

Малыха В. В.

АВТОМАТИЗИРОВАННЫЕ МЕТОДЫ ПОИСКА В НТБ

В опорной НТБ Волгоградского государственного научно-исследовательского и проектного института нефтяной промышленности около полутора лет ведется автоматизированный поиск НТИ.

В марте 1990 г. библиотека приступила к автоматизированным методам поиска научнотехнической литературы и документации с помощью ЭВМ по автоматизированной документальной информационно-поисковой системе (АДИПС) «Нефть-4». В 1989 г. был заключен договор с отраслевым информационным институтом на 20 тыс. документов в год. Он включает:

- методику индексирования запросов;
- готовые поисковые предписания;
- тезаурус (алфавитный указатель дескрипторных статей и синонимов);
- частотный словарь из пяти фасетов (вещества, идентификаторы, объекты, процессы, характеристики).

Информационно-поисковый язык АДИПС «Нефть-4» — искусственный язык на сложной лингвистической основе, предназначенный для автоматизированного поиска информации с помощью ЭВМ.

Первое время применялись готовые поисковые предписания, составленные по ключевым словам и весьма неоднозначно. Самостоятельное составление поисковых предписаний — целая наука, предполагающая использование указателей роли, контекста, расширения, отрицания, поиска по тексту документа.

В течение полутора лет проведено около 40 сеансов автоматизированного поиска по базе данных 240—277 тыс. документов за 1988—1990 гг. Среди них книги, статьи из тематических сборников и журналов, депонированные рукописи, информационные материалы, патентные документы, неопубликованные материалы. Нормативно-техническая документация в АДИПС «Нефть-4» вводится в незначительном объеме.

К достоинствам АДИПС «Нефть-4» можно отнести быстрый поиск и воспроизведение документов; наличие реферата на каждом документе; извлечение сведений по теме из кажущегося неидентичным документа; комфортную информационную среду. Однако новые документы вводятся в систему с большим опозданием. РЖ ВИНТИ «Горное дело», «Геология» публикуют вторичные документы ранее АДИПС «Нефть-4». К сожалению, теряется много релевантных документов; система не выдает ни готовой обзорной информации, ни фактографических данных. Кроме того, в неопубликованных документах недостает сведений: в отчетах среди данных, относящихся к заглавию, часто не указывается вид отчета (промежуточный, заключительный); в сведениях об ответственности иногда не приводится название организации, дата выпуска отчета, количественная характеристика (однако это происходит не по вине отраслевого института, так как он сам закупает базы данных). В депонированных рукописях отсутствуют сведения об организации-депоненте, дате депонирования, даже если документы хранятся в отраслевом институте или ВНИИ экономики, организации

Вид документа	Количество выявленных документов	
	АДИПС «Нефть-4»	ГСК, систематический каталог
Книги, монографии	—	3
Статьи из журналов, тематических сборников;	3 (1 дублетный)	28 (1 дублетный)
из них иностранных	—	2
Депонированные рукописи	—	2
Авторские свидетельства	2	2
Отчеты	6	—
Информационные листки	5	—
Всего	16	37

производства и технико-экономической информации в газовой промышленности.

Таким образом, комфортная среда сводится к нулю и начинается поиск традиционными методами, но самым отрицательным является то, что утрачен поиск по тематике. Приведем пример.

В 1991 г. проведен автоматизированный поиск по АД ИПС «Нефть-4» на ЭВМ и ручной поиск по ГСК, систематическому каталогу за 1988—1990 гг. по теме «Технология бурения и крепления скважин в соленосных отложениях» (см. табл.).

Из таблицы видно, что отчеты и информационные листки в ГСК не вводятся. Основные материалы, раскрывающие тему, содержатся в статьях. Из-за несовершенства АД ИПС «Нефть-4» ЭВМ выдала только 3 документа, тогда как в ГСК их в 9 раз больше. Поэтому можно сделать вывод о том, что необходимо дополнительный поисковый признак по УДК, нужны разные методы поиска, взаимно контролирующие и помогающие раскрытию научной темы.

Еще один пример неувязки с АД ИПС «Нефть-4».

В 1991 г. состоялся сеанс автоматизированного поиска на ЭВМ по теме «Бурение, разработка и добыча с помощью горизонтальных скважин». Вот один из релевантных документов из иностранного журнала.

Giannesini J.-F. Production technology for horizontal wells takes new direction (Технология эксплуатации скважин с горизонтальным стволом) // *World Oil*.— 1989.— Vol. 208.— № 5.— P. 46—48.

В том же 1989 г. эта статья как одна из актуальных переведена на русский язык и опубликована в журнале «Нефть, газ и нефтехимия за рубежом» и в «Экспресс-информации» (серия «Бурение, разработка и эксплуатация газовых и газоконденсатных месторождений в зарубежных странах»). Дважды эта статья фигурировала на русском языке в переводном журнале и информационном издании, но таких сведений в автоматизированной базе данных не оказалось.

Если одновременно проводится поиск по нескольким темам, то у оператора ЭВМ имеется рулон на несколько сот документов. Для того, чтобы понять, раскрыта ли тема, нужно все разрезать, пользуясь указателем

ограничения (из шести цифр), разложить по номерам запроса и обращения и отсеять «шум». Встречаются и чисто технические сложности: документы разного формата по сравнению с классической карточкой 7,5×12,5 см; печатающее устройство иногда выдает «слепое» воспроизведение документов.

Анализ работы на ЭВМ показал, что выдается много «шумовых» документов, при этом теряются релевантные. Причины в несовершенстве поискового образа запроса, информационно-поискового языка АД ИПС «Нефть-4», поискового образа документа, в отсутствии информационно-поискового языка УДК как связующего звена в реализации поиска.

Разработчики АД ИПС «Нефть-4» явно проигнорировали информационно-поисковый язык УДК как основной поисковый признак. Утрачен принцип поиска по тематике и заменен малопонятным лингвистическим обеспечением. На каждом документе АД ИПС «Нефть-4» стоит классификационный индекс УДК, которым нельзя пользоваться как дополнительным поисковым признаком. Даже при беглом знакомстве с последним (не превышающим двадцати символов) напрашивается вывод, что нет поиска по конкретной тематической направленности. Можно проводить поиск по автору, году издания, типу документа и т. п., но самым главным остается тематическая направленность.

Если бы информационно-поисковый язык нефтяной промышленности был достаточно разработан, то на местах шла бы стабильная стратегия поиска с минимальным «шумом». Для применяемой в повседневной работе «Рабочей схеме для систематизации по УДК литературы по нефтяной, газовой и смежным отраслям промышленности» необходимо продумать автоматическое присвоение индексов. В мировой практике подобный опыт есть [1]. Нужна единая методика индексирования по классификационным индексам УДК, усечение индексов УДК справа и слева, применение операторов и команд.

Вместо этого накопленный опыт и знания ликвидируются и создается с нуля новая сложная для понимания лингвистическая теория. При рыночной экономике будут покупаться и потребляться только те услуги, что соответствуют требованиям мировой практики.

СПИСОК ЛИТЕРАТУРЫ

1. Кильдюшевская З. В. Универсальная десятичная классификация в автоматизированных информационно-поисковых системах // НТИ. Сер. 1.— 1990.— № 2.— С. 16—17.