Беларуси. В процессе подготовительной работы к удалению этих КС из СЭК специалисты ЦНБ НАН Беларуси проработали список из 49 200 терминов, который включал КС, содержащие латинские гласные буквы. Были отредактированы термины на русском языке, в которых оказались ошибочно внесенные буквы на латинице, и иноязычные термины с включением букв на кириллице. Выявлены и исправлены ошибочные формы терминов, часть КС заменена на АЗ.

Чтобы КС на иностранных языках отражались исключительно в ЭК библиотеки, в 610-е поле заносится код 8 «Термин, не поступающий в СЭК» (рис. 2). При этом фильтр, установленный для удаления полей 610 с кодом 8, отсекает КС из экспортируемых в СЭК новых БЗ.

Мета	Словарь	Наименование
606	\$3 BY-NLB-ar30238 \$a СИСТЕМАТИКА (биол.) \$2 DVNLB	Наименование темы, используемое как предмет
606	\$3 BY-NLB-ar2805484 \$a КОСТИСТЫЕ РЫБЫ \$2 DVNLB	Наименование темы, используемое как предмет
606	\$3 BY-NLB-ar3292208 \$a ГОЛЫЦЫ (посоковые) \$2 DVNLB	Наименование темы, используемое как предмет
607	\$3 BY-NLB-ar10168 \$a Европа \$2 BY-auth	Географическое наименование, используемое как предмет
610	8 \$a Teleostei	Неконтролируемые тематические термины
610	8 \$a Pisces	Неконтролируемые тематические термины
610	8 термин, не поступающий в СЭК	Неконтролируемые тематические термины
675	8 термин, не поступающий в СЭК	Универсальная десятичная классификация (УДК)
	1 первичный термин (более существенный)	
К.. Пк2	вторичный термин (менее существенный)	Наименование
		Тематический термин

Рис. 2. Пример 610-го поля с кодом 8 «Термин, не поступающий в СЭК»

В процессе индексирования документов на иностранных языках с 2015 г. фактически прекращено создание новых КС на языке оригинала. При формировании ПОД индексаторами определяются АЗ, а также КС, составленные на русском языке. Лишь в редких случаях, когда новый научный термин еще не вошел в русскоязычный терминооборот и индексатор не может при переводе точно передать на русском языке понятийный аппарат иностранных терминов, создаются КС на языке оригинала. Отказаться от составления иноязычных КС с целью унификации ПОД позволила существующая в словарях АЗ система гиперссылок, которая не допускает потери информации пользователями.

В настоящее время для формирования ПОД активно используются АЗ, которыми в значительном объеме и пополняются тематические словари ЭК. Поэтому первоочередной задачей при ведении словаря «Неконтролируемые тематические термины» яв-

ляется его редактирование и замена КС на АЗ. В процессе редактирования осуществляется смысловая обработка, при которой определяется семантическое значение термина, выявляются синонимы, более широкие и узкие понятия, устанавливаются связи термина во всех предметных областях. При редакции иноязычных КС проводится семантико-сопоставительный анализ иностранной и русской терминологии. На завершающем этапе выполняется отбор АЗ, соответствующих смысловому содержанию редактируемого КС, и замена КС на выбранные записи (рис. 3).

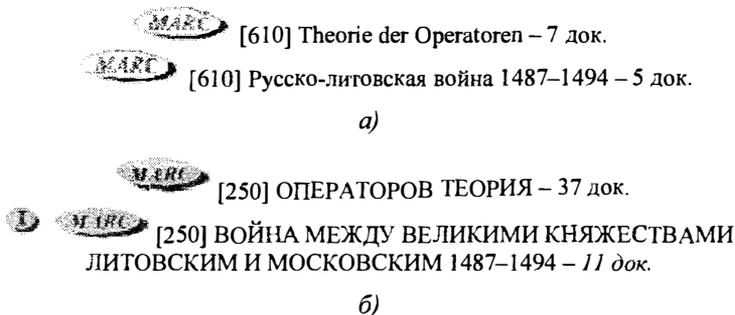


Рис. 3. Пример замены КС (а) на АЗ (б)

Для замены КС на АЗ предусмотрена возможность использования программных средств, обеспечивающих достаточно быстрое выполнение процесса. Однако если возникает необходимость проверить правильность использования конкретных КС, при сплошном просмотре БЗ анализируется каждый ПОД, где представлены данные термины.

При редактировании словаря ведется работа по выявлению КС и составлению списка терминов, научность и достоверность которых подтверждается авторитетными источниками, а также созданию на их основе новых АЗ. Выполняется и лексическо-грамматическая обработка КС, которые остаются в составе словаря «Неконтролируемые тематические термины», приведение их к единой для КС форме согласно принятым решениям для библиотек – участниц СКК. Обязательно проверяется правильность применения КС в ПОД и осуществляется исправление неверного использования терминов, в процессе чего выполняется переиндексирование БЗ.

В начале 2020 г. была завершена редакция словаря КС на русском и белорусском языках, составляющего основную часть словаря «Неконтролируемые тематические термины». На данный момент в этом словаре присутствуют КС, которые прошли смысловую обработку и доработаны до принятой формы в соответствии с методическими решениями для ЭК и СЭК. Продолжается плановая работа по редактированию словаря КС на иностранных языках.

Несмотря на преобладающее применение АЗ, КС все же не потеряли свою актуальность и словарь «Неконтролируемые тематические термины» продолжает пополняться новыми терминами.

Формируются АЗ и КС после оценки перспективности использования терминов, заявленных на создание АЗ, в поисковом образе на новые документы, поступающие в библиотеку. Если термин введен в научный терминологический оборот, но в дальнейшем его использование в БЗ маловероятно и поэтому нецелесообразно создавать АЗ, то составляются новые КС. Создаются КС и на понятия, которые не представлены в авторитетных источниках, но необходимы для выражения содержания документов. Новые КС заносятся в БЗ и отражаются в словаре «Неконтролируемые тематические термины». В дальнейшем КС в составе БЗ поступают в СЭК и заимствуются библиотеками – участницами СКК при индексировании документов.

Для лингвистического обеспечения ЭК библиотеки КС по-прежнему имеют существенное значение. Составленные в большей степени на узкоспециальные термины они дополняют поисковый образ, сформированный из АЗ, и обеспечивают полноту раскрытия содержания документов.

Ведение и редактирование КС направлены на упорядочение логико-семантической структуры словаря «Неконтролируемые тематические термины», позволяют избежать дублирования терминов в тематических словарях ЭК и исключить субъективизм при выборе из словарей понятий по определенной тематике как при индексировании документов библиотекарями, так и при проведении информационного поиска в каталоге пользователями. Поскольку КС из локальных ЭК отражаются и в СЭК, то приведение КС к унифицированной форме в процессе редактирования словаря дает возможность использовать их и библиотекам, работающим со словарями СЭК.

Список литературы

1. Астапович, Л. Л. Проблемы и тенденции формирования средств лингвистического обеспечения в ЦНБ им. Я. Коласа Национальной академии наук Беларуси / Л. Л. Астапович, О. М. Дрозд // Корпоративные библиотечные системы: технологии и инновации : тр. VI науч.-практ. конф., 23–29 июня 2008 г., СПб. (Россия), Хельсинки, Тампере (Финляндия), Стокгольм, Упсала (Швеция) / СПб. гос. политехн. ун-т, Федеральное агентство по культуре и кинематографии Рос. Федерации, Ком. по культуре СПб., Ассоц. регион. библ. консорциумов (АРБИКОН). – СПб., 2008. – С. 127–132.

2. Ловейкина, Л. В. Новые компьютерные технологии в практике каталогизации Центральной научной библиотеки имени Я. Коласа НАН Беларуси / Л. В. Ловейкина, Г. О. Лаужель // Библиотеки национальных академий наук: проблемы функционирования, тенденции развития : науч.-практ. и теорет. сб. – Киев, 2010. – Вып. 8. – С. 264–270.

3. Астапович, Л. Л. Редактирование словарей электронного каталога / Л. Л. Астапович, О. И. Старовойтова // Развитие информатизации и государственной системы научно-технической информации (РИНТИ-2017) : докл. XVI Междунар. конф., Минск, 16 нояб. 2017 г. – Минск : ОИПИ НАН Беларуси, 2017. – С. 285–289.