

УДК 025.4+ 025.34:004

Г. А. Скарук,кандидат педагогических наук,
старший научный сотрудник ГПНТБ СО РАН**ЯЗЫКИ ТЕМАТИЧЕСКОГО ПОИСКА
В ЭЛЕКТРОННЫХ КАТАЛОГАХ НАУЧНЫХ БИБЛИОТЕК:
ТЕНДЕНЦИИ, ПРОБЛЕМЫ, ПОДХОДЫ К РЕШЕНИЮ**

В статье описаны тенденции и проблемы развития информационно-поисковых языков в электронных каталогах научных библиотек.

Ключевые слова: информационно-поисковые языки, электронные каталоги, научные библиотеки.

Под статус *научных* подпадают библиотеки различного уровня, величины, ведомственной принадлежности: от крупнейших федеральных, областных и краевых до библиотек отдельных научно-исследовательских учреждений системы РАН. Соответственно различаются объем и содержание их фондов, контингент читателей/пользователей (от массового читателя до исследователей высшей научной квалификации), характер читательских запросов. Как известно, именно эти характеристики определяют требования к лингвистическому обеспечению (ЛО) их электронных каталогов (ЭК). Между тем принадлежность к категории научных подразумевает приоритет интересов именно научной аудитории перечисленных библиотек. Это предполагает ряд общих подходов к составу и свойствам ЛО ЭК.

В Государственной публичной научно-технической библиотеке Сибирского отделения Российской академии наук (ГПНТБ СО РАН) было проведено исследование требований к процессу и результатам поиска в карточных и электронных каталогах различных категорий пользователей [2, 8]. Особое внимание уделялось мнениям научных работников. Необходимо отметить, что, по результатам опроса, именно и только при научных целях поиска его сфера в любом случае не может быть ограничена ЭК одной библиотеки. Поэтому для библиотечных учреждений, имеющих статус научных, чрезвычайно важно предоставить своим читателям доступ к профильным базам данных (БД), сводным ЭК и каталогам других

библиотек в дистанционном режиме. В этих поисковых системах могут использоваться как языковые средства разных типов, так и различные варианты одних и тех же информационно-поисковых языков (ИПЯ). Очевидно, что лингвистические средства специальных, узкотематических БД (особенно содержащих сведения о статьях из журналов и сборников) будут более специфичны, чем ЛО ЭК универсальных библиотек. В каталогах библиотек конкретных научных учреждений могут применяться самостоятельно разработанные (или доработанные) языковые средства, например, локальные рубрикаторы. Следовательно, в состав нормативно-справочной базы ЛО ЭК научных библиотек обязательно должны быть включены словари различных поисковых языков (не только тех, которые применяются в этой системе), инструкции по их использованию, а по возможности – и таблицы сопряжения ИПЯ. Большую помощь окажут навигаторы, обеспечивающие доступ к ЭК библиотек, содержание фондов которых может заинтересовать читателей, внутрибиблиотечные и виртуальные консультации по языкам и стратегиям систем удаленного поиска. Целесообразно приводить всплывающие подсказки о перечисленных возможностях, информацию о системах онлайн-консультаций и виртуальных справок ведущих научных библиотек непосредственно на собственных библиотечных сайтах. Очевидно, что если подобная работа (за исключением последнего пункта) проводится в крупных библиотеках на регулярной основе, то в небольших академических библиотеках такая помощь читателям предоставляется достаточно редко.

Отдельную и существенную проблему представляет собой *отсутствие единого общепризнанного ИПЯ, обязательного для использования в ЭК всех российских библиотек*. А значит, поиск для конечных пользователей как в ЭК, представленных на сайтах отдельных библиотек, так и в распределенных корпоративных каталогах, затруднен ввиду того, что в каждой библиотеке применяется собственный набор поисковых языков.

В этом направлении развернута серьезная работа, но усилия центральных библиотек не всегда поддерживаются на местах. Среди самых значимых шагов можно назвать создание и поддержание Единого авторитетного файла Российской национальной библиотеки [3], представленного на сайте информационно-библиотечного центра ЛИБНЕТ и используемого в Сводном каталоге библиотек России (СКБР), разработка единой методики предметизации [4], близящийся к завершению выпуск десятитомных Средних таблиц Библиотечно-библиографической классификации (ББК) [1].

Следующий момент нашего опроса, который необходимо выделить в контексте темы – *требования читателей к разнообразию средств и стратегий поиска в библиотечных каталогах при решении научных задач*. Научные работники лидировали по этой позиции, выбирали наиболее разнообразные и сложные поисковые языки и стратегии, часто при поиске использовали несколько ИПЯ. В связи с этим хочется обсудить две взаимосвязанные тенденции развития ЛО ЭК:

▣ *применение в ЭК широкого набора различных по структуре, лексическому наполнению и свойствам ИПЯ;*

▣ становление комплексного подхода в ведении и использовании языков тематического поиска.

Об использовании в составе ЛО электронных каталогов научных библиотек нескольких разноплановых лингвистических средств с уверенностью можно говорить как об устойчивой тенденции, возникшей еще на этапе становления российских ЭК. Отметим наиболее характерные черты:

◎ преимущество в выборе языковых средств между карточными и электронными каталогами;

◎ расширение состава ЛО ЭК по сравнению с ЛО системы карточных каталогов;

◎ совместное использование ИПЯ разных типов, как правило, иерархических классификаций и вербальных поисковых языков;

◎ наличие в составе ЛО нескольких ИПЯ одного типа, чаще нескольких иерархических классификаций.

Некоторое сокращение числа используемых ИПЯ в последнее время можно интерпретировать как стремление к формированию состава ЛО по принципу оптимального минимума. Это подразумевает выбор набора ИПЯ, необходимых и достаточных для решения полного комплекса задач ЭК.

Проведенный в нашей библиотеке эксперимент по сопоставлению возможностей ИПЯ разных типов (иерархических классификаций, языков предметных рубрик и ключевых слов) при поиске в ЭК [6] продемонстрировал целесообразность:

◆ использования таблиц иерархических классификаций для уточнения запроса;

◆ применения полного комплекса ИПЯ для реализации поиска с научными целями;

◆ использования ключевых слов из заглавий с последующим обращением к авторитетным файлам контролируемых предкоординатных ИПЯ (иерархических классификаций, языка предметных рубрик) при поиске с учебными целями;

◆ поиска посредством иерархической классификации для обеспечения наибольшей полноты выдачи, а также по запросам социально-гуманитарной тематики, запросам, содержащим сложные / множественные понятия;

◆ поиска на языке предметных рубрик с целью обеспечения наибольшей точности, а также по запросам в области техники.

В ведении и использовании каждого из этих языковых средств ЭК на современной стадии развития наметился ряд проблем [9].

Для *иерархических классификаций* одной из проблем поиска в ЭК остается *сложность тематического структурирования больших массивов библиографических записей*, выданных в ответ на широко-тематические запросы. Такой подход обеспечивался структурой карточных систематических каталогов. А значит, снижается обзорность представленного пользователю информационного массива. Структурирование выдачи по иерархическому принципу гипотетически возможно, но пока такая возможность нигде не реализована.

Практики указывают на то, что без специальных знаний или вспомогательных средств пользователи ЭК часто затрудняются оценить содержание буквенно-цифровых кодов классификационных ИПЯ, их бывает *трудно использовать*. Поэтому в авторитетных файлах классификационных ИПЯ разрабатывается система словарного доступа к классификационным индексам – «предметный вход».

Кроме того, в этом случае важно подчеркнуть, что разработчики часто путают простоту самой классификационной системы с простотой ее использования. В электронном каталоге, в отличие от карточных, появляется возможность создания различных путей поиска для пользователей с разным уровнем библиографической и профессиональной подготовки. Каждый пользователь карточного систематического каталога в любом случае должен был выполнять один и тот же набор шагов. В ЭК же для неопытных пользователей часть этих процедур может осуществляться «по умолчанию». В то же время высококвалифицированные специалисты получают возможность просматривать классификационные таблицы в полном объеме и строить сложные поисковые предписания.

Иерархические классификации имеют наиболее жестко структурированный по ограниченному числу оснований и связей нормативный словарь. Поэтому могут возникнуть *затруднения при поиске по произвольному сочетанию признаков, по узкопредметным и межотраслевым запросам.*

Но применение этого языка в ЭК в режиме посткоординатной системы, подразумевающее использование всех фрагментов полного индекса как независимых поисковых элементов (с возможностью усечения каждого элемента) в принципе дает возможность преодолеть ограничения, связанные с линейностью сложных индексов.

Еще одна проблема – полная *неразработанность* в России *методики систематизации для электронного каталога*, которая в связи с изменением условий и возможностей поиска должна существенно отличаться от традиционной [7].

Ряд специфических проблем существует и для *языка предметных рубрик*. Среди проблем, унаследованных от карточных каталогов, назовем *затруднения при использовании* этого языка для *поиска информации об общих понятиях и по широким темам.*

Если пользователь в процессе поиска по словам из предметных рубрик в ЭК приводит какое-то широкое понятие, то более узкие, подчиненные ему понятия не войдут в выдачу, если для их выражения в нормативном словаре существуют самостоятельные более точные предметные рубрики. Например, задав поисковое предписание *«цветны\$» * «металл\$»* в результате поиска в ЭК пользователь получает только описания книг, посвященных цветным металлам в целом. Книги только о медных, никелевых, серебряных рудах и т. д. в выдачу не войдут, так как для описания их содержания существуют самостоятельные рубрики, которые и были приведены в поисковых образах документов.

Для такого поиска более целесообразно применять иерархические классификации. Преодоление же названных затруднений при использовании предметизационного ИПЯ связывают с развитием системы ссылок в авторитетных файлах предметных рубрик.

Еще одна часто возникающая трудность связана с *определением адекватности предметных рубрик* для каталогов, описывающих фонды разных объемов. Так, в ЭК, содержащих небольшое количество библиографических записей, часто используют более широкие рубрики, чем для описания тех же документов в каталогах большого объема. Поэтому часто

возникают проблемы при формировании сводных и распределенных каталогов, а также при поиске в них. Преобладающая сегодня тенденция – ужесточение требований к адекватности предметных рубрик, их точному соответствию содержанию документов.

Обратимся к *языку ключевых слов*. Основной проблемой для этого языка слов считают его *ограниченные возможности в плане помощи читателю при поиске*. Поиск по ключевым словам предполагает большую интеллектуальную нагрузку на пользователя и применение более сложной поисковой логики. Эти свойства языка очень затрудняют самостоятельное индексирование запросов неподготовленными пользователями, на которых ориентированы современные ЭК. Это означает, что обеспечить удовлетворительную полноту поиска способен только пользователь, хорошо владеющий терминологией отрасли.

Однако практики говорят о предпочтении читателями этого языка. В то же время следует учитывать, что пользователь выбирает язык общения с ЭК, но никак не стратегию поиска в каталоге (результаты использования которой он не всегда может предвидеть). Имея в виду, что в ЭК язык индексирования запроса может отличаться от языка индексирования документа, а также существует возможность автоматического выбора поисковой стратегии, приведенные данные нельзя рассматривать как однозначный аргумент в пользу вербальных ИПЯ. Решающую роль при выборе ИПЯ должны играть требуемые результаты поиска. Но на текущей стадии развития лингвистического и программного обеспечения ЭК близость плана выражения естественного языка и ИПЯ, безусловно, имеет большое значение. Поэтому необходимо выявлять для пользователей специфические свойства ключевых слов или в процессе обучения, или в процессе диалога с поисковой системой.

Серьезные проблемы вызывает и *использование* этого языка *в реальной практике индексирования*. Попытки исследователей формализовать методические приемы индексирования на языке ключевых слов зачастую не берутся в расчет практиками, заимствующими методические положения из других ИПЯ. Соответственно, при поиске обостряются все названные выше трудности, связанные с отсутствием нормализации лексических единиц.

Рассмотрим теперь проблемы тематического поиска в свете *тенденции комплексного использования ИПЯ в ЭК*.

Различные языки обнаруживают разную эффективность при

реализации различных типов запросов. Часто недостатки ИПЯ являются оборотной стороной их достоинств. Комплексное использование лингвистических средств дает возможность полнее реализовать преимущества и компенсировать недостатки языков. Но этот подход не всегда реализуется на практике, зачастую ЛО ЭК воспринимается как набор изолированных языков.

Комплекс – (от лат. *complexus* – *связь, сочетание*) это *совокупность* объектов, связанных *между собой*, образующих единую систему.

Основное требование к комплексу ИПЯ еще на начальных этапах становления концепции лингвистического обеспечения автоматизированных библиотечно-информационных систем это *целостность*, которая заключается в максимально возможном взаимодействии и связанности всех ИПЯ и сохранении такой связанности при их использовании.

Помимо этого комплекс характеризуют такие свойства, как структурность, функционирование как способ взаимодействия с другими системами и средой. С этих позиций ЛО представляется как совокупность объектов, взаимодействие которых обуславливает появление новых интегративных качеств, не свойственных образующим ее компонентам.

При этом необходимо учитывать, что при поиске в ЭК возможно не только комплексное, но и самостоятельное применение отдельных языковых средств, то есть пользователь каталога может искать только на одном языке (например, предметных рубрик) или использовать несколько ИПЯ (например, классификационный ИПЯ в сочетании с ключевыми словами).

В процессе поиска важно иметь в виду два основных принципа: *принцип дифференциации функций и принцип интеграции функций ИПЯ*. Эти и другие принципы комплексного использования языков тематического поиска в ЭК подробно рассмотрены в публикации во втором номере журнала «Научные и технические библиотеки» (2014).

Последнее из определяющих пути развития лингвистических средств ЭК на современном этапе тенденция – *стремление к упрощению ИПЯ и процесса поиска с их использованием*. Чтобы проиллюстрировать разброс мнений по этому поводу, приведем две точки зрения о комфортной работе в информационной среде, элементом которой является электронный каталог. Обе позиции основаны на подходе к ИПЯ как к средствам организации знаний в информационной среде. Одна из них ориентирована

на «сопровождение» процесса интеллектуальной деятельности, другая – на его активизацию.

Ю. А. Шрейдер выделял следующие условия комфортной работы в информационной среде: возможность корректировки информационной потребности и легкость использования, способность «не предъявлять существенных и неожиданных трудностей». «Удобно – это значит, что предлагаемые пользователю средства легко им осваиваются и не требуют от него смены привычных установок» [10, с. 17].

Другое мнение, немного забытое сегодня, принадлежит Т. Б. Соколовской [5]. Она характеризует комфортность как «такой способ организации информационной среды, который создает условия для наиболее эффективной переработки информации», «наличие информационных предпосылок, стимулирующих творческое мышление». Отсюда следует, что информационная среда должна быть организована таким образом, чтобы не столько фиксировать привычные, но и способствовать выявлению новых связей между объектами, принципов их размежевания и подобия, тенденций развития, представлять информацию с разных точек зрения, в разных контекстах. Предложенный подход вновь возвращает нас к тезису об эффективности комплексного использования ИПЯ. С этой точки зрения требования «легкости использования», привычности, безусловно, не самые основные.

Сегодня первая доминирует позиция. Тенденция упрощения, отошедшая на второй план в начале 90-х годов XX в., вновь обрела актуальность в связи с повсеместным распространением персональных компьютеров и интернет-технологий и соответственным ростом числа самостоятельно проводимых поисков.

В США, на родине OPAC, проблемы доступности ЭК для читателей обсуждались в 80–90-х годах прошлого века. Тогда высказывалось мнение, что применение простейших средств и стратегий поиска приводит к серьезному снижению его качества. А значит, читатель не получит доступ к полной информации по интересующей его теме. Поэтому простота и демократичность поиска – совсем не тождественные понятия.

Выход видится в обеспечении простоты, базирующейся на применении знаний экспертов в области информационного поиска. Такие знания должны составить основу «базы знаний» ЭК – экспертной посреднической системы.

В частности, появляется возможность:

♦ включения в состав ЛО как «сложноорганизованных» комбинационных, так и более простых и менее мощных по структуре и лексическому наполнению ИПЯ;

♦ обеспечения различного уровня экспликации структуры и лексики ИПЯ в зависимости от квалификации пользователя и параметров запроса;

♦ автоматизированного управления рядом процедур, обеспечивающих построение и корректировку поискового предписания, как на базе одного, так и нескольких ИПЯ (в том числе автоматического перевода запроса с естественного языка на ИПЯ);

♦ автоматического выбора ИПЯ, а также перехода с одного поискового языка на другой в зависимости от параметров запроса и удовлетворенности пользователя результатами поиска. Соответственно, обеспечиваются разные уровни (от простого – до сложного) использования ИПЯ.

Следует отметить, что для всех указанных в статье проблем и противоречий существуют варианты решений. По всей видимости, дело теперь в доброй воле библиотечных специалистов и разработчиков программных продуктов для библиотек, в консолидации усилий библиотечного сообщества. И тогда тенденция к максимальному учету при создании ЛО ЭК интересов и потребностей его пользователей, давно ставшая приоритетной за рубежом, возобладает и в России.

Список использованных источников

1. Библиотечно-библиографическая классификация : сред. табл. / [гл. ред. Сукиасян Э.Р.] ; Рос. нац. б-ка, Б-ка Рос. акад. наук. – М., 2001. – Вып. 1 – 5. – 2012.
2. Бусыгина Т. В. Языковые средства доступа к информации: от каталога к Интернету / Т. В. Бусыгина, Л. А. Жарикова, Г. А. Скарук // Электронные ресурсы библиотек региона : материалы регион. науч.-практ. конф. (г. Новосибирск, 24–28 сент. 2007 г.). – Новосибирск, 2008. – С. 212–216.
3. Национальный авторитетный файл : рук. по созданию, ведению и использованию / Рос. нац. б-ка, Нац. информ.-библ. центр «ЛИБНЕТ». – М. : Фаир-Пресс : центр «ЛИБНЕТ», 2005. – Вып. 1 : Наименование организации. Предметные рубрики. – 184 с.
4. Руководство по методике предметизации. Опыт Российской национальной библиотеки / Рос. нац. б-ка, Нац. информ.-библ. центр «ЛИБНЕТ» ; [авт.-сост. : Ю. Г. Селиванова и др.]. – М. : Фаир-Пресс, 2005. – 407 с.
5. Соколовская Т. Б. О комфортности информационной среды / Т. Б. Соколовская // Науч.-техн. информация. Сер. 2. – 1980. – № 11. – С. 5–10.

6. Скарук Г. А. Комплексное использование лингвистических средств тематического поиска в электронном каталоге / Г. А. Скарук // Библиотековедение. – 2010. – № 2. – С. 34–40.

7. Скарук Г. А. Систематизация документов : учеб.-практ. пособие / Г. А. Скарук. – СПб. : Профессия, 2014. – 112 с.

8. Скарук Г. А. Пользователь у каталога: результаты социологического исследования / Г. А. Скарук, Л. А. Жарикова, Т. В. Бусыгина // Науч. и техн. б-ки. – 2009. – № 6. – С. 5–18.

9. Скарук Г. А. Поисковые языки электронных каталогов : конспект лекции по курсу «Справ.-поисковый аппарат» / Г. А. Скарук, Л. А. Жарикова, А. А. Стукалова ; Гос. публ. науч.-техн. б-ка Сиб. отд-ния РАН, Сиб. регион. библ. центр непрерыв. образования. – Новосибирск : ГПНТБ, 2009. – 43 с.

10. Шрейдер Ю. А. Категория комфортности как основной критерий организации знаний / Ю. А. Шрейдер // Междунар. форум по информ. и документации. – 1993. – № 3–4. – С. 17–18.

UDC 025.4+025.34:004

C. A. Skaruk,

Ph. D., Senior Researcher, SPSTL SB RAS

LANGUAGES OF THE SUBJECT SEARCH

IN ONLINE PUBLIC ACCESS CATALOGUES OF SCIENTIFIC LIBRARIES:

TRENDS, PROBLEMS, APPROACHES TO THE SOLUTION

The paper describes trends and problems of the development of the information retrieval languages (IRL) in Online Public Access Catalogues (OPAC) of scientific libraries.

Keywords: information retrieval languages, Online Public Access Catalogue, scientific libraries.