

ЛИНГВИСТИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ЭЛЕКТРОННОГО КАТАЛОГА



■ **Ирина СИЗЫХ,**
главный библиотекарь отдела научной и технической
обработки Научной библиотеки Государственного
педагогического университета

Создать условия, при которых читатель может получить доступ к информационно-поисковым системам библиотек и вести эффективный поиск в них, помогает лингвистическое обеспечение и его основная составляющая — информационно-поисковые языки. Лингвистическое обеспечение — это комплекс мероприятий, направленных на разработку языковых средств автоматизированной системы, а также сами средства. Данный термин стал применяться для обозначения всего комплекса языковых средств, используемых в информационно-поисковой системе: собственно информационно-поискового языка; методики индексирования документов и запросов на них; методики составления тезаурусов, различных правил, вспомогательных средств создания и ведения информационно-поискового языка.

Задача библиотек состоит не только в том, чтобы собрать в своих фондах как можно больше документов по своему профилю, но и сделать их доступными для пользователя. Этому способствуют каталоги, базы данных, библиографические издания. Информация в них должна быть систематизирована и представлена в таком виде, который позволяет осуществлять быстрый поиск.

Средством свертывания информации и смысловой обработки документов является информационно-поисковый язык. Практически все технологические процессы и операции, выполняемые в библиотеках, базируются на использовании информационно-поисковых языков и

В современных условиях не вызывает сомнений необходимость формирования информационной культуры у будущих специалистов. Задача электронного каталога библиотеки — предоставить пользователю максимально благоприятные условия для поиска необходимой информации. Она может быть решена при существовании развитой системы лингвистического обеспечения электронного каталога.

других лингвистических средств. Следовательно, качество лингвистического обеспечения самым прямым образом влияет на качество производимых библиотекой информационных продуктов и услуг.

Лингвистическими средствами реализуется одна из важнейших функций библиотеки — обеспечение доступа потребителей к информационным ресурсам. Стремительное внедрение в практику нашей работы новых технологий не могло не повлиять на развитие лингвистического обеспечения, породив целый комплекс проблем, требующих своего решения.

В настоящее время быстрый рост количества, сложности и срочности информационных запросов, вызванный непрерывным увеличением численности специалистов и сложности решаемых ими научно-технических задач, а также быстрый рост фонда, в котором необходимо производить информационный поиск, делает задачу автоматизации процедуры информационного поиска всё более актуальной. Своевременная, точная и полная выдача учебных и научных документов в ответ на многочисленные и разнообразные информационные

запросы приобретает характер задачи массового обслуживания, которая может быть успешно решена лишь путем применения соответствующих средств механизации и автоматизации. Для того чтобы процедуру сопоставления поисковых образов документов с информационным запросом можно было выполнить механически, не вникая в их смысл, необходимо их выразить на определенном языке, слова и фразы которого обладали бы смысловой однозначностью. Таковым является информационно-поисковый язык.

Информационно-поисковый язык должен:

1) полно и точно передавать содержание, а при необходимости — форму и назначение документа, отражаемого в данной информационно-поисковой системе;

2) обеспечивать однозначное толкование терминов индексирования;

3) допускать внесение изменений (дополнений и исправлений);

4) обеспечивать достаточную простоту и удобство индексирования, информационного поиска и ведения данной информационной поисковой системы;



Коллектив отдела научной и технической обработки

5) отражать современное состояние терминосистемы данной области знания.

Библиотека Новосибирского государственного педагогического университета ведёт электронный каталог с 2000 г., и начинали мы работу в программе «Библиотека 4.0», причём никаких методик заполнения полей у нас не было. В «Библиотеке 4.0» работа велась в том же коммуникативном формате MARC-21, поэтому конвертация библиографических записей в программу «MARC SQL» прошла относительно безболезненно. Но в связи с новыми потребностями (вступление в корпорацию «MARC», связь с другими библиотеками и необходимость в дальнейшем переходе на электронный заказ изданий) появилась необходимость расширить библиографическое описание изданий по сравнению с описанием в программе «Библиотека 4.0» (дополнить: персоналии, содержание, сведения о произведениях, упомянутых в издании, и т. д.). Поэтому после конвертации, пересматривая все записи, дополняем их необходимыми сведениями.

Мы стремимся иметь развитую структуру лингвистического обеспечения, включающую в себя: классификационные информационно-поисковые языки, вербальные информационно-поисковые языки, инструктивно-методическую документацию.

Классификационные информационно-поисковые языки предназ-

начены для индексирования документов и информационных запросов посредством понятий и кодов какой-либо классификационной системы (например, ББК, УДК). Как правило, они используются не столько для формирования поискового образа запроса, сколько для определения местонахождения документа в библиотеке.

Следующим элементом лингвистического обеспечения являются вербальные информационно-поисковые языки. К ним относят язык предметных рубрик и язык ключевых слов. Так, при ведении электронного каталога неизбежно возникает вопрос, каким информационным поисковым языком воспользоваться для раскрытия предметного содержания документов: языком предметных рубрик или языком ключевых слов, или тем и другим одновременно?

Язык предметных рубрик предназначен для индексирования документов и информационных запросов посредством предметных рубрик. В его основе лежит алфавитный перечень предметных рубрик, представляющий собой краткую формулировку темы на естественном языке.

Язык ключевых слов служит для координатного индексирования документов и запросов посредством ключевых слов.

Лингвистические средства по праву относятся к числу важнейших системообразующих элементов любой информационно-библиотечной системы. Но существуют и пробле-

мы. Одна из них — проблема выбора информационно-поискового языка.

Наиболее эффективным среди вербальных информационно-поисковых языков является язык ключевых слов, свободно выбираемых из лексикона специалистов-пользователей и вводимых в базы данных в нормализованной лексикографической форме. Нужно сказать, что вид поиска по ключевым словам — один из наиболее оцененных не только индексаторами, но и, конечно, пользователями.

Пользователи заинтересованы оперировать ключевыми терминами, которые для них являются наиболее понятным поисковым средством. При поиске по языку ключевых слов от читателя не требуется знаний таблиц классификации. Практика показывает, что при поиске с помощью ключевых слов удается найти большее число релевантных документов, чем при обращении к предметным рубрикам.

К сожалению, в настоящее время не существует надежной нормативно-справочной базы лингвистического обеспечения баз данных на государственном уровне. В частности, существует ГОСТ 7.59—2003 «Индексирование документов. Общие требования к систематизации и предметизации», который устанавливает общие правила индексирования документов. А те вопросы, которые актуальны для многих библиотек, в том числе и для нашей, связанные с обеспечением однообразия индекс-

сирования, не отражены в нормативных документах. Для обеспечения высокого качества индексирования литературы и, вследствие этого, лингвистического обеспечения баз данных необходимо большее внимание уделять информационным потребностям и запросам пользователей. Следует изучать результаты информационного поиска, в первую очередь — анализировать неудовлетворенные запросы читателей.

Результаты индексирования существенным образом связаны с качеством поиска в любой информационно-библиотечной системе и в значительной мере определяют перечень и качество предлагаемых библиотекой услуг. Поэтому повышение эффективности индексирования и, следовательно, качества лингвистического обеспечения было и остается первоочередной задачей каждой библиотеки.

Ситуация усугубляется и тем, что число библиографических записей, создаваемых в электронных каталогах, растёт с каждым днём. Нужно заметить, что чем больше проходит времени, тем масштабнее требуется редактирование. О его необходимости свидетельствует хотя бы то, что в данный момент при индексировании документа по ключевым словам большую роль играет субъективный фактор, то есть полнота и точность раскрытия содержания документа прямо зависят от уровня квалификации библиотечного работника, его знакомства с проблематикой отрасли.

Показателем высокого качества индексирования является составление словарей ключевых слов по единой методике, при этом учитываются особенности самого документа, так как он предназначен в первую очередь для потенциальных потребителей конкретной библиотеки. Для более полного отражения содержания документов, как правило, достаточно 10—12 ключевых слов, не являющихся синонимами, в зависимости от документа их может быть как меньше, так и больше.

Мы до сих пор спорим о количестве ключевых слов. Иногда наши ключевые слова не вмещаются на карточку. Полагаю, в данной ситуации мы должны быть ориентированы прежде всего на электронный каталог и на потребности читателей.

Пользователи электронного каталога также должны выполнять процедуру координатного индексирования, формируя свой запрос на основе ключевых слов. И если биб-

лиотекари имеют определённые знания в данной области, то пользователи при проведении информационного поиска обречены на «свободное плавание» в огромном массиве данных без специальных средств навигации. При этом им не гарантировано получение искомой информации.

Любой словарь ключевых слов со временем нуждается в редактировании, потому что обрастает большим количеством синонимов, омонимов, нарушений правил при составлении ключевых слов, введением понятий в единственном и множественном числе. Мы записываем ключевые слова во множественном числе, так как состоим в корпорации MARC. Но при этом есть много примеров тому, что одно и то же слово во множественном и единственном числе имеет совершенно разные значения. Например, рассмотрим произведение Бахтина М.М. «Эпос и роман» и роман «Война и мир». В первом случае роман в ключевом слове будет определять «жанр литературы», а во втором случае мы уже напишем «романы как произведения художественной литературы»; «гражданское право» и «гражданские права» — в одном случае речь идёт об отрасли права, в другом — о правах граждан. Иногда один и тот же термин может употребляться только в единственном числе либо только во множественном. Существительные, употребляемые в разных отраслях знания то в единственном, то во множественном числе, приводятся в той форме, которая соответствует принятой в отрасли. «Вода» как химическое вещество; «воды» как природные воды. Эти нюансы сложно выявить и исправить при редактировании словаря «Ключевых слов», так как книги уже нет в руках, поэтому приходится просматривать каждую запись и исправлять ошибки, что занимает очень много времени.

Необходимость в редактировании словарей у нас возникла уже в процессе работы в программе. Объём работы увеличивается с каждым годом, нагрузка возрастает, пользование словарями значительно сокращает время на обработку информации, но при этом необходимо, чтобы в словарях был порядок: если кто-либо пропустит какой-то знак или делает лишний пробел (это самое малое), то может случиться, что запись будет копировать другой сотрудник, и ошибок будет появляться всё больше и больше. Созданные автоматически словари поисковых элементов



Логотип Библиотеки

нуждаются в редактировании ещё по нескольким причинам: в связи с внесимыми дополнениями и изменениями при вводе библиографической записи, из-за допущенных орфографических ошибок и при всех изменениях, вносимых в ГОСТы.

В программе «MARC SQL» есть большое количество словарей: автор, заглавие, место издания, персоналии, гриф высшей школы и др. Все словари, создаваемые и используемые библиотекарями в программе, относятся к информационно-поисковому языку. При работе мы используем внутренние и внешние словари. Внутренние словари создаются в момент образования самого библиографического описания, и они постоянно расширяются, каждый новый термин вводится в словарь в виде новой дополнительной строчки, то есть если у нас создается новый документ, мы формулируем заглавие, значит, это заглавие заносится в словарь заглавий. То же самое и со словарем авторов: каждый будет занесён в словарь авторов; если данный автор был ранее занесен, мы можем использовать его из словаря и не писать заново. При этом исключается вероятность ошибки и увеличивается скорость библиографического описания.

Рассмотрим словарь серий. При работе в программе «Библиотека 4.0» в нем пользовались сокращениями согласно ГОСТу, но опыт работы и практика показали, что многие читатели не знают правил сокращений, это затрудняло им поиск информации. Было принято решение отредактировать данный словарь: не применять сокращения, убрать кавычки и опускать слово «серия».

В связи с изменениями таблиц классификации или появлением новой информации по индексированию документа возникает необходимость редактирования поля каталожных индексов, то есть замены старых каталожных индексов на новые. Всё это

выполняется очень быстро, применяется групповая корректировка, которая проводится по мере необходимости.

Новые изменения коснулись полного написания всех названий городов: если раньше в описании, согласно ГОСТу, сокращали название таких городов, как Москва, Санкт-Петербург, Ростов-на-Дону, Нижний Новгород, то в новых Российских правилах каталогизации рекомендовано названия городов писать полностью.

При работе в программе «MARC SQL-1.9», а теперь и в 1.10 версии этой программы мы избавили словарь ключевых слов от персоналий, появилось поле «Персоналии». За время работы в это поле было внесено очень много изменений: помимо того, что в него входят фамилия, имя, отчество, годы жизни, мы пробовали дополнить его справочными данными, в связи с чем словарь стал «раздуваться». Возьмём один из примеров: так, про Михаила Васильевича Ломоносова в «Персоналии» библиотекарь мог написать: «русский просветитель, русский поэт, переводчик и русский учёный» и т. д., в зависимости от того, с каким источником работает, и при этом он будет прав. Поэтому было принято решение писать только годы жизни, главное для нас — единообразие в записях.

Кроме того, ещё мы столкнулись с тем, что в источниках один и тот же автор пишется по-разному: например, Гете и Гёте, Белль и Бёлль, Мёрдок и Мэрдок, Шекспир Вильям и Шекспир Уильям, поэтому решено было от менее используемой, неправильной или устаревшей формы фамилии делать ссылку на другую: так, «Шекспир Вильям см. Шекспир Уильям», «Гете см. Гёте». Если раньше при переводе фамилии автора писали Фильдинг, то сейчас пишут Филдинг. При этом в процессе поиска читатель найдёт все записи по данному автору. Такие ссылки всегда были в алфавитном карточном каталоге, мы пришли к необходимости их и в словаре «Персоналии». Далее мы будем продолжать работу по более корректному заполнению полей, и, как следствие, мы надеемся получить более актуальное лингвистическое обеспечение, делающее поиск более эффективным.

Но недостатком таких изменений является уменьшение скорости работы программы, так как в момент сохранения документа создаётся новая запись практически по каждому полю. А в нашем каталоге более

70 тысяч записей на отдельные издания, и при работе задержка в 2—3 секунды очень чувствительна. Поэтому с целью ускорения работы программы мы начали более активно использовать внешние словари.

Внешние словари создаются программистами, для которых отдел обработки готовит информацию о том, какие наиболее типовые формулировки следует в них включать. После того как словарь создан, его можно использовать для формирования соответствующих полей новых библиографических записей. Эти словари используются только для заполнения полей библиотекарями, поэтому скорость работы программы резко увеличивается. Кроме того, использование этих словарей сводит к минимуму возможность появления грамматических ошибок. Мы эти словари часто используем как примеры правильного заполнения полей. Таким образом, у нас были созданы внешние словари: грифы высшей школы, библиография, примечание, код отношений, основные сведения об издании, так как мы выбрали те поля, в которых ограниченное количество значений.

При использовании внешних словарей мы выиграли не только во времени, но и в качестве. Когда записи создаются во внутренних словарях, в них не исключены ошибки, которые могут распространяться дальше при использовании данных записей. А при использовании внешних словарей, если и допущена где-то неточность, то она не будет распространяться.

К лингвистическому обеспечению относят также инструктивно-методическую документацию, которая, к сожалению, очень плохо разработана на государственном уровне. В нашей библиотеке есть «Положение об электронном каталоге». При работе в программе мы придерживаемся ГОСТа и «Методики заполнения полей». После вступления в корпорацию «Межрегиональная аналитическая роспись статей» мы познакомились с документацией, и нам понравилась «Методика заполнения полей» для аналитической росписи статей, на её основе мы разработали свою методику. Методика заполнения полей предназначена для служебного пользования, в ней фиксируются принятые нами решения по сложным случаям. Она очень необходима, так как индексированием документов занимается

не один сотрудник. Она обеспечивает стабильность и единообразие принятых для библиографического описания решений. Если бы данной методики не существовало, то новому сотруднику пришлось бы в каждом случае обращаться к каталогу, чтобы выяснить, как работали предшественники. Мы вынуждены самостоятельно разрабатывать данную методику и с нетерпением ждём печатного варианта Российских правил каталогизации, с которыми мы уже ознакомились в Интернете. И пусть наши решения позже оказывались неверными, мы в процессе работы путём проб и ошибок приходим к верному решению и, применяя групповую корректировку, ошибки ликвидируем, главное для нас — это единообразие во всех описаниях нашего каталога.

Есть много факторов, тормозящих разработку лингвистического обеспечения, а, следовательно, и автоматизацию библиотек. Среди них:

- 1) отсутствие концепции лингвистического обеспечения;
- 2) чрезвычайная сложность и трудоемкость этой работы;
- 3) слабая координация (скорее, полное отсутствие таковой) на межбиблиотечном уровне.

Влияние новых информационных технологий становится всё более определяющим для всех сфер человеческой деятельности. Библиотеки как организации, неразрывно и органично связанные с процессами генерации, накопления и хранения знаний, находятся на переднем крае внедрения новых технологий.

ЛИТЕРАТУРА

Гендина, Н.И. Подготовка библиотечных кадров в сфере лингвистического обеспечения библиотечной технологии: проблемы совместимости и преемственности содержания среднего, высшего и дополнительного образования / Н.И. Гендина, А.Ш. Меркулова // Научные и технические библиотеки. — 2005. — № 4. — С. 24—30.

Скипор, И.А. Лингвистическое обеспечение корпоративных библиотечно-информационных систем и сетей / И.А. Скипор, Е.А. Сбитнева // Научные и технические библиотеки. — 2004. — № 7. — С. 28—40.

Сукиасян, Э.Р. Требования к ИПЯ / Э.Р. Сукиасян // Библиотека. — 2004. — № 4. — С. 73—75.

Шаронова, Н.В. Проблема индексирования полнотекстовых документов по ключевым словам / Н.В. Шаронова, О.В. Канищева // Библиотеки учебных заведений. — 2007. — № 23. — С. 23—32.