

СМЫСЛОВАЯ ОБРАБОТКА ДОКУМЕНТНОГО ПОТОКА НББ. СОВРЕМЕННОЕ СОСТОЯНИЕ И ПЕРСПЕКТИВЫ РАЗВИТИЯ

*Пугачева Светлана Александровна, зав. отделом индексирования
документов НББ*

Прогресс неуклонно идет вперед, выдвигая свои требования. Подчиняясь им, происходят изменения во всех сферах деятельности человека. Стремительное развитие современных информационно-коммуникационных технологий внесло существенные изменения и в область создания информационных ресурсов.

Уже начиная с 60-х гг. XX в. в мире наблюдается нарастание потока научной и культурной информации. Оно происходит такими темпами, что работа по систематизации, логическому и любому другому упорядочиванию знаний становится жизненно важной для создания условий, необходимых для дальнейшего существования науки и культуры.

Крупные институты, в том числе и библиотеки, вынуждены сами отображать развитие общества и культуры. Так, например, индексирование, библиографическое описание, аннотирование книг в библиотеках, казалось бы, только с целью облегчения нахождения необходимой информации, по сути, служит делу управления культурой. А значит тот, кто составляет каталоги и картотеки, осуществляет тем самым организацию самих знаний, в них заключенных, и формирует, пусть в упрощенной форме, их структуру.

Благодаря продвижению прогресса в 1992 г. в Национальной библиотеке Беларуси (НББ) стал зарождаться новый, соединяющий в себе всю точность алфавитного, доступность предметного и обстоятельность систематического, электронный каталог (ЭК). Принятие решения о создании ЭК привело к активной работе по выбору средств лингвистического обеспечения.

В нашей стране, как известно, отсутствует единый научно-методический и информационный центр по разработке и ведению лингвистических средств. Национальная библиотека Беларуси, стоя у истоков создания ЭК, столкнулась с проблемой выбора информационно-поискового языка (ИПЯ). НББ всегда стремилась иметь развитую структуру лингвистического обеспечения, которая включала бы ИПЯ различных типов: язык библиографического описания, классификационные, ИПЯ ключевых слов и предметных рубрик со своими методиками индексирования. Это связано с тем, что в универсальной публичной библиотеке, какой мы являемся, входной документный поток имеет широкий видовой диапазон, состав читателей разнообразен и соответственно ему широк спектр поступающих запросов. Кроме того, общеизвестна зависимость между качеством применяемых в библиотеке лингвистических средств и уровнем обслуживания читателей.

Поэтому при выборе ИПЯ специалисты библиотеки исходили из того, что потребителю информации необходимо предоставить широкие

возможности поиска. В связи с этим было решено использовать следующие поисковые средства:

- классификационные языки (УДК, ББК, ГР НТИ);
- вербальные языки (дескрипторный язык с контролируемой лексикой и язык предметных рубрик). Язык предметных рубрик приняли только для пополнения карточного каталога.

Создавая ЭК, каталогизаторы не сомневались в выборе классификационных ИПЯ. Решение вести классификационный язык УДК было принято без колебаний. Важную роль в этом сыграло наличие в библиотеке карточного систематического каталога, организованного по таблицам УДК. ББК решено было использовать в отношении индексирования картографических и нотных изданий. В качестве поискового языка предложили использовать ГР НТИ, учитывая заложенные в нем возможности проводить многоаспектное индексирование путем присвоения документу нескольких равноправных рубрик, а также – возможное использование рубрикатора в международном обмене информацией.

Встал вопрос еще об одном – вербальном языке. После изучения вопроса выбор был сделан в пользу дескрипторного ИПЯ. Предпочтение было отдано ему в связи с тем, что в отличие от языка предметных рубрик, он позволяет осуществлять детальный поиск информации на определенную тему в машиночитаемых библиографических базах данных.

Для индексаторов библиотеки абсолютно новыми инструментами индексирования были два: рубрикатор ГР НТИ и дескрипторный язык. Освоение методики индексирования по рубрикатору ГР НТИ не вызвало особых трудностей. Наибольшие сложности возникли при освоении дескрипторного языка.

Что касается классификационных языков, то работа с ними регламентирована ГОСТами 7.49–84 и 7.59–2003, таблицами классификации и вызывает затруднения лишь в принятии классификационного решения. При условии соблюдения принципов и методики индексирования, регулярного знакомства с изменениями и дополнениями к таблицам, индексатор почти всегда без колебаний может точно отразить содержание документа с помощью этих ИПЯ. Кроме того, большую помощь и поддержку в решении сложных вопросов индексирования оказывает коллектив, занимающийся ведением и совершенствованием Универсальной Десятичной Классификации (УДК) на русском языке в структуре ВИНТИ РАН. Последние полтора года отдел индексирования документов Национальной библиотеки Беларуси регулярно задает вопросы по индексированию документов и сообщает о замеченных ошибках в русском варианте УДК на форуме УДК.

Иначе обстоит дело с дескрипторным ИПЯ. При его использовании также существуют свои правила, действует ГОСТ 7.66–92, но работу с дескрипторами и ключевыми словами трудно регламентировать, тем более выдвинуть к ней требования стандарта. Здесь незаменимо особое языковое чутье, лексическая интуиция и уровень интеллекта индексатора. Потребовались определенные знания и немалые профессиональные усилия

со стороны индексаторов, чтобы овладеть процессом координатного индексирования документов, именно так называется процесс выражения содержания документов с помощью ключевых слов и словосочетаний. Учитывая, что при координатном индексировании используются преимущественно слова и словосочетания из словарей нормализованной лексики, предстояло создать дескрипторный словарь. В поисках готовых словарей специалисты (НББ) посетили многие крупные библиотеки и информационные центры на территории бывшего СССР. После долгих размышлений решено было создать свой словарь ключевых слов.

Существует несколько необходимых для построения словаря способов получения начального массива ключевых слов:

- 1) накопление ключевых слов и словосочетаний при свободном индексировании документов;
- 2) лексикографическая обработка таблиц классификации;
- 3) лексикографическая обработка готовых словарей.

Был выбран первый способ. За 1992–1994 гг. наш словарь насчитывал около 101 тыс. ключевых слов, соответствующих содержанию документов, однако содержал множество ненормализованных терминов, встречались орфографические ошибки, одни и те же понятия были представлены в разном словесном выражении. Происходило это по причине клавиатурного набора ключевых слов индексаторами и отсутствия инструктивно-методических материалов по работе с ключевыми словами. Первоначально при формировании дескрипторного словаря использовалась методика индексирования документов ИНИОН, но она не давала решений по нормализации ключевых слов, извлекаемых из содержания документов.

Все это вызвало необходимость разработать свою, библиотечную технологическую инструкцию. В 1994 г. в первой редакции вышла *“Инструкция по индексированию входного потока документов для формирования массивов электронного каталога”*. Ее составитель – является кандидат филологических наук, доцент Белорусского университета культуры и искусств С.В. Ильина. В инструкции подробно отражены правила описания основного содержания документов с помощью информационно-поисковых языков, принятых в библиотеке. Особое внимание обращено на методику координатного индексирования. Эта инструкция является руководством для НББ и по сей день. Необходимые методические решения, отсутствующие в инструкции, принимаются индексаторами совместно со службой ведения ИПЯ и вносятся в картотеки методических решений.

Производя автоматический ввод в процессе координатного индексирования, индексатор извлекает термины, необходимые для отражения содержания данного документа, из массива уже имеющихся в локальной базе данных авторитетных записей и вносит их в поисковый образ документа (ПОД). В случае отсутствия требуемого термина в базе индексатор выписывает его на рабочий листок и вместе с документом передает для дальнейшей нормализации в службу ведения ИПЯ редактору-лингвисту. Если термин принимается для внесения в базу данных, то он с

соответствующей пометой возвращается индексатору. Решение редактора-лингвиста можно опротестовать, внося веские аргументы в пользу утверждения данного термина. В случае принятия решения о необходимости включения нового термина он вводится в базу в виде заголовка авторитетной записи или в вариантной форме. Таким образом, благодаря этой методике, ежегодно локальная база данных увеличивается на 3,5–4 тыс. новых лексических единиц.

Постепенно в результате осуществления всех вышеперечисленных видов индексирования обозначился образ ЭК, содержательная часть которого представлена индексами УДК, рубриками ГР НТИ и дескрипторами.

Национальная библиотека Беларуси стоит на пороге освоения новой автоматизированной информационно-библиотечной системы (АБИС). Для ЭК индексируются книги, авторефераты диссертаций, ноты, изодокументы, аудио- и видеодокументы, картографические издания, документы международных организаций, периодические издания, электронные ресурсы, статьи.

Большие требования в наше время предъявляются к качеству информации, повышается интеллектуальный уровень информационных ресурсов. Важно сделать информацию точной и исчерпывающей, а это невозможно без повышения качества индексирования.

В НББ документный поток обрабатывается с помощью нескольких ИПЯ и имеется возможность сравнить преимущества и недостатки этих языков и их влияние на информационный поиск. Поиск осуществляется преимущественно по дескрипторам. Происходит это, на наш взгляд, по причине кажущейся простоты обращения с данными ЛЕ. На самом деле этот вид поиска ограничивает диапазон охвата документов, так как состав дескрипторов, соответствующий основному содержанию конкретного документа, не позволяет осуществить более широкий поиск. Например, возьмем для поиска тему “Библиотеки”:

Пользуясь дескрипторами как средством поиска при данном запросе, мы не найдем полный состав документов, касающихся конкретных библиотек, к примеру, Российской национальной библиотеки. Документы, по содержанию относящиеся к этому учреждению, в ЭК будут находиться за словосочетанием, отражающим полное наименование данной библиотеки.

Как видим, для того, чтобы получить ответ, релевантный запросу, необходимо предусмотреть поиск по всем возможным типам и видам библиотек. Это связано с тем, что, руководствуясь ГОСТом 7.66–92 по координатному индексированию, вышестоящие понятия в отношении смыслового анализа документов применяются только в исключительных случаях: если термин очень специфичен и не дает полного представления об основном содержании документа.

Например, для ПОДа с термином “экзогенные ферменты” индексатор обязан дать дополнительно вышестоящий термин – “комбикорма”; для термина “открытое овальное отверстие” – “сердце”; для термина

“зернопропашные севообороты” необходим термин “земледелия системы” и т. д.

Попробуем произвести поиск предложенной выше темы “Библиотеки” по индексу УДК – 027.

Как видно, на экране появляется информация о документах, многоаспектно отражающих данную тему, – сведения о библиотеках всех видов и типов.

По рубрике ГР НТИ лучше всего проводить поиск массива документов по определенному направлению, без учета нюансов темы. Например, получить информацию об общем количестве документов по теме “Экология”, полученных библиотекой за определенный период, разумнее всего будет по рубрике ГР НТИ 34.35 (Экология) и 87 (Охрана окружающей среды. Экология человека).

В то время как по ключевым словам и индексам УДК эту задачу выполнить проблематично, по рубрикам ГР НТИ поиск осуществляется вполне доступно. Поиск по рубрикам ГР НТИ и индексам УДК предполагает хорошее знание таблиц УДК и рубрикатора ГР НТИ и наличие баз данных УДК и ГР НТИ на компьютерах в зоне обслуживания. Таким образом, для получения релевантной информации необходимо использовать результаты всех видов индексирования.

Новые возможности для осуществления качественной смысловой обработки документов и проведения тематического поиска предоставляет национальная база данных авторитетных записей, созданная в НББ.

Большие перспективы открывает перед библиотечным миром в плане информационного развития мирового сообщества программа ЮНЕСКО “*Информация для всех*”, ориентируя на “новые глобальные цели и вполне определенные действия – способствовать повышению информационной культуры, строить национальную информационную политику и инфраструктуру так, чтобы свести к минимуму опасный разрыв между информационно богатыми и информационно бедными” и продвигая сетевое взаимодействие в сфере информации и знаний на местном, национальном, региональном и международном уровнях.

В последнее время много говорится об обеспечении доступа к достоверной информации, об опасности, грозящей информационному наследию ввиду большого количества недоброкачественной информации, вольных пересказов научных изданий и т. д. В свете этих сообщений весьма актуальной становится работа библиотек по предоставлению исчерпывающей и точной информации. И в этом деле смысловая обработка документов играет немаловажную роль. Глубина и детальность индексирования при этом становятся главными составляющими полной и достоверной информации о содержании документов, предоставляемой потребителям.