

СТЕНДОВЫЕ ДОКЛАДЫ

Проблемы и тенденции формирования средств лингвистического обеспечения в ЦНБ им. Я. Коласа Национальной академии наук Беларуси

*Астапович Л.Л., зав. отделом научной каталогизации,
Центральная научная библиотека НАН Беларуси*

*Дрозд О.М., заместитель директора по научной работе,
Центральная научная библиотека НАН Беларуси*

В докладе рассматривается поэтапное формирование лингвистических средств для электронного каталога Библиотеки и дальнейшее взаимодействие с библиотеками республики.

Центральная научная библиотека им. Я. Коласа Национальной академии наук Беларуси (ЦНБ НАН Беларуси) в 1994 г. приступила к созданию электронного каталога (ЭК). Сразу же возник вопрос о формировании средств лингвистического обеспечения (ЛО), играющего значительную роль в функционировании ЭК. Надо сразу отметить, что здесь мы рассматриваем классический подход к понятию ЛО, при котором лингвистическим обеспечением называют комплекс информационно-поисковых языков (ИПЯ), прежде всего, классификационных и вербальных.

Формирование средств ЛО явилось проблемным и трудоемким процессом. Планируя выбор средств ЛО, мы исходили из того, что они предназначены обеспечить индексирование документов и эффективный поиск в ЭК по тематическим запросам. Для выполнения этой задачи лингвистические средства должны быть сформированы при помощи классификационной системы, т.е. ИПЯ классификационного типа, и ИПЯ вербального типа.

Классификационная система, по которой организован систематический каталог (СК) ЦНБ НАН Беларуси – Универсальная десятичная классификация (УДК). Использование УДК как одного из лингвистических средств расширяет возможности поиска (по запросам отраслевого характера) в ЭК и, благодаря своей универсальности, обеспечивает лингвистическую совместимость ЭК ЦНБ НАН Беларуси с ЭК других библиотек.

Решение о ИПЯ вербального типа во многом зависело от выбора коммуникативного формата. Так как в тот период отсутствовала государственная политика в области автоматизации библиотечных процессов, то выбор каждой библиотеки зависел от ее технической,

программной и финансовой обеспеченности. Учитывая все перечисленные факторы, Библиотека выбрала USMARC-формат.

В данном формате для заполнения поля «Тематическая рубрика» необходимо было бы использовать предметизационный ИПЯ, в основе которого лежит алфавитный перечень предметных рубрик (ПР). Такой ИПЯ должен быть построен единообразно, стандартизировано.

Ввиду отсутствия возможностей у Библиотеки приобрести готовые информационно-поисковые тезаурусы, и отсутствия опыта по составлению предметных (тематических) рубрик (за исключением алфавитно-предметного указателя (АПУ) к СК), было решено заполнять поле «Ключевые слова». В качестве ИПЯ вербального типа был выбран дескрипторный язык – язык ключевых терминов. Началась работа по составлению методики координатного индексирования документов и составление перечня КС в виде словаря. Исходили из того, что основное смысловое содержание документа и информационной потребности может быть достаточно полно и точно выражено списком КС, которые явно или в скрытом виде содержатся в тексте. Под КС в нашем случае понимаются слова или словосочетания из названия или текста документа, которые несут в данном тексте существенную смысловую нагрузку с точки зрения информационного поиска. КС брались из заглавия и оглавления документа, аннотации или реферата, введения и заключения, а также из текста документа.

В первое время при координатном индексировании документов контроль за КС полностью отсутствовал. Систематизаторы, не обладавшие опытом определения КС и, не имея методики по их составлению, выбирали ключевые слова непосредственно из текста документа без учета того, какие КС уже использовались ранее для индексирования таких же или близких по смыслу документов. В данном случае не устранялась синонимия, полисемия и омонимия КС, не приводились к нормальному виду их грамматические формы. Недостатки такой работы очень быстро проявились при формулировании запросов для поиска в ЭК Библиотеки. Для выработки единого подхода к составлению КС обращались за помощью в другие библиотеки. Некоторые из них работали с приобретенными тезаурусами по своей тематике (медицинская и сельскохозяйственная библиотеки), другие, из-за отсутствия опыта, так же как и мы искали выход из создавшейся ситуации.

Используя методики других библиотек по составлению КС, применяя рубрики из таблиц УДК и АПУ к СК, термины из различных словарей и тезаурусов был выработан подход к составлению КС, разработана и утверждена «Методика составления КС» для ЭК ЦНБ НАН Беларуси на документы, изданные на русском и белорусском языках.

В 1997-2003 гг. в рамках программы по развитию научно-информационной компьютерной сети Республики Беларусь была создана интегрированная система автоматизированной библиотечной деятельности – система БИТ – 2000и. В качестве Национального обменного формата был определен формат BELMARC, который соответствует Международному

коммуникативному формату UNIMARC, а также Российскому формату RUSMARC. Основные преимущества формата BELMARC:

- разработан на основе международного формата UNIMARC;
- введение в него полей и подполей национального использования, отражающих практику каталогизации в библиотеках Беларуси, а также некоторые требования к полноте описания национального документа, не представленные международным стандартом;
- совместимость с Российским коммуникативным форматом RUSMARC и обеспечение взаимной конвертируемости с форматом USMARC;
- возможность устанавливать связи между записями, созданными как на подобные, так и на различные виды изданий.

Система БИТ-2000и полностью совместима с Национальным форматом BELMARC и предусматривает интерактивную работу с онлайн-каталогом Библиотеки и со Сводным электронным каталогом (СЭК) библиотек Беларуси. В 2003 г. ЦНБ НАН Беларуси перешла на новую АБИС БИТ-2000и и соответственно на BELMARC- формат.

В отконвертированной базе данных (БД) ЭК индексы УДК были занесены в поле 675 (Универсальная десятичная классификация), а ключевые слова – в поле 610 (Неунифицированные тематические термины). Составление КС осуществлялось по ранее разработанной методике. Продолжилась работа со словарем КС, который определяется как «Тематический словарь» ЭК : велся контроль состава словаря, проводилась нормализация КС, устранение омонимии и полисемии осуществлялось при помощи помет (лексиграфически).

В результате нормализации и устранения омонимии и полисемии образовался словарь ключевых слов, представляющий собой алфавитный перечень КС, в котором применяется система ссылок и отсылок.

После многочисленных доработок и изменений, проведения полной редакции словаря КС, в 2001 г. была утверждена переработанная «Методика составления ключевых слов для электронного каталога» (и их ввода в ЭК) на документы на русском и белорусском языках.

КС на документы, изданные на белорусском языке, составляются на белорусском и русском языках. При невозможности проверить правильность перевода некоторых фамилий, названий растений и т.д. с белорусского языка на русский, КС даются только на белорусском языке.

ЦНБ НАН Беларуси располагает достаточно обширным фондом естественнонаучной литературы на иностранных языках, который постоянно пополняется новыми документами. При составлении методики координатного индексирования документов на иностранных языках учитывалось, что задачей ЦНБ НАН Беларуси является обеспечение потребности ученых в научной информации, которая освещает передовые достижения и новые направления в современной науке. Для многоаспектного выражения смыслового содержания документов по научной тематике,

изданных на английском, французском, немецком и польском языках, КС приводятся на русском языке и языке оригинала.

В иностранных документах зачастую содержатся описания новых явлений и понятий, употребляются новые научные термины, эквивалента которым еще нет в русском языке. В случае отсутствия точного перевода научного термина в отраслевых словарях, КС даются только на языке оригинала.

На документы, изданные на других иностранных языках, КС даются на русском языке. Эти решения отражены в «Методических рекомендациях по определению и введению ключевых слов в электронный каталог документов на иностранных языках».

В настоящее время ведется работа по созданию системы корпоративной каталогизации (СКК) изданий и ведения сводного электронного каталога (СЭК) Беларуси. Информационную основу СКК составляет СЭК, который представляет собой базу данных (БД), содержащую библиографические (БЗ) и авторитетные (АЗ) записи, представленные в BELMARC-формате. Записи должны формироваться всеми участниками СКК на основе принятых соглашений. В технологию корпоративного взаимодействия вовлекаются локальные информационные ресурсы библиотек-участниц, в т.ч. и ЦНБ НАН Беларуси, как участницы СКК. Структура для поддержки функционирования СКК и СЭК создается на базе Национальной библиотеки Беларуси (НББ), поэтому в основу регламентирующих документов СЭК заложены нормативные документы НББ. В связи с тем, что в методике координатного индексирования ЦНБ НАН Беларуси и НББ существуют значительные различия, специалисты двух библиотек в настоящее время пытаются выработать новую методику, которая позволит обеспечить точное и полное описание документного потока посредством КС и АЗ-дескрипторов в удобной для пользователей форме. Это большая и сложная работа, при которой необходимо учесть все произошедшие изменения, упорядочить состояние тематического словаря ЭК. Вновь необходима редакция «Тематического словаря» ЭК Библиотеки, т.к. в результате объединения локальных информационных ресурсов библиотек (ЦНБ и НББ) появились синонимы, выраженные КС и АЗ-дескрипторами, различные варианты названий организаций, родовых имен, географических наименований и т.д.

Из тематического словаря выводятся КС-Персоналии, которые заменяются АЗ и, соответственно, заносятся в поле 600 Имя лица, используемое как предмет.

Такую же работу ведем и с КС на наименование организации, родовое имя, географическое наименование, наименование торговой марки, заглавие, используемое как предмет. Эти КС, при наличии АЗ в СЭК или тематическом словаре ЭК Библиотеки, удаляются. Заменяющие их АЗ заносятся в соответствующие поля 6-Блока определения тематики.

Возникают некоторые сложности по заполнению поля 606 Наименование темы, используемое как предмет. Перестраиваясь, в целях унификации БЗ, под методику координатного индексирования для СКК, мы сталкиваемся с

разрозненностью использования АЗ-дескрипторов при индексировании схожих документов. Нередко встречается использование общего понятия для обозначения частного, игнорируя имеющийся конкретный дескриптор.

Замена КС на АЗ проводится, как программно, так и вручную. Программно – когда к КС привязано более 15 БЗ.

При этом проверяется наличие АЗ в ЭК Библиотеки и, чтобы избежать ошибок, в заявке указываются идентификатор удаляемого КС (*u133716*) и заменяющей ее АЗ (*BY-NLB-ar2456*).

Для унификации БЗ НББ предложила ЦНБ НАН Беларуси отказаться от составления КС на иностранных языках (оставив только имена лиц). Но, учитывая, что отказ от КС, сформулированных на языке оригинала, на иностранные научные документы может ограничить поисковые возможности пользователей, такое решение не было принято и КС на иностранных языках составляются на все наименования предметов.

В настоящее время ЦНБ НАН Беларуси проходит довольно сложный этап – выработка оптимальных решений, устраивающий всех участников СКК.

Несмотря на возникающие сложности при формировании лингвистических средств ЭК, корпоративное взаимодействие библиотек – участниц СКК, способствует развитию и совместимости ЛО. Формирование по единым принципам лингвистических средств (особенно дескрипторных ИПЯ) позволяет пользователям довольно легко ориентироваться как в локальных ЭК, так и в СЭК, вести более успешно поиск и получать максимум интересующей информации, а также оптимизирует работу сотрудников библиотеки.

Библиография:

1. Антопольский, А. Б. Лингвистическое обеспечение электронных библиотек / А. Б. Антопольский // Электронные библиотеки [Электронный ресурс] : Рос. науч. электрон. журн. – 2002. – Т. 5, вып. 2. – Электрон. тект. дан. (1 файл : 47218 байт). – Библиогр.: 10 назв. – Загл. с экрана. – Режим доступа: <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2002/part2/antopolskii>.
2. Астапович, Л. Л. Создание библиографических записей в электронном каталоге Центральной научной библиотеки им. Я. Коласа НАН Беларуси / Астапович Л. Л., Ловейкина Л. В., Прудникова Л. И. // Б-ки нац. акад. наук: проблемы функционирования, тенденции развития : науч.-практ. и теорет. сб. / Междунар. ассоц. акад. наук, совет директоров науч. б-к и информ. центров. – Киев, 2007. – Вып. 5. – С. 196-203.
3. Дрозд, О. М. Внедрение национального коммуникативного формата BELMARC в работу ЦНБ им. Я. Коласа Национальной академии наук Беларуси / Дрозд О. М., Астапович Л. Л., Грядовкина М. Ю. // Науч. и техн. б-ки. – 2007. – №7. – С. 39-42.

4. Зайцева, Е.М. Лингвистическое обеспечение автоматизированных библиотечно-информационных систем: современные требования и направления развития / Е.М. Зайцева // Науч. и техн. б-ки. – 2000. - № 3. – С.54-57
5. Методика составления ключевых слов для электронного каталога / Нац. акад. наук Беларуси, Центр. науч. б-ка им. Я. Коласа ; сост.: Л.Ф. Кравцова, Л.Л. Астапович, Л.А. Авгуль ; науч. ред. Н.Ю. Березкина. – Минск, 1999. – 10 с.
6. Методических рекомендациях по определению и введению ключевых слов в электронный каталог документов на иностранных языках каталога / Нац. акад. наук Беларуси, Центр. науч. б-ка им. Я. Коласа ; сост.: Солодова Л.С., Шилова Л.А., Шанина Я.Е.; науч. ред. Березкина Н.Ю. – Минск, 1998. – 14 с.
7. Щербинина, Г.С. Координатное индексирование в свете интеграционных процессов / Г.С. Щербинина. – Науч. и техн. б-ки. – 2002. - № 10. – С. 72 – 76