

Э. Р. Сукиасян

РГБ

**Классификационная система
как инструмент автоматизированного поиска.**

**Часть 1. Почему классификационный поиск
эффективнее вербального?**

Проведено сравнение возможностей принципиально различных видов поиска – вербального и классификационного.

Ключевые слова: автоматизированный поиск, вербальный поиск, классификационный поиск, тематический поиск, содержательный поиск.

UDC 025.4.03

Eduard Sukiasyan

Russian State Library

Classification system as a tool for automated search.

**Part 1. Why is classification search more effective than
a verbal one?**

The functionalities of principally different types of search – verbal and classification search are compared.

Keywords: automated search, verbal search, classification search, subject search, content search.

В одной из последних статей о ББК я пообещал написать о классификационном поиске, как говорят, «в автоматизированном режиме». Тема очень интересная, многое забыто: в последние десять лет о поэлементном поиске в печати ничего не найти. Очень жаль: это ведь наша, отечественная, разработка. Приоритет закреплён первыми публикациями в печати, рядом диссертационных работ. Почему же мы не знаем, не стараемся применить на деле то, что придумано у нас в стране?

Сначала рассмотрю «теорию» вопроса, возможности вербального и классификационного поиска. Надо объяснить, почему в зарубежных системах почти всегда используется вербальный поиск. К сожалению, так же обстоит дело и в отечественных. Во второй части статьи (будет опубликована в № 4 «НТБ») я расскажу о поэлементном классификационном поиске, покажу его преимущества, раскрою технологию, рассчитанную на компьютерное использование. Попутно опишу и некоторые другие забытые сегодня возможности ББК (например, классификационно-дескрипторный язык). Придётся сказать правду о том, как и почему были отложены в сторону авторские проекты, преданы забвению результаты экспериментов.

Вопрос об инструментах поиска имеет свою историю. В 1992–1993 гг. в журнале «Научные и технические библиотеки» был опубликован цикл статей А. И. Вислого и В. В. Мосягина о компьютеризации библиотеки. Много не потеряло своего значения. Тогда я написал полемические заметки [1], через десять лет вновь вернулся к тому же вопросу [2]; позже эти статьи вошли в сборник [3]. Говоря о содержательном, тематическом поиске, я писал: «Правильно отмечают А. И. Вислый и В. В. Мосягин: наибольшая эффективность будет достигнута с помощью классификационных индексов, меньшая – с помощью ПР или ключевых слов. Не надо проводить параллели: “классификационные индексы” вовсе не обязательно определяют аналогию с традиционным карточным СК. Так же, как и в случае с АК, возможности компьютеризации здесь огромны. В карточном СК поиск идёт слева направо, а индекс воспринимается так же, как любое порядковое число. Возможности развитой классификационной системы (УДК, ББК) в карточном каталоге раскрываются не более чем на 15%. Ведь индекс состоит из значимых элементов, и только машина может вести глубокий многоаспектный (правильнее: многоэлементный) поиск. Больше того: компьютер может по каждому из элементов вести поиск по иерархии («выше – ниже»), а поисковое предписание определит то сочетание признаков, ту их комбинацию, которая полностью раскроет тему. Можно также комбинировать тематические (содержательные) признаки и элементы БО: в этом случае возможности поиска конкретного издания существенно увеличиваются. Умышленно избегаю термина “булева алгебра”, который не всегда понятен библиотекарям-гуманитариям» [Там же. С. 438].

Этим, собственно говоря, всё сказано. Вербальный поиск (по ПР и ключевым словам) – это поиск по внешнему образу слова (давайте признаем: это поиск по буквам, которые образуют слово данного языка). Классификационный поиск имеет отношение к сущностям, семантике, смыслам, которые обозначают элементы классификационных индексов (не только

цифры и буквы, как можно подумать, но и вся система применяемых опознавательных знаков – пунктуационных и математических). Например, в ББК сочетание цифр с предшествующим знаком равенства относится к таблице языковых типовых делений, поставленное в круглые скобки – уже к этническим типовым делениям, например: = 632.94 Крымско-татарский язык, но (=632.94) Крымские татары.

В большинстве случаев, если мы имеем дело с тематическим, содержательным поиском, нам совершенно всё равно, каков будет «внешний образ», какие слова (знаковые образы) будут напечатаны на титульном листе, какие термины выбраны индексирующим лицом для характеристики содержания документа, релевантного нашему запросу. Например, если мы вписываем в «рамочку» запроса в ЭК название города, то уверены: хоть что-нибудь получим – план, карту, справочник, атлас, альбом, путеводитель! Более того, мы надеемся получить библиографические сведения о тех источниках, в которых город был назван до переименования. Запрашиваем, например, «Сергиев Посад». В «нормальном» ЭК нам дадут справку: «Назв. с сент. 1991 г., в 1930–1991 – Загорск, до 1930 – Сергиев». И выдадут всю литературу.

Спросите, какие ЭК я отношу к «нормальным»? Те, которые сохранили в новой электронной форме культуру обслуживания читателей средствами карточных каталогов, выработанную десятилетиями библиотечной практики. Например, имеют ссылочно-справочный аппарат, с помощью разделителей «разговаривают» с читателями. Хороший карточный каталог обязательно свяжет все варианты переименований организации (например, МГБИ – МГИК – МГУК – МГУКИ, с указанием дат).

Отказавшись от карточных каталогов в пользу методически безграмотного ЭК, мы ставим читателя в такое положение, когда он часто получает ответ на экране – «Результат – 0 документов». В карточном систематическом каталоге, с которым десятилетиями работали квалифицированные специалисты, читатели хорошо ориентировались. Знали: если ничего не нашли, надо посмотреть разделитель «впереди», на ступень выше. Для дилетанта систематический каталог оставался «вещью в себе», непреодолимой преградой. До сих пор не могу понять, почему некоторые специалисты с инженерным (или математическим) дипломом оказались дилетантами на поприще библиотечной автоматизации. Неужели они ничего не искали в каталогах в студенческие годы? Как же они писали дипломы и диссертации – ведь в те времена не было ЭК.

Говоря об эффективности вербального поиска, мы имеем дело с мифологией, верой. Научных доказательств никто привести не может, поэтому, наверное, и исследования не проводятся: их результаты всегда очевидны.

Семантический уровень поиска (т.е. по смыслам, а не по словесным образам) достигим только в работе с тезаурусом, в котором в процессе разработки обеспечиваются выявление и учёт самых разных парадигматических отношений (выше – ниже, целое – часть, система – элемент, причина – следствие и т.д.). Но тезаурус – это сложно. Проще поставить «ключевое слово».

Если мы имеем дело с вербальным поиском, то должны понимать объёмные ограничения. Во-первых, они обусловлены известными «вредностями» естественного языка, такими, например, как синонимия и омонимия. Специалисты считают, что с помощью автоматизации словарей эти недостатки можно преодолеть. Но смысл (семантику) запроса можно выразить фразой, составленной из слов, не являющихся ни синонимами, ни омонимами. Иначе говоря, синонимичными могут оказаться не только слова, но и фразы.

Чтобы было понятно, приведу пример из собственной практики. Теорией поиска я никогда серьёзно не занимался, темой моей диссертации была централизованная классификация (сегодня мы говорим о централизованной систематизации). В 1970-х гг. по Москве «гремела слава» автоматизированной ИПС, работающей в ВИНТИ. Авторы – коллектив лингвистов, математиков и программистов, в который входили Э. С. Бернштейн, Д. Г. Лахути и В. С. Чернявский, – выступали с докладами, публиковали статьи. Становилось понятно: эра карточных (систематических и предметных) каталогов кончилась, завтра можно будет купить комплект машин, ввести туда информацию, и каталогизаторам делать будет нечего.

Я часто бывал в ВИНТИ, однажды зашёл в комнату с «волшебной машиной» и попросил провести поиск. Мне сказали: назовите тему диссертации. Тут я схитрил: вместо «Централизованная классификация» (эти слова встречались в названиях книги, автореферата, дюжины статей) произнёс: «Публикация индексов в изданиях». По сути – тот же смысл. Но для «первой в СССР и в мире автоматизированной документальной информационно-поисковой системы с автоматическим (пословным) индексированием текстов (русскоязычных) документов» подвох был непреодолим: в выдаче оказалось... «0 документов».

Проверили. Слова из «параллельного мира» были введены с моей фамилией в поле автора. Но произошло странное: мне было сказано, что я «неправильно сформулировал запрос». С тех пор мои дискуссии со сторонниками вербального подхода закачиваются одинаково: оказывается, что я чего-то «недопонимаю», а объяснить суть никто не берётся. Справедливости ради скажу: был человек, с которым можно было дискутировать – Софья Константиновна Виленская (1918–2000), прекрасно разбиравшаяся в классификационных системах (знала и ББК, и УДК), предметизации, коор-

динатном индексировании. Полемизировать же со специалистами (лингвистом, математиком, программистом), отвергающими классификационный подход «с порога», без аргументов, непродуктивно.

Другие ограничения связаны с тем, что при вербальном поиске применяется тот или иной естественный язык. С. К. Виленская однажды в разговоре со мной заметила: многообразие языков в совокупности со всеми особенностями их синтагматики (наличие категорий рода, числа, склонений, спряжений и пр.) мешает переводу зарубежных практических пособий по предметизации на русский язык. Действительно, не владеющий языком предметного каталога специалист не может (даже с переводчиком) построить запрос на «чужом» для него языке. В грузинском, например, нет предлогов. Их функцию выполняют окончания. Попробуйте представить себе архитектуру грузинского предметного каталога.

Язык систематического каталога интернационален. Мне не нужно знать, как произносится индекс 025.45. Достаточно того, что я знаю классификационный индекс. Поэтому в полутора сотнях стран для меня поиск доступен. И не только в каталоге, но и в фонде, на полке. Хотя каталожные и полочные разделители будут непонятны для меня (так же, как и язык ЭК, – свой в каждой уважающей себя стране).

Наши ЭК – на русском языке, богатышем, но сложном. Далеко не все россияне могут правильно сформулировать запрос. В советскую пору мне пришлось разговаривать с читателем, который свой запрос описал словами «сетка» или «решётка». Но с помощью таблиц ББК мы довольно быстро выяснили: речь идёт об автомобильном транспорте (39.3), двигателях (39.35), карбюраторах (39.353), топливных фильтрах. Попробуйте мне подсказать, каким должен быть алгоритм обслуживания такого читателя, если никаких инструментов, кроме ключевых слов, в его распоряжении нет?

Почему же все отечественные АБИС используют вербальный поиск? Это вопрос, на который давно дан ответ. И не один. Но складывается впечатление, что литературу у нас читают не все! Во всяком случае, разработчики, программисты, те, кого мы называем компьютерщиками, считают, что чтение профессиональных библиотечных изданий ниже их достоинства. У них «свои» компьютерные журналы, свои конференции. На наших конференциях они просто «упиваются» собственными проектами и достижениями (пробовал пойти с ними на контакт – не получилось).

Для проверки своей рабочей гипотезы (о том, что не читают библиотечную литературу), провёл пару раз «неприличный» эксперимент: позволил себе в статье назвать «нехорошими словами» нескольких специалистов. Был уверен: прочитают, не сами, так их сотрудники, передадут. Готовился к

тому, что руки больше не подадут, был готов даже судиться за оскорбление. Но ничего не произошло! Профессиональную библиотечную периодику даже руководители не всегда читают.

В октябре 2006 г. участники Всероссийского совещания библиотек – методических центров подготовили «Обращение к руководителям разработки ведущих АБИС, применяемых в библиотеках Российской Федерации». Этот небольшой документ был опубликован во многих профессиональных изданиях [4]. Кроме того, мы разослали его текст по адресам всех отечественных АБИС. Мы предлагали сотрудничество. В ответ – молчание.

Не секрет, что большинство АБИС, разработанных в нашей стране, не оригинальны – у каждой из них (особенно в начальных версиях) чётко угадывается первоисточник (как правило, американский или европейский).

Особенность американской библиотеки – «предметный вход» в единственном каталоге, который так и называется – *Catalog*. Сами американцы, не всегда знающие историю, не могут объяснить, «какой» это каталог (а всем это очень хочется узнать!). В едином алфавите стоят карточки на авторов, заглавия и предметные рубрики. Такой каталог, появившийся в середине XIX в., стал называться «Словарным», а первое руководство по его ведению («Правила словарного каталога») было издано в 1876 г. Чарльзом Кеттером (1837–1903). Сегодня американцы уже не помнят, как этот каталог называется, и отвечают русским посетителям: «У нас один каталог, который отвечает на все вопросы. А у вас в стране несколько? Можем вас пожалеть».

Однако, например в испаноязычном мире, даже алфавитные каталоги делятся на «Каталоги по именам» и «Каталоги по заглавиям» – и читатели считают это удобным. В Европе и Америке нет ограничений на отражение документов по их заглавию. У нас, как известно, есть: на заглавие карточки создавались в тех случаях, когда автор на обложке не указан, а также при отражении учебной и справочной (в том числе словарной) литературы.

Теперь понятно: «впечатывая» в рамочку ЭК слово, мы оказываемся на месте американского читателя, который в карточном каталоге привык искать «слово» (каталогу всё равно, будет ли это имя автора, заглавие или предметная рубрика). Выражение «поиск осуществляется по любому слову библиографической записи» придумано в России. Лучше бы не было такой «услуги», которая ведёт к столь великому «шуму».

Зачем же США поддерживает две национальные классификационные системы?

Статус национальной системы имеет Десятичная классификация Дьюи (ДКД), разработанная великим американским библиотекарем, основателем нашей профессии Мелвилем Дьюи (1831–1931). Начиная с первого издания

(1876) ДКД постоянно совершенствовалась, росла в объёме, в 13-м издании (1932) превратилась из перечислительной в комбинационную систему. В 2011 г. вышло в свет 23-е издание в 4-х томах, и сразу же были объявлены основные направления пересмотра в следующем 24-м издании.

С 2013 г. пост главного редактора ДКД занимает Михаэль Панцер (Кёльн, ФРГ). Впервые на эту должность назначен специалист не из США. ДКД – частная коммерческая система, принадлежащая издательству *Forest Press*, основанному М. Дьюи. Для получения прав на перевод, издание и применение приобретаются лицензии. ДКД поддерживает и развивает *OCLC* (с 1988 г.), оплачивающий деятельность небольшого штата. Принципиальные решения принимаются международным правлением (10 человек), собирающимся дважды в год.

Вторая национальная классификационная система США – Классификация библиотеки Конгресса (КБК). Условной датой рождения КБК считается 1912 г., хотя предложение о разработке системы высказано в конце XIX в. КБК – система, специально предназначенная для расстановки многомиллионных библиотечных фондов. Основные классы и подклассы обозначены прописными буквами от А до Z, к которым присоединяются как цифровые, так и буквенные обозначения, как правило, перечислительные (номерные, порядковые). Таблицы КБК издаются отдельными выпусками (по классам) по мере необходимости. Поэтому, если для ДКД обязательный опознавательный признак – это номер издания (например, ДКД20), то для КБК эту функцию выполняет год публикации.

ДКД – наиболее распространённая в мире система (первое место по числу переводов и количеству стран). КБК используется в университетских библиотеках США и Канады, некоторых других стран. Обе системы применяются по основному предназначению (для систематической расстановки фондов). Увидеть систематический каталог по ДКД мне не удалось. Такие каталоги организуются на основе таблиц УДК – в европейских странах, Японии (где сильны традиции УДК). Свою, весьма оригинальную, практику применения УДК имеет ФРГ.

Мифом является распространяемое дилетантами от классификации мнение о схожести ДКД и УДК. Да, УДК (тогда она называлась Руководством Международного библиографического института) была разработана П. Отле (1868–1944) и А. Лафонтеном (1854–1943) в 1895–1905 гг. на основе ДКД (разрешение было передано лично М. Дьюи). Они предложили таблицы типовых подразделений, которые в России назвали определителями УДК. Позже – на протяжении десятков лет – таблицы развивались автономно. Например, в ДКД, в отличие от УДК, сохранился класс 400, не применяются известные в УДК опознавательные знаки. Типовые подразделения

присоединяются к основным индексам с помощью тире, следующая за ним цифра показывает, из какой именно таблицы взято деление. Вспомогательные таблицы называются *Tables*, основные – *Schedules*. Порядок их применения строго регулируется правилами. Алфавитно-предметный указатель в ДКД по традиции, идущей от М. Дьюи, называется *Relative index* (Указатель родственных отношений).

Сделаем вывод: читателям США (и многих других стран) предлагается ознакомиться с фондами, организованными в систематическом порядке. Возможности классификационного «входа» реализуются с помощью открытия фондов для читателей. Классификационные индексы знают даже школьники. В лифте многоэтажного университетского книгохранилища указаны не этажи, а классификационные индексы. На этаже – полумрак, но у входа – огромное табло с планом размещения фондов. Достаточно прикоснуться к индексу, чтобы в зале осветился тот сегмент, который вам нужен. Берите кресло на колёсиках (оно снабжено полкой для сумки или рюкзака, поднимающимся сбоку столиком, автономной системой освещения, крючками для одежды) и направляйтесь к стеллажу.

С найденной книгой совсем не обязательно спускаться в зал – можно работать на этаже. Книги, взятые с полок, не рекомендуется возвращать на место. Их кладут на «красную полку» (у самого пола), потом сотрудники библиотеки считают и расставляют книги. Всё это я видел своими глазами ещё в «том веке». Потом об этом писал, рассказывал. Почему сегодня за границу ездят целыми группами, но многого не видят? В командировки нужно ездить в качестве библиотекаря, а не гостя.

Не нужен читателям в США систематический каталог! Универсум знаний раскрывается в фонде, а качественный поиск обеспечивает «предметный вход». Почему же вербальный поиск во многих странах использует предметные рубрики, а не примитивные, лишённые парадигматики ключевые слова?

Теория предметизации в США начала развиваться на полвека раньше, чем в нашей стране. Но это не помешало отечественным специалистам внести свой вклад в развитие теории и практики предметизации, несмотря на то, что начиная с 1920-х гг. Н. К. Крупская выступала против предметного каталога (не обеспечивающего, по её мнению, необходимый «уровень партийной пропаганды»).

В 1950-х гг. А. Я. Кушultz (1907–1985) убедительно доказала, что дискуссия «Систематический или предметный каталог?» беспочвенна, так как используемые инструменты индексирования раскрывают различные стороны семантического содержания документа. С. К. Виленская внесла в теорию предметизации бесценный вклад: ввела понятие предметного комплексиро-

вания, обосновала наличие узких и широких предметных комплексов, а предметные рубрики охарактеризовала как адекватные и обобщающие. За рубежом до сих пор говорят о широких (*broad*) и узких (*narrow*) предметных рубриках, а предметных комплексов не знают.

В США пошли по грамотному пути создания единого для всех библиотек Списка предметных рубрик. Авторитетный *LCSH* (Список предметных рубрик Библиотеки Конгресса) приобрёл нормативные функции. На его основе была подготовлена сокращённая версия для небольших библиотек. В структуре и содержании *LCSH* отразились достижения теории координатно-индексирования. Сегодня это ещё не совсем тезаурус, но уже и не список. Сохранились ссылочные и справочные указания, но введена и парадигматика (показаны родовидовые связи, а во многих случаях и отношения: целое – часть, система – элемент и др.).

В последние годы в нашей стране по инициативе ЛИБНЕТа функции единого списка предметных рубрик (с правами авторитетного файла) приобрёл аппарат, который многие годы отработывался в Российской национальной библиотеке. Казалось бы, пора прекратить порочную практику индексирования ключевыми словами: открыт доступ к грамотному аппарату, эффективность вырастет во много раз. Но мы как работали, так и работаем.

Массовое внедрение АБИС с примитивными по функциям и возможностям ЭК нанесло колоссальный ущерб развитию библиотечного обслуживания. Читатели остались наедине с компьютерами – в их программах нет диалога, нет никаких ссылок и справок. Исчез карточный алфавитно-предметный указатель (его можно было бы вести в виде электронного файла), благодаря которому раскрывалась многоаспектность литературы.

По моим оценкам, основанным на собственных впечатлениях и литературных источниках, в 1980-х гг. мы отставали от США на 20–30 лет. В 1990-х гг. там запустили в работу новые системы с огромными возможностями. Прошло два десятка лет. У них и каталогов не осталось – в нашем традиционном понимании. А что происходит у нас? Вот как мне ответили на этот вопрос в одной из библиотек: работаем над новыми версиями, развиваем быстродействие. А как эти «достижения» оценили ваши читатели? Мне отвечают: это сложно, практически невозможно – читатели у нас проходят по одной «линии», автоматизация – по другой.

Чтобы двигаться вперёд, надо хотя бы знать направление, вектор развития. Об этом – во второй части статьи.

СПИСОК ИСТОЧНИКОВ

Сукиасян Э. Р. Компьютеризация библиотеки: с чего начать и чем закончить? // Науч. и техн. б-ки. – 1993. – № 12. – С. 19–30.

Sukiasyan E. R. *Kompyuterizatsiya biblioteki: s chego nachat i chem zakonchit?* // *Nauch. i tehn. b-ki.* – 1993. – № 12. – S. 19–30.

Сукиасян Э. Р. Компьютеризация библиотеки: с чего начиналось, тем и закончилось // Там же. – 2003. – № 9. – С. 17–28.

Sukiasyan E. R. *Kompyuterizatsiya biblioteki: s chego nachinalos, tem i zakonchilos* // *Tam zhe.* – 2003. – № 9. – S. 17–28.

Сукиасян Э. Р. Каталогизация. Классификация. Электронные каталоги и автоматизированные библиотечные системы : избр. статьи / Э. Р. Сукиасян. – С.-Петербург : Профессия, 2009. – 535 с.

Sukiasyan E. R. *Katalogizatsiya. Klassifikatsiya. Elektronnye katalogi i avtomatizirovannye bibliotchnye sistemy : izbr. stati / E. R. Sukiasyan.* – S.-Peterburg : *Professiya*, 2009. – 535 s.

Обращение к руководителям разработки ведущих АИБС, применяемых в библиотеках Российской Федерации // Информ. бюл. РБА. – 2007. – № 44. – С. 92–100; Библиотека. – 2006. – № 11. – С. 63–65; Науч. и техн. б-ки. – 2007. – № 4. – С. 100–114; Унив. кн. – 2006. – № 12. – С. 50–53.

Obrashchenie k rukovodityam razrabotki vedushchih AIBS, primenyaemyh v bibliotekah Rossiyskoy Federatsii // *Inform. byul. RBA.* – 2007. – № 44. – S. 92–100; *Biblioteka.* – 2006. – № 11. – S. 63–65; *Nauch. i tehn. b-ki.* – 2007. – № 4. – S. 100–114; *Univ. kn.* – 2006. – № 12. – S. 50–53.

Eduard R. Sukiasyan, *Cand. Sci (Pedagogical)*; Assistant professor, Sector Head of Chief editorial board of Library Bibliographical Classification Research Center, Russian State Library

sukiasyaner@rst.ru

3 Vozdvizhenka street, Moscow, 119019 Russia