

ИРИНА НИКОЛАЕВНА ВОЛКОВА

## БАНКИ ТЕРМИНОВ И ИХ ПРАКТИЧЕСКОЕ ИСПОЛЬЗОВАНИЕ

Научно-техническая революция привела к возросшему значению проблем научно-технической терминологии для народного хозяйства. Создание и внедрение автоматизированных управляющих и информационных систем, переход к программно-целевым методам планирования, развитие международной кооперации и интеграции при решении научно-технических и экономических задач предъявляют все более жесткие требования к языку науки и техники с точки зрения однозначности терминологии, используемой в документации различных видов.

Следствием этого явилась значительная активизация работ по упорядочению и стандартизации научно-технической терминологии на национальном и международном уровнях, что привело к созданию большого количества стандартов и рекомендаций по терминологии (международных, региональных, национальных, фирменных и т. д.).

Советом Экономической Взаимопомощи (СЭВ) принято 160 стандартов на термины и определения, Международной организацией по стандартизации (ИСО) опубликовано 180 международных терминологических стандартов, Международной электротехнической комиссией выпущено около 60 глав Международного электротехнического словаря (МЭС) и специальных терминологических публикаций (имеют статус стандартов МЭК).

Национальными организациями по стандартизации 40 стран принято свыше 15 тыс. стандартов на термины и определения.

Большие успехи в области упорядочения и стандартизации терминологии достигнуты в Советском Союзе, где действуют свыше 700 государственных и отраслевых стандартов на термины и определения и издано более 100 сборников рекомендуемых терминов КНТТ АН СССР.

Массивы стандартизованных и рекомендованных терминов исчисляются сотнями тысяч. Необходимым условием для использования этих массивов в практической и научной работе, быстрого поиска нужных терминов является представление стандартизованной и рекомендованной терминологии в виде определенной организационной структуры. Одним из распространенных способов структурной организации больших терминологических массивов является создание банков терминологических

данных, в которых хранение, обработка и выдача терминологической информации осуществляется с использованием ЭВМ.

Под **банками данных** обычно понимается совокупность наборов данных большого объема и сложной структуры, а также комплекса технических средств и программного обеспечения для накопления, обработки и выдачи необходимой информации.

Совокупность наборов данных принято называть **базой данных**. В банках терминов базой данных является совокупность содержащихся в них терминов и сопровождающих их элементов данных. Эти совокупности называют **терминологическими статьями**, иногда в этом значении используют также термины — **терминологическая запись** и **словарная статья**.

По сведениям, опубликованным Международным информационным центром по терминологии (Инфотермом), в мире функционируют или находятся в процессе разработки несколько десятков терминологических банков<sup>1</sup>. Такие банки имеются в Великобритании, Венесуэле, ГДР, Испании, Канаде, Люксембурге, СССР, США, Франции, ФРГ, Швеции и ряде других стран.

Следует отметить, что службы, при которых организованы банки терминов, как и лексические массивы, явившиеся основой для их формирования, достаточно разнообразны.

Ряд банков терминов был создан на базе тех терминологических массивов, которые сформировались в период активных исследований по машинному переводу: система TEAM лингвистической службы фирмы «Сименс» (ФРГ), ЕВРОДИКАУТОМ (Люксембург), системы ТЕРМИУМ и ТЕРМИНОК (Канада) и ряд других. Система ТЕРМДОК шведского центра технической терминологии создана на базе лексикографической службы этого центра, на протяжении многих лет активно занимающегося изданием терминологических словарей. Банки терминов НОРМА-ТЕРМ (Франция), ТЕРМ (ФРГ) созданы при службах стандартизации на основе массивов стандартизованных терминов.

Массовое создание и внедрение информационно-поисковых систем потребовали создания для них лингвистического обеспечения, что сопровождалось организацией массивов терминов, также сыгравших свою роль в процессе формирования банков терминов.

Для характеристики банков терминов используются такие показатели, как состав и объем базы данных (охват тематических областей, набор элементов данных терминологической статьи, количество языков, на которых представлена терминологическая информация), нормативность терминологической информации, категории пользователей.

Большое значение для функционирования банка терминов имеет состав его базы данных. Чем большее количество элементов данных со-

<sup>1</sup> World Guid to Terminological Activities (Organizations, Terminology Banks, Committees).— Infoterm Series A. München: Verlag Dokumentation, 1974.— 312 p.

держит терминологическая статья, тем шире возможности использования этого банка пользователями.

Например, в банке терминов TEAM каждая терминологическая статья может содержать до 99 элементов данных (разработчики этого банка называют их **категориями**). В их числе основной термин, его синонимы, одно или несколько определений или контекст, библиографические сведения (источник, автор, дата публикации и др.), грамматические параметры термина, характеристика сферы его применения, а также эквивалентные термины на английском, французском, испанском, русском, итальянском, португальском и голландском языках (эквивалентные термины могут быть представлены не на всех языках).

Из перечисленных выше элементов данных для обработки немецкой терминологии практически используется около 25 элементов<sup>2</sup>.

Терминологическая статья банка ТЕРМДОК предусматривает большое количество как лингвистических, так и экстралингвистических элементов данных.

Терминологическая статья может содержать до 100 элементов данных, относящихся к лингвистической информации для 26 разных языков. В настоящее время лингвистическая информация охватывает 10 языков: шведский, английский, французский, испанский, немецкий, датский, норвежский, финский, русский, японский. Наиболее часто используются 10 лингвистических элементов данных: 1) основной термин; 2) синонимы и квазисинонимы; 3) антонимы; 4) произношение, грамматические характеристики, сокращения, определения, пометы, касающиеся использования, примечания, перекрестные ссылки<sup>3</sup>.

Из экстралингвистических элементов данных наиболее часто используются около 10: обозначение документа, номер термина, классификационные коды, дата принятия (издания), указание на эксперта, ссылки на иллюстрации.

По степени нормативности терминологической базы данных существующие банки терминов можно разделить на две группы: банки, основу которых составляет стандартизованная терминология и которые являются как бы нормативными словарями, и банки, которые формируются и пополняются не только из нормативных источников. Напр., банк терминологических данных Французской ассоциации по стандартизации (АФНОР) — НОРМАТЕРМ включает термины, установленные французскими терминологическими стандартами; термины, содержащиеся в тер-

<sup>2</sup> Brinkmann K. H. Use of the TEAM Terminology Data Bank for the Terminology Work of the Deutsches Institut für Normung.— Infoterms Series 6. München—New York—London—Paris, 1981, p. 444.

<sup>3</sup> The TERMDOK system.— TNC, 1978.— 45 p.

минологических приложениях к промышленным стандартам, а также в международных стандартах ИСО на термины и определения<sup>4</sup>.

Банк терминологических данных ТЕРМ, который создается в рамках взаимосвязанных банков данных Института по стандартизации ФРГ, также носит нормативный характер, т. е. включает только стандартизованную терминологию.

Такие банки терминов, как TEAM фирмы «Сименс», Федерального управления по вопросам языкоznания (ФРГ), ТЕРМИУМ и ТЕРМИНОК (Канада), ТЕРМДОК (Швеция) и ряд других, не являются полностью нормативными и содержат не только стандартизованную терминологию. Поэтому информация, выдаваемая ими, является достоверной (с точки зрения ее документальности), но не всегда устанавливает официальную предпочтительность употребления того или иного термина.

В процессе формирования баз данных этих банков специалисты «оценивают» каждый термин и присваивают ему так наз. «весовой индекс». Напр., в банке терминологических данных Монреальского университета весовой индекс 0 означает, что данный термин неприемлем для употребления, 1 — возможность использования термина сомнительна, 2 — термин может быть использован, 3 — термин стандартизован.

По составу языков выделяются банки: **одноязычные**, т. е. такие, в которых все элементы данных терминологической статьи представлены только на одном языке (напр., банк неологизмов во Франции включает терминологическую информацию только на французском языке); **частично дву- и многоязычные**, т. е. такие, в которых приводятся эквивалентные термины (а иногда и определения, и ряд других элементов терминологических данных) на нескольких языках. К числу таких банков относится TEAM, база данных которого включает кроме термина и определяющих его элементов данных на немецком языке также термины (а иногда и определения) на английском, французском, испанском, итальянском, русском и голландском языках. Банк ЕВРОДИКАУТОМ включает терминологию на английском, немецком, датском, французском, итальянском и голландском языках; **полностью двуязычны** банки **терминов**. Такими являются банки терминов Канады, которые в силу демографических особенностей являются полностью англо-французскими, т. е. каждая терминологическая статья в их базе данных состоит из двух идентичных блоков: на английском и французском языках.

Цель многих банков терминов — оказание помощи, в первую очередь, переводчикам. К числу таких банков можно отнести TEAM Лингвистической службы фирмы «Сименс» (ФРГ), ЕВРОДИКАУТОМ Европейского экономического сообщества (Люксембург), ТЕРМИУМ Монреальского университета (в настоящее время официальный банк терминов

<sup>4</sup> Laurent J. La banque automatisée de données terminologiques de l'AFNOR NORMATERM.— Courier de la Normalisation, 1979, N 245, p. 46—51.

правительства Канады), ТЕРМИНОК — банк терминов провинции Квебек (Канада), ТАРГЕТ — банк университета Карнеги-Мэллон (США).

Одной из задач создаваемого банка технических терминов Великобритании является контроль за правильностью употребления английских терминов, включенных в уже действующие терминологические банки (напр., Швеции, ФРГ).

К числу задач, решаемых с помощью банка TEAM фирмы «Сименс», относится и лексикографическая работа — подготовка и издание различных словарей. Для этого используется фотонаборное устройство ДИГИСЕТ.

Технологический процесс автоматизированной подготовки словарей может быть представлен следующим образом: выбор определенного множества терминов из автоматизированной базы терминологических данных; сортировка и представление данных в форме, обычной для словарей; подготовка страницы набора (т. е. кадра фотопленки с текстом, сверстанным в формы, обычные для словарей), тиражирование.

Преимуществами машинных словарей является экономное, с точки зрения места, размещение накопленной информации; возможность быстрой актуализации массива и, вследствие этого, максимальная оперативность в подготовке соответствующих современному уровню словарей; оперативная подготовка словарей с различным набором элементов данных.

В Советском Союзе имеется несколько банков терминов. Один из них функционирует во Всесоюзном научно-исследовательском институте технической информации, классификации и кодирования (ВНИИКИ) Госстандарта — это Автоматизированная система информационно-терминологического обслуживания (АСИТО), в которой аккумулированы данные по стандартизованной терминологии. Основными задачами этой системы является повышение эффективности внедрения стандартизованных терминов, оперативности и качества информационно-терминологического обеспечения народного хозяйства, содействие взаимосогласованию разрабатываемых и стандартизуемых терминологий, реализация обязательств СССР в ИСО, МЭК, МОЗМ, ЕОКК и других международных организациях при подготовке русских версий международных терминологических стандартов и словарей.

АСИТО разработана на базе фонда терминологических документов и системы ручных библиографических и терминологических карточек Справочного банка терминов (СБТ) ВНИИКИ. АСИТО — развивающаяся система. Первая очередь этой системы — АСИТО-1 с 1981 г. находится в промышленной эксплуатации, одновременно ведутся работы по совершенствованию и модернизации системы — созданию АСИТО-2. В АСИТО-2 планируется расширение состава обрабатываемых документов и элементов данных, реализация новых видов поиска, напр., по определениям, расширение набора выходных форм и т. д.

В АСИТО-1 обрабатываются данные из государственных стандартов, стандартов СЭВ и ИСО на термины и определения. Как и в других автоматизированных терминологических системах, основной единицей хранения является терминологическая статья с набором элементов данных, содержащих терминологическую и библиографическую информацию о конкретном стандартизованном термине.

В АСИТО-1 каждая терминологическая статья включает заглавный — стандартизованный термин; допустимые и недопустимые синонимы стандартизованного термина; краткую форму (для терминов-словосочетаний); эквиваленты стандартизованного термина на английском, немецком и французском языках; определение понятия; буквенные обозначения величин; обозначение и наименование стандарта, которым установлен термин; номер термина в стандарте; дату утверждения и введения в действие стандарта; классификационный код стандарта. Всего в автоматизированной базе данных АСИТО-1 представлено около 70 тыс. терминологических статей.

Программные средства обеспечивают ведение БД и различные виды поиска. Основной вид поиска — диалоговый, с использованием дисплея. Поиск может осуществляться по любым элементам данных терминологической статьи и их сочетаниям. Реализована возможность поиска с учетом словоизменительных вариантов слов и синтаксических вариантов словосочетаний. Предусмотрена возможность поиска и выдачи информации при неполном лексическом совпадении запроса с термином, введенным в систему, а также усечений концов словоформ запроса. Так, при наличии в банке термина *дискретная случайная последовательность* информация о нем должна быть выдана и в том случае, если в запросе будет другой порядок терминоэлементов: *случайная дискретная последовательность*. Точно также иное окончание термина в запросе по сравнению с окончанием термина, представленного в АСИТО, не должно влиять на полноту ответа. Таким образом, система позволяет вводить запрос на естественном языке без предварительного индексирования, дает возможность ограничить поиск определенной предметной областью, задавать категорию нормативно-технических документов и другие параметры.

Система позволяет подготавливать различные указатели: сводные и систематические, простые и пермутационные (типа KWOC), по категориям нормативно-технических документов, терминов на русском языке и иноязычных эквивалентов и др.

Предусмотрена возможность обмена информацией на магнитных лентах на основе единого коммуникативного формата для терминологических данных, структура которого унифицирована с коммуникативным форматом библиографических данных в соответствии с международным стандартом ИСО 2709, а содержание соответствует проекту формата

MATER (ISO 6156). Пользователь системы может получать на магнитной ленте как весь массив стандартизованной терминологии, так и его часть в соответствии с запросом.

Терминологический банк АСИТО используется для информационно-терминологического обслуживания научно-исследовательских и проектно-конструкторских организаций, издательств, промышленных предприятий (в настоящее время АСИТО обслуживает 400 коллективных абонентов); обмена информацией с банками терминов (как внутри страны, так и с зарубежными); подготовки статистико-аналитических данных о стандартизованной терминологии.

Учитывая, что количество банков терминов, без сомнения, будет увеличиваться, необходимо разработать целевую комплексную программу создания общегосударственной системы терминологических банков, которая включала бы, наряду с банками стандартизованных терминов, также и банки нестандартизованных лексических единиц.

Система терминологических банков должна быть способна выполнять функцию межнациональной терминологической службы, гарантирующей точность и строгость терминоупотребления во всех документах в научно-исследовательском, административно-управленческом, общественно-историческом, культурно-историческом аспектах, охватывать все отрасли науки и техники и включать русские термины и иноязычные (и национальные) эквиваленты, употребляемые в терминологических системах национальных языков союзных республик и в международном профессиональном общении.

Общегосударственная система всех терминологических банков, видимо, должна включать банк (или банки) терминов фундаментальных наук, банки общетехнических и отраслевых терминов, банки терминов на национальных языках союзных республик.

Наличие уже в настоящее время большого числа банков терминов, в которых хранение, обработка и выдача терминологической информации осуществляется с помощью ЭВМ, поставило проблему их взаимодействия, организации обмена между ними терминологическими данными. Создание международной сети взаимосвязанных банков терминов — системы ТЕРМНЕТ — одна из задач Инфотерма. Для реализации этой задачи ведется разработка международного коммуникативного формата для обмена терминологической информацией на магнитных носителях. Эту работу в течение ряда лет ведет специально созданная в рамках ИСО/ТК 37 «Терминология. Принципы и координация» рабочая группа 4 (РГ 4) «Использование средств вычислительной техники в терминологии и лексикографии». Этой рабочей группой подготовлен проект международного стандарта 6156 «Формат для обмена терминологическими данными на магнитной ленте» (сокращенно формат MATER). Цель разработки этого стандарта — установление требований к формату и по-

рядку размещения данных на магнитных носителях. Он предназначен для обмена одно- и многоязычной терминологической информацией<sup>5</sup>.

Факт повсеместного распространения терминологических банков говорит о многом. В настоящей статье дано только самое общее представление о терминологических банках, о мировой терминологической сети, которая находится в стадии создания.

Интерес к банкам терминов, как уже отмечалось, проявляют за рубежом не только службы стандартизации и информационные центры, но и правительственные учреждения и крупные фирмы. В создание банков терминов вкладываются значительные средства.

Это объясняется тем, что терминологии фиксируют современный уровень научного и технического знания. Дальнейшее развитие наук, научная коммуникация, вопросы государственного планирования, статистического учета, процесса преподавания, информационная деятельность невозможны без принятия единых, одинаково всеми понимаемых терминологий, номенклатур и марок изделий, буквенных обозначений величин и разных сокращений.

---

<sup>5</sup> Перстнев И. П., Волкова И. Н. Коммуникативный формат для обмена терминологическими данными на магнитной ленте.— Научно-техническая терминология, 1980, № 9, с. 22—24.