

ИНФОРМАЦИОННО-ПОИСКОВЫЕ ЯЗЫКИ И СИСТЕМЫ

УДК 025.4

Э. Р. Сукиасян

Российская государственная библиотека

Тезаурофасет – ИПЯ XXI века

Рассказано о сущности понятия «тезаурофасет», особенностях системы и технологии работы, обеспечивающей высокую эффективность поиска. Освещены перспективные направления модернизации поисковых систем.

Ключевые слова: информационно-поисковые языки, тезаурофасет, тезаурус, классификации библиотечно-библиографические, Ш. Р. Ранганатан, М. Дьюи.

UDC 025.4

Eduard Sukiasyan

Russian State Library, Moscow, Russia

Thesaurofacet – the information retrieval language of the 21st century

The term of “thesaurofest” is scrutinized; the characteristics of the system, the technologies of efficient information retrieval are analyzed. Prospects for retrieval systems modernization are discussed.

Keywords: Information retrieval languages. Thesaurofacet, thesaurus, library and bibliographic classifications, Melvil Dewey, Shiyali Ramamrita Ranganathan.

Термин *тезаурофасет* появился на несколько десятилетий позже понятия, раскрывающего сущность нового поискового инструмента. Выдающийся индийский учёный Ш. Р. Ранганатан (1892–1972) не разделял информационный поиск на вербальный (словесный, выраженный элементами естественного языка) и классификационный (использующий собственный искусственный язык). В своих лекциях Ранганатан объяснял: слова и клас-

сификационные индексы – это элементы единой системы, отображающие две её стороны в единой плоскости. Затем в его руках появлялась полоска бумаги, на обеих сторонах которой было что-то написано. Профессор объяснял красиво и понятно: с одной стороны все видели надпись на понятном естественном языке, могли её прочитать; на обороте – классификационный индекс. Если мы его понимаем, то можем перевести в словесные образы того же естественного языка. И если все операции совершены правильно, то образы совпадут. Обратите внимание: образы, а не слова.

Образы, понятия, сущности в информационном поиске – *онтологии*. Термин, часто употребляемый Ранганатаном, взят непосредственно из средневековой философии: он предложен в 1613 г. Рудольфом Гоклениусом (1527–1628), впервые использовавшим также термин *психология*. Мы оперируем сегодня понятием *термин индексирования (indexing terms)*. Ключевые слова, предметные рубрики и дескрипторы – термины индексирования вербального ИПЯ, а классификационные индексы – термины индексирования классификационного ИПЯ.

У Ранганатана вы не найдёте понятие *термин индексирования*. Его придумали наши современники за рубежом. Ранганатан пользовался интересным понятием *device*, которое в словарях определяется как схема, план, проект, приём, механизм, аппарат, прибор, эмблема, девиз, способ, средство.

Вернёмся к лекции Ранганатана. Говоря, что в принципе все образы лежат на одной поверхности, он останавливался, брал в руки полоску бумаги, на которой подряд, друг за другом были записаны все девайсы придуманной темы – на разных поисковых языках. Незаметное для аудитории движение – и концы полоски соединялись, в руках оказывалось кольцо Мёбиуса. Вспомним: Ранганатан был ещё и математиком, профессором топологии. Кольцо Мёбиуса, «открытое» в 1875 г., не имеет двух поверхностей. Все девайсы оказывались рядом, слившись в единую *цепь (chain)* – ещё одно любимое слово Ранганатана), *кольца (ring)* – термин тоже взят из топологии) которой не могут быть первыми и последующими...

Итог рассуждений: нельзя знать предметизацию (*subject indexing*), не владея методами классификационного анализа. Но у нас, как известно, можно: ограничив свои навыки «теорией ключевых слов» (а есть ли она?), исключив все другие девайсы из собственного мира практического индексирования.

Читать работы Ш. Р. Ранганатана сложно, но ещё сложнее понимать скрытые в них «смыслы». Он считал своего предшественника (тоже математика!) М. Дьюи (1851–1931) гениальным уже потому, что тот назвал алфавитно-предметный указатель к своей Десятичной классификации *Relative index* (мы переводим: *указатель родственных отношений*). Мы так и не

поняли, в чём же «гениальность» Дьюи: указатели к классификационным таблицам делались и раньше, назывались просто указателями. Трудно ли расставить все «слова» по алфавиту? Сегодня это может сделать машина.

В нашей стране систематический каталог – отражение реального фонда в классификационной системе – мог существовать (боюсь, что существует кое-где и сегодня) без всякого указателя. Только в 1984 г. обязательность указателя была регламентирована ГОСТом 7.84–84 (сейчас это требование отражено в ГОСТе 7.59–2003). Подробное обоснование изложено в [1]. Тем не менее большинство отечественных ИПС располагает одним «входом» – чаще вербальным (в самой примитивной его форме – ключевых слов). Систематические карточные каталоги, имеющие, безусловно, более высокие поисковые характеристики, во многих библиотеках ликвидируют. При этом решение администрации носит всегда субъективный характер: нам неизвестно ни одного случая, когда ему предшествовало исследование, проведённое с привлечением читателей.

Одним из первых термин *тезаурофасет* применил в своей статье ученик Ранганатана – Т. Раджан (Бангалор, Индия) в 1976 г. [2]. Появление термина в мировой литературе обычно связывают с именем Джин Эйтчисон (Великобритания), статья которой была опубликована в 1970 г. [3]. К этому времени Д. Эйтчисон уже была хорошо известна благодаря своим практическим работам (тезаурусам), участвовала, например, в разработке Тезауруса ЮНЕСКО. За несколько лет ей удалось опубликовать ряд статей (некоторые – с соавторами), выступить на ряде международных конференций.

И всякий раз Д. Эйтчисон стремилась подчеркнуть главную мысль: тезаурофасет – это не публикация в одном томе (как можно подумать) фасетной классификации и тезауруса, пусть даже в границах одной предметной области, хотя бы и семантически связанных друг с другом. Это нечто иное, использующее лучшие приёмы и высшие достижения теории и практики индексирования. Это особый ИПЯ, стоящий выше отдельно взятых тезауруса и систематического указателя к нему и классификационной системы со словарным указателем. Он более сложный, как можно подумать, а, напротив, удивительно простой, лёгкий в использовании. И уже потому более эффективный.

Согласимся с этими положениями. Современные комбинационные классификационные системы достигли своего апогея. Их структура позволяет оперировать тысячами комбинированных индексов (сложных и составных), их семантика обладает огромной мощностью (сотни тысяч комбинаций), детализация доходит до 12–14 ступеней. Поисковые операции требуют знания особых правил.

Не менее сложны по своей структуре тезаурусы. Создание тезауруса принято считать крайне трудоёмкой операцией. Кто-нибудь пробовал читать универсальный или многоотраслевой тезаурус или просто работать с ним? Спрошу иначе: как вы считаете, можно ли найти в тезаурусе ошибку или пропуск дескриптора? На этот вопрос Д. Эйтчисон отвечает так: я сама делаю тезаурусы и уверена, что их можно развивать бесконечно. Но этот труд должен быть коллективным. А это значит: тот, кто пользуется тезаурусом, должен его пополнять. Он должен и хотеть это делать, и понимать, как это просто сделать.

В 1982 г. был предложен адекватный термин: *классаурус* (это то же самое, что тезаурофасет) [4]. Между прочим, второй термин отражает суть дела правильнее. Ведь сначала поиск ведётся по классификационной системе, выбирается предметная область. В рамках установленных пользователем содержательных границ поиск продолжается по тезаурусу, отражающему только понятия конкретной предметной области.

Авторы, имена которых я называю, являются (или являлись) членами ИСКО – Международного общества по организации знаний. В справочнике ИСКО, которым я располагаю как член Общества, названы их имена. Т. Раджан родился в 1928 г., Д. Эйтчисон – в 1925. Предпринятые попытки связаться с ними к успеху не привели.

Я вспоминаю слова Ранганатана, произнесённые им на одном из юбилейных заседаний: «Молодёжь часто забывчива». Наша молодёжь, к тому же, не любит трудностей, связанных с языком, не увлекается изучением теории ИПЯ и ИПС. Посмотрите, сколько диссертаций защищается в стране, например, по истории библиотечного дела. Наверное, хорошо работают преподаватели! Поэтому мы знаем, что и как у нас было, но не можем представить себе, что будет, что для этого надо предпринять.

Создать систему, используя технологию тезаурофасета, оказывается, проще и дешевле. Мы же любим говорить о грандиозных замыслах будущей новой универсальной классификационной системы (или о проектах разработки универсального тезауруса). Как правило, все эти проекты остаются на бумаге.

С тезаурофасетом всё обстоит проще, к тому же он дешевле. Совсем не нужна глубоко детализированная классификационная система. Не нужен всеобъемлющий тезаурус с многочисленными пересечениями парадигматики. Достаточно элементарной, но очень грамотной фасетной классификации. Обычно спрашивают: сколько уровней? Я нашёл ответ: должны найти место все существующие, известные специальности, по которым ведётся

подготовка кадров. Медицина, например, это очень широко, офтальмология – достаточно, астигматизм – слишком узко. Важно, чтобы в классификационной части медицины были, по возможности, отражены все фасеты.

Какие в медицине фасеты? Любой специалист за рубежом сразу же начинает перечисление, так как его учили проводить фасетный (у нас принят термин *категориальный*) анализ: а) анатомические понятия – части (голова, руки, ноги) и их подразделения; функции – кровообращение, дыхание, ...; б) болезни (есть принятые в мировой практике классификаторы – их и надо взять); в) методы диагностики и лечения – лекарственная терапия, физиотерапия, ...; г) отрасли медицины; д) пациенты (здоровые и больные) – по разным признакам; е) фармакология (огромный класс химических средств). Это – далеко не все категории медицины. Собственный ряд категорий имеет, например, организация здравоохранения или такая сфера деятельности, как гигиена и санитария. Конечно, должны быть раскрыты все общие категории (в применяемых у нас системах мы называем их определителями или типовыми делениями).

Для каждой науки, области практической деятельности, отрасли уже давно определены их категории. То, что они «складированы» у нас в пассивной части памяти (если назову – вспомните сразу), не украшает наше образование. Нас больше учат описанию, в то время как за рубежом – анализу.

Процесс индексирования в системе тезаурусасет определён правилами и не допускает субъективизма. Определяем предметную область путём просмотра таблиц основных делений фасетной классификации, находим перечень фасетов (категорий). Последовательно фиксируем все выраженные в содержании категории. Некоторые из них могут быть скрыты. Затем работаем в границах каждой категории по предлагаемому тезаурусу. Термины отбираем в тексте источника. Технология координатного индексирования не ограничивает число дескрипторов. Но все они должны быть представлены в тезаурусе. Некоторые могут иметь статус неиспользуемого с указанием принятого. Такое решение действует в отношении синонимов или устаревших терминов.

ПОД (поисковый образ документа) имеет вид сочетания цифрового обозначения класса категории (это то же, что и классификационный индекс) и словесно (вербально) выраженного дескриптора. Работа в автоматизированной системе предполагает, что код дескриптора выводится в форме, приемлемой для машины. Сам по себе (например, по своей структуре) он ничего не обозначает. Это может быть, например, порядковое число. Давно замечено: это выгодно – последние номера (коды) новее. А в запросах часто (примерно в 70% случаев) задаётся критерий новизны.

Подчеркнём: в отличие от предкоординации в классификационной части системы, тезаурусная часть – живая, постоянно развивающаяся. Появился в источнике новый термин? Он вводится в систему. Конечно, есть свои правила: далеко не все термины сразу становятся дескрипторами. Сначала надо провести терминологический анализ, найти для нового термина (понятия) все парадигматические связи. Тезаурус – не список ключевых слов.

Функция поиска принадлежит пользователю (читателю, абоненту), по логике он же и должен полностью вести поисковые операции. Все алгоритмы и рекомендации выводятся на экран монитора, при этом автоматически контролируется как последовательность, так и правильность выполнения каждой операции. Надо понимать и внимательно выполнять каждое указание. Система, как бы она ни была запрограммирована, делает лишь то, что ей приказали. Конечно, многое зависит от диалога. В нашей стране практически все ИПС немые и глухие. Когда в одной из библиотек Швеции я ошибся при поиске в 18-й раз, раздался сигнал, и электронный каталог что-то произнёс по-шведски (система была двуязычной, я работал в английском интерфейсе). Ко мне сразу подошли читатели, помогли разобраться.

Почему у нас ЭК не общается с пользователем? Почему бы не сказать (написать на экране): вам следует сначала прочитать то и то. Или: ваш запрос не требует поиска в ЭК, пройдите в зал и возьмите книгу на полке. Почему у нас нет ссылок, не объясняется, что и в какой ситуации необходимо сделать. Мне всегда отвечают: всё сказано, просто вы не там посмотрели. Оказывается, я должен сначала потратить пару часов на изучение инструкции, написанной «странным» языком программиста, никогда в жизни не видевшего живого читателя...

Поиск проводится в два этапа. На первом определяется предметная область: по правилам иерархической систематики, как в любой фасетной (аналитико-синтетической) классификационной системе. Пользователь должен только выбрать область, в рамках которой будет вести поиск; сначала «кликнуть» на первой странице, появится вторая, третья... На экране появится предупреждение: вы можете вернуться и выбрать другую область. На втором этапе пользователь работает с тезаурусом, в котором для дескриптора указаны связи «выше – ниже», ассоциативные и прочие.

Тезаурус (мне очень жаль, что многие читатели никогда не видели настоящий тезаурус) – это мощный информационный ресурс. С ним интересно работать, приятно вести поиск. Отраслевой тезаурус (тезаурофасет, по сути дела, – совокупность отраслевых тезаурусов) «разговаривает» с пользователями, например химиками или металлургами, на их профессиональном языке. Потому что этот блок тезауруса подготовлен специально обученными химиками (или металлургами), а не лингвистами или программистами.

стами.

Полагаю, многие читатели хотят узнать, как обстоит дело с социальными (общественными) и гуманитарными науками. Мы считаем, что сфера гуманитарного знания не фасетизируется. За рубежом – иное мнение, накоплен огромный опыт. У тезаурофасета нет ограничений. Более того: те, кто знаком с поиском на деле, хорошо знают, как трудно порой отнести то или иное явление к конкретному классификационному делению. А нужно ли это делать? До определённых границ мы всё делили, разбивали, систематизировали. ...Всегда ли можно понять, почему мы отнесли то или иное явление к классу конкретного деления? Не проще ли искать по известному термину (слову, названию, фамилии, наконец), внешний образ которого всем известен?

Нашли поисковый дескриптор? «Положите в корзину» (все знают, как это делается в интернет-магазине). Можете сразу же просмотреть массив источников. Уже сегодня есть примеры, когда система сразу предлагает полный текст. Статью можно пробежать глазами. Если она нужна вам, система по запросу сделает копию. Книга? Посмотрите её титульные элементы, оглавление, указатели. Хотите прочитать полностью? Система закажет для вас книгу по межбиблиотечному абонементу или сообщит: книга есть в продаже (выдаст реквизиты).

Если вы хотите уточнить параметры поиска, найдите другую область и новые дескрипторы. Результаты можно увидеть на любом этапе. Хотите их на время сохранить? Дайте системе команду. Если вы продолжите работу по конкретизации (детализации) поиска, то каждый следующий шаг будет количественно уменьшать объём выдачи. В конце концов вы получите одно, два, три... названия, точно отвечающих запросу. И если будете последовательны, то обнаружите, что автор многих публикаций – одно лицо, которым можете быть... вы сами. Адекватная запросу выдача – небольшая область на плоскости. Вам стоит просмотреть все источники, которые вы последовательно относили к нежелательным. Там, быть может, вы найдёте самую ценную информацию.

Фантастика? Но я же пишу о системе XXI века. Да, у нас сегодня нет ничего подобного. Мы очень медленно «прорастаем» идеями. Нам нужно не заниматься «совершенствованием» и «модернизацией» того наследия, которое нам досталось от XX века, а посмотреть «за горизонт», выделить средства, подобрать и обучить кадры.

В нашей стране был (возможно, ведётся и до сих пор) оригинальный каталог, называвшийся комплексно-системным, который похож на тезаурофасет. *Комплексно-системный каталог* – это библиотечный каталог, в котором библиографические записи располагаются по крупным комплексам и

их подразделениям в систематическом порядке, а внутри их – по комплексным предметным рубрикам в алфавитном порядке. Идея тезаурофасета была воплощена в карточном виде.

Я об этом писал: «В 1930-е гг. во Всесоюзном институте растениеводства (Ленинград) работал талантливый библиотекарь – Георгий Валентинович Гейнц (1894–1942). Он имел дело с литературой по сельскому хозяйству; вёл систематический каталог, в котором найти нужную книгу порой было трудно, что его, конечно, не устраивало. Всё чаще библиотекарь, прекрасно знающий фонд, не обращался к каталогу, а сразу доставал с полки нужную книгу. Читателям это нравилось, а Георгий Валентинович понимал: когда фонд будет большой, находить книги станет сложно. А если придёт новый библиотекарь? Сколько ему понадобится времени, чтобы изучить фонд? ... Г. В. Гейнц придумал новый вид каталога, назвал его комплексно-системным. Внешне он был вроде бы систематическим: индексы на первых ступенях делений. Современники писали: “внешне всё как везде”. Но не совсем. На определённом этапе вместо детализации – той, которая принята в систематическом каталоге, появляются предметные слова (автор называл их и рубриками, и ключевыми словами). В “таблицах классификации” следом за классификационными делениями шли алфавитные списки понятий» [5. С. 79, 80]. О Г. В. Гейнце и комплексно-системном каталоге писали мало. «Схема» комплексно-систематического каталога издавалась дважды, опыт распространялся по стране.

Более близкая нам реальность – классификационно-дескрипторный язык (КЛАДЕС), который разрабатывался в начале 1960-х гг. Попытки его практического использования предпринимались в «Ленинке» на заре развития «Информкультуры» (руководителем работы был А. В. Соколов). Так что приоритет в разработке этого направления принадлежит нашей стране.

Убеждён: будущее – за системой тезаурофасет. Мы это поймём не завтра и не послезавтра. Пока будем работать с классификационными системами, искать по ключевым словам.

СПИСОК ИСТОЧНИКОВ

1. Сукиасян Э. Р. Систематический каталог : практ. пособие / ГБЛ; сост. Э. Р. Сукиасян. – Москва : Кн. палата, 1990. – 182 с.
Sukiasyan E. R. Sistematicheskiy katalog : prakt. posobie / GBL ; sost. E. R. Sukiasyan. – Moskva : Kn. palata, 1990. – 182 s.

2. **Rajan T.** Dewey's relative index – a precursor to Thesaurofacet / T. N. Rajan // *Annals of Library Science & Documentation*. – 1976. – Vol. 23, № 4. – P. 270–274.

3. **Aitchison J.** The Thesaurofacet: a multipurpose retrieval language / Jean Aithison // *Journal of Documentation*. – 1970. – Vol. 26, № 3. – P. 187–203.

4. **Bhattacharyya G.** Classaurus : its fundamentals, design and use / G. Bhattacharyya // *Universal classification : subject analysis and ordering systems : proceeding of the 4th International Study Conference on Classification Research, 6th Annual Conference of Gesellschaft für Klassifikation, Augsburg, 28 June – 2 July 1982. – Frankfurt : Indeks Verlag, 1982. – Vol. 1. – P. 139–148.*

5. **Сукиасян Э. Р.** Практика каталогизации: повседневные проблемы и пути их решения. Статья 3 / Э. Р. Сукиасян // *Науч. и техн. б-ки*. – 2011. – № 2. – С. 79–83.

Sukiasyan E. R. *Praktika katalogizatsii: povsednevnye problemy i puti ih resheniya. Statya 3 / E. R. Sukiasyan // Nauch. i tehn. b-ki. – 2011. – № 2. – S. 79–83.*

Eduard Sukiasyan, *Cand. Sc. (Pedagogy); Assistant Professor, Sector Head of the Chief Editorial Board of Library Bibliographical Classification Research Center, Russian State;*

sukiasyaner@rsl.ru

3 Vozdvizhenka str., Moscow, 119019 Russia