

Подготовка баз данных и информационных изданий по естественным и техническим наукам: современный взгляд и перспективы*

А. И. ЧЕРНЫЙ

Всероссийский институт
научной и технической информации,
Москва, Россия

Несмотря на появление и беспрецедентно быстрое развитие новой информационной технологии, научная литература в современной науке служит — и еще долго будет служить — главным источником и каналом получения и распространения результатов научных исследований, а также главным хранилищем научных знаний как упорядоченной совокупности идей, фактов, гипотез и теорий, главным средством закрепления научного приоритета. Эта роль научной литературы не изменяется несмотря на то, что в настоящее время большая ее часть выходит не только в печатной, но и в электронной форме. Каждый ученый, чтобы сохранить свою квалификацию, должен следить за содержанием научной литературы по своей отрасли, предмету или проблеме. Для помощи ученым в этом деле, которое становится для них все более трудным из-за быстрого роста количества и многообразия научной литературы, были созданы специальные информационные издания — реферативные журналы (1880 г.) и бюллетени сигнальной информации, а в 60-е гг. нынешнего века — базы данных. Значение этих инструментов для развития науки и техники с каждым днем увеличивается. С появлением компьютеров и других новейших средств информационной техники для подготовки информационных изданий и баз данных, а также для облегчения их использования учеными и специалистами начали создаваться автоматизированные информационные системы (АИС). От эффективности этих систем в возрастающей степени зависят темпы развития науки и техники, сроки внедрения их достижений в производство. К таким системам относится АИС интегрального типа "Ассистент", которая была разработана в ВИНТИ в 1972–1985 гг. и уже более 20 лет находится в промышленной эксплуатации. В настоящее время в ВИНТИ ведется разработка усовершенствованной версии АИС, в которой используются все основные принципы построения "Ассистент".

ВВЕДЕНИЕ

За последние 15–20 лет в области информационной технологии произошли большие изменения: появились мощные и доступные по цене компьютеры, в том числе персональные; созданы недорогие запоминающие устройства огромной емкости, из которых особого упоминания заслуживают лазерные компакт-диски; разработаны и производятся сканеры с высокой разрешающей способностью и распознаванием образов — для автоматизации ввода текстов и рисунков в компьютеры, а также лазерные и другие принтеры — для автоматизации вывода данных из компьютеров; выпускаются разнообразные аналоговые и цифровые копировальные аппараты; все шире используются быстродействующие телефаксы для дистанционной передачи изображений; появились и быстро развиваются цифровые сети передачи данных, среди которых центральное место заняла глобальная сеть Интернет.

Эти технологические достижения позволили перейти к подготовке и выпуску научно-технических изданий не только в печатной, но и в электронной форме. Они также существенно расширили для ученых и специалистов возможности по распространению и использованию информационных ре-

сурсов. Под информационными ресурсами здесь понимаются имеющиеся в наличии запасы материалов и средств, которые необходимо использовать для поиска и получения нужной информации. К категории информационных материалов и средств относятся все виды научной литературы, в том числе и выходящие в электронной форме, базы данных, а также надлежащим образом организованные собрания документов, текстов и данных.

Однако новейшие достижения в области информационной технологии пока мало затронули интеллектуальные процессы аналитико-синтетической переработки научных документов и других письменных источников научной информации. В обстановке технологической эйфории было упущено из виду, что информационные ресурсы как таковые появляются именно благодаря выполнению этих интеллектуальных процессов, что именно они определяют качество информационных ресурсов, от которого в решающей степени зависит эффект их использования.

Возможно, что качественная сторона аналитико-синтетической переработки письменных источников научной информации оказалась отодвинутой на второй план из-за появления широких возможностей машинного манипулирования данны-

*Статья подготовлена автором на основе диссертации в виде доклада на соискание ученой степени д-ра техн. наук, ВИНТИ, 1999 г.

ми, их быстрого перебора. За счет такого манипулирования и перебора многие специалисты по информатике пытались и продолжают пытаться компенсировать отсутствие моделей, которые описывают действительно те или иные информационные процессы, а не их имитации.

Этим можно объяснить тот парадоксальный факт, что на практике использование новой информационной технологии пока не привело, как ожидалось, к существенному повышению уровня информационного обеспечения ученых и специалистов, а, напротив, породило новые серьезные проблемы. Например, теперь на информационный запрос поисковые системы обычно выдают огромное количество источников информации, из которых многие являются нерелевантными, т. е. не соответствующими запросу. Это создает ситуации информационной перегрузки, что затрудняет для потребителей принятие правильных решений.

В прежние годы, когда информационная технология была развита значительно меньше, при создании автоматизированных информационных систем (АИС) существенно больше внимания уделялось содержательным аспектам и качеству обработки документов и информации. Поэтому опыт разработки и эксплуатации таких АИС сохраняется немалое значение и в настоящее время. К их числу можно отнести и интегральную информационную систему "Ассистент", которая была разработана в ВИНТИ для подготовки Реферативного журнала и баз данных по естественным и техническим наукам в 1972–1985 гг.

НАУКА И НАУЧНАЯ ЛИТЕРАТУРА

Научная литература является основной формой существования науки как системы знаний о природе, обществе и мышлении; она также служит основным средством передачи научных знаний в пространстве и времени, т. е. выполняет функцию социальной памяти. Под научной литературой здесь понимаются произведения письменности, содержащие результаты научных исследований и разработок. В содержании и структуре научной литературы быстрее всего отражаются все изменения, происходящие в науке.

Подготовка и использование (просмотр, чтение) научной литературы являются составной частью науки как вида деятельности. Поэтому от состояния и использования научной литературы в значительной, если не в решающей, степени зависит эффективность научных исследований, темпы их развития и использование достижений науки в производстве. Через научную литературу ученые осуществляют обмен информацией о полученных ими результатах, без чего наука как вид коллективной деятельности не может существовать.

Общий смысл сказанного можно выразить в следующей максиме: "Без научной литературы нет и не может быть науки".

В течение последних трех веков в мире сложилась устойчивая система научной коммуникации, основанная главным образом на использовании научной литературы. Были также разработаны средства, помогающие ученым и специалистам своевременно следить за содержанием мировой научной литературы и выявлять в ней материалы, представляющие для них непосредственный интерес. Это потребовалось из-за экспоненциального роста количества публикуемых научных документов и углубляющейся дифференциации (специализации) и интеграции в современной науке.

Мощный толчок развитию науки дало появление в 1665 г. научных журналов, которые быстро стали основным средством фиксации и распространения результатов научных исследований, особенно в области точных и естественных наук, а также в медицине. В 1800 г. в мире выходило уже около 700 научных и медицинских журналов, а к началу 1900 г. их

численность возросла до 10 тыс. В современном мире выходит около 60 тыс. научных журналов по всем отраслям науки и техники, медицине и сельскому хозяйству. В них ежегодно публикуется примерно 2,5–3 млн статей.

Уже в первых научных журналах стали публиковаться сообщения о новых книгах и рецензии на них. Это должно было помочь читателям лучше ориентироваться в современной научной литературе. В дальнейшем такая практика привела к появлению специальных периодических изданий, целиком посвященных обзору новых научных публикаций. Такие периодические издания получили название информационных.

ИНФОРМАЦИОННЫЕ ПОТРЕБНОСТИ УЧЕНЫХ

Информационная потребность — одно из центральных понятий информационной науки и практики. Было показано, что конкретная информационная потребность у исследователя появляется только тогда, когда перед ним возникает соответствующая задача, и он начинает осмысливать сущность этой задачи и искать пути ее решения. Был предложен механизм возникновения и уточнения информационной потребности у ученого-исследователя, для чего было использовано учение физиолога А. А. Ухтомского (1875–1942) о доминанте и концепция психолога Д. Н. Узнадзе (1886/7–1950) о психологической установке. По А. А. Ухтомскому доминанта — это временно господствующий рефлекс, который в текущий момент трансформирует и направляет другие рефлексы и работу всего рефлекторного аппарата в целом. Постановка задачи формирует в сознании человека определенную доминанту, которая в дальнейшем — вплоть до полного решения этой задачи — контролирует его поведение при сборе информации. А психологическая установка Д. Н. Узнадзе — это общее психологическое состояние личности, которое служит как предпосылкой к действию, направленному на удовлетворение какой-либо потребности, так и регулятором данного действия (без участия сознания).

Особенностью поведения ученых-исследователей при удовлетворении своих информационных потребностей является то, что они предпочитают сами, без помощи посредников, проводить поиск нужных им публикаций по соответствующим информационным изданиям и заказывать копии тех публикаций, которые представляются им наиболее интересными. Такое поведение можно объяснить тем, что исследователю обычно бывает труднее, чем инженеру или другому специалисту, четко сформулировать свою информационную потребность: он сам не всегда ясно осознает, какая информация ему действительно нужна. Кроме того, иногда он не хочет раскрывать интересующую его тему возможным конкурентам. Именно для ученых-исследователей главным образом и предназначается сигнальная информация.

Таким образом, никакие информационные издания не освобождают ученого-исследователя от необходимости читать научную литературу: они лишь помогают ученому находить и выбирать для прочтения или просмотра те публикации, которые вызывают у него интерес. Это означает, что любое информационное издание или база данных будет для ученого-исследователя малополезной, если у него не будет возможности быстро получать копии заинтересовавших его публикаций.

В отличие от ученых-исследователей инженеры и другие специалисты обычно более четко формулируют свои информационные потребности и для их удовлетворения с большей готовностью прибегают к помощи посредников — информационных и библиотечных работников. Они не всегда нуждаются в копиях первоисточников. Их больше интересуют конкретные факты, а не идеи. Эти различия в информационных

потребностях ученых-исследователей и специалистов необходимо обязательно учитывать при разработке систем их информационного обеспечения.

ИНФОРМАЦИОННЫЕ ИЗДАНИЯ

Как уже отмечалось, информационные издания предназначены для оказания помощи ученым и специалистам в слежении за содержанием мировой научной литературы, количество которой непрерывно возрастает. Основными видами информационных изданий являются реферативные журналы и бюллетени сигнальной информации. К категории информационных изданий относятся также указатели цитирования научной литературы.

В настоящее время все крупные информационные издания выходят не только в печатной, но и в электронной форме. Электронные версии информационных изданий часто называются библиографическими базами данных. Но от изменения физической формы сущность и роль информационных изданий не становится иной. Важнейшие проблемы построения и использования информационных изданий рассмотрены в публикациях.

Реферативные журналы

Первый специальный реферативный журнал появился в Германии (1830 г.). Это был журнал *Pharmaceutisches Central-Blatt*, который выходил под названием *Chemisches Zentralblatt* до 1969 г.

В настоящее время в мире выходит не менее 2 тыс. разнообразных реферативных журналов, которые продолжают служить для ученых и специалистов главным средством поиска нужной им информации в мировой научной литературе. Поэтому когда в нашей стране потребовалось создать мощное информационное обеспечение для ускоренного развития науки и техники, то в 1952 г. было принято решение начать подготовку и издание в АН СССР многоотраслевого Реферативного журнала (РЖ). Первые выпуски этого журнала — по астрономии, математике, механике и химии — вышли в свет уже в 1953 г.

Потребовалось 10 лет, чтобы этот журнал сложился как ведущий многоотраслевой реферативный журнал мира: в 1963 г. РЖ издавался уже в виде 23 основных томов и 123 входивших в него выпусков, а также 31 отдельного выпуска, не входившего в сводные тома. В них за год было отражено почти 690 тыс. различных научных документов, в том числе статей из 11 тыс. отечественных и зарубежных журналов. Назрела необходимость в повышении качества РЖ, особенно в сокращении сроков его подготовки, посредством применения новой информационной технологии, а также в расширении возможностей использования РЖ для повышения уровня информационного обеспечения ученых и специалистов.

С этой целью в 1969 г. было решено создать в ВИНТИ большую автоматизированную информационную систему, основанную на использовании новейших достижений в информационной технологии. Эта система получила название "Ассистент" (Автоматизированная Справочно-информационная СИСТЕМА по Науке и Технике). Такое название было выбрано с намерением подчеркнуть, что создаваемая автоматизированная информационная система должна стать помощником для ученых и специалистов в решении ими исследовательских и проектно-конструкторских задач.

К этому времени в мире были достигнуты большие успехи в развитии вычислительной техники и электронного фотонабора. Это создало новые возможности для подготовки информационных изданий и их более широкого использования учеными и специалистами. Началась разработка соответствующих АИС в США (Medlars, Chemical

Abstracts Service), в Великобритании (INSPEC), в Нидерландах (Excerpta Medica). Достижения в создании этих систем тщательно изучались и использовались при разработке системы "Ассистент".

Основной формой отражения любого документа в реферативном журнале является реферат. В международном стандарте *ISO 214-1976 Documentation — Abstracts for Publications and Documentation* термин "реферат" (abstract) определяется как текст, который точно, но в сокращенном виде представляет содержание какого-либо документа без дополнительной его интерпретации или критических замечаний и без указания того, кем был подготовлен этот текст. Реферат сопровождается библиографическим описанием документа.

Обычно реферат подготавливается специалистом, не являющимся автором исходного документа. Для обозначения подобного текста, подготовленного самим автором документа, ранее использовались термины "синопсис" (synopsis) или "резюме" (résumé).

Реферат не предназначается для замены исходного документа: он должен лишь привлечь внимание ученого или специалиста, сообщить о появлении данного документа в мировой научной литературе и кратко описать его тематику. По этому сообщению ученый или специалист должен определить, интересуется ли его данный документ и нужно ли прочитать его. Для этой же цели служат авторские и редакционные резюме, публикуемые при статьях в современных научных журналах.

Из сказанного следует, что реферативный журнал — как и бюллетень сигнальной информации — не выполняет в полной мере своей информационной функции, если читатели такого журнала не имеют возможности быстро получить доступ к полным текстам заинтересовавших их документов.

Объем реферата зависит от объема исходного документа, но не должен превышать 3% от последнего. Как показывает многолетний опыт разных реферативных служб мира, средний объем реферата без библиографического описания исходного документа составляет 1000–1200 знаков или 170–200 слов.

При оценке возможностей централизованной подготовки многоотраслевого реферативного журнала необходимо было получить ответ на принципиально важный вопрос: допустимо ли использование одного и того же реферата в реферативных журналах по разным отраслям и проблемам? Данный вопрос может быть сформулирован иначе: возможно ли подготовить такой реферат документа, который был бы пригоден для помещения в реферативных журналах по разным отраслям и проблемам? Такая возможность до сих пор оспаривается несмотря на то, что специальные исследования и обширный практический опыт дают на такой вопрос однозначно утвердительный ответ.

Обычно каждый научный документ посвящен рассмотрению одной, центральной темы, которая выражается в его заглавии. Информация, относящаяся именно к этой центральной теме, и должна быть отражена в реферате, если он предназначается для широкого круга пользователей.

Обычно в научном документе может содержаться — и содержится — сопутствующая информация, которая лишь косвенно относится к его центральной теме. Эта информация может представлять большой интерес для пользователей, занятых решением соответствующих задач или проблем. Поэтому для научного документа можно составить так называемый объектно- или проблемно-ориентированный реферат, отражающий не основную, а сопутствующую информацию, которая относится к конкретному объекту или проблеме. Но для этого референт должен знать, что именно интересуется пользователя реферативного журнала, для которого подготавливается объектно- или проблемно-ориентированный реферат.

Более того, на один и тот же научный документ можно составить несколько разных объектно- или проблемно-ориентированных рефератов, отражающих разную сопутствующую информацию.

При централизованной подготовке реферативных журналов по отраслям науки, техники и народного хозяйства референты не знают — и не могут знать — разнообразные и изменяющиеся информационные потребности пользователей. Поэтому они обязаны отражать только центральную тему или предмет документа, который не зависит от конкретных информационных потребностей пользователей, т. е. составлять *документно-ориентированные рефераты*.

Необходимо отметить, что для подготовки объектно- или проблемно-ориентированного реферата требуется затратить значительно больше интеллектуального труда, чем для составления документно-ориентированного реферата. Соответственно, использование объектно- и проблемно-ориентированных реферативных журналов позволяет обеспечивать значительно более эффективное обслуживание ученых и специалистов, занятых решением конкретных задач и проблем. Однако, когда задача или проблема решена, то создававшийся для этого объектно- или проблемно-ориентированный реферативный журнал становится архивным материалом.

Итак, каждый научный документ обычно имеет одну центральную тему или предмет. Поэтому на него в содержательном плане можно составить только один документно-ориентированный реферат. В крупнейших зарубежных реферативных службах мира — как в одно-, так и в многоотраслевых — (INIST во Франции, INSPEC в Великобритании, CAS и BIOSIS в США, JICST в Японии и др.) осуществляется однократное реферирование документов.

Вариант технологии автоматизированного формирования Реферативного журнала и баз данных ВИНТИ при одноразовой подготовке рефератов и многократном их использовании для разных выпусков РЖ и баз данных разработан и проверен на практике в Производственно-издательском комбинате ВИНТИ (г. Люберцы, Московская область) в 1998 г.

К проблеме однократного реферирования тесно примыкает проблема однократного индексирования научных документов. Аргументы, приведенные в обоснование допустимости однократного реферирования научных документов, в полной мере пригодны и для обоснования возможности однократного индексирования.

В связи с этим следует отметить, что — как показали специальные исследования — доля одних и тех же документов, отраженных в реферативных журналах и базах данных по разным отраслям науки и техники (следовательно, и дублирование в подготовке рефератов этих документов) очень велика: в некоторых случаях она достигает 45%. Поэтому даже частичное сокращение такого дублирования может быть важным средством экономии затрачиваемых на это ресурсов.

При подготовке любого реферативного журнала решаются, по крайней мере, три важные задачи.

- Все документы по определенной отрасли или проблеме, которые рассеяны в мировой научной литературе и написаны на разных языках, собираются в один массив. Это особенно касается журнальных статей, рассеяние которых отдельно рассмотрено в следующем разделе.

- Каждый документ представляется в виде короткого текста — реферата, в котором отражается центральная тема или предмет этого документа.

- Все рефераты даются на одном языке, хотя документы написаны на разных языках.

Благодаря этому пользователь (читатель) реферативного журнала получает возможность быстро просматривать рефераты и правильно определять релевантность документов-источников.

В последнее десятилетие все большее число научных изданий выходит не только в печатной, но и в электронной

форме. Например, в каталоге *Directory of Electronic Journals, Newsletters and Academic Discussion Lists* (7th ed., 1999) перечислено уже более 3,4 тыс. изданий, выходящих в электронной форме — в два раза больше, чем было приведено в предшествующем выпуске этого каталога (1996 г.). В каталоге *Ulrich's International Periodicals Directory* (37th ed., 1999) было включено 10 332 периодических издания, которыми можно пользоваться исключительно в режиме интерактивного теледоступа или дополнительно получать их в печатном виде. Это означает, что данные издания имеют электронные версии. 3451 периодическое издание выходит на компакт-дисках. Ныне лишь одно англо-голландское издательство Reed Elsevier выпускает более 1,2 тыс. научных журналов не только в печатной, но и в электронной форме.

Из-за появления все большего числа научных журналов в электронной форме, а также благодаря совершенствованию программных средств для автоматизированного поиска документов непосредственно по их полным текстам создается впечатление, что реферативные журналы отживают свой век. Но такой вывод является неверным. Ибо электронные публикации в мировой научной литературе так же, как и прежде, рассеяны по многочисленным изданиям и так же даются на разных языках. Поэтому автоматизированный поиск релевантных публикаций в электронных изданиях является весьма трудоемким и к тому же характеризуется низкой точностью.

Трудно сомневаться в том, что дальнейшее развитие информационной технологии позволит автоматизировать процессы сбора, индексирования и реферирования печатных и электронных публикаций, включая возможно и автоматический перевод их заглавий и авторских резюме на другие языки. Но это едва ли приведет к скорому отмиранию реферативного журнала как важного средства информационного обеспечения ученых и специалистов. При этом не имеет решающего значения то, в какой форме будет выходить реферативный журнал, — в печатной или электронной. Например, известные французские реферативные журналы *PASCAL* (точные, естественные и технические науки, медицина) и *FRANCIS* (гуманитарные и общественные науки, экономика) с 1995 г. перестали выходить в печатном виде.

Необходимо также отметить, что реферативные журналы помогают решать не только информационные, но и важные общенаучные задачи: они способствуют сохранению единства науки и ее интеграции, помогают в выработке и утверждении единой научной терминологии, облегчают построение научных классификаций и т. п. Почти за 170 лет своего существования (с 1830 г.) реферативные журналы хорошо вписались в сложившуюся систему научной коммуникации и до сих пор не имеют более или менее эквивалентной замены.

Важной составной частью любого информационного издания являются указатели — типовые (алфавитно-предметные, именные, источников) и специальные. Особенно большое значение имеют алфавитно-предметные указатели, которые позволяют читателю быстро находить в реферативном журнале или в бюллетене сигнальную информацию публикации по их предметам, а не только по принадлежности к тому или иному классу (разделу) рубрикатора.

Для автоматического составления алфавитно-предметных указателей к реферативным журналам и бюллетеням сигнальной информации широко используется технология KWIC (KeyWord-in-Context = ключевое слово в контексте) или KWOC (KeyWord-out-of-Context = ключевое слово вне контекста). Эта технология позволяет формировать алфавитно-предметные указатели как по заглавиям документов, так и по их дескрипторным поисковым образам.

При составлении алфавитно-предметных указателей по дескрипторным поисковым образам документов каждая цепочка дескрипторов и ключевых слов, образующих поисковый образ документа, рассматривается как его "квазизаглавие". Компьютер поочередно выводит дескрипторы (но не ключевые слова!) в поисковую колонку указателя, сохраняя их последовательность в поисковом образе (т. е. как бы производится их цепная перестановка, или пермутация). Поэтому такие указатели были названы "квазипермутационными". Специальные исследования показали, что по эффективности такие указатели не уступают традиционным алфавитно-предметным указателям. При проектировании системы "Ассистент" предполагалось, что эти указатели будут переводиться на английский и другие иностранные языки. Это должно было существенно облегчить для ученых и специалистов зарубежных стран использование Реферативного журнала ВИНТИ. С середины 1997 г. двуязычными алфавитно-предметными указателями (на русском и английском языках) снабжается РЖ "Информатика".

Издания сигнальной информации

Сигнальной информацией (СИ) называются сведения о всех статьях, публикуемых в научных журналах по какой-либо отрасли, группе отраслей или по всем отраслям науки, техники, народного хозяйства. Эти сведения должны быстро — за 30–40 дней со времени их опубликования — доводиться до потребителей с тем, чтобы потребители могли заказать и быстро получить копии заинтересовавших их конкретных статей. В некоторых изданиях сигнальной информации отражаются не только журнальные статьи, но и другие виды научной литературы.

Известны следующие основные формы распространения сигнальной информации.

- Бюллетени и указатели, в которых приводятся оглавления определенных журналов (обычно с переводом на базовый язык заглавий статей, опубликованных на иностранных языках). Некоторые из таких бюллетеней снабжаются алфавитно-предметными указателями, составленными автоматически по заглавиям публикаций (по технологии KWOC).

- Бюллетени и указатели, в которых приводятся библиографические описания отражаемых публикаций, располагаемые в соответствии с определенной рубрикой.

- Службы избирательного распространения информации о документах, которые тем или иным путем регулярно оповещают каждого потребителя-абонента о всех новых публикациях по интересующему его предмету.

В настоящее время издания сигнальной информации подготавливаются с помощью компьютерных технологий и распространяются как в виде печатных бюллетеней, так и на электронных носителях (дискеты, компакт-диски). Когда сигнальная информация распространяется в виде печатных бюллетеней, то последние снабжаются авторским и предметным указателями. Электронные бюллетени сигнальной информации могут содержать сведения, не попадающие в их печатные версии (например, авторские резюме статей). Электронными массивами сигнальной информации (базами данных) можно пользоваться и в режиме интерактивного теледоступа.

В библиотечной практике такой вид текущего информационного обслуживания применяется давно (например, в форме издания бюллетеней новых поступлений для читателей библиотеки). Но в больших масштабах он стал осуществляться лишь в конце 50-х годов, когда в США стал выходить бюллетень *Current Contents* по биологическим наукам, который в дальнейшем получил широкое распространение и высокую цену мирового научного сообщества.

В настоящее время бюллетень *Current Contents* выходит — в печатной и электронной форме (на дискетах и компакт-дисках) — по семи разделам науки и техники: биологическим наукам; сельскому хозяйству, биологии и охране окружающей среды; физике, химии и наукам о Земле; клинической медицине; машиностроению и компьютерной технологии; общественным и поведенческим наукам; искусству и гуманитарным наукам. В нем оперативно воспроизводятся оглавления около 6,5 тыс. научных журналов.

В числе первых печатных бюллетеней СИ, которые подготавливались автоматизированным путем, необходимо назвать двухнедельный бюллетень *Chemical Titles*, который с 1961 г. выпускается американским информационным центром *Chemical Abstracts Service* (г. Колумбус, шт. Огайо). Этот бюллетень выходит и в настоящее время. В нем отражается содержание около 800 важнейших журналов по химии (225 тыс. статей в год). Центральной частью бюллетеня *Chemical Titles* является алфавитный указатель ключевых слов, извлеченных компьютером из заглавий статей и представленных по технологии KWOC. До 1998 г. для подготовки бюллетеня использовалась технология KWIC.

Из других крупномасштабных изданий СИ необходимо назвать библиографические бюллетени *Current Papers in Physics*, *Current Papers in Electrical and Electronics Engineering* и *Current Papers on Computers and Control*, которые выпускаются информационной службой INSPEC (Великобритания) с 1966 г. по настоящее время.

С появлением новой информационной технологии широкое распространение получили базы данных, в которых оперативно воспроизводятся оглавления журналов по всем отраслям науки, техники и народного хозяйства. Такие базы данных распространяются на компакт-дисках, но могут также использоваться в режиме интерактивного теледоступа. Крупнейшими из них являются службы *UnCover* (1988–, США, св. 17 тыс. журналов); *Current Citation* (1993–, США, св. 14 тыс. журналов); *ContentsFirst* (1993–, США, св. 11,5 тыс. журналов); *Inside* (1993–, Великобритания, св. 14 тыс. журналов) и *SweetScan* (1993–, Нидерланды, св. 14 тыс. журналов). Все эти службы СИ обеспечивают быстрое изготовление и доставку потребителям копий заинтересовавших их статей. Для этого используются электронная почта, факс и другие средства. При необходимости копии могут быть доставлены через два часа после заказа и даже быстрее.

ВИНТИ начал издавать бюллетени СИ в 1967 г. (5 серий). Для автоматизированной подготовки этих бюллетеней была применена технология KWIC, реализованная на ЭВМ Gamma-3 (Франция).

В концепции системы "Ассистент" предусматривалась подготовка и выпуск триады информационных изданий, состоящей из бюллетеней СИ, Реферативного журнала и обзоров "Итоги науки и техники". Эти информационные издания отличались друг от друга по полноте охвата мировой научной литературы, по степени интеллектуальной обработки и "сжатия" исходных публикаций, а также по срокам подготовки и выпуска. Бюллетени СИ должны быстро (за месяц-полтора) оповещать потребителей о всех опубликованных в мире документах по точным, естественным и техническим наукам; в Реферативном журнале должно раскрываться содержание наиболее значимых, по мнению его научных редакторов, из опубликованных документов, но с задержкой в 3–4 месяца; в "Итогах науки и техники" должны даваться аналитические обзоры наиболее важных научных достижений, но через полтора—два года со времени их опубликования.

Итак, при проектировании системы "Ассистент" подготовка бюллетеней СИ была определена как ее регулярная функция. В этих бюллетенях было решено отражать все без исключений статьи, указанные в оглавлениях журналов, т. е. использовать формальный признак. Такое решение позволяло

компенсировать субъективизм, который неизбежно проявляется научными редакторами при отборе публикаций для отражения в Реферативном журнале. Этот субъективизм выражается в суждениях типа “не представляет интереса”, “уже отражалось ранее” и т. п. Благодаря тотальному охвату должна была фиксироваться и доводиться до сведения пользователей информация о всех публикациях в мировой научной литературе, значительная часть которой при ином подходе была бы практически потеряна навсегда.

Уже в 1977 г. система “Ассистент” обеспечивала выпуск 104 серий бюллетеней СИ (по химии и химической технологии — 41, по физике — 13, биологии — 16, автоматике и радиоэлектронике — 13, металлургии и сварке — 21). В этих бюллетенях приводились библиографические описания научных документов, расположенные в соответствии с Рубрикаторм информационнх изданий ВИНТИ. Все бюллетени СИ были обеспечены средствами для быстрого заказа и получения копий первоисточников, в том числе на микрофишах.

В 1976 г. осуществлялось регулярное копирование на микрофиши 848 важнейших журналов по химии и химической технологии. В дальнейшем число научных журналов, копируемых на микрофиши, возросло до 1,5 тыс. наименований. Эти микрофиши служили для быстрого ознакомления с текстами статей и изготовления их копий непосредственно на месте использования. Сочетание выпуска бюллетеней СИ с микрофишами было весьма популярным видом информационного обеспечения ученых АН СССР. Оно сыграло важную роль в период становления Дальневосточного научного центра АН СССР.

Однако из-за непродуманных решений администрации ВИНТИ бюллетени СИ стали выпускаться в качестве побочного продукта подготовки Реферативного журнала: они стали выходить после РЖ и не охватывали более широкий круг первоисточников, чем РЖ. Такого рода бюллетени не имели никаких преимуществ перед РЖ (кроме цены) и были его библиографическим вариантом. Интерес потребителей к бюллетеням СИ упал, и с 1980 г. их выпуск (кроме 13 серий по физике) был прекращен.

Указатели цитирования научной литературы

Указатели цитирования научной литературы начали выпускаться в США с 1964 г. Их разработка, связанная с именем Ю. Гарфилда, стала практически возможной благодаря появлению ЭВМ.

В указателе цитирования научной литературы фиксируются библиографические ссылки на статьи, опубликованные в наиболее важных журналах, которые сделаны в других статьях и иных публикациях. При этом указываются авторы, заглавия и выходные данные всех цитируемых и цитирующих публикаций, а также их тематика (по ключевым словам из заглавий). Эти сведения позволяют быстро находить все основные публикации по какой-либо теме, появившиеся в мировой научной литературе. Для этого нужно знать имя хотя бы одного автора, занимающегося данной темой.

Но указатель цитирования научной литературы является не только весьма эффективным инструментом информационного поиска; он служит также важным средством оценки относительной научной значимости каждой попавшей в него публикации (по частоте ее цитирования), а также ценным источником другой наукометрической и информетрической информации.

В настоящее время выпускаются три указателя цитирования научной литературы: *Science Citation Index (SCCI)* — по точным, естественным и техническим наукам (выходит с 1964 г.); *Social Science Citation Index (SCI)* — по общественным наукам (выходит с 1973 г.); *Arts & Humanities Citation*

Index (A&HCI) — по искусству и гуманитарным наукам (выходит с 1977 г.). Для подготовки этих указателей используются около 6,5 тыс. тщательно отобранных научных журналов.

ВИНИТИ рассматривал возможность выпуска указателей цитирования научной литературы, которые не имели бы столь явного англо-американского акцента, какой проявляется в названных указателях, и более объективно отражали бы вклады разных стран в развитие мировой науки. Однако в то время (вторая половина 60-х гг.) это намерение не было осуществлено из-за отсутствия в АН СССР необходимых ресурсов. С появлением все большего числа электронных версий научных журналов технология подготовки указателей цитирования научной литературы может быть существенно упрощена. Это позволяет по-новому рассмотреть вопрос о создании указателей цитирования по более широкому кругу научных журналов, чем это делается в США.

ОТБОР НАУЧНОЙ ЛИТЕРАТУРЫ ДЛЯ ОТРАЖЕНИЯ В ИНФОРМАЦИОННЫХ ИЗДАНИЯХ

Главное назначение информационных изданий и особенно реферативных журналов заключается в быстром и адекватном отражении — в виде систематизированных рефератов или библиографических описаний — предметно-тематического содержания публикаций и документов по какой-либо отрасли или проблеме, появляющихся в мировой научной литературе. Важнейшим условием эффективного выполнения информационными изданиями этой функции является возможно более полный охват научной литературы. В решении этой задачи наибольшие трудности возникают с периодическими изданиями (журналами), которые в естественных науках служат главным источником новой научной информации для ученых-исследователей: в них появляется не менее 70% всех научных публикаций.

Трудности с отбором и приобретением научных журналов для их отражения в информационных изданиях вызваны следующими основными причинами:

— число научных журналов велико — их около 60 тыс., и оно непрерывно возрастает;

— подписные цены на научные журналы ежегодно возрастают на 10–15%, и это происходит, по крайней мере, в течение последних 30 лет;

— в журналах статьи по любой отрасли, предмету или проблеме “рассеяны”: 1/3 релевантных статей содержится в журналах, не имеющих никакого отношения к данной отрасли, предмету или проблеме.

Первые две причины достаточно просты и понятны. Однако третья — предметно-тематическое рассеяние статей в журналах — нуждается в специальном обсуждении.

Если рассмотреть предметно-тематическую структуру всей мировой научной литературы с точки зрения какой-то одной науки, отрасли или межотраслевой проблемы, то эта литература разделится на три части: 1) источники, прямо и безусловно относящиеся к данной науке, отрасли или проблеме; 2) источники, лишь частично относящиеся к этой науке, отрасли или проблеме; 3) источники, ни по каким внешним признакам — названиям и другим — не относящиеся к выбранной науке, отрасли или проблеме. Такое разделение всей научной литературы на три части можно считать результатом проявления трехзначности логики (ДА, МОЖЕТ БЫТЬ (ВОЗМОЖНО), НЕТ), которая присуща логике мышления человека.

Однако, если более глубоко, на уровне конкретных публикаций проанализировать содержание литературы, которая внешне совсем не относится к данной науке, отрасли или проблеме, то окажется, что в этой литературе содержится

много релевантных публикаций. Обычно в таких публикациях описываются достижения данной науки с точки зрения интересов другой науки или отрасли. Это является одним из проявлений взаимосвязанности всех наук, которые изучают лишь разные стороны одних и тех же объектов, относящихся к природе, обществу и мышлению.

Как уже отмечалось ранее, в любой отрасли точных и естественных наук основным источником публикаций, подлежащих отражению в информационных изданиях, являются научные журналы. Для отбора журналов в целях их отражения в каком-либо реферативном журнале или базе данных необходимо использовать два основных показателя:

— продуктивность журнала, которая определяется числом ежегодно публикуемых в нем статей по соответствующей отрасли, предмету или проблеме;

— используемость журнала учеными или специалистами, которая определяется частотой цитирования статей из этого журнала, частотой запросов на копии помещенных в нем статей и некоторыми другими показателями.

Ясно, что наилучшие результаты отбора могут быть достигнуты при сочетании обоих показателей — количественного и качественного (с привлечением экспертов на заключительном этапе такого отбора).

Основой для отбора научных журналов по их продуктивности должен служить эмпирический “закон рассеяния” журнальных статей, который был открыт английским химиком и библиографом С. Брэдфордом (1878–1948) в 1934 г. Задача, стоявшая перед С. Брэдфордом, заключалась в том, чтобы в условиях нехватки ресурсов расширить полноту охвата журнальных статей в реферативных журналах за счет выписки для них наиболее важных научных журналов и отказа от журналов, которые можно считать второстепенными.

Проведенный С. Брэдфордом анализ распределения статей в журналах по прикладной геофизике (326) и смазочным материалам (164) позволил ему предложить следующую формулировку “закона рассеяния”:

“Если расположить научные журналы в порядке убывания продуктивности в отношении статей по данному предмету, то их можно разделить на ядро периодических изданий, которые непосредственно посвящены данному предмету, и несколько групп или зон, содержащих такое же число статей, что и ядро, тогда как числа периодических изданий в ядре и в последующих зонах будут соотноситься как $1 : n : n^2 \dots$ ” (Bradford S. C. *Documentation*.— London: Crosby Lockwood & Son, 1953, p. 154).

В принципе число зон в таком распределении может быть любым, причем число релевантных статей в каждой зоне будет равно частному от деления общего числа таких статей на число зон. При использовании многозонной модели рассеяния статей в научных журналах оно может быть представлено как вариант распределения ципфовского типа.

В своих исследованиях С. Брэдфорд основывался на использовании простейшей, трехзонной модели рассеяния. Из полученных им эмпирических данных следовало, что в соотношении $1 : n : n^2 \dots$, где $n = const.$ и $n \square 5$. Если опытным путем определить, что при использовании трехзонной модели рассеяния $1/3$ всех релевантных статей содержится, например, в 25 журналах, то тогда во второй зоне должно насчитываться $25 \times 5 = 75$ журналов, а в третьей зоне — $25 \times 5^2 = 25 \times 25 = 625$ журналов.

Справедливость “закона рассеяния” С. Брэдфорда подтверждена многочисленными эмпирическими исследованиями, которые были проведены за последние 50 лет. И хотя аналитическое описание этого закона неоднократно уточнялось (Vickery B. C., 1948; Brooks V. C., 1977; Chen Y.-S. and Leimkuhler F. F., 1986; Egghe L., 1986 и др.), есть серьезные основания считать, что первоначальная формулировка “закона рассеяния”, предложенная С. Брэдфордом, является не только самой простой, но и наиболее разумной и практичной.

Из рассмотрения трехзонной модели рассеяния С. Брэдфорда с экономической точки зрения следует, что относительные затраты на приобретение каждой релевантной статьи из журналов, входящих во вторую зону, будут примерно в 5 раз выше, чем релевантной статьи из журналов ядра, и в 25 раз выше из журналов третьей зоны. Поэтому при приобретении (комплектовании) журналов для подготовки реферативного журнала по какой-либо одной отрасли или предмету обычно приходится ограничиваться подпиской на журналы, образующие первую (ядерную) и вторую зоны рассеяния. А многочисленные журналы, относящиеся к третьей зоне рассеяния и содержащие $1/3$ всех релевантных статей, не приобретаются и соответственно не отражаются в данном реферативном журнале.

Необходимо отметить, что закон рассеяния статей описывает отображение какой-то одной отрасли, предмета или проблемы на всем множестве выходящих в мире научных журналов. Это означает, что “ядерные” журналы по одной отрасли, предмету или проблеме не являются таковыми для другой, а относятся ко второй или третьей зоне рассеяния.

Из сказанного следует, что для обеспечения исчерпывающе полного охвата и отражения журнальных статей по любой отрасли, предмету или проблеме необходимо приобретать и “просеивать” все выходящие в мире научные журналы по всем отраслям и предметам. Поэтому наиболее рациональным подходом, требующим относительно наименьших затрат на приобретение и “просеивание” (просмотр и тематическую разметку) научных журналов, является централизованная подготовка многоотраслевого реферативного журнала.

Тогда задача отбора и приобретения научных журналов для подготовки серий и выпусков реферативного журнала, отражающих соответственно отрасли и предметы $S_1, S_2, S_3, \dots, S_k$, сводится к составлению списков журналов, относящихся к “ядру” и второй зоне для каждой S_i ($i = 1, 2, 3, \dots, k$), а затем к объединению полученных списков в единый. Такой подход должен обеспечить достаточную полноту охвата журнальных статей по любой S_i (не менее 70%) при наименьших относительных затратах.

Для объективного отбора важнейших научных журналов, относящихся к точным, естественным и техническим наукам (тематический профиль РЖ ВИНТИ), была разработана методика, основанная на использовании электронной версии каталога *Ulrich's International Periodicals Directory* (33rd ed., 1994–1995) на компакт-диске. В этом справочнике каждый журнал описан 42 стандартными признаками, которые включают индекс Десятичной классификации Дьюи, предметную рубрику (их насчитывается 973), указание типа журнала (научный, деловой и т. п.), сообщение о рецензировании помещаемых в журнале статей, об отражении журнала в разных реферативных журналах и другие признаки. По этим признакам — путем их “взвешивания” и суммирования “весов” — можно судить о научной значимости каждого журнала.

По этой методике из 147 тыс. журналов, включенных в каталог *Ulrich's International Periodicals Directory* (33rd ed., 1994–1995), были отобраны иностранные журналы по тематическому профилю РЖ ВИНТИ, которые отмечены в данном каталоге как научные (*academic/scholarly*), — их было 4461. После этого полученный список был дополнен иностранными журналами, которые отражаются в указателях цитирования *Science Citation Index* и *Social Science Citation Index* (информатика, психология и экономика), а также журналами, которые указаны как важнейшие членами и членами-корреспондентами РАН. В результате этого итоговый список расширился до 5576 наименований. Имеются серьезные основания считать, что суммарная продуктивность этих журналов

составляет не менее 650–700 тыс. авторских статей. В настоящее время работа в этом направлении продолжается.

Список важнейших иностранных журналов, отобранных таким образом, следует ранжировать по их научной значимости. Для этого можно использовать индексы их относительной цитируемости (*impact factors*), которые публикуются в указателе *Journal Citation Reports* (выходит на компакт-дисках). *Impact factor* журнала — это отношение числа всех библиографических ссылок на статьи, опубликованные в данном журнале за предшествующие два года, к числу статей, опубликованных в нем за то же время.

ИНФОРМАЦИОННЫЙ ПОИСК

Практический опыт показывает, что при автоматизированном поиске информации наилучшие результаты достигаются тогда, когда поиск проводится по поисковым образцам документов и/или по их рефератам. Это означает, что необходимо уделять особое внимание повышению качества координатного индексирования документов дескрипторами и ключевыми словами, а также качества составления текстов рефератов.

Тексты рефератов должны быть написаны ясно и лаконично; в них желательно применять короткие предложения с прямым порядком слов и не использовать причастных и деепричастных оборотов, а также эллиптических конструкций. В текстах рефератов не следует применять произвольных (окказиональных) сокращений, что нередко делается в целях уменьшения знакового объема рефератов. В тех редких случаях, когда заглавие первоисточника не отражает его центральной темы или предмета, оно должно заменяться смысловым заглавием, которое дается в квадратных скобках.

Такое редактирование не только улучшает читаемость и восприятие рефератов потребителями, но также существенно облегчает их автоматическое индексирование и автоматический перевод на другой язык — один или несколько.

За последние 20 лет развитие новой информационной технологии весьма существенно облегчило выполнение поиска научной информации. Широкодоступными стали технические средства, о которых в начале 70-х гг. можно было только мечтать. Здесь имеются в виду персональные компьютеры с высоким быстродействием и огромной памятью, сканеры с распознаванием образов, компакт-диски, лазерные принтеры, быстродействующие цифровые линии и сети связи, особенно сеть Интернет, телефакс и т. п. Эти технические средства — благодаря своей высокой производительности и сравнительно небольшой стоимости — позволяют более или менее удовлетворительно решать многие информационные задачи путем простого перебора элементов информации, основанного на использовании упрощенных, а иногда и примитивных моделей.

Появление полнотекстовых баз данных, казалось бы, открыло широкие возможности для повышения эффективности автоматизированного поиска информации. Однако, проведение такого поиска в полнотекстовых базах данных по методике, разработанной в расчете на использование дескрипторных языков, дает — как уже отмечалось ранее — очень низкую точность выдачи. Это приводит к тому, что при поиске во все возрастающих по объему базах данных на каждый запрос выдается слишком много формально релевантных документов, которые потребителю становится все труднее отфильтровывать по фактической релевантности и использовать для решения своей задачи. Из-за этого у потребителя возникают «информационные перегрузки», которые становятся все более серьезной проблемой. На неизбежность возникновения этой проблемы нами было указано еще в 1975 г.

Специальные исследования показали, что при проведении автоматизированного информационного поиска по ключевым

словам в полных текстах документов обычно достигается большая полнота поиска, но точность является низкой. Это объясняется тем, что в полном тексте документа ключевое слово, указанное в запросе, может относиться не к центральной теме или предмету документа, а к сопутствующей, которая не интересует потребителя. Поэтому факт наличия такого ключевого слова в тексте документа не обязательно говорит о его релевантности запросу.

Однако при проведении информационного поиска по полным текстам документов имеется возможность повысить точность поиска путем использования — в дополнение к ключевым словам — других характеристик искомых документов (например, библиографических ссылок, имен и т. п.). Такой возможности нет при информационном поиске по поисковым образцам документов. Естественно, что это усложняет процесс информационного поиска, особенно когда он проводится через посредника.

Для повышения точности информационного поиска по ключевым словам в полных текстах документов представляется целесообразным проверить следующий подход. Предположительно релевантные документы, выданные в результате поиска по их полным текстам, подвергаются автоматическому реферированию — например, по методу Х.-П. Луна. Полученные рефераты (точнее — экстракты из документов) используются в качестве поисковых образов этих документов. Повторный поиск по этим рефератам предположительно должен существенно повысить точность выдачи. Современные компьютеры позволяют осуществлять такой двухэтапный поиск.

Из сказанного следует, что развитие новой информационной технологии требует разработки и применения более адекватных моделей информационного поиска и других информационных процессов. А для этого необходимо продолжение и углубление научного изучения этих процессов, особенно их интеллектуальных аспектов.

Особенно большого внимания заслуживает разработка эффективных методов автоматического индексирования и реферирования документов (текстов), а также методика составления информационно-поисковых тезаурусов как одного из способов организации знаний. Данная методика широко использовалась специалистами по информатике в нашей стране и за рубежом. Об этом свидетельствует тот факт, что в конкурсе, который был проведен Всесоюзной ассоциацией «Союзинформ» и Ассоциацией информационных работников России в декабре 1991 г. — марте 1992 г., работа по частоте цитирования статей, опубликованных в сборнике *Научно-техническая информация* за 30 лет (1961–1991 гг.), заняла первое место.

В основу этой методики создания информационно-поисковых тезаурусов была положена идея построения многомерных сетей дескрипторов, связанных друг с другом ассоциативными отношениями, которые играют наиболее важную роль при информационном поиске. Таких отношений было выбрано пять: соподчинение, сходство (функциональное), вид—род (род—вид), причина—следствие (следствие—причина), часть—целое (целое—часть). Относительная значимость разных видов ассоциативных отношений — а всего их насчитывается не менее двух десятков — определялась по опубликованным в литературе результатам экспериментов на появление свободных ассоциаций.

В методиках создания информационно-поисковых тезаурусов, которые были предложены западно-европейскими и американскими специалистами и закреплены в рекомендациях МОС и ЮНЕСКО, принципы отбора ассоциативных отношений для их отображения в информационно-поисковом тезаурусе не определяются; ассоциативные отношения обычно не различаются по видам, и их отбор производится в основном интуитивно. Это делает процедуру составления информаци-

онно-поискового тезауруса в значительной степени “кулинарной”.

В последние два десятилетия исследования фундаментальных проблем индексирования, реферирования, информационного поиска и других информационных процессов сократились до минимума. И это произошло не только в России, но и в других развитых странах. В настоящее время все еще используются результаты исследований, которые проводились в 60-е и в первой половине 70-х гг. Ясно, что одной из причин недооценки важности научных исследований информационных процессов является эйфория, которая порождена огромными достижениями в области информационной технологии за последние два десятилетия.

ИНТЕГРАЛЬНЫЕ ИНФОРМАЦИОННЫЕ СИСТЕМЫ

Анализ разных видов информационных изданий и услуг показывает, что все они осуществляются на основе использования — в разных сочетаниях — преимущественно одних и тех же структурных элементов аналитических описаний документов. Это имена авторов, заглавия документов, выходные данные этих документов, их классификационные индексы, дескрипторы и ключевые слова, рефераты или аннотации и некоторые другие. Поэтому если однократно ввести в компьютер АИС все неповторяющиеся элементы информации, представляющие (описывающие) каждый научный документ и необходимые для выполнения функций, которые заданы этой информационной системе, и в течение установленного времени хранить эти элементы внутри системы, то такая АИС будет способна эффективно выполнять функции нескольких простых, специализированных информационных систем. При этом удельная стоимость выполнения каждой функции в такой многофункциональной системе будет заметно ниже, чем в специализированной, однофункциональной системе. Последнее обусловлено тем, что исключаются многократное повторение дорогостоящего ввода одних и тех же элементов информации в разные специализированные системы, а также параллельное хранение их в этих системах. Такая АИС дает особенно большой положительный эффект, когда она является многоотраслевой.

Автоматизированная информационная система, построенная на указанном принципе — как детский конструктор, в функциональном плане представляет собой как бы объединение из нескольких специализированных систем. Поэтому такие комплексные системы были названы интегральными информационными системами. Их создание стало возможным благодаря появлению компьютеров, имеющих достаточно большое быстродействие и большую внутреннюю и внешнюю память, а также быстродействующих фотонаборных машин, способных работать в качестве выводных устройств компьютеров.

Итак, под интегральной информационной системой (ИИС) здесь понимается совокупность методов и средств, которая позволяет на основе однократной исчерпывающей аналитико-синтетической переработки научных документов высококвалифицированными специалистами, однократного ввода результатов такой переработки в ЭВМ многократно использовать эти результаты для решения разнообразных информационных задач — подготовки баз данных, формирования информационных изданий разных типов и тематики, избирательного распространения информации о документах, ретроспективного поиска документов и информации. Слово “однократный” здесь не следует понимать в том смысле, что на входе ИИС каждый документ полностью перерабатывается одним специалистом и за один прием. Это слово означает, что при аналитико-синтетической переработке документов исключается повторное выполнение одних и тех же

операций: в ЭВМ разрешается вводить только такие сведения (элементы информации), которые еще не содержатся в ней (т. е. не были введены в ЭВМ ранее) и не могут быть автоматически извлечены непосредственно из уже имеющихся. Аналитико-синтетическая переработка документов включает следующие операции: полное библиографическое описание каждого документа, его классификация (присвоение индексов применяемой рубрикации), индексирование ключевыми словами, шифрами УДК, а также шифрами МКИ и НКИ (для описаний изобретений), составление реферата или аннотации документа.

В ИИС должны быть введены, по крайней мере, следующие сведения о каждом научном документе:

- сведения о типе документа (журнальная статья, книга, описание изобретения, диссертация, карта и т. п.);
- заглавие документа на языке оригинала и на русском языке (для документов на иностранных языках);
- имя или имена авторов;
- сведения о месте работы авторов (если такие сведения приведены в документе);
- выходные данные документа;
- указание о языке, на котором написан документ;
- рубрикационные шифры;
- шифры УДК;
- шифры МКИ и НКИ (для описаний изобретений);
- шифры специальных классификаций — биологических, химических или других (при необходимости);
- ключевые слова и дескрипторы, выражающие центральную тему или предмет документа;
- реферат или аннотация.

Перечисленных сведений достаточно для автоматической подготовки почти любых документальных баз данных и информационных изданий, а также для удовлетворения большинства информационных запросов ученых и специалистов. Такой подход стал возможным на практике только благодаря появлению компьютеров и других новейших средств информационной техники. Он позволил создавать АИС, которые обеспечивают максимальную экономию интеллектуального труда, затрачиваемого на аналитико-синтетическую переработку научных документов, а также экономию ручного труда, затрачиваемого на ввод результатов этой переработки в ЭВМ.

Практическое создание ИИС в ВИНТИ было начато в 1969 г. Эта ИИС не имела аналогов в мировой практике и отличалась следующими особенностями:

- ИИС проектировалась как многоотраслевая система, охватывающая все точные, естественные и технические науки и научные дисциплины;
- ИИС создавалась как полная система, обеспечивающая выполнение всех основных функций по информационному обеспечению ученых и специалистов;
- ИИС должна была стать самой большой в мире автоматизированной информационной системой, перерабатывающей 2,5–3 млн документов в год, т. е. в 5–8 раз больше, чем крупнейшие зарубежные системы такого типа.

ИИС создавалась для выполнения следующих основных функций:

- подготовка библиографических бюллетеней сигнальной информации (как в печатной, так и в электронной форме) по разным отраслям, предметам и проблемам, которые бы отражали все статьи, упомянутые в оглавлениях соответствующих научных журналов, а также все другие издания и документы по тематическому профилю РЖ ВИНТИ;
- подготовка и выпуск печатных РЖ с необходимыми указателями по разным отраслям, предметам и проблемам;
- формирование баз данных по разным отраслям, предметам и проблемам и их распространение на магнитных лентах для использования в целях информационного обслуживания пользователей;

— избирательное распространение информации о публикуемых в мире документах по тематическому профилю ВИНТИ;

— ретроспективный поиск документов и информации по разовым запросам в любой части или во всем накопленном массиве данных;

— выдача по запросам пользователей полных копий научных документов, введенных в систему и отраженных в ее информационных изданиях и базах данных.

Для выполнения этих функций в ИИС было предусмотрено создание четырех подсистем:

— подсистемы аналитико-синтетической переработки документов;

— подсистемы полиграфического размножения и рассылки информационных изданий;

— подсистемы справочно-информационного обслуживания;

— подсистемы выдачи копий документов.

При проектировании ИИС было также предусмотрено ее взаимодействие с другими органами информации, входившими в то время в Государственную систему научно-технической информации СССР, в которой ВИНТИ выполнял функции головной организации.

Краеугольным камнем в концепции многоотраслевой ИИС было предположение, что каждый научный документ может и должен быть представлен только одним рефератом, который был бы пригоден для реферативного журнала по любой отрасли, предмету или проблеме.

ЗАКЛЮЧЕНИЕ

За последние годы научная литература под воздействием достижений информационной технологии, особенно создания сети Интернет, а также появления возрастающего числа электронных изданий сильно изменилась на физическом уровне. Но она мало изменилась в своей сущности и продолжает служить главным средством научной коммуникации между учеными и специалистами, каким она была в течение всей истории науки.

Однако появление сети Интернет и Web-технологии создает в сфере научной коммуникации — если уже не создало — еще более сложную ситуацию, чем та, которая сложилась в середине XVII века и вызвала появление научных журналов (1665 г.), а затем и специальных реферативных журналов (1830 г.). Как было показано Д. Прайсом, рост численности реферативных журналов происходил примерно такими же темпами, что и численности научных журналов.

В какой бы физической форме ни появлялись научные публикации, для ученого-исследователя неизменной остается необходимость следить за научной литературой, знакомиться с содержанием конкретных публикаций и читать их. Только теперь это становится все более трудным делом, так как из-за полной открытости сети Интернет для опубликования любых сообщений предметно-тематическое рассеяние публикаций в мировой научной литературе быстро увеличивается. Поэтому не уменьшается, а даже возрастает значение информационных изданий, которые собирают и представляют в одном

месте все доступные публикации по той или иной отрасли, предмету или проблеме и дают их в виде рефератов или библиографических описаний на одном языке.

Здесь мало помогает то новое обстоятельство, что все большее число первичных изданий выходит также в виде полнотекстовых электронных версий, которые могут быть подвезгнуты компьютерной обработке. Вообще новая информационная технология, если говорить о первичных научных изданиях, помогает в основном при редакционно-издательской подготовке публикаций, при их хранении, беглом просмотре и быстрой доставке публикаций потребителям. Но эта технология пока мало помогает человеку в интеллектуальной переработке публикаций — в их реферировании, индексировании и переводе на тот или иной язык, а также в проведении более точного поиска информации.

Из сказанного можно сделать следующие основные выводы-рекомендации.

1. Информационные издания и особенно реферативные журналы еще долго — по крайней мере, не менее одного-двух десятилетий — будут служить для ученых главным путеводителем по мировой научной литературе. Поэтому необходимо сохранять и развивать специальные информационные службы, занимающиеся подготовкой реферативных журналов и других информационных изданий, а также соответствующих баз данных. Для этого очень важно изучать и использовать огромный практический опыт, накопленный крупнейшими научно-информационными центрами мира, в том числе ВИНТИ.

2. Для эффективной автоматизации подготовки информационных изданий и баз данных, а также использования их в целях информационного обеспечения ученых и специалистов необходимо осуществлять компьютерное моделирование важнейших информационных процессов аналитико-синтетической переработки научных документов, особенно процессов реферирования, индексирования и перевода с одного языка на другой.

3. Все более важное значение приобретает разработка эффективной компьютерной технологии информационного поиска по полным текстам публикаций научных документов. До сих пор это делается по технологии, ориентированной на использование заглавий документов, их поисковых образов и/или рефератов, и поэтому дающей низкую точность.

4. Нужно искать пути международной кооперации по распределенной аналитико-синтетической переработке научной литературы. Каждая участвующая в такой кооперации страна должна собирать всю выходящую в ней научную литературу, перерабатывать ее по согласованным правилам и обмениваться результатами такой переработки с другими участвующими странами. Это позволило бы значительно повысить полноту охвата мировой научной литературы, значительно сократить сроки ее аналитико-синтетической переработки и уменьшить расходы на подготовку соответствующих международных и национальных баз данных и реферативных журналов.

5. Необходимо изучать и разрабатывать методы и технологию индивидуального информационного обслуживания ученых-исследователей РАН на их рабочих местах.