

Сукиасян Э.Р.

### **Новые технологии в эпоху коммерческих отношений. Заметки с выставки "ONLINE INFORMATION-94"**

В последние годы на страницах зарубежных периодических изданий все чаще можно встретить информацию о проходящих в конце каждого года в Лондоне выставках-конференциях "ONLINE INFORMATION". Это мероприятие понемногу стало смотрам новейших достижений технологии on-line (практика написания термина "on-line" предполагает, что во всех случаях, когда речь идет о прилагательном, например "online information", дефис опускается). Как это часто бывает, по публикациям складывается лишь условный, "справочно-статистический" образ. На многие вопросы ответов просто нет. Меня, например, интересует техника и технология поиска в базах данных. Но познакомиться с поиском, тем более в режиме on-line, заочно трудно. Этому вопросу, как правило, почти не уделяется внимания в описаниях систем; в лучшем случае перечисляются лишь категории поисковых признаков — "точек входа" в базу данных.

Поиск является одним из важнейших параметров любой информационно-поисковой системы. Только в поиске выявляется ее подлинная эффективность, нередко и сама целесообразность ее разработки. Суть дела состоит ведь не только в том, чтобы изменить форму каталога, например заменить каталожные шкафы персональным компьютером. Новая форма должна предоставлять более широкий спектр возможностей при поиске. Однако на деле чаще всего принимаются во внимание технические детали: быстродействие, возможность одновременной работы пользователей в удаленных от базы данных помещениях. Во всем остальном электронный каталог, как ни печально это наблюдать, остается традиционным. Например, в режиме тематического поиска при наличии в записи классификационных индексов поиск по индексу идет "слева направо", так как машина воспринимает индекс как элементарную последовательность знаков. Но точно так же ведется поиск и в карточном систематическом каталоге.

Функциональная эффективность поиска в электронном каталоге должна на порядок превосходить возможности карточного каталога — только в этом случае можно говорить об экономической

эффективности дорогостоящей автоматизации. Пока в наших библиотеках карточные каталоги и их электронные собратья стоят рядом, у нас есть возможность сравнивать их в работе, а не только радоваться светящимся экранам дисплеев. И если наш электронный каталог "нашел" что-то такое, чего мы не обнаружили в карточном, автоматизация имеет смысл.

Работа в режиме on-line (синонимы: в интерактивном каталоге, реальном режиме времени) предполагает подключение терминала потребителя к "живой" базе данных, пополняемой в это же время новой информацией. В точном переводе выражение on-line означает "подключенный". Технология on-line рано или поздно приведет к развитию телекоммуникационных связей: эксплуатация антенны ("тарелки") на крыше здания библиотеки обойдется на порядок дешевле, чем использование телефонных каналов. Раньше других это поняли наши американские коллеги. В Библиотечной школе при университете г. Питсбурга (Пенсильвания) уже несколько лет работает отделение телекоммуникации.

Принятое в зарубежном мире название электронного каталога — OPAC (On-line public access catalogue). Если мы располагаем каталогом на оптическом диске (CD-ROM), магнитной ленте, диске, дискете, то должны называть такой каталог правильно — машиночитаемым. Электронный каталог — вид машиночитаемого каталога, имеющий два существенных отличия: он подключен к "живой" (постоянно пополняемой) базе данных и предоставлен в непосредственное распоряжение читателей ("public access").

Возможность побывать на очередной, 18-й выставке-конференции "ONLINE INFORMATION-94" представилась благодаря поддержке Программы модернизации Российской государственной библиотеки (РГБ) со стороны Комиссии Европейского союза, а непосредственное финансирование осуществлялось Британским советом. Регистрация в качестве официального участника позволила нам получить полный комплект документов и собрать интересный материал.

### **"ONLINE INFORMATION-94": только факты**

Выставка-конференция проходила на протяжении трех дней в "Олимпии 2" — комплексе, специально предназначенном для проведения подобных мероприятий, расположенном в престижном районе Лондона. В шестиэтажном здании масса аудиторий, экспозиционные помещения, все необходимые вспомогательные службы. (Требовалось накормить 10 тыс.

участников и гостей, установить связь, подключить тысячи энергоемких приборов, не считая осветительных и информационных, и при этом обеспечить абсолютную пожарную безопасность.)

В "ONLINE INFORMATION-94" участвовали специалисты из 49 стран, всего 3867 человек из 1293 разного рода организаций. Самой большой была делегация страны-организатора — Великобритании — 2186 человек, затем шли делегации США (337), Франции (229), Германии (227), Нидерландов (143); из остальных стран — меньше чем по 100 человек. Наименьшие делегации приехали из Чехии (8), Польши (7), Новой Зеландии и Словении (по 6), Австралии и Хорватии (по 5), Кувейта и Малайзии (по 3), Канады, Египта, Фарерских островов, Исландии, Кении, Украины (по 2), Аргентины, Бразилии, Гонконга, Индии, Ирана, Литвы, Маврикии, Саудовской Аравии, Сингапура, Словакии, Турции (по одному специалисту). Россию представляли 29 человек из 12 организаций, еще примерно 10 специалистов посетили выставку-конференцию без официальной регистрации.

На 18 заседаниях конференции заслушаны 63 доклада, при этом 5 заседаний были целиком дискуссионными. На одном из заседаний с лекцией "Информация и демократия" выступил депутат английского парламента Тони Билл. Параллельно в нескольких аудиториях велась демонстрация программных продуктов (запланировано более 140).

Рабочим языком конференции был английский, но 5 заседаний проведено на других европейских языках (немецком, испанском, французском, итальянском, нидерландском). Специально для библиотекарей организованы 4 лекции, спонсируемые крупнейшими международными и национальными организациями (одна из лекций прочитана президентом ФИД).

На стендах выставки были представлены 311 организаций, в том числе 15 служб реферирования и обработки данных, 53 производителя и распространителя CD-ROM, 21 организация, создающая и эксплуатирующая базы данных в режиме on-line, 40 издающих организаций, 14 фирм, разрабатывающих программные продукты и т. д. Были широко представлены рекламные, брокерские, торговые организации, но производителей электронной техники здесь не было вовсе. Единственное техническое устройство, показ и рекламу которого разрешили организаторы, предназначалось для хранения и поиска оптических дисков (карусельный магазин с электромеханическим приводом), фирмы Kubik Enterprises, Inc. (США, FAX (604) 737-0791).

Конечно, послушать все доклады, посетить демонстрационные заседания, осмотреть все стенды за три дня практически невозможно. Многие приезжают сюда ежегодно: обновляется до 40% информации. Пришлось с сожалением игнорировать заседания конференции: доклады опубликованы, а самое интересное часто можно услышать во время дискуссий и внеплановых выступлений.

**On-line и CD-ROM:  
сосуществование  
или противостояние?**

В трех томах документов само мероприятие названо традиционно: "ONLINE INFORMATION-94", но в пригласительном билете, напечатанном накануне выставки-конференции, появилось существенное дополнение: "ONLINE/CD-ROM INFORMATION-94". Судя по всему, сами организаторы на последнем этапе подготовки увидели, какое место будут занимать проблемы CD-ROM, и внесли существенную поправку в название мероприятия.

Демонстрировались около тысячи CD-ROM различного характера и содержания, причем почти все можно было либо приобрести, либо заказать, произвести подписку. Судя по каталогу, можно было познакомиться с 2500 базами данных, "обычно работающих в режиме on-line". Однако при попытках поработать с такими базами данных выяснилось, что демонстрация осуществляется с какого-либо носителя (чаще всего с CD-ROM), что резко ограничивало возможности поиска. Практически в режиме on-line работали всего 20 зарубежных (для Великобритании) систем, но и здесь просьбы с нашей стороны "пообщаться" с базой данных не находили понимания до тех пор, пока разговор не касался деталей поиска.

Интересное наблюдение: один и тот же запрос, обращенный к "живой" базе данных и ее отражению на CD-ROM, реализуется часто совершенно по-разному — и по скорости выдачи, и по объему полученной информации. Однозначного вывода мы не получили. Хозяева системы не смогли объяснить нам "случайности" с выпадением из памяти той или иной записи (или, напротив, внезапным появлением "еще одной", которой раньше не было). А ведь часто нужная информация может оказаться в единственной, ранее неизвестной публикации.

Считается, что в режиме on-line ты как бы общаешься с машиной. Но оказалось, что иногда машина хочешь говорить с тобой на своем языке, а в твоём распоряжении нет никаких возможностей изучить, понять ее язык. То же можно сказать и об оптических

дисках. Это ведь просто книга в особой форме. Разве мы не знаем, что некоторые книги бывают приятными во всех отношениях, удобными в пользовании, богатыми по содержанию? Но многие не имеют никаких указателей. Горестно видеть такое "изделие" в форме красивого оптического диска<sup>1</sup>.

Мне кажется, что в информационном обществе границы применения режима on-line и использования информации в виде CD-ROM будут четко обозначены. Сегодня они размыты. Для библиотеки или информационного учреждения CD-ROM – только форма хранения информации. Не сегодня, так завтра мы будем выдавать оптические диски читателям на дом по абонементу. Другое дело – технология on-line. Она открывает перед библиотеками поистине безграничные возможности. Мне помнится, с каким удивлением участники пленарного заседания 49-й сессии ИФЛА в Мюнхене (1983 г.) слушали доклад со странным названием "Колодец или водопроводный кран". Прошло больше десяти лет, но и сегодня мы черпаем информацию в значительной мере каждый из своего "колодца".

### **Поиск: ожидания и реальность**

Как продиагностировать возможности любой системы? В работе, стало быть, в поиске. Надо только проявлять изобретательность в придумывании запросов: они должны быть не слишком простыми, но и не ставить систему в тупик.

На соседних стендах всегда можно посмотреть какой-нибудь указатель и взять пример из него. В адресных базах можно спросить адрес Британской библиотеки (у нее их несколько) или, что сложнее – офис ALA (Американской библиотечной ассоциации) в Вашингтоне (штаб-квартира ALA находится в Чикаго). Хороший запрос для любой базы – "Российская государственная библиотека", так как проверяется сразу много параметров, в том числе важнейший: оперативность обновления информации (у нас и адрес сменился, и наименование, и первое лицо, обычно указываемое в справке). В библиографических базах по заголовку "Российская государственная библиотека" проверяется полнота и качество "Authority file" – важнейшего вспомогательного элемента системы (в Нью-Йорской публичной библиотеке машина напечатала

---

<sup>1</sup> Эффективность поиска информации на оптическом диске во многом определяется имеющейся в распоряжении пользователя СУБД (системой управления базой данных, англ. DBMS—Data-base management system).

несколько страниц прежних и иноязычных заголовков РГБ). В больших ретроспективных каталогах интересно дать запрос на свою фамилию и посмотреть, вытянет ли машина все записи на автора, соавтора, составителя, редактора. Нашим зарубежным коллегам обычно доставляет радость показать на экране произведения Л.Н. Толстого или Ф.М. Достоевского. Тоже интересно, если попросить машину выдать книги на шведском языке, опубликованные в Финляндии.

Надо сказать, что 90% посетителей обычно удовлетворяются справочной аннотацией и проходят мимо, взяв с собой рекламный материал. Серьезного разговора не получается даже в том случае, когда каждого ожидает подарок: фирменный пакет, авторучка или блокнот, а то и оптический диск. Но когда начинается настоящая работа, хозяева раскрываются: одни перестают расхваливать свою систему и начинают работать с вами, в то время как другие нервничают и мешают. Приглашенных девушек-моделей в Лондоне не было — на стендах работали профессионалы — сами разработчики, администраторы, маркетологи, изучавшие рынок сбыта.

К сожалению, поиск информации в подавляющем большинстве баз данных (как в режиме on-line, так и на CD-ROM) примитивен и не идет дальше самых элементарных характеристик: по имени (автора, редактора), по первому слову заглавия, по ключевому слову, часто по любому слову в библиографической записи. Непонятное для меня умиление на лице: в ответ на введенное слово "Classification" машина выдала больше сотни строчек, которые далее надо просматривать сплошь. Спрашиваю, нет ли в записи ДК М. Дьюи или УДК — можно было бы ввести самый общий индекс 020 или 02 и сразу отсеялись бы совсем не нужные мне "классификации". Нельзя: никакого систематического поиска нет.

Слишком часто машина испытывает наше терпение: выдает ответ по одному признаку, потом ищет по второму, а если что-то осталось в выдаче, то может искать и по третьему. Каждый раз сначала появляется на экране одно и то же меню. Спрашиваю: а нельзя ли сразу выдать две-три характеристики (например, автора, первое слово названия и год издания или язык). И слышу в ответ: а зачем? Иногда, правда, интересуются: а как это можно сделать? Объясняю, что для этого на экране должна появиться маска (бланк, формат), и сразу убеждаюсь, что мне просто не верят. Между тем рядом, на стендах Британской библиотеки, где демонстрируется BLAISE-LINE (the British Library Automated Information Service), такой поиск ведется, хотя и не во всех базах.

Очень дорого, объясняют мне. Время читателя, затраченное на сплошной просмотр, ни в долларах, ни в фунтах никто не оценивает.

Интересно, что многие разработчики программ в библиотеках никогда не работали. На мое замечание, не стоило ли предварительно собрать и проанализировать реальные запросы читателей, мой американский коллега спросил, представляю ли я, во что обойдется в США проведение такого социологического исследования.

Нас, естественно, больше всего интересовал тематический поиск. Увы, в лучшем случае "систематический вход" представлен одним рядом делений (не больше 70), которые отсекают часть базы в определенном отраслевом разрезе. Между тем почти все признают, что этот вид поиска наиболее эффективен, если он ведется последовательно. Причем эффективность резко возрастает при сочетании систематического поиска с предметным или поиском по формальным признакам. Что же мешает двигаться по этому пути?

Общая точка зрения: этот вид поиска слишком дорог и технологически сложен. Нужно ведь заложить в программу диалог с читателем, сама машина должна заставить его раскрыть особенности запроса (все понимают, что систематический поиск требует не только движения мысли "выше—ниже", но и заходов в смежные, пересекающиеся области, выявления общих и специфических категорий — без самого читателя здесь не обойтись). Но есть и другая причина: колоссальных затрат требует разработка классификационной системы, а затем ее поддержание up-to-date. Процесс старения классификационных систем объективен и обусловлен развитием науки и общественной практики.

На мой вопрос об использовании широко распространенных универсальных классификационных систем, таких, как ДК М. Дьюи или УДК, следовал правильный ответ о нецелесообразности применения универсальной системы в отраслевой или специальной базе данных. Но существует немало и специальных классификаций. Однако информации о них слишком мало.

**"Предметный вход": язык и терминология** Говорить об интернационализме в мире современных технологий можно лишь с большой долей условности. Систем, работающих на двух языках, слишком мало. Сегодня господствует монолингвистика: наверное, не ошибусь, если скажу, что 90% показанных в Лондоне баз данных на английском языке. По-

нятно, что мы имеем в виду рабочий язык системы. К сожалению, слишком часто он совпадает с языком информационных данных, введенных в базу. С нашей точки зрения, не совсем этичным выглядит лозунг "У нас лучшая в мире база данных!" (а таких на выставке было несколько), если минутой позже выясняется, что опубликованной в России литературы нет вовсе. Как бы между прочим хозяева этой системы замечают, что там, в России, книг на эту тему просто не печатают. Приходится доказывать, вооружившись справочниками ЮНЕСКО (славу Богу, они оказались на одном из стендов), что упущен немалый массив информации.

Мы не имели бы ничего против английского языка, если бы базы данных располагали систематическим поиском. Но если его нет, приходится формулировать запрос на чужом языке, подбирать ключевые слова. Надо сказать, что английский язык богат синонимами. Так, "первое" слово в названии документа (а это очень важный элемент поиска) может иметь множество "планов выражения" при одинаковом "плане содержания" (термины из лингвистики, освоенные нашими теоретиками предметизации). Огромное количество научных публикаций начинается, например, со слов "Analysis", "Examination", "Exploration", "Investigation", "Research". Какой уж тут поиск!

Ясно, что помочь может контролируемая лексика, лучше всего тезаурус. Конечно, его не вложить в CD-ROM вместе с базой данных, но сделать тезаурус элементом базы данных, работающей в режиме on-line, можно элементарно. Подходим к одному стенду, другому — нет тезауруса. И снова те же аргументы: дорого, нужно следить за терминологией, нужно все время обновлять лексику, вносить исправления и дополнения. Наконец, находим базу данных, в которой есть какой-то словарь. Рядом с термином — цифры, показывающие количество документов в поисковом массиве. Отвергнутые термины (аскрипторы) есть? Нет, все действующие. Не без помощи специалиста, свободно владеющего языком, делаю выборку участвующих в поиске терминов: "Appliance", "Arrangement", "Contrivance", "Device", "Equipment", "Facilities", "Gadget", "Mechanism", "Outfit". Все эти английские слова объединяет одно русское, понятное каждому электронщику — "железки".

Несколько тезаурусов, опубликованных в виде книг, нашлись на стендах выставки. Это были хорошие примеры минимизации труда и расходов "на единицу продукции". Издания содержали всего одну часть: лексико-семантический перечень дескрипторов с указанием родовидовых связей. Других видов парадигматических



отношений не показано. Дескрипторы не имеют уточнений. Отсутствуют систематические и пермутационные указатели, нет развернутых предисловий.

**Информация и коммерция. Кто кого побеждает?**

Складывается мнение, что производителей баз данных сегодня в первую очередь интересует возможность продать свой продукт, выиграть время в соревновании с конкурентами. При этом используются всевозможные честные, а нередко и нечестные приемы рекламы и маркетинга, скрытого управления потребителем. И в Лондоне, как и у нас, встретились нам фирмы, которые готовы приложить к своей программной продукции или к производимым CD-ROM то или иное оборудование, сначала совсем бесплатно или со скидкой. И без глубокого анализа видно: купил, значит, попал в долгосрочную кабалу к фирме-производителю. Не пройдет и недели, как выяснится, что какой-нибудь элемент системы "недодали" и за него надо платить дополнительно. Только в эксплуатации выясняются ранее скрытые недостатки системы. Поэтому до поры детали аккуратно скрываются. Например, информации о границах обрабатываемой продукции по видовому признаку нет вовсе, указываются количественные параметры ("свыше 6 тысяч журналов"), без уточнений о хронологических, языковых и других критериях отбора.

Производитель базы данных практически никогда не дает сравнительной информации, не показывает преимуществ своего продукта в сопоставлении с имеющимися на рынке аналогами. Есть, например, более десятка баз данных по лекарственным препаратам (плюс-минус смежные вопросы фармакологии, фармации, лекарственной химии, фармакотерапии и пр.). Как быть потребителю? Сравнить что-то можно только после покупки. И вот рождается новый вид услуг: заключите договор с фирмой, обозначьте свои потребности, мы проведем для вас аналитическое исследование рынка и дадим рекомендации. Стоимость таких рекомендаций весьма прилична и составит от 25 до 40% стоимости базы данных. Но как ни стараются скрыть свои связи производители и консультанты, видно, что все эти "аналитические" службы – дочерние предприятия производителей.

Приходится признавать тот факт, что информация стала измеряться в долларах и фунтах. Но почему развитие рыночных отношений должно приводить к примитивизму в способах и методах обработки и поиска информации? Скажем прямо: в современной

"передовой" практике, продемонстрированной в Лондоне, почти ничего нет от теории 1950—1970 гг., с которой мы познакомились по известной монографии А.И. Михайлова, А.И. Черного и Р.С. Гиляревского, по многочисленным переводам зарубежных авторов, изданным в нашей стране (укажем, в частности, Дж. Солтона, Ф. Ланкастера, Ч. Мидоу, И. Ложе; этот перечень можно было бы продолжить).

Нам кажется, что мы нашли ответ на этот вопрос в томе трудов конференции "ONLINE INFORMATION-94". Два доклада непосредственно связаны с качеством индексирования в целях эффективного поиска (Paul Oven, Structured for success: the continuing role of quality indexing in intelligent information-retrieval system; Peter J. Vicil. Associative and Bodlean retrieval: a comparative analysis). Оба докладчика работали в одной и той же организации Information Access Company (США), представленной на трех стендах выставки. С ними удалось поговорить и оказалось, что их фирма обрабатывает, создает и распространяет (в режиме on-line и CD-ROM) информацию для товаропроизводителей и финансовых организаций. Возможности созданных здесь баз данных (как по объему информации, так и по номенклатуре поисковых категорий) на порядок выше лучших библиотечных и библиографических систем. Поразительна сама технология: мне предложили, например, перечислить все желаемые параметры какого-либо промышленного изделия на естественном языке. Всю дальнейшую работу, связанную с анализом понятий, поиском в тезаурусе, категоризацией и ранжированием поисковых признаков, ведет машина. Будем откровенны: названная нам стоимость реализации одного запроса была фантастически высокой. А как же, пояснили мне, ведь приходится содержать огромный штат научных сотрудников, работающих над перспективными технологиями. И при таких ценах находятся и покупатели, и абоненты, подключенные к базе on-line? Да, но это не библиотеки! Так и получается: если платят, то и "музыку заказывают" соответствующего качества.

Можно ли библиотекам подняться на нужный уровень платежеспособности? Конечно, но только в одном случае: им надо тоже стать товаропроизводителями (а информация — это товар, и притом весьма дорогостоящий). Пока же и у нас, и во всем мире многие библиотеки хладнокровно позволяют коммерческим фирмам собирать, обрабатывать и продавать библиографическую информацию. Успех в Лондоне сопутствовал российской (московской) фирме Andrigal Ltd. ("Андригал"). Наши коллеги обрабатывают

ежегодно свыше 250 научных журналов на русском языке, публикуют в нескольких сериях более 70 тыс. рефератов, готовят и издают прекрасные справочные пособия. Сложился свой круг подписчиков и покупателей. "Андригал" не экономит на рекламе; фирма не первый раз в Лондоне, на выставке имеет свой стенд и ведет активную работу. Грустно: работа эта, по сути дела, библиотечная. Но библиотеки не очень-то торопятся стать "товаропроизводителями". Единственное исключение — ГПНТБ России, у которой был и свой стенд, и добротный товар, и доклад на конференции.

### **Два слова об организаторах "ONLINE/CD—ROM INFORMATION"**

Заканчивая, хотелось бы рассказать о компании "Learned Information Ltd" — организаторе выставки-конференции. Перевести на русский язык ее название сложно.

"Learned Information Ltd" — крупнейший издательский дом, специализирующийся в области информации об информации. Здесь публикуются книги (в том числе на CD—ROM), материалы научных конференций, журналы ("On-line and CD—ROM Review", "The Electronic Library", "Monitor", "Expert Systems", "Electronic Documents", "School Library 2000", "Library Manager"), ежемесячная газета "Information World Review". С этими изданиями можно было познакомиться на стенде выставки.

Кроме "ONLINE INFORMATION" здесь организуются и другие международные форумы. Чтобы представить объем работы, связанной со сбором заявок, формированием программы, проведением самого мероприятия, изданием материалов конференций и каталогов выставок, перечислим запланированные на первую половину 1995 г.: 28 февр. — 2 марта "Electronic Books International", международный форум по электронным книгам, проводимый с 1993 г. (2 пленарных, 7 секционных заседаний, в первый день 7 предварительных выступлений, выставка); 7—9 марта 9-я ежегодная конференция "Компьютеры в библиотеках" (4 параллельных заседания секций ежедневно, выставка); 10—11 апр. "ECLIPS-95" (выставка-конференция с демонстрацией информационно-библиотечных технологий и программных продуктов, 5 пленарных заседаний, 8 семинаров); 16—18 мая — "Internet World International-1995" (выставка-конференция по международным сетям обмена данными).

"Learned Information", как можно было убедиться на ее стенде, организация с глобальными интересами. Пока мы говорили только

о ее европейской части, расположенной в Оксфорде (Великобритания). Другая часть — в г. Медфорде (штат Нью-Джерси, США). Конференций и выставок, организуемых в западном полушарии, ничуть не меньше. Конечно, принять участие во всех конференциях, даже просто посетить каждую выставку для нас пока слишком дорого. Но надо стараться хотя бы получать информацию — доклады конференций и семинаров, каталоги выставок, чтобы представлять себе тенденции развития мировой информационной технологии.