

Учреждение Российской академии наук  
Государственная публичная научно-техническая библиотека  
Сибирского отделения РАН

Сибирский региональный библиотечный центр  
непрерывного образования

**Г. А. Скарук, Л. А. Жарикова, А. А. Стукалова**

## **Поисковые языки электронных каталогов**

*Конспект лекции по курсу «Справочно-поисковый аппарат»*



Новосибирск, 2009

УДК 025.4+025.34:004  
ББК 73+78.37  
С42

**Скарук Г. А.**

С42 Поисковые языки электронных каталогов : конспект лекции по курсу «Справочно-поисковый аппарат» / Г. А. Скарук, Л. А. Жарикова, А. А. Стукалова ; отв. за вып. Е. Б. Артемьева ; Гос. публич. науч.-техн. б-ка Сиб отд-ния Рос. акад. наук. – Новосибирск : ГПНТБ СО РАН, 2009. – 44 с.

ISBN 978-5-94560-181-9

Характеризуются структура и свойства комплекса информационно-поисковых языков современных электронных каталогов, содержание, причины и следствия изменения их характеристик в электронной среде. Даются рекомендации по эффективному использованию лингвистических средств. Вводится общее представление о форматах библиографических и авторитетных данных.

Для слушателей системы дополнительного образования библиотечных специалистов.

УДК 025.4+025.34:004  
ББК 73+78.37

**ISBN 978-5-94560-181-9** © Учреждение Российской академии наук  
Государственная публичная научно-техническая  
библиотека Сибирского отделения РАН  
(ГПНТБ СО РАН), 2009

## Введение

Эффективность использования информационных ресурсов библиотек и качество библиотечного сервиса как на локальном уровне, так и в режиме сети во многом определяют информационно-поисковые языки (ИПЯ) электронных каталогов (ЭК), обеспечивающие доступ к документным массивам. Обоснованный выбор и эффективное использование лингвистического обеспечения (ЛО) – задачи, без решения которых создание локальных и распределенных электронных каталогов теряет смысл.

Для эффективного поиска в электронных каталогах потребовалось преобразование всех основных элементов ИПЯ и правил их использования, заданных в традиционных концепциях для карточных каталогов и проблемно-ориентированных баз данных. Это диктуется, во-первых, необходимостью приспособить языковые средства к использованию в автоматизированном режиме. Во-вторых, необходимо обеспечить пользователю возможность применять ИПЯ самостоятельно (без участия информационного посредника), в том числе в условиях дистанционного доступа.

## Основные понятия

**Информационно-поисковый язык (ИПЯ)** – формализованный искусственный язык, предназначенный для индексирования документов, информационных запросов и описания фактов для последующего хранения и поиска (ГОСТ 7.74–96<sup>1</sup>).

Его основные составляющие – алфавит, лексика, грамматика. *Лексика* ИПЯ – это совокупность входящих в его состав *лексических единиц* (ЛЕ), то есть обозначений отдельных понятий, принятых в данном языке и неделимых в этой функции. В разных ИПЯ это могут быть слова естественного языка, устойчивые словосочетания, аббревиатуры, символы, даты, цифровые или буквенно-цифровые коды и т. п.

Грамматика ИПЯ позволяет отразить связи (отношения) между лексическими единицами. Эти связи существуют в виде парадигматических и синтагматических отношений.

*Парадигматические отношения* – логические и ассоциативные отношения между ЛЕ (тождества, иерархии, пересечения, психологические ассоциации).

Отношения, устанавливаемые в результате объединения ЛЕ во фразы (предложения) называются *синтагматическими*.

Для характеристики совокупности применяемых в ЭК информационно-поисковых языков, правил и методов их ведения и использования и регламентирующей эти правила нормативно-методической и справочной документации используется понятие *лингвистическое обеспечение* (ЛО). Согласно современным представлениям к ЛО относятся и форматы представления библиографических или авторитетных данных.

В состав ЛО электронных библиотечных каталогов входят все основные типы информационно-поисковых языков. Основу ЛО составляют ИПЯ, ранее применявшиеся в карточных каталогах – язык библиографического описания, язык предметных рубрик и иерархические классификации.

## Язык библиографического описания

**Язык библиографического описания** – это информационно-поисковый язык, основными лексическими единицами которого являются элементы, реализующие коммуникативную функцию между запросом

---

<sup>1</sup> ГОСТ 7.74–96. Информационно-поисковые языки. Термины и определения // Стандарты по библиотечному делу. – СПб., 2000. – С. 376–398.

пользователя и содержанием библиографической информации. Такими элементами являются, например: заголовок, автор, заглавие, год издания и другие данные<sup>2</sup>. Алфавит языка библиографического описания совпадает с алфавитом естественного языка. Кроме того, он включает в свой состав арабские и римские цифры, знаки пунктуации.

Лексической единицей языка библиографического описания является обозначение минимального отдельного понятия, имеющего значения. Это принятые в естественном языке слова и словосочетания, аббревиатуры, символы, даты, общепринятые сокращения и т. п.

Таким образом, язык библиографического описания позволяет реализовать поиск документов определенного автора, конкретных глав произведений, а также найти документ или группу документов интересующей тематики по отдельным словам и словосочетаниям, выступающим в качестве ключевых слов по всему тексту библиографического описания. Именно машиночитаемая библиографическая запись может обеспечить максимальное количество точек доступа. Поэтому, в целях обеспечения адекватности поиска, язык библиографической записи должен регулироваться специальными правилами.

Свод «Правил составления библиографического описания» был создан в СССР в 1986 г. Практика работы по этим правилам показала их высокую эффективность и гибкость, значительную степень соответствия международным нормативам. Однако для более эффективного функционирования библиографической записи в интерактивном каталоге появилась необходимость пересмотра российских правил каталогизации, которые должны учитывать не только действующие, но и перспективные международные рекомендации в сфере научной обработки информации<sup>3</sup>. Поэтому в 2004 г. на базе Российской государственной библиотеки (РГБ), Российской национальной библиотеки (РНБ), Российской книжной палаты (РКП) в сотрудничестве с центром Либнет были разработаны «Российские правила каталогизации» в целях согласования российских и международных правил для обмена библиографическими записями без дополнительной доработки.

---

<sup>2</sup> Справочник библиографа / науч. ред. : А. Н. Ванеев [и др.]. – СПб. : Профессия, 2002. – 527 с.

<sup>3</sup> *Каспарова Н. Н.* Концепция пересмотра российских «Правил составления библиографического описания» // Девятая научная конференция по проблемам книговедения : тез. докл. – М., 2000. – С. 200–205.

Чтобы более четко представить позиции правил в отношении языка библиографической записи, следует отдельно рассмотреть элементы библиографической записи:

1. *Область заглавия и сведений об ответственности* относится к числу основных поисковых признаков и должна быть доступна для интернационального пользователя. Из этого следует, что язык основного заглавия в библиографической записи должен соответствовать языку выходных данных документа. В случае, когда принимается решение о конверсии языка в интересах национального пользователя или других целях в данной области, эти данные приводят в качестве дополнительной точки доступа.

2. *Область издания, выходных данных* и др. Сведения об издании, выходные данные, область физической характеристики и область серии приводятся на том языке, на котором они приведены на титульном листе документа.

3. *Область примечания*. Независимо от языка документа или его выходных данных язык, на котором формируются дополнительные примечания, соответствует государственному языку той или иной страны. Следовательно, эти точки доступа ориентированы на национального пользователя. Например, для России – это русский язык, для Германии – немецкий и т. д.

4. *Язык данных* содержательной каталогизации – предметных рубрик, ключевых слов.

Точки доступа предметного или содержательного характера ориентируются на национальные классификации или международные (в переводе на национальный язык страны, если это не англоязычные страны), а также национальные рубрикаторы или тезаурусы. Во всех случаях лингвистическая часть этих данных ориентирована на язык национального потребителя. В том случае, если в системе есть возможности для их перевода в классификационные (содержательные) данные на другом языке, либо функционируют нормативные многоязычные файлы предметных рубрик, то это значительно обогащает поисковые возможности информационно-библиографической системы.

5. *Заголовок библиографической записи*. Более сложно решается проблема выбора языка заголовка библиографической записи, хотя общий подход соответствия языку выходных данных сохраняется, за исключением идентифицирующих признаков (указание духовного сана и т. п.), которые формируются не с документа, а, как правило, из других источников и соответствуют языку национального пользователя.

Проблему выбора языка этих данных помогают решить многоязычные нормативные / авторитетные файлы имен, наименований учреждений, анонимных классических произведений, унифицированных заглавий, географических наименований. Однако не всегда страна или библиотека имеет доступ к нормативному файлу, в связи с чем, вынуждена делать более сложный выбор, учитывая статус заголовка библиографической записи и его функции.

5.1. *Имя лица.* Имя индивидуального автора или другого лица в заголовке библиографической записи может быть дано на том языке, на котором автор преимущественно публикует свои работы.

Но эту логику не удастся сохранить, поскольку при выборе разных языковых форм имени автора, форма имени на языке национального пользователя более известна, и правила отдают предпочтение языку национального пользователя.

Например, библиографические записи на переводы произведений А. С. Пушкина, изданные на иностранных языках, в Российском каталоге будут иметь в заголовке русскоязычную форму имени, но другие языковые формы должны быть зафиксированы в нормативном / авторитетном файле. Можно поступить и по-другому – в заголовке дать имя на языке документа, а другие формы, в том числе русскоязычную, привести в нормативном / авторитетном файле.

Аналогичная двойственная позиция сохраняется в отношении языка наименования коллектива в заголовке библиографической записи.

5.2. *Унифицированное заглавие* анонимного классического произведения в заголовке библиографической записи – это устоявшаяся форма названия произведения, не имеющего автора, однако имеющего национальное происхождение и потому известного национальному пользователю именно на языке своей страны, то есть той страны, которой принадлежит данное анонимное классическое произведение.

Наиболее известное российское анонимное классическое произведение – «Слово о полку Игореве», российский пользователь знает на русском языке, хотя имеются и другие языковые версии, но они другие и переводиться могут иначе. Например, французский вариант – «Le Dit de la campagne d'Igor». Дословный перевод на русский язык – «Слово о походе Игоря». Близко, но французская версия дана не как самостоятельная, а как перевод с русского языка. Российские правила каталогизации отдают предпочтение в качестве основной точки доступа национальной языковой форме той страны, которой принадлежит данное анонимное классическое произведение. При наличии соответствующего нормативного

файла, по любой другой языковой форме анонимного классического произведения оно будет найдено, а в заголовке может быть та, которая представляется более известной национальному пользователю.

5.3. *Географическое название* в заголовке библиографической записи. Язык этих данных в заголовке библиографической записи соответствует языку выходных данных документа, но в качестве дополнительной точки в формате или нормативном файле географических названий необходимо привести соответствующую форму географического названия на языке национального пользователя<sup>4</sup>.

Итак, новые «Российские правила каталогизации» дают более четкий алгоритм действий в отношении языка библиографической записи, а именно основные точки доступа в максимальной степени должны быть приведены на языке, отражающем язык выходных данных документа.

Интересы национального пользователя реализуются в языке примечаний, дополнительных точках доступа во всех элементах записи, в полях предметного или содержательного блока, а также на уровне многоязычного авторитетного файла различных данных – имен лиц, наименований коллективов, унифицированных заглавий, географических названий и т. д.

Элементам и областям в библиографическом описании предшествуют знаки предписанной пунктуации. В отличие от обычных грамматических знаков они выполняют опознавательные функции областей и элементов.

Заголовок от описания отделяют точкой. Области описания отделяют друг от друга точкой и тире. При повторении отдельных областей повторяют точку и тире, за исключением области серии. При повторении области серии сведения о каждой серии заключаются в отдельные круглые скобки без знаков. Это единственное изменение, которое касается правил приведения предписанной пунктуации.

В ГОСТ 7.1–2003 впервые подчеркнуто, что для различения предписанной и грамматической пунктуации применяют пробелы в один печатный знак до и после предписанного знака. Исключение составляет точка с запятой – пробелы оставляют только после них.

---

<sup>4</sup> *Каспарова Н. Н.* Язык библиографической записи в многоязычной коммуникации «документ-потребитель» : позиции новых «Российских правил каталогизации» // Библиотеки и ассоциации в меняющемся мире: новые технологии и новые формы сотрудничества. Электронные информационные ресурсы и социальная значимость библиотек будущего : тр. Девятой междунар. конф. «Крым-2002» (Украина, Авт. Респ. Крым, 8–16 июня, 2002 г.). – М., 2002. – Т. 1. – С. 379–383.



В новом ГОСТе изменены правила употребления строчных и прописных букв: с прописной буквы следует приводить только первое слово области описания, а в элементах строчные и прописные буквы применяются в соответствии с нормами языка, на котором составлено описание. Например, первые слова сведений, относящихся к заглавию и сведений об ответственности, записываются со строчной буквы, если они не являются именами собственными, первыми словами заглавий или цитатами<sup>5</sup>.

Все это позволит преодолеть существующий разноречивой в подходах различных национальных правил каталогизации к выбору языка библиографической записи и тем самым устранить конфликт между интересами национального и интернационального пользователя.

### **Форматы представления библиографических данных**

**Формат** – способ расположения и представления данных на носителе информации (ГОСТ СИБИД 7.0-99<sup>6</sup>).

Формат определяет:

- структуру – что (какие данные) должно быть и где (в каких местах, в каком порядке) расположено;
- способ представления данных – как что указывается (в какой форме и по каким правилам вносится информация).

**Библиографический формат** – это совокупность взаимосвязанных элементов данных, которая служит для создания, хранения и обмена библиографическими записями в машиночитаемой форме<sup>7</sup>. Элементом данных является наименьшая единица библиографической записи, подлежащая точной идентификации. Он служит «точкой доступа» при поиске.

Элементы данных, содержащиеся в записи, можно разделить на собственно данные (заглавие, автор, примечание и т. д.) и сведения, необходимые для выполнения операций с записью (номер записи в системе, дата создания записи, дата и время последнего обращения к записи, язык каталогизации и т. д.).

---

<sup>5</sup> Калинина Г. П. Новые правила библиографического описания в ГОСТ 7.1–2003 // Библиографическая запись : основные стандарты / Б-ка Рос. кн. палаты. – М., 2006. – С. 99–128.

<sup>6</sup> ГОСТ 7.0–99. Библиотечно-информационная деятельность. Библиография. – Минск, 1999. – 23 с.

<sup>7</sup> Скворцов В. В. Форматы MARC21, UNIMARC, RUSMARC – их настоящее и будущее // Библиотечные компьютерные сети: Россия и Запад. – М., 2003. – С. 75–98.

Место, предназначенное для элементов данных, называется *полем* или *подполем*. Поле – определенная строка символов, идентифицированная меткой, содержащая одно или более подполей. Поле – часть записи, соответствующая одной области библиографического описания, одной предметной рубрике, одному шифру хранения и т. д. Оно может быть фиксированной или переменной длины. Подполе – определенная единица информации в пределах поля.

В формате обозначены обязательные и необязательные (факультативные) поля и подполя. Например, для поля «Индивидуальный автор» подполе «Имя автора» является обязательным к заполнению в любом случае, а заполнение подполя «Язык произведения» может быть обязательным или необязательным в зависимости от принятых правил каталогизации, требований к полноте библиографической записи и наличия соответствующих данных. Поля и подполя могут быть повторяющимися и неповторяющимися, в зависимости от характера содержащихся в них данных.

Чтобы компьютер мог считывать и интерпретировать библиографическую запись, оперировать с данными, разработана система символов – меток, индикаторов, идентификаторов, разделителей полей.

Трехзначная метка поля показывает, к какой категории относятся данные в поле.

*200 1# \$aПроблемы космоноосферной футурологии\$еобзор\$fB. П. Казначеев, А. Н. Дмитриев*

Идентификатор подполя – это код, состоящий из одного или нескольких символов, идентифицирующий отдельные подполя внутри переменного поля. Предположим, что идентификатор состоит из двух символов. Первый символ – разделитель, всегда один и тот же уникальный символ, установленный по ISO 2709, второй символ – код подполя, который может быть цифровым или буквенным.

*200 1# \$aПроблемы космоноосферной футурологии\$еобзор\$fB. П. Казначеев, А. Н. Дмитриев*

Во многих форматах для отражения дополнительной информации о данных, содержащихся в поле, связи данного поля с другими полями записи, специфических видах обработки данных используются индикаторы, следующие в записи непосредственно за меткой поля.

*200 1# \$aПроблемы космоноосферной футурологии\$еобзор\$fB. П. Казначеев, А. Н. Дмитриев*

Метки полей, индикаторы и идентификаторы подполей называются определителями содержания.

На рис. 1 приведена общая схема поля библиографического формата.

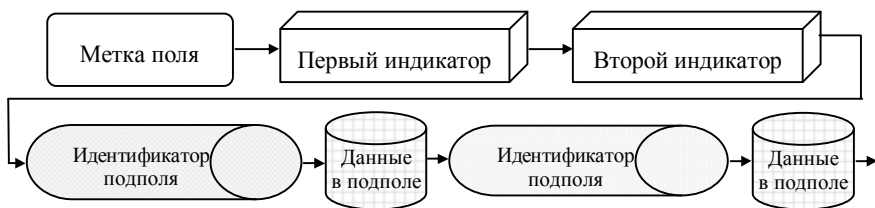


Рис. 1. Общая схема поля библиографического формата

Способ представления библиографической записи в машиночитаемой форме унифицирован специально разработанным форматом MARC. MARC – это акроним названия «Machine Readable Catalog». Работы по созданию этого формата, отражающего расположение, содержание и систему кодирования информации для отображения стандартной библиографической каталожной карточки в виде, удобном для обработки на ЭВМ, начала Библиотека Конгресса США в 1964 г.

Большинство информационно-развитых стран имеют свои машиночитаемые форматы: USMARC, CANMARC, UKMARC, FINMARC и т. д., расхождения в которых вызваны различиями в национальных правилах и практике каталогизации. Поэтому при обмене библиографическими записями появляется необходимость редактирования некоторых полей.

Для преодоления несовместимости форматов Международной федерацией библиотечных ассоциаций (IFLA) был разработан международный коммуникативный формат MARC (UNIMARC), который призван обеспечивать международный обмен библиографической информацией в машиночитаемой форме между национальными библиографическими службами.

Мировая практика показывает, что в большинстве случаев делается различие между коммуникативным форматом и форматом каталогизации. Главным отличием формата каталогизации от коммуникативного формата является их различное предназначение.

**Коммуникативным форматом** называется формат для обмена готовыми библиографическими записями между библиотеками.

Основным требованием к коммуникативному формату является то, что единственная его основа – набор используемых данных: автор, заглавие, место издания, дата издания, ISBN и т. д. Коммуникативный формат в своей обязательной части должен быть однозначно определен, или, во всяком случае, допускать минимум альтернатив. Ни один из факторов, влияющих на различие внутрибиблиотечных форматов,

не может отражаться ни на содержании, ни на структуре коммуникативного формата. В то же время это требование не применимо к формату, используемому внутри конкретной библиотеки.

*Формат каталогизации* представляет собой рекомендации по наполнению машиночитаемых записей в базах данных для их соответствия правилам составления библиографической записи, методикам индексирования и коммуникативным форматам. Рекомендации могут быть не очень жесткими, учитывающими специфику внутрибиблиотечных форматов.

*Внутрибиблиотечные форматы* помимо общих сведений о документе включают локальные данные: сведения о местонахождении, инвентарные номера, различное лингвистическое обеспечение. Такие форматы имеются практически во всех крупных библиотеках и, как правило, довольно серьезно отличаются друг от друга даже тогда, когда основываются на одном и том же прототипе. Эти отличия обусловлены разной практикой каталогизации в отдельных библиотеках, технологией обработки поступлений, различным используемым программным обеспечением<sup>8</sup>. В этих условиях обмен информацией возможен при наличии коммуникативного формата.

Пока создаваемые каждой из российских библиотек записи использовались только ею самой, проблема отсутствия единого формата не стояла остро. Но как только появилась возможность и необходимость обмениваться данными, в конце 1970 – начале 1980-х гг. был разработан обменный – Государственный коммуникативный формат (ГКФ), затем несколько усовершенствованный – Международный коммуникативный формат (МЕКОФ). Этот формат определял структуру и наполнение каждой записи, описывающей издание или его часть, и позволял передавать библиографические данные практически на все виды документов – книги, статьи из журналов и сборников, авторефераты, патенты, депонированные рукописи и т. д. Однако МЕКОФ забыли, а нового обменного формата не разработали.

Лишь в 1998 г. издательством Российской национальной библиотеки (РНБ) был опубликован формат RUSMARC<sup>9</sup>. RUSMARC – это российская версия Международного коммуникативного формата UNIMARC,

---

<sup>8</sup> Цветкова И. Б., Скворцов В. В. Коммуникативный формат представления библиографических записей в машиночитаемой форме для библиотек России // Библиотечные компьютерные сети: Россия и Запад. – М., 1998. – С. 73–77.

<sup>9</sup> Российский формат машиночитаемой каталогизации (RUSMARC) / М-во культуры Рос. Федерации, Рос. библ. ассоц. – СПб. : Изд-во Рос. нац. б-ки, 2000.

в трактовке и категориях действующих в России ГОСТов и Правил каталогизации, базирующаяся на выборе наиболее общих схем представления элементов данных. С практической точки зрения это означает, что любая запись, переданная в Российском коммуникативном формате, должна адекватно восприниматься любым программным обеспечением, о котором заявлено, что оно работает с форматом UNIMARC.

Российский коммуникативный формат состоит из следующих блоков:

0 – Блок идентификации;

1 – Блок кодированной информации;

2 – Блок описательной информации;

3 – Блок примечаний;

4 – Блок связи записей;

5 – Блок взаимосвязанных заглавий;

6 – Блок анализа содержания;

7 – Блок интеллектуальной ответственности;

8 – Блок международного использования;

9 – Блок локального использования.

Библиографическая запись книги одного автора будет выглядеть в RUSMARC следующим образом:

Казначеев, В. П.

Проблемы космоносферной футурологии : обзор / В. П. Казначеев, А. Н. Дмитриев. – Новосибирск, 2005. – 292 с.; 20 см.

ISBN 5-7014-0375-0

И. Дмитриев А.Н.

E081.1+C555.0

1. Биосфера

**маркер** поз. 5 – п (новая запись);

поз. 6 – а (текстовые материалы, кроме рукописных. В том числе печатные текстовые материалы, микроформы, а также электронные текстовые материалы);

поз. 7 – m – (монографический – документ, представляет собой физически единое целое или издается в заранее определенном количестве частей);

поз. 8 – 0 – (иерархическая связь отсутствует);

поз. 17 – 1 – [запись, составленная на основе каталожной карточки (ретроконверсия) или импортированная из другого формата, не предоставляющего достаточно данных для корректного заполнения всех обязательных (в том числе условно обязательных) элементов формата, и неоткорректированная по документу];

- поз. 18 – # – (запись составлена по правилам ISBD).
- 001** EL/0000171523 – Идентификатор записи;
- 010** ## \$a5-7014-0375-0 – Международный стандартный номер книги (ISBN);
- 035** ## \$aRU\NSKC\63011034\CAT\471357 – Другие системные номера;
- 100** ## \$a20070502d2005 |||y0gusy02##### – Данные общей обработки;
- 101** 0# \$aрус – Язык документа;
- 102** ## \$aRU – Страна публикации или производства;
- 105** ## \$aу z 000#у – Поле кодированных данных: текстовые материалы, монографические;
- 200** 1# \$aПроблемы космооосферной футурологии\$еобзор\$fВ. П. Казначеев, А. Н. Дмитриев – Заглавие и сведения об ответственности;
- 210** ## \$aНовосибирск\$c[б.и.]\$d2005 – Публикация, распространение и др.;
- 215** ## \$a292 с.\$d20 – Количественная характеристика;
- 606** 0# \$aБиосфера – Наименование темы как предмет;
- 686** ## \$2rubbk\$aE081.1\$vLBC/SL – Индексы других классификаций;
- 700** #1 \$aКазначеев\$bВ.П.\$gВлаиль Петрович – Имя лица – первичная интеллектуальная ответственность;
- 701** #1 \$aДмитриев\$bА.Н.\$gАлексей Николаевич – Имя лица – альтернативная интеллектуальная ответственность.

Благодаря созданию Российского коммуникативного формата стало возможным осуществление обмена библиографическими записями, что способствовало решению следующих задач:

- улучшение доступности библиографической информации;
- создание сводных каталогов;
- сокращение затрат при каталогизации.

## Языки тематического поиска

Одна из основных характеристик ИПЯ – принадлежность их к определенному типу.

Для классификации языков поиска по теме (содержанию) документа используют несколько признаков:

### 1. Стадия интеграции лексики

Связанность (сопряжение) ЛЕ *до* процесса индексирования, в нормативных словарях ИПЯ – *предкоординация*. Таким образом понятия – «слова» как бы заранее уже объединены в «предложения». Например,

*Астма бронхиальная – Лекарственная терапия;*  
*10.21.51 Правовое регулирование государственных расходов и бюджетного финансирования;*

*Ч 736.232 Описание статей из периодических изданий и сборников.*

К предкоординатным языкам относятся язык предметных рубрик, ряд классификационных систем – Библиотечно-библиографическая классификация (ББК), Универсальная десятичная классификация (УДК), Десятичная классификация Дьюи (ДКД), Государственный рубрикатор научно-технической информации (ГРНТИ) и др.

Выбор не связанных между собою слов *на* стадии формирования поискового образа документа (ПОД) или поискового образа запроса (ПОЗ) – **посткоординация (координация)**. Слова, как правило, не связаны в «предложения», а существуют в виде набора отдельных лексических единиц. Например,

*Астма, лекарство, терапия;*

*Ключевые слова, рубрикаторы, электронные каталоги.*

Языки координатного индексирования – дескрипторный ИПЯ, язык ключевых слов.

2. План выражения (а отчасти и содержания) ИПЯ

Языки, в которых для обозначения понятий применяются слова естественного языка, именуются **вербальными**. Например,

*Астма бронхиальная – лекарственная терапия;*

*Информационно-поисковые языки.*

К вербальным ИПЯ относятся язык предметных рубрик, дескрипторный ИПЯ, язык ключевых слов.

Языки, где в качестве лексических единиц используются специальные символы, получили название **семантических кодов**. Например,  
*10.21.51 Правовое регулирование государственных расходов и бюджетного финансирования;*

*Ч 736.232 Описание статей из периодических изданий и сборников;*

*65.5 Мировая экономика.*

К кодовым ИПЯ относятся ББК, УДК, ДКД, ГРНТИ и другие иерархические классификации.

3. Способ задания лексики (с использованием словаря или нет).

Языки, где каждая ЛЕ по регламентируемому правилам вносится в специальный контролируемый словарь, – **контролируемые ИПЯ**.

Языки, не имеющие словаря, – **неконтролируемые**.

Характеристики каждого ИПЯ по названным позициям описывает табл. 1.

Представление о том, на основе каких ИПЯ возможен поиск по теме в ЭК, представленных на сайтах крупнейших российских и некоторых новосибирских библиотек, дает табл. 2.

Очевидно, что главная устойчивая тенденция – применение в ЭК комплекса различных по структуре, лексическому наполнению и функциональным возможностям ИПЯ. Это позволяет качественно расширить возможности поиска.

Таблица 1

### Основные характеристики информационно-поисковых языков

ИПЯ	Алфавит	Лексическая единица	План выражения	План содержания	Стадия интеграции лексики	Способ задания лексики	Грамматика	
							синтагматика	парадигматика
Иерархическая классификация	Буквы, цифры, знаки пунктуации естественного языка	Деление классификации	Условные коды	Специфический, близок к естественному языку	Предкоординатный	Контролируемый	Грамматические знаки	Иерархический способ построения таблиц. Ссылочно-справочный аппарат
Язык предметных рубрик	Алфавит естественного языка, знаки пунктуации	Предметная рубрика	Естественный язык	Специфический, близок к естественному языку	Предкоординатный	Контролируемый	Знак «--»	Предметные комплексы. Ссылочно-справочный аппарат
Дескрипторный ИПЯ	Алфавит естественного языка	Дескриптор	Естественный язык	Специфический, близок к естественному языку	Координатный	Контролируемый	Указатели роли Указатели связи	Ссылки в словарных статьях тезауруса
Язык ключевых слов	Алфавит естественного языка	Ключевое слово	Естественный язык	Естественный язык	Координатный	Неконтролируемый	—	—



Таблица 2

## Информационно-поисковые языки электронных каталогов

Библиотека	Карточные каталоги		Электронные каталоги									
	ИПЯ, используемые при поиске <sup>10</sup>		ИПЯ, используемые при поиске					индексы, присутствующие в БЗ				
	ИК	ЯПР	ИК	ЯПР	ДИПЯ	ЯКС	СЗ	ИК	ЯПР	ДИПЯ	ЯКС	СЗ
Российская государственная библиотека	+		+	+		+	+	+	+		+	+
Российская национальная библиотека	+	+	+	+		+		+	+		+	+
Государственная публичная научно-техническая библиотека России	+	+		+		+	+	+	+			+
Государственная публичная научно-техническая библиотека СО РАН	+		+	+		+	+	+	+		+	+
Библиотека по естественным наукам РАН			+				+	+				+
Всероссийская государственная библиотека иностранной литературы им. М. И. Рудомино				+					+			
Библиотека МВТУ им. Н. Э. Баумана	+		+					+				+
Фундаментальная библиотека СПбГТУ	+	+	+	+				+	+			
Научная библиотека Новосибирского государственного технического университета	+	+		+	+	+	+	+	+	+	+	+
Новосибирская государственная областная научная библиотека	+			+		+	+					

<sup>10</sup> ИК – иерархическая классификация, ЯПР – язык предметных рубрик, ДИПЯ – дескрипторный информационно-поисковый язык, ЯКС – язык ключевых слов, СЗ – слова из заглавия документа.

В рамках тенденции отметим следующие особенности:

- прее́мственность в выборе языковых средств между карточными и электронными каталогами;
- преимущественная ориентация на использование ИПЯ, разработанных для карточных каталогов: цифровых и / или буквенно-цифровых иерархических классификаций и языка предметных рубрик;
- расширение состава ЛО ЭК по сравнению с ЛО системы карточных каталогов;
- совместное использование ИПЯ разных типов, как правило, иерархических классификаций и вербальных поисковых языков;
- включение в состав ЛО нескольких ИПЯ одного типа, чаще нескольких иерархических классификаций.

В России существует несколько координационно-методических центров, отвечающих за поддержание нормативных словарей и развивающих самостоятельные теоретические направления по использованию в ЭК отдельных типов языков тематического поиска. В области теории и практики предметизации таким центром является Российская национальная библиотека. Идеи автоматизированного поиска с использованием ББК развиваются специалистами Российской государственной библиотеки. Приоритет в разработке ЛО ЭК на базе языка ключевых терминов и ГРНТИ принадлежит Государственной публичной научно-технической библиотеке России.

Основная предпосылка изменения свойств ИПЯ в электронной среде – изменение принципов организации поискового массива и реализации поисковых процедур. К наиболее важным следствиям можно отнести множественность «точек доступа» к библиографическим записям и другие возможности: логического сочетания поисковых признаков, использования в качестве поисковых реквизитов как полных индексов, так и отдельных элементов комбинированных индексов предкоординатных ИПЯ, применения критериев релевантности на неполное соответствие, усечения индексов, отделение поискового словаря от поискового массива, упорядочения информационного массива в зависимости от поисковой задачи по любому заданному признаку. В связи с этим изменяются как свойства ИПЯ, так и требования к ним и способы реализации этих требований. Однако последствия «компьютерной революции», особенно в сфере использования предкоординатных ИПЯ традиционных каталогов, на наш взгляд, и сегодня недооцениваются российскими специалистами, поэтому значительная часть преимуществ автоматизированного поиска по сей день остается гипотетической, методически и программно необеспеченной.

При общих достаточно активных усилиях по развитию ИПЯ электронных каталогов в библиотеках разного уровня, в этой области еще остается ряд проблем. К таковым можно отнести: некоторую избыточность состава ЛО; недоиспользование поисковых возможностей ИПЯ (причем не только тех новых, которые обеспечивает автоматизация, но и части тех, которые предлагали традиционные карточные каталоги); неиспользование общепризнанных версий авторитетных файлов, отсутствие единых подходов к оценке и выбору направлений развития ЛО ЭК; слабое осознание специфики и отсутствие единых принципов комплексного использования ИПЯ в ЭК; недостаточность средств помощи конечным пользователям ЭК в использовании поисковых языков; недостаточная адаптация методик индексирования документов для ЭК.

### Язык предметных рубрик

*Язык предметных рубрик (ЯПР)* – информационно-поисковый язык, предназначенный для индексирования документов (частей документов) и информационных запросов посредством предметных рубрик (ГОСТ СИБИД 7.74–96). Лексическая единица ЯПР – предметная рубрика. *Предметная рубрика* – элемент информационно-поискового языка, представляющий собой краткую формулировку темы на естественном языке.

Предметные рубрики (ПР) наряду с ключевыми словами имеют самую длительную историю использования в ЭК. В автоматизированном режиме стали очевидными многие ограничения ЯПР, которые в карточных каталогах были не так заметны. Особенно острая критика в адрес этого языка звучала на начальных этапах создания ЭК. Тогда ЭК воспринимался как «наследие системы карточных каталогов». Вопрос о бесперспективности системы предметных рубрик Библиотеки Конгресса США ставился даже в 1996 г.<sup>11</sup>, по истечении более, чем двадцатипятилетнего периода развития ОРАС. Поэтому подверглись изменениям практически все элементы этого поискового языка.

W. E. Studwell<sup>12</sup> так обозначила основные этапы развития ЯПР в США: конец XIX в. – 1970-е гг. – период следования традиционным

---

<sup>11</sup> Studwell W. The three camps: the present and possible future status of subject access in online catalogs // Behav. a. Soc. Sci. Libr. – 1996. – Vol. 14, № 2. – P. 51–54.

<sup>12</sup> Studwell W. Cutterism, codism and consumerism: the past, present and future of LC subject access // Techn. Serv. Quart. – 1991. – Vol. 9, № 1. – P. 19–25.

правилам каталогизации, разработанным Ч. Кеттером; в 1970–1980-е гг. (с началом эпохи электронных каталогов) основное внимание уделялось созданию полного универсального свода правил по формулированию ПР; 1990-е гг. – стремление к максимальному учету интересов пользователей.

К настоящему моменту основная задача в области предметизации – адаптация ЯПР к различным видам информационной среды, в том числе электронной. По всей видимости, в создании универсального общероссийского списка предметных рубрик нам предстоит пройти все эти этапы, и сейчас мы находимся на втором из них. Хотя имеются рекомендации использовать в качестве общесетевого лингвистического средства Список предметных рубрик Библиотеки Конгресса США<sup>13</sup>, не подтвержденные, впрочем, серьезным анализом рекомендуемой системы.

Еще одно направление – исследование возможности применения существующих словарей предметных рубрик в рамках набора метаданных Dublin Core для индексирования электронных ресурсов, представленных в глобальных компьютерных сетях. Показательно, что мнения по поводу направлений преобразования предметизационного ИПЯ, предназначенного для содержательного индексирования ресурсов Интернета, диаметрально противоположны. Предлагается как упрощение структуры и содержания индексов ИПЯ (в соответствии с концепцией метаданных Dublin Core)<sup>14</sup>, так и усложнение (в соответствии с задачей индексирования распределенных гипертекстовых документов)<sup>15</sup>. С этих позиций рассматривается, например, применение предметных рубрик Библиотеки Конгресса США в фасетном режиме с упрощением синтаксиса.

Основные изменения поисковых возможностей ЯПР связаны с усечением терминов и с поиском по каждому отдельному элементу индекса.

В электронном каталоге поиск с усечением на ЯПР позволяет включить в выдачу все слова, имеющие общую основу. Например, при поиске с усечением по рубрике «Информат\$» в выдаче будут документы, содержащие в поле предметных рубрик слова «информатика», «информатике», «информативность», «информатизация», «информатизацией» и т. п.

---

<sup>13</sup> Шрайберг Я. Л. Основные положения и принципы разработки автоматизированных библиотечно-информационных систем и сетей: главные тенденции окружения. Основные положения и предпосылки, базовые принципы. – М., 2000. – 130 с.

<sup>14</sup> A faceted approach to subject date in Dublin Core metadata record / L. M. Chan [et al.] // J. Internet Catal. – 2001. – Vol. 4, № 1/2. – P. 35–47.

<sup>15</sup> Кунер И. Р. Гипертекст как новый объект индексирования // Библиография. – 2001. – № 2. – С. 72–74.

Поиск с усечением по заголовку предметной рубрики обеспечит выдачу всех документов с рубриками, начинающимися с данного заголовка.

В современных ЭК возможен поиск либо по каждому подзаголовку, либо по каждому отдельному слову предметной рубрики. Так, документ с предметной рубрикой «*Питание спортсменов – гигиена*» будет выдан в ответ как на отдельные запросы «*Питание*», «*Спортсмен\$*», «*Гигиена*», так и на объединенный запрос «*Питание*» \* «*Спортсменов*» \* «*Гигиена*». Таким образом, читателям становится доступен поиск по произвольному сочетанию признаков.

Возможность доступа к каждому слову предметной рубрики в авторитетном файле и непосредственно в поисковом образе документа изменила и подходы специалистов к оптимизации структуры сложных рубрик. В методической литературе предлагаются следующие шаги: отказ от описательных формулировок в пользу сложных рубрик<sup>16</sup>, ужесточение правил принятия решений о применении инверсии<sup>17</sup>, использование в структуре ПР реляторов<sup>18</sup>.

Это приводит к сближению свойств ЯПР и дескрипторного поискового языка, о котором мы поговорим позже. Менее значимыми становятся и отличия ЯПР от иерархических классификаций. Это касается прежде всего локализации отдельных поисковых признаков в линейных структурах ЛЕ, в том числе отмеченная Е. М. Ручимской<sup>19</sup> множественность локализации отраслей по отношению к одному предмету в ЯПР в противовес множественности локализации предметов в иерархических классификациях.

Но в результате теряется характерное отличие языка предметных рубрик – жесткая связанность терминов (заголовков и подзаголовков) в текстах ПР. Принятая в карточных предметных каталогах форма ПР,

---

<sup>16</sup> Голубцов С. Б., Кузьмина И. Д., Лория Е. Л. Предметизация с учетом требований автоматизированного поиска и коммуникативных форматов : материалы к методике // Предметный поиск в традиционных и нетрадиционных информационно-поисковых системах. – СПб., 2000. – Вып. 14. – С. 145–163.

<sup>17</sup> Войскунский В. Г. Об одном подходе к предметному индексированию // Предметный поиск в традиционных и нетрадиционных информационно-поисковых системах. – СПб., 1999. – Вып. 13. – С. 26–33.

<sup>18</sup> Голубцов С. Б. Реляционные модели в языке предметных рубрик. Общие принципы и перспективы использования // Предметный поиск в традиционных и нетрадиционных информационно-поисковых системах. – СПб., 2000. – Вып. 14. – С. 164–167.

<sup>19</sup> Ручимская М. Е. К вопросу о сходстве и различии языка предметных рубрик и ИПЯ классификационного типа // Науч. и техн. б-ки СССР. – 1983. – № 1. – С. 12–16.

во-первых, исключает ложную координацию при поиске и, во-вторых, позволяет пользователям «увидеть сочетаемость терминов»<sup>20</sup> и уточнить контекст темы уже при составлении поискового предписания.

Возможность сохранить все перечисленные преимущества, увеличив в то же время гибкость поиска, дает изменение способа организации машиночитаемого нормативного словаря – авторитетного файла ПР. Появляются разнообразные дополнительные способы доступа к нормативным записям. Возможен доступ пользователя:

- к ПР, где названное пользователем слово является ведущим (способ поиска в карточных предметных каталогах);
- ко всем ПР словаря, содержащим любое из слов запроса;
- к ПР, максимально близким к полному тексту запроса.

Все перечисленные возможности реализованы в Авторитетном файле предметных рубрик РНБ через функции «пермутационный поиск» и «алфавитный» поиск (рис. 2).

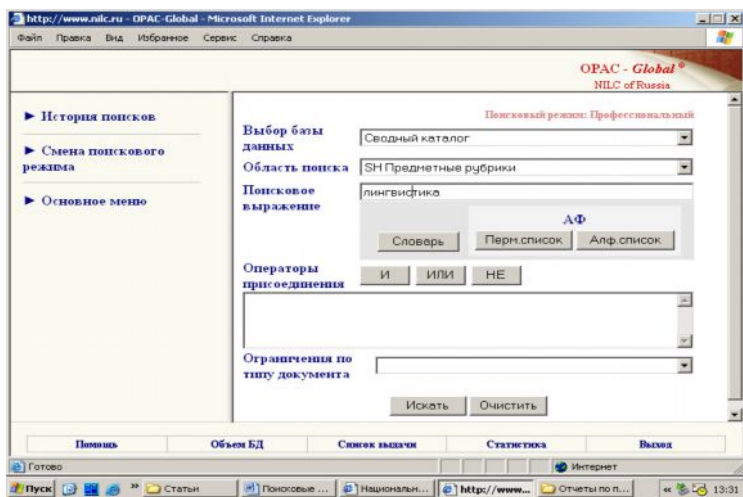


Рис. 2. Возможности поиска в Авторитетном файле предметных рубрик РНБ

Используя все эти возможности, читатель может уточнить контекст запроса и затем уже выполнить поиск.

<sup>20</sup> Чернозатонская Е. В. Предметный поиск и поиск по дескрипторам в электронных каталогах и библиографических базах данных // Библиотеки и ассоциации в меняющемся мире: новые технологии и новые формы сотрудничества : материалы 3-й Междунар. конф. «Крым-96», Форос, Ялта, Авт. Респ. Крым, Украина, 1–9 июня 1996 г. – М., 1996. – Т. 2. – С. 621–624.

Для авторитетного файла предметных рубрик разработан Российский коммуникативный формат для представления авторитетных / нормативных записей<sup>21</sup>, согласованный с первой версией Российского коммуникативного формата представления библиографических записей. Этот формат может применяться для записи следующих видов заголовков: имени лица, наименования организации, родового имени, географического названия, унифицированного заглавия, обобщающего унифицированного заглавия, тематической предметной рубрики.

Формат разработан с учетом национальной методики предметизации. В нем предусмотрена возможность создавать и поддерживать нормируемый список сложных ПР. Машиночитаемая авторитетная запись содержит принятый заголовок, параллельные заголовки, ссылки, справки, примечания каталогизатора.

Кроме того, формат включает ссылочные и справочные записи. Исходным элементом справочной записи является пояснительный заголовок, от которого делается ссылка к определенному классу или категории заголовков (сродни традиционному примечанию или методическому указанию). Ссылочная запись – машиночитаемая запись, исходным элементом которой является вариантный или принятый заголовок, от которого формируются ссылки или от вариантного заголовка к соответствующему принятому заголовку (то есть ссылка «См.»), или от принятого заголовка к взаимосвязанным заголовкам (то есть перекрестная ссылка «См. также»). Структура всех видов записей показана на рис. 3.

Формат включает следующие функциональные блоки:

*Маркер записи;*

*0 – Блок идентификации;*

*1 – Блок кодированной информации;*

*2 – Блок заголовков;*

*3 – Блок справок и примечаний;*

*4 – Блок формирования (трассировки) ссылок «см.»;*

*5 – Блок формирования (трассировки) ссылок «см. также»;*

*6 – Блок классификационных индексов;*

*7 – Блок связанных заголовков;*

*8 – Блок источников информации;*

*9 – Блок локального использования.*

---

<sup>21</sup> Российский коммуникативный формат представления авторитетных / нормативных записей. – СПб., 1998. – 224 с.



Рис. 3. Структура записей в авторитетном файле предметных рубрик

Приведем пример авторитетной записи<sup>22</sup>:

000 01244пх 22002053 450;

001 RUNLRAUTH\66440586 – Идентификатор записи;

005 20030709120205.0 – Идентификатор версии;

100 ##\$a20030709arusy0189 са – Данные общей обработки;

152 ##\$bnlr\_sh – Коды правил предметизации и системы классификации;

215 ##\$aСанкт-Петербург, город – Заголовок – Географическое наименование как ПР;

305 1#\$асм. также под названием отдельных частных вопросов или отраслевых понятий с подзаголовком – Санкт-Петербург, г. Например \$bМузыкальная культура – Санкт-Петербург, г.\$bКоммунальное хозяйство – Санкт-Петербург, г.\$bМедицина – История – Санкт-Петербург, г. – Ссылочное примечание «См. также»;

330 1#\$aДокументы о городе за весь период его существования см. под рубрикой «Санкт-Петербург, г.» □ \$a1\$a1\$a0□ Документы о конкретных «географических объектах» на территории Санкт-Петербурга – Примечание об области применения;

330 1#\$aДокументы о конкретных «географических объектах» на территории Санкт-Петербурга см. под названием этих объектов. Например, Садовая ул. (Санкт-Петербург, г.); Грибоедова канал (Санкт-Петербург, г.) – Примечание об области применения;

<sup>22</sup> Пример из Авторитетного файла предметных рубрик Российской национальной библиотеки.



330 1#*Документы о городе с 1924 по 1991 г. см. под рубрикой «Ленинград, г.»*;

330 1#*Под этой рубрикой собираются документы о городе общего характера, например, «Санкт-Петербург, г. – Путеводители» с 1703 по 1914 г. – Примечание об области применения;*

330 1#*Документы о городе с 1914 по 1924 г. см. под рубрикой «Петроград, г.» – Примечание об области применения;*

415 ##*Петербург, город – Формирование ссылки «См.»*;

686 ##*Д890(2P31-4Je-2C)\$vLBC\SL\$2tubbk – Индекс ББК;*

801 #0*RU\$bNLR\$c20030709 – Источник записи.*

Далее остановимся коротко на основных проблемах, которые сегодня выделяются при использовании ПР в ЭК.

*Поиск по широким темам.* Среди проблем, унаследованных от карточных каталогов, F. Calderon называла затруднения при использовании ЯПР для поиска информации об общих понятиях и по широким темам<sup>23</sup>.

Если в процессе поиска по словам из предметных рубрик приведено 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 получим только описания книг, посвященных цветным металлам в целом. Книги только о медных рудах, никелевых рудах, 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. Поэтому, если нужна полная информация обо всех цветных металлах, в поисковом предписании (ПП) необходимо перечислить названия каждого из них: «цветн\$» \* «металл\$» + «медн\$» + «никел\$» + «кобальт\$» + «цинк\$» + «серебр\$» и т. д. Для такого поиска более целесообразно использовать иерархические классификации.

*Адекватность терминов индексирования.* Р. С. Гиляревский, А. И. Михайлов, А. И. Черный в конце 1970-х гг. высказывали мнение, что меньшая по сравнению с дескрипторными ИПЯ глубина индексирования (во многом определяющая свойства ЯПР и как «поискового ключа», и как «аннотации»), хотя и соответствует сложившейся практике, в принципе обусловлена линейным способом представления индексов в традиционных предметных каталогах. При значительном

---

<sup>23</sup>Calderon F. Library of Congress Subject Hearings. Vested interest versus the real needs of information society // Catal. and Classif. Quart. – 1990. – Vol. 11, № 1. – P. 85–94.

увеличении глубины индексов становится неудобно использовать карточные каталоги<sup>24</sup>. Но и в 1995 г. по результатам оценки существующих перечней предметных рубрик Z. Khurshad делает вывод об их непригодности для описания узкоспецифичных и новых тем<sup>25</sup>. Обилие публикаций, так или иначе касающихся проблемы адекватности ПР, свидетельствует о том, что и сегодня она стоит достаточно остро<sup>26</sup>.

С учетом этой проблемы нужно проводить поиск в ЭК. Формулируя ключевую тему запроса, можно начать поиск с конкретного узкого понятия (наименования лекарства, вещества, болезни, фамилии конкретного лица и т. п.), а при недостаточном или нулевом итоге поиска нужно постараться найти обобщающую, более широкую формулировку (название класса, вида, подраздела и т. п.), например:

«Подсолнечное масло» – 0 документов;

«Растительные масла» – 3 документа.

Сегодня усилия направлены на решение этой проблемы в электронной среде. Один из распространенных подходов – смягчение условий присвоения документам обобщающих рубрик. Например, при индексировании книги

**Филиппов, Михаил Михайлович.**

Леонардо да Винчи / [М. М. Филиппов].

Микеланджело / [С. М. Бриллиант]. Рафаэль / [С. М. Бриллиант]. Рембрандт / [А. Калинина]. Александр Иванов : Биограф. повествование / [А. И. Цомакион ; Сост., общ. ред. Н.Ф. Болдырева ; Послесл. Андрея Северского]. – 2-е изд. – Челябинск : Урал LTD, 1998.

обобщающую рубрику рекомендуется заменить набором адекватных<sup>27</sup>:

*Рафаэль Санти (1483–1520);*

*Микеланджело Буонаротти (1457–1564);*

*Рембрандт, Харменс Ван Рейн (1606–1669);*

*Леонардо да Винчи (1452–1519).*

---

<sup>24</sup> Михайлов А. И., Черный А. И., Гиляревский Р. С. Основы информатики. – М., 1968. – 756 с.

<sup>25</sup> Khurshid Z. Improvisations in cataloging of thesis and dissertations // Catal. a. Classif. Quart. – 1995 – Vol. 20, № 2. – P. 51–59.

<sup>26</sup> Clack D. H. Subject access to African American studies resources in online catalogs: issues and answers // Catal. a. Classif. Quart. – 1994. – Vol. 19, № 2. – P. 46–49; Mann T. «Cataloging must change!» and indexing consistency studies: misreading the evidence at our peril // Catal. a. Classif. Quart. – 1997. – Vol. 23, № 3–4. – P. 3–45.

<sup>27</sup> Руководство по методике предметизации. Опыт Российской национальной библиотеки. – СПб., 2005. – 407 с.

## Иерархические классификации

**Иерархическая классификация** – классификационная система, в которой отношения классов образуют иерархическую классификационную структуру (ГОСТ 7.74–96). Иерархическая классификационная структура предполагает отношения между лексическими единицами, где один класс является подклассом другого, более широкого класса ЛЕ.

Иерархические классификации имеют обширную и длительную практику применения в карточных каталогах. В электронной среде классификации приобретают действительно уникальные свойства. Их возможности принципиально отличаются от свойств различных типов вербальных ИПЯ.

На ранних стадиях автоматизации библиотек высказывались мнения о том, что классификации устарели. Но эти доводы сегодня признаны несостоятельными<sup>28</sup>. Более того, очевидно, что возможности классификационных ИПЯ в электронной среде намного расширились.

В отличие от языка предметных рубрик, иерархические классификации позволяют создавать и просматривать обширные тематические массивы библиографических записей по несколько иному (отраслевому) принципу. При этом, меняя длину индекса путем усечения в ту или иную сторону, мы либо сужаем, уточняем, либо расширяем тематические рамки обзора. Подобными возможностями в таком объеме не располагает больше ни один поисковый язык ЭК.

Поиск по усеченным классификационным индексам обеспечивает дополнительные (по сравнению с карточными каталогами) и действительно уникальные возможности. Эта процедура предполагает автоматическое включение в выдачу всех лексических единиц, иерархически подчиненных приведенным в запросе. Таким образом, с помощью классификационного ИПЯ мы имеем возможность в кратком поисковом предписании объединить большое число понятий по отраслевой теме.

Например, если пользователя интересуют книги о болезнях органов дыхания (а это бронхиальная астма, пневмонии, бронхиты, плевриты и т. д.), нет необходимости перечислять их все. Достаточно привести в поисковом предписании индекс ББК с усечением «Р412\$»\* и в поле поиска будут включены индексы:

*Р 412.2 Болезни бронхов;*

*Р412.201.1 Бронхиты;*

---

<sup>28</sup> *Sapiie J.* Reader–interest classification: the use–friendly schemes // *Catal. a. Classif. Quart.* – 1995. – Vol. 19, № 3–4. – P. 143–155.

\* Примеры здесь и далее приводятся по полным таблицам ББК.

*Р 412.202 Бронхиальная астма;*

*Р 412.3 Болезни легких;*

*Р 412.301.1 Пневмонии и т. д.*

Проиллюстрируем разницу поиска в ЭК с усечением на классификационном и предметизационном ИПЯ на конкретных примерах.

В таблицах ББК тема «Генетика» представлена множеством разных индексов, в том числе:

*Е04 Общая генетика;*

*П836.15 Ветеринарная генетика;*

*Е34 Генетика вирусов;*

*Е64 Генетика животных;*

*П345.1 Генетика лесных культур;*

*Р252.215 Медицинская генетика;*

*Е54 Генетика растений и т. д.*

Если же в поисковом предписании записать предметную рубрику «генети\$», то в ответ на запрос в ЭК ГПНТБ СО РАН будут выданы документы, посвященные и общим проблемам генетики, и генетике животных и растений, медицинской генетике и т. д.

Если нужна литература о погоде, то, задав поисковое предписание на языке предметных рубрик «погод\$», в ЭК ГПНТБ СО РАН получим 113 описаний документов, в поле предметных рубрик которых встречаются слова «погода», «погоды», «погодный» и т. п. При этом проблемы погоды в них могут рассматриваться не только с точки зрения метеорологии, но и сельского хозяйства, астрологии. Кроме того, в выдачу попадут и нерелевантные документы, например, об историке М. О. Погодине. Выбрав же индекс ББК «Д 246 Погода. Прогнозы погоды» и задав поисковое предписание «Д 246\$», можно просмотреть 85 документов, в описаниях которых, кроме рубрики «погода» встречаются и другие, например «Атмосферные осадки, прогнозы», «Температура воздуха, прогнозы», «Облачность, прогнозы», «Гидрометеорологическая служба» и т. п. Но проблема в этом случае будет рассматриваться только с точки зрения метеорологии и климатологии. Общими же для обеих выдач окажутся только 63 документа.

В поисковом предписании с помощью предметной рубрики и семантического кода выражено одно и то же понятие «Млекопитающие»:

*ПР: Млекопитающие\$;*

*ББК: Е693.36\$.*

В обоих случаях применено правостороннее усечение.

В результате поиска (рис. 4) на языке предметных рубрик выдано 129 документов, и все они посвящены только млекопитающим вообще.

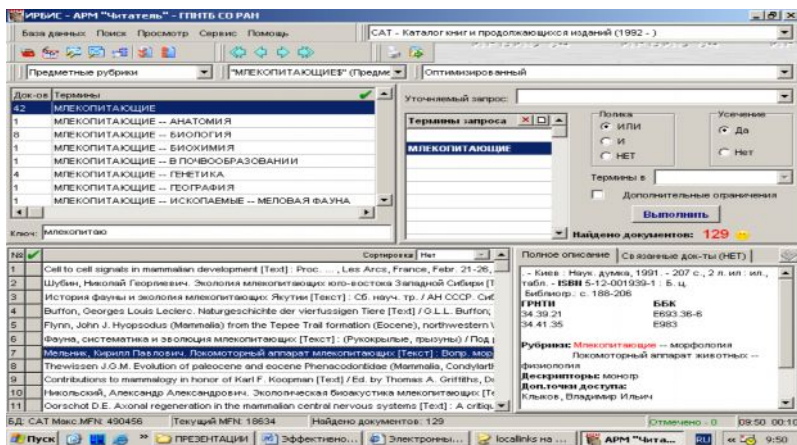


Рис. 4. Результат поиска в ЭК ГПНТБ СО РАН на ЯПР

При поиске (рис. 5) по ББК получено 335 записей, посвященных не только млекопитающим в целом, но и отдельным представителям класса. В принципе вполне возможно составить полное поисковое предписание на ЯПР, однако для этого потребуется более 200 рубрик.

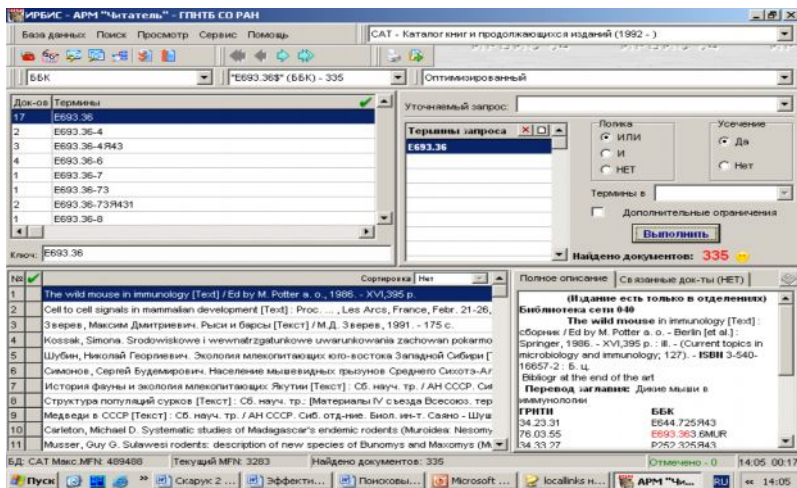


Рис. 5. Результат поиска в ЭК ГПНТБ СО РАН на языке ББК

Причина – в плане выражения и способе структурной организации поисковых словарей (рис. 6). В результате усечения предметной рубрики в выдачу включаются все документы с общим заголовком и различными подзаголовками. При поиске с усечением по ББК в выдачу входят все понятия, иерархически подчиненные данному. Это два разных вида поиска, в каждом из которых решаются свои задачи.



Рис. 6. Правостороннее усечение терминов в языках предметных рубрик и ББК

Итак, такой поиск дает возможность, не изменяя объема отраженного в индексе понятия, существенно расширить поле обзора и объем выдачи. А значит, в иерархических классификациях изначально решена типичная для поиска на вербальных ИПЯ «проблема общих и частных терминов». Поиск с усечением по одному широкому термину индексирования предполагает одновременный поиск по всему множеству иерархически подчиненных ему частных.

В ряду преимуществ цифровой нотации – независимость от конкретного естественного языка и системы письма. Это особо актуально в многоязычной сетевой электронной среде. Уникальность каждого цифрового индекса исключает возможность поискового шума из-за ложной координации или полисемии лексических единиц.

Особая функция классификации в том, что она представляет своего рода «карту знаний». Этим обусловлены ее возможности помощи пользователям.

Авторитетные файлы этого типа ИПЯ обеспечивают определенную целостность и обозримость представления знаний. Поэтому пользователю проще ориентироваться и уточнять запрос при составлении поискового предписания. Иерархию принято считать одной из наиболее простых и удобных форм представления знаний, соответствующих естественным структурам их организации в речемыслительной деятельности.

Для эффективного использования в ЭК необходима *модификация лексического состава и грамматики* классификационных ИПЯ. Можно выделить следующие тенденции: унификацию структуры и лексики классификационных таблиц, оптимизацию структуры систем классификации с использованием принципов фасетизации, формализацию использования грамматических средств для обеспечения однозначности их интерпретации в ЭК, существенное расширение ссылочно-справочного аппарата.

В то же время, в отличие от ситуации с ЯПР, практически не разработана методика индексирования на этом ИПЯ для ЭК. Между тем, методики индексирования для систематического и электронного каталогов должны существенно различаться.

Рассмотрим основные проблемы поиска с использованием иерархических классификаций.

*Иерархическое структурирование выдачи.* Одной из проблем поиска в ЭК остается сложность тематического структурирования больших массивов библиографических записей, выданных в ответ на широкотематические запросы. Такой подход обеспечивается структурой карточных систематических каталогов. А значит, снижается обозримость представленного пользователю информационного массива. Структурирование выдачи по иерархическому принципу гипотетически возможно, но пока такая возможность нигде не реализована.

*Понимание смысла индексов.* Практики указывают на то, что без специальных знаний или вспомогательных средств пользователи ЭК часто затрудняются оценить содержание буквенно-цифровых кодов.

*Недостаточная гибкость.* 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 использование этого языка в ЭК в принципе позволяет преодолеть ограничения, связанные с линейностью сложных индексов.

Предлагаются различные методы реализации классификационных ИПЯ в ЭК, а именно:

1. Использование полных индексов классификации как единой совокупности знаков, максимально отражающих содержание документа, в том числе поиск по индексам с различной степенью усечения в зависимости от поисковой задачи.

2. Применение иерархической классификации как посткоординатной системы, подразумевающее использование всех фрагментов полного индекса как независимых поисковых элементов, с возможностью усечения каждого элемента.

3. Использование словесных эквивалентов к цифровым индексам:

3.1. Замена каждого индекса в таблицах классификации соответствующей вербальной формулировкой (тезаурус на базе классификации);

3.2. Связывание классификационных индексов с соответствующими терминами «предметного входа» (тезаурус для классификации);

3.3. Связывание классификационных индексов с лексическими единицами других вербальных ИПЯ.

Возможности поэлементного поиска по индексам комбинационных классификаций были описаны еще в 1976 г. В. А. Мишиным<sup>29</sup>, однако в отечественных ЭК данная стратегия до сих пор на практике не реализована.

Продемонстрировать гипотетические возможности такого поиска можно на примере. Читателю необходимы книги по истории российского казачества в XVII–XVIII вв. Тему адекватно отражает индекс ББК «ТЗ(2)46-294», но документы с этим индексом в систематическом каталоге не найдены. Используя вышестоящий индекс «ТЗ(2)4-294» читатель получил две книги, содержащие в числе прочей и информацию об указанном в запросе историческом периоде. В ЭК можно еще более эффективно (если бы в современных ЭК была возможность поиска по элементам индексов) построить поисковое предписание, используя усечение: «ТЗ(2\$) \* «\$-294\$». При такой формулировке в выдачу войдут документы не только обо всех более широких исторических периодах, но и об истории казачества различных регионов и населенных пунктов России.

Пока же электронные каталоги в основном допускают поиск только по полным индексам с правосторонним усечением.

Аргументацию критиков применения иерархических классификаций в ЭК по причине их громоздкости и сложности для восприятия оп-

---

<sup>29</sup> Мишин В. А. Возможности машинного поиска информации по классификационным индексам // Организация систематических и предметных каталогов научных библиотек. – Вып. 2. – М., 1976. – С. 49–55.



ровергает N. Williamson<sup>30</sup>. Она указывает на то, 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. Возможности раскрытия структуры и содержания ИПЯ в авторитетных файлах гораздо шире, чем в печатных таблицах классификации.

В авторитетных файлах иерархической классификации особо важно обеспечить словарный доступ к классификационным индексам – «предметный вход». Поэтому нормативная запись обязательно должна содержать вербальную формулировку индекса, ссылки и отсылки («см.» и «см. также»), а также дополнительные ключевые слова. Предполагается, что поэлементный доступ окажется возможен не только к полным индексам, но и к их элементам, и к отдельным словам вербальных формулировок в составе авторитетных записей.

Проект Российского коммуникативного формата представления классификационных данных приведен на сайте Национальной службы развития системы форматов RUSMARC. Назначение данного формата – представление информации о классификационных индексах и связанных с ними заголовках, которые формируются в соответствии с определенной системой классификации. В формате RUSMARC для классификационных данных определены три вида классификационных записей. Запись основных таблиц – запись, в которой классификационный индекс или ряд индексов относится к основным таблицам классификации; классификационный индекс может быть простым или сложным. Запись вспомогательных таблиц – запись, в которой классификационный индекс относится к вспомогательным таблицам классификации. Запись термина индексирования – запись, в которой поле 254 Общий пояснительный термин индексирования содержит термин индексирования, который выражает понятие и не может быть связан с одним

---

<sup>30</sup> *Williamson N. J.* The role of classification in online systems // *Catal. a. Classif. Quart.* – 1989. – Vol. 10, № 1/2. – P. 95–104.

классификационным индексом. Предполагается, что записи в этом формате будут выглядеть следующим образом:

**Запись основных таблиц:**

*T3(2) История СССР*

*100/08 а (индекс основных таблиц)*

*250##\$aT3(2)\$hИстория. Исторические науки\$hИстория\$jИстория*

*СССР*

**Запись термина индексирования:**

*Автобиография писателя*

*См. в персоналии соответствующего писателя с типовым делением 12.*

*Например: Ш5(2=P)51-4 Пушкин 12 Автобиографии, дневники, воспоминания Пушкина*

*100/08 с (термин индексирования)*

*254##\$aАвтобиография писателя*

*753##\$aСм. в персоналии соответствующего писателя с типовым делением 12. Например:*

*Ш5(2=P)51-4 Пушкин 12 Автобиографии, дневники, воспоминания Пушкина.*

## Дескрипторные языки

**Дескрипторный информационно-поисковый язык** – это ИПЯ, предназначенный для координатного индексирования документов и информационных запросов посредством дескрипторов.

Алфавит дескрипторного ИПЯ совпадает с алфавитом естественного языка. Лексика дескрипторного ИПЯ включает дескрипторы и аскрипторы. Дескриптором называется лексическая единица, выраженная информативным (несущее в данном тексте наибольшую смысловую нагрузку, с точки зрения информационного поиска) словом (вербально) и являющаяся именем класса синонимичных или близких по смыслу ключевых слов. Аскрипторы (недескрипторы) – ЛЕ, подлежащие замене на дескрипторы при индексировании.

Возможность применения координатных ИПЯ появляется только в условиях автоматизированного поиска. Дескрипторные поисковые языки изначально разрабатывались с учетом требований алгоритмизации процесса поиска в автоматизированных информационно-поисковых системах. В интерактивных каталогах достаточно часто лексические единицы ЯПР преобразуются в дескрипторы.

Основное отличие дескрипторного языка от языка предметных рубрик состоит в посткоординации ЛЕ. Этим обусловлена большая гибкость дескрипторного языка при описании содержания документов и запросов. Позволяя осуществить поиск по любому, заранее не заданному сочетанию признаков, ДИПЯ, как писал М. А. Пробст, «дает потребителю средства индивидуализировать запрос»<sup>31</sup>.

Но в электронной среде отличия такого рода между ИПЯ постепенно стираются. В то же время в результате поиска на ДИПЯ может возникнуть ложное сочетание поисковых признаков и выдача, в результате поиска, документов, в которых тема запроса имеет второстепенное значение.

О том, что последняя проблема достаточно актуальна для ЭК, свидетельствуют материалы заседания специальной группы по интерактивным каталогам на конференции Американской библиотечной ассоциации, где предлагалось сократить количество «второстепенных элементов» в поисковом образе документа во избежание чрезмерно большой выдачи<sup>32</sup>.

В качестве основного средства помощи пользователю в поисковых системах, реализованных на базе ДИПЯ, рассматриваются информационно-поисковые тезаурусы (ИПТ). ИПТ – нормативный словарь дескрипторного ИПЯ с зафиксированными в нем парадигматическими отношениями ЛЕ (ГОСТ 7.74–96). Парадигматические отношения здесь указывают на общность или противопоставление значений и использование лексических единиц. В соответствии с ГОСТ 7.25–2001<sup>33</sup> в тезаурус должны входить вводная часть, лексико-семантический указатель и дополнительные (факультативные) указатели: систематический, пермутационный, иерархический (рис. 7).

ИПТ облегчает отбор соответствующих запросу ЛЕ. Как правило, в этом качестве тезаурусы сравниваются с иерархическими классификациями. Как преимущество выделяется полииерархичность тезауруса, который «содержит в себе множество классификаций» и позволяет обеспечить значительно большее количество разнообразных семантических взаимосвязей между ЛЕ. С другой стороны, это обстоятельство снижает обозримость тезауруса, так как «установить взаимосвязи гораздо

---

<sup>31</sup> Пробст М. А. Тезаурус и информационный поиск // Науч.-техн. информ. Сер. 2. – 1979. – № 9. – С. 14–20.

<sup>32</sup> Engle M. LITA Online Catalog Interest Group meeting, Los Angeles, Febr., 1994 // Techn. Serv. Quart. – 1995. – Vol. 12, № 3. – P. 43–46.

<sup>33</sup> ГОСТ 7.25–2001. Тезаурус информационно-поисковый одноязычный. Правила разработки, структура, состав и форма представления. – М., 2001. – 13 с.

легче внутри строгой механики классификации, чем посредством перекрестных ссылок, которыми связаны тезаурусы»<sup>34</sup>.



Рис. 7. Структура информационно-поискового тезауруса

### Язык ключевых слов

**Язык ключевых слов** – ИПЯ, предназначенный для индексирования документов и информационных запросов посредством ключевых слов. **Ключевое слово** – это информативное слово (или словосочетание), приведенное к стандартной лексикографической форме и используемое для координатного индексирования (ГОСТ 7.74–96).

Судя по публикациям, под поиском по ключевым словам подразумеваются достаточно разнородные процессы: поиск по ключевым словам, отобранным из текста документа на основании различных методик<sup>35</sup>, при этом поисковый образ документа может быть дополнительно обо-

<sup>34</sup> Maniez J. A classifications still relevant in databases? // Doc. Languages and Databases. – Frankfurt/M., 1991. – P. 120–129.

<sup>35</sup> Воройский Ф. С. Индексирование документов в АБИС // Библиотека. – 1996. – № 9. – С. 42–44 ; Ханжин А. Г. Разработка методов координатного индексирования документов. Ч. 2 // Науч.-техн. информ. Сер. 2. – 1995. – № 9. – С. 14–19.

гашен либо индексатором, либо с помощью автоматических процедур; поиск по полным текстам аннотаций и рефератов<sup>36</sup>; по полным текстам заглавий и по наиболее информативным словам заглавий<sup>37</sup>; а также по словам из традиционных предметных рубрик в режиме посткоординации<sup>38</sup>. Каждый из перечисленных методов поиска имеет свои особенности. В современных ЭК чаще всего функция поиска по ключевым словам предполагает поиск по отдельным словам из всех текстовых полей библиографических записей: заглавия, предметных рубрик, ненормируемых ключевых слов, аннотации или реферата (если они есть).

Мнения по поводу роли языка ключевых слов в составе ЛО ЭК очень разнообразны: от высказываний о целесообразности придания языка ключевых слов статуса основного языка каталога крупной библиотеки<sup>39</sup> до предложений об использовании его в качестве дополнительного поискового средства<sup>40</sup>. Подводя итоги эксперимента по поиску в ЭК на ЯКС, И. П. Никольцева, О. А. Степанова делают вывод, что этот язык будет интересен, прежде всего для узкоспециальных, профильных библиотек, не имеющих в составе ЛО языков с развитым нормативным словарем, а также специальных баз данных в составе ЭК (например, краеведческих)<sup>41</sup>. Столь же разнообразны

---

<sup>36</sup> Белоногов Г. Г. Компьютерная лингвистика в ВИНТИ // ВИНТИ – Москве. – М., 1997. – С. 140–146.

<sup>37</sup> Sparck J. K. Summary performance comparisons TREC–1, TREC–2, TREC–3, TREC–4, TREC–5, TREC–6 // NIST Spec. Publ., Pap. 6<sup>th</sup> Text Retrieval Conf (TREC–6). – 1998. – № 500–240. – P. 131–138.

<sup>38</sup> Ханжин А. Г. Разработка методов координатного индексирования документов. Ч. 2 // Науч.-техн. информ. Сер. 2. – 1995. – № 9. – С. 14–19.

<sup>39</sup> Зайцева Е. М. Лингвистическое обеспечение автоматизированных информационно-библиотечных систем: современные требования и направления развития // Науч. и техн. б-ки. – 2000. – № 3. – С. 54–57; Фуралев О. А., Ручинская Е. М. Соотношение традиционного каталога с электронным каталогом и перспективы использования языка предметных рубрик в электронном каталоге // Библиотеки и ассоциации в меняющемся мире: новые технологии и новые формы сотрудничества : материалы 3-й Междунар. конф. «Крым-96», Форос, Ялта, Авт. Респ. Крым, Украина, 1–9 июня 1996 г. – М., 1996. – Т. 2. – С. 187–189.

<sup>40</sup> Разработка автоматизированных информационно-библиотечных систем : (метод. рек.) / Гос. б-ка СССР им. В. И. Ленина. – М., 1990. – 77 с.

<sup>41</sup> Никольцева И. П., Седьшева О. А. О возможности содержательного индексирования документов ключевыми словами (по материалам эксперимента, проведенного в РНБ на базе комплекса «Охрана окружающей среды» // Предметный поиск в традиционных и нетрадиционных информационно-поисковых системах. – СПб., 1998. – Вып. 12. – С. 60–73.

и высказывания различных авторов по поводу функций ключевых слов в автоматизированном поиске по отношению ко всем типам АИПС. Некоторые специалисты говорят о том, что дескрипторные ИПЯ уступают лидирующие позиции в области автоматизированного поиска языку ключевых слов, хотя отмечают в этой связи снижение показателей качества поиска<sup>42</sup>. В то же время Н. И. Гендина пишет об ошибочности тезиса об «отмирании» в условиях развития новых информационных технологий всех иных ИПЯ, кроме языка ключевых слов<sup>43</sup>. J. Rowley называет в качестве преобладающей тенденцию сочетания при поиске естественных и нормализованных языков индексирования<sup>44</sup>.

Общим признаком всех перечисленных методов является значительно меньшая степень лексического контроля при формировании поисковых словарей или же его отсутствие. Поэтому при составлении поискового образа запроса рекомендуется перечислять все возможные синонимы каждого слова в наименовании темы. Так происходит, например, когда пользователю ЭК нужна информация о психологии детей в возрасте от 14 до 18 лет. Педагоги, как правило, называют их старшеклассниками, врачи и психологи могут выбрать другие варианты. Поэтому в запросе указывается несколько наименований:

*(«старшеклассни\$» + «дет\$» + «подростк\$» + «юноши\$») \* «психологи\$»;*

или еще один пример:

*(«ЭВМ» + «электрон\$» + «автоматиз\$» + «компьютер\$») \* «библиоте\$».*

Как и для ДИПЯ, основным достоинством считается возможность поиска по любому сочетанию признаков, по новым проблемам, узко-

---

<sup>42</sup> Блюменау Д. И., Афанасова Л. Н. Индикаторный метод компьютерного свертывания в процессе обучения аналитико-синтетической переработке информации // Науч. и техн. б-ки. – 2001. – № 12. – С. 29–42.

<sup>43</sup> Гендина Н. И. Состояние теории, практики и подготовки кадров в сфере лингвистического обеспечения информационно-библиотечной технологии: стимулы и препятствия // Библиотеки и ассоциации в меняющемся мире: новые технологии и новые формы сотрудничества [Электронный ресурс] : материалы 10-й Междунар. конф. «Крым-2003», Судак, Авт. Респ. Крым, Украина, 7–15 июня 2003 г. – М., 2003. – Т. 2. – Режим доступа: <http://www.gpntb.ru/win/inter-events/crimea2003/index2.htm>.

<sup>44</sup> Rowley J. The controlled versus natural indexing languages debates revisited: a perspective in retrieval practice and research // J. Inf. Sci. – 1994. – Vol. 20, № 2. – P. 108–119.

тематическим запросам, терминология которых не отражена в контролируемых словарях.

*Проблемы поиска.* В плане помощи пользователю возможности ЯКС очень ограничены. Поиск по ключевым словам предполагает большую интеллектуальную нагрузку на пользователя и применение более сложной поисковой логики. Эти свойства очень затрудняют использование ключевых слов при самостоятельном индексировании запросов неподготовленными пользователями, на которых ориентированы современные ЭК. Это означает, что обеспечить удовлетворительную полноту поиска способен только пользователь, хорошо владеющий терминологией отрасли.

Однако практики<sup>45</sup> говорят о предпочтении пользователями этого языка. В то же время следует учитывать, что пользователь выбирает язык формулировки поискового предписания, но никак не стратегии 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<sup>46</sup>, 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 совпадение плана выражения естественного языка и ИПЯ, безусловно, имеет большое значение. Поэтому необходимо выявлять для пользователей специфические свойства ЯКС 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

---

<sup>45</sup> Гальцева Т. Е. Электронный каталог в библиотеке вуза // Науч. и техн. б-ки. – 1996. – № 8. – С. 6 – 11 ; Зайцева Е. М. Лингвистическое обеспечение автоматизированных информационно-библиотечных систем: современные требования и направления развития // Науч. и техн. б-ки. – 2000. – № 3. – С. 54–57 ; Мухина С. Кто ищет, тот найдет // Библиотека. – 2000. – № 12. – С. 24–25.

<sup>46</sup> Фуралев О. А., Ручимская Е. М. Соотношение традиционного каталога с электронным каталогом и перспективы использования языка предметных рубрик в электронном каталоге // Библиотеки и ассоциации: новые технологии и новые формы сотрудничества : материалы 3-й Междунар. конф. «Крым-96», Форос, Ялта, Авт. Респ. Крым, Украина, 1–9 июня 1996 г. – М., 1996. – Т. 2. – С. 187–189.

слов<sup>47</sup> зачастую не берутся в расчёт практиками, заимствующими методические положения из других ИПЯ. Соответственно, при поиске обостряются все названные выше проблемы, связанные с отсутствием нормализации лексики.

### Комплексный поиск

Под *комплексным поиском* будем понимать поиск с применением нескольких поисковых языков в рамках одной поисковой задачи. Как правило, в таком случае в одном поисковом предписании сочетаются ИПЯ разных типов, например, классификационный и вербальный.

Логическое сочетание любых поисковых признаков (полных индексов либо их элементов) стало возможно лишь в электронной среде. На базе комплекса ИПЯ могут быть получены качественно новые результаты поиска. Иногда они существенно превосходят результаты использования каждого ИПЯ в отдельности. Проанализируем несколько конкретных ситуаций.

Самый распространенный вариант – сочетание классификационных индексов с предметными рубриками и / или словами из заглавий документов. Такой способ задания поискового предписания удобно применять для того, чтобы выделить конкретные аспекты в рамках широкой отраслевой темы. Рассмотрим это на примере. Читателя интересуют книги о применении ультразвука в медицине. Чтобы получить поисковое предписание на вербальном поисковом языке (предметных рубрик, ключевых слов), нужно перечислить все возможные сферы применения ультразвука и все названия заболеваний, для лечения которых он используется. Сделать это с исчерпывающей полнотой очень трудно. Поэтому сочетание термина «ультразвук» с наиболее широким классификационным индексом окажется более удобным. Правостороннее усечение кода ГРНТИ позволит включить в ПП все подчиненные основному индексу темы и понятия (рис. 8):

*«PR=ультразвук\$» \* «76\$»;*  
*ключевые слова из код ГРНТИ;*  
*заглавий и ПР.*

---

<sup>47</sup> *Воройский Ф. С.* Индексирование документов в АБИС // Библиотека. – 1996. – № 9. – С. 42–44 ; *Воройский Ф. С.* Некоторые пути повышения качества поисковых характеристик электронных каталогов // Библиотеки и ассоциации в меняющемся мире: новые технологии и новые формы сотрудничества : материалы 3-й Междунар. конф. «Крым-96», Форос, Ялта, Авт. Респ. Крым, Украина, 1–9 июня 1996 г. – М., 1996. – Т. 2. – С. 286–289 ; *Ханжисин А. Г.* Разработка методов координатного индексирования документов. Ч. 2 // Науч.-техн. информ. Сер. 2. – 1995. – № 9. – С. 14–19.





## Заключение

В электронной среде возможности поисковых языков существенно обогащаются за счет поэлементного поиска и усечения терминов индексирования. Это приводит к изменению свойств ИПЯ, специфических для каждого языка.

Изменение принципов реализации поисковых процедур и возможностей доступа к элементам сложных индексов в условиях автоматизации приводит к сокращению различий между предкоординатными и посткоординатными ИПЯ. Порядок следования отдельных элементов индексов становится неважен. В то же время большую значимость приобретает близость плана выражения ИПЯ к естественному языку. Не менее значимы и принципы структурной организации нормативных словарей, характерные для цифровых иерархических классификаций и вербальных поисковых языков. Обусловленные этими свойствами функциональные различия на данный момент являются определяющими при использовании в ЭК различных лингвистических средств.

Каждый из рассмотренных нами ИПЯ имеет ряд устойчивых свойств, которые в одних вариантах поиска могут проявиться как несомненные преимущества, а в других – привести к неудаче. Соответственно, и сторонники, и критики применения того или иного поискового языка в ЭК находят в каждом случае веские аргументы. Для реализации различных типов поисковых запросов будут эффективны различные ИПЯ. Наиболее эффективный вариант – применение в ЭК комплекса взаимодополняющих по своим свойствам и функциональным возможностям поисковых языков.

В то же время очевидно, что существенная часть возможностей поисковых языков в отечественных ЭК пока не реализована. Основные направления развития ЛО интерактивных каталогов видятся в использовании единых ИПЯ и единых форматов представления информации, унификации структуры и лексики поисковых языков, стремлении к корректировке свойств ИПЯ программными средствами, развитии в ЭК функции помощи пользователям в использовании лингвистических средств.

## Содержание

Введение .....	3
Основные понятия.....	4
Язык библиографического описания .....	4
Форматы представления библиографических данных.....	9
Языки тематического поиска .....	14
Язык предметных рубрик .....	19
Иерархические классификации .....	27
Дескрипторные языки .....	34
Язык ключевых слов .....	36
Комплексный поиск .....	40
Заключение .....	42

*Учебное издание*

*Скарук Галина Анатольевна  
Жарикова Людмила Анатольевна  
Стукалова Анна Александровна*

## **Поисковые языки электронных каталогов**

*Конспект лекции по курсу «Справочно-поисковый аппарат»*

Редактор *Р. К. Суханова*  
Корректор *А. В. Овечкина*  
Верстка *Н. А. Айгаровой*

Подписано в печать 25.03.09. Формат 60x84/16. Бумага писчая. Гарнитура «Таймс».  
Печать офсетная. Усл. печ. л. 2,6. Уч.-изд. л. 2,5. Тираж 400 экз. Заказ № 102.

Редакционно-издательский отдел ГПНТБ СО РАН.  
630200, Новосибирск, ул. Восход, 15. E-mail: rio@spsl.nsc.ru.  
Полиграфический участок ГПНТБ СО РАН. 630200, Новосибирск, ул. Восход, 15.