

*Сизых Ирина Николаевна,
гл. библиотекарь отдела
научной и технической обработки
научной библиотеки
Новосибирского государственного
педагогического университета*

Лингвистическое обеспечение электронного каталога научной библиотеки НГПУ

В условиях перехода человечества к информационному обществу, компьютеризации учебного процесса, увеличения в числе и объеме сетевых источников информации и создания корпоративных информационных баз данных возникает необходимость формирования у будущего специалиста информационной культуры, заключающейся, в частности, в активном и эффективном взаимодействии с современными информационными технологиями. Это является одним из традиционно сложившихся направлений информационно-библиографической работы библиотеки высшего учебного заведения. Задача электронного каталога библиотеки – предоставить пользователю максимально благоприятные условия для поиска необходимой информации. Эта задача может быть решена при существовании развитой системы лингвистического обеспечения электронного каталога.

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 – 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-поисковый язык, методики индексирования 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 сделать их доступными для пользователя, дать информацию о них. Всему этому способствуют каталоги, базы данных, библиографические издания. Информация в них должна быть систематизирована и представлена в таком виде, который позволяет осуществлять быстрый поиск. Понятие лингвистического обеспечения шире понятия информационно-поискового языка.

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 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 информации к информационным ресурсам. Стремительное внедрение новых информационных технологий в практику работы в последнее десятилетие 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, а также количества и сложности решаемых ими научно-технических задач и быстрый рост фонда документов, в котором необходимо производить информационный поиск, делает задачу автоматизации процедуры информационного поиска все 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. 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. Это 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. обеспечить 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;
5. отражать современное состояние терминосистемы в данной области знания.

Электронный каталог библиотека Новосибирского государственного педагогического университета ведет с 2000 года, и начинали мы работу в программе «Библиотека 4.0», при

этом 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 «МАРК SQL» прошла относительно безболезненно. Но в связи с новыми потребностями (вступление в корпорацию «МАРС», связь с другими библиотеками и необходимость в дальнейшем перехода на электронный заказ изданий) появилась необходимость расширить библиографическое описание изданий по сравнению с описанием в программе «Библиотека-4» (дополнить: персоналии, содержание, сведения о произведениях, упомянутых в издании, и т.д.). Поэтому после конвертации, пересматривая все записи, дополняем их необходимыми сведениями. Для своих читателей мы расширили библиографическое описание. Мы стремимся иметь развитую структуру лингвистического обеспечения.

В состав лингвистического обеспечения включают: классификационные информационно-поисковые языки, вербальные информационно-поисковые языки, инструктивно-методическую документацию.

Классификационные информационно-поисковые языки предназначены 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 относятся 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 проблемы. К ним относится 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 удастся 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 индексирования, 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, 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 первостепенной задачей каждой библиотеки.

Ситуация усугубляется и тем, что число библиографических записей, создаваемых в электронных каталогах, растет с каждым днем. Нужно заметить, что чем больше проходит времени, тем масштабнее будет редактирование, без которого не обойтись. О его необходимости свидетельствует хотя бы то, что в данный момент при индексировании документа по ключевым словам большую роль играет субъективный фактор, т. е. полнота и

точность раскрытия содержания документа прямо зависит от уровня квалификации библиотечного работника, его знакомства с проблематикой отрасли.

Успехом высокого качества индексирования является составление словарей ключевых слов по единой методике, при этом учитываются особенности самого документа, т. к. он предназначен в первую очередь для потенциальных потребителей конкретной библиотеки. Для более полного отражения содержания индексирования 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, прежде всего мы должны быть ориентированы на электронный каталог и на наших пользователей.

Пользователь электронного каталога также должен выполнять процедуру координатного индексирования, формируя свой запрос на основе ключевых слов. И если библиотекарь имеет определенные знания в данной области, то пользователь при проведении информационного поиска обречен на «свободное плавание» в огромном массиве данных без специальных средств навигации. При этом пользователю 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 в словарь понятий в единственном и множественном числе. Мы записываем ключевые слова во множественном числе, т.к. состоим в корпорации МАРС, а в ней было принято решение писать ключевые слова во множественном числе. Но при этом есть много примеров тому, что одно и то же слово во множественном и единственном числе имеет совершенно разное значение. Например, рассмотрим произведение Бахтина М. М. «Эпос и роман» и художественный 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 идет 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», т. к. книги уже нет на 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. Объем работы увеличивается с каждым годом, нагрузка возрастает, пользование словарями значительно сокращает время на обработку информации, но при этом необходимо, чтобы в словарях был порядок: стоит пропустить какой-то знак или сделать лишний пробел (это самое малое), в спешке это будет копировать другой сотрудник, и ошибок в записях будет появляться все больше и больше. Созданные автоматически словари поисковых элементов нуждаются в редактировании еще по нескольким причинам: в связи с вносимыми дополнениями и изменениями при вводе библиографической записи, из-за допущенных орфографических ошибок и при всех изменениях, вносимых в ГОСТы.

В программе «МАРК 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, т.е. если у нас создается новый документ, мы формулируем заглавие, значит, это заглавие заносится в словарь заглавий. То же самое и со словарем авторов: каждый автор будет занесен в словарь авторов; если данный автор был ранее занесен в этот словарь, мы можем использовать его из словаря и не писать заново. При этом исключается вероятность ошибки и увеличивается скорость библиографического описания.

Рассмотрим один из словарей - словарь серий. При работе в программе «Библиотека 4.0» в этом словаре 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». Часто библиографическое описание имеет две и более серии, или серия делится на подсерии; в этом случае решено каждую серию заносить в повторяемое поле «серия», а каждую подсерию в свое поле, что значительно улучшает поиск информации по данному полю. Например, у серии «Высшее профессиональное образование» есть несколько подсерий: «Педагогические специальности», «Иностранные языки», «Менеджмент» и т.д. И по каждому из данных заглавий можно вести поиск.

В связи с изменениями таблиц классификации или появлением новой информации по индексированию документа возникает необходимость редактирования поля каталожных индексов, т.е. замены старых каталожных индексов на новые (например, в 63 отделе, 63.52

(2=Рус) меняли на 63.521 (=411) (этнология, русские)). Все 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. В новых Российских правилах каталогизации рекомендовано названия городов писать полностью. Групповой корректировкой все наименования городов были приведены в полной форме.

При работе в программе «МАРК 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 вносятся фамилия, имя, отчество, годы жизни, мы пробовали дополнить его справочными данными, в связи с этим словарь стал «раздуваться». Возьмем один из примеров: так, про Ломоносова Михаила Васильевича в «Персоналии» библиотекарь мог написать: русский просветитель, русский поэт, переводчик и русский ученый и т.д., в зависимости от того, с каким источником работает библиотекарь, и при этом он будет прав. Поэтому было принято решение писать только годы жизни, главное для нас - единообразие в записях.

Кроме того, еще 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 найдет все записи по данному автору. Такие ссылки всегда были в алфавитном карточном каталоге, мы пришли к необходимости таких ссылок и в словаре «Персоналии». Далее мы будем продолжать работу по более корректному заполнению полей, и, как следствие, мы надеемся получить более актуальное лингвистическое обеспечение, дающее лучшие результаты поиска и решение многих проблем.

Но недостатком такой формы работы является уменьшение скорости работы программы, т.к. в момент сохранения документа создается практически по каждому полю новая запись. А в нашем каталоге более 70 000 тыс. записей на отдельные издания, и при работе задержка в 2-3 сек. очень чувствительна. Поэтому с целью ускорения работы программы мы начали более активно использовать внешние словари.

Внешние словари создаются специальным образом программистами, для них отдел обработки готовит информацию, 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 из программы 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, т.е. мы выбрали те поля, в которых ограниченное количество значений.

Так, например, раньше словарь «Библиография» мы заполняли так, как было написано в источнике: библиография, краткая библиография, список использованной литературы, список литературы, литература, цитируемая литература и т.д. При этом словарь стал просто огромным. Обсудив все примеры, мы приняли решение писать просто «библиография». При этом из внешнего словаря копируется запись, и дописываются только нужные страницы.

При использовании внешних словарей мы выиграли не только во времени, но и в качестве. Когда создаются записи во внутренних словарях, в них не исключаются ошибки (невнимание), и они распространяются дальше при использовании данной записи из словаря. А при использовании внешних словарей эта ошибка исключается, если и будет допущена неточность в какой-то записи, то она не распространяется.

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 ее основе мы разработали свою методику. Методика заполнения полей – методика служебного пользования, в которой фиксируются принятые нами решения по сложным случаям. Она очень необходима, так как индексированием документов занимается не один сотрудник. Такая методика необходима, она обеспечивает стабильность и единообразие принятых для библиографического описания решений. Все методические решения, принятые нами в процессе работы, мы вносим в данную методику. И если бы данной методики не существовало, то новому сотруднику пришлось бы в каждом случае обращаться в каталог, чтобы выяснить, как работали предшественники. Мы вынуждены самостоятельно

разрабатывать данную методику и с нетерпением ждем печатного варианта Российских правил каталогизации, с которыми мы уже ознакомились в Интернете. И пусть наши решения позже оказывались неверными, мы в процессе работы путем 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 все 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. От того, как мы сможем внедрить и использовать новые информационные технологии, зависит наше будущее.

Список используемой литературы:

1. 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.
2. Скипор И. Л. Лингвистическое обеспечение корпоративных библиотечно-информационных систем и сетей / И. Л. Скипор, Е. А. Сбитнева // Научные и технические библиотеки. – 2004. - № 7. – С. 28-40.
3. Шаронова Н. В. Проблема индексирования полнотекстовых документов по ключевым словам / Н. В. Шаронова, О. В. Канищева // Библиотеки учебных заведений. – 2007. - № 23. – С. 25-32.
4. Сукиасян Э. Р. Требования к ИПЯ / Э. Р. Сукиасян // Библиотека. – 2004. - № 4. – С. 73-75.