

## СОЗДАНИЕ ПРОБЛЕМНО-ОРИЕНТИРОВАННЫХ БАЗ ДАННЫХ

ЗЕНОНА АТКОЧЮЕНЕ, РОМУАЛЬДАС БРОНЮКАИТИС

В настоящее время применяется программно-целевой метод планирования, который на базе Комплексного системного подхода обеспечивает поступательное развитие общества в соответствии с требованиями основного закона социализма и нацелен на всестороннюю интенсификацию производства, дальнейшее повышение эффективности и качества работы, внедрение достижений научно-технического прогресса. Это определяется такими причинами, как динамика современного производства, сокращение жизненного цикла продукции, усиление роли специализации, превращение науки в непосредственную производительную силу нашего общества, большая взаимозавязка всех сфер хозяйственной жизни общества.

Целевые комплексные программы это объективная реальность, требующая соответствующей перестройки и информационной деятельности. Она обусловила появление в Государственной системе научно-технической информации (ГСНТИ) нового направления работы — сопровождения научно-технических программ. «В сущности, постановка этого вопроса перед информационными центрами — говорит председатель ГКНТ академик Г. И. Марчук на всесоюзном совещании по сопровождению научно-технических программ (Москва, 11 ноября 1981 г.) — не должна рассматриваться как принципиально новая. Проблемы информационного обеспечения всегда решались и будут решаться на основе информационных потребностей: чем

заныты предприятия и организации на местах, какие планы они реализуют, какие задачи решают, на какие конечные результаты нацелены. Это — основа всей дальнейшей информационной работы, неотъемлемая ее часть. Значит, наблюдение за ходом выполнения программ позволяет знать тонкости решаемых на местах задач, а следовательно, целенаправленно, точно и своевременно обеспечить исполнителей отдельных заданий и этапов программ необходимой научно-технической информацией. Таким образом, задачи сопровождения и информационного обеспечения программ органично связаны, и нельзя отрывать процесс сопровождения программ от процесса их информационного обеспечения...».

Однако знать только информационные потребности специалистов явно недостаточно для эффективного информационного обеспечения программ. Необходимо иметь и достаточно хорошую информационную базу, в которой осуществляется поиск и выявляются релевантные по показателям полноты и точности источники информации.

Бывшие до недавнего времени традиционными способы организации данных в виде отдельных массивов, содержащих данные, нужные для решения конкретных задач, в настоящее время не являются эффективным средством обеспечения информацией исследовательского процесса или процесса принятия решений. Из-за прикрепления к конкретным программам они статичны, для решения различных

задач данные приходится многократно дублировать, что приводит к неоправданным экономическим затратам (система большую часть времени и ресурсов тратит на внутреннюю работу, а не на обслуживание потребителей).

«Принципиально новый подход к организации данных заключается в создании баз и банков данных, представляющих собой проблемно-ориентированные информационно-справочные системы коллективного пользования»<sup>1</sup>.

Однако, создание проблемно-ориентированных баз данных само по себе является комплексной проблемой. Поэтому для ее анализа используем системный анализ, который составляет методологическую основу комплексного подхода к решению экономических и научно-технических проблем<sup>2</sup>.

Проблемно-ориентированной базой данных (ПОБД) будем называть совокупность взаимосвязанных массивов, обеспечивающих хранение сведений о некоторой комплексной или научно-технической программе с такой минимальной избыточностью, которая позволяет их оптимальное использование во всех практических случаях, связанных с решением задач по этой программе.

Для создания и последующего использования ПОБД необходимо наличие информационных ресурсов, программного обеспечения создания базы данных (БД), технических средств обработки и хранения данных, программного обеспечения для поиска требуемой информации.

**Информационные ресурсы.** В настоящее время в рамках Государственной системы научно-технической информации (ГСНТИ) организован обмен информацией на магнитных лентах, что обеспечивает значительную экономию затрат при вводе информации, создает реальные предпосылки для формирования распределенного автоматизированного банка данных по различным видам и тематике опубликованных и неопубликованных документов. «Подготовку и распространение баз данных на магнитных лентах осуществляют 36 всесоюзных, центральных отраслевых, республиканских и межотраслевых территориальных органов НТИ. Общий объем выпускаемых ими баз данных на магнитных лентах достиг в 1983 г. 3 млн. описаний документов, а в конце одиннадцатой пятилетки базы данных будут формироваться по всему годовому потоку документов, обрабатываемому в ГСНТИ»<sup>3</sup>.

В указанные объемы баз данных могут включаться конкретные фактические данные об объектах, событиях, явлениях, сами документы и информация о документах.

Фактические данные получают в результате обработки системой большого количества поступающих в нее документов. Иногда потребителей интересуют сами документы, а не отдельные факты. Это обычно бывает в тех случаях, когда по тем или иным причинам отсутствуют необходимые фактические данные или потребитель точно не знает, какие факты ему нужны.

<sup>1</sup> Мамиконов А. Г. Принятие решений и информация.— М.: Наука, 1983.— С. 168.  
<sup>2</sup> Джалалов Ю. О целевых комплексных программах в технике.— М.: Знание, 1982.— 64 с.

<sup>3</sup> Арутюнов Н. Б. О дальнейшем повышении эффективности Государственной системы научно-технической информации, как важного условия ускорения научно-технического прогресса в народном хозяйстве.— М.: Гос. ком. по науке и технике, 1984.— С. 26.

В подобных ситуациях требуются целые документы или их копии.

Чтобы эффективно вести поиск и отбор нужных документов из больших информационных массивов, целесообразно предварительно получить сигнальную информацию — информацию о документах, т. е. библиографические данные: автор, заглавие, год и место издания, реферат или аннотация, место хранения документа и т. д. Подобные сведения позволяют потребителю быстро оценить целесообразность непосредственного обращения к соответствующим документам.

Наряду с обменом информацией на машинночитаемых носителях осуществляются работы по созданию сети автоматизированных центров научно-технической информации (САЦТИ) <sup>4</sup>.

Информационные вопросы САЦТИ являются совокупностью первичных и вторичных документов, составляющих единый справочно-информационный фонд сети (СИФ).

Создание единого справочно-информационного фонда в САЦТИ осуществляется в условиях, характерных:

— большим объемом хранимой научно-технической информации сложной лингвистической структуры;

— большим расходом памяти из-за неизбежного дублирования НТИ при формировании массивов документов / данных;

— необходимостью регулярного проведения коррекции и реорганизации информационных массивов.

Кроме того, при проектировании подсистемы накопления и хранения НТИ необходимо считаться с требованиями устойчивости подсистемы к изменению ее функционирования.

В связи с этим при создании САЦТИ целесообразно использовать методы и средства хранения и ведения информационных массивов, которые ориентировались бы на интегрированную обработку информации и коллективное использование массивов документов / данных.

Создание сети преследует цель автоматизировать не только внутренние процессы прохождения информации в автоматизированных информационных центрах (АИЦ), но также и взаимодействие АИЦ, что связано с распределением и поиском информации по сети, поэтому справочно-информационный фонд сети должен состоять из:

— распределенной базы документов / данных;

— распределенного документально-го фонда.

Каждый из этих массивов в свою очередь предполагается подразделить на основной массив, обеспечивающий максимальную полноту СИФ по охвату мирового документального потока, и дубль-массив, тематический состав которого определяют сами АИЦ.

При такой организации справочно-информационного фонда сети дублирование информационных массивов сохраняется, однако повышается эффективность обслуживания потребителей.

Дублирование и объем хранения научной и технической информации в САЦТИ определяется технико-экономическими затратами на хранение документов и на передачу их по каналам связи.

Распределенная база документов / данных, представляющая собой массив вторичных документов в машинно-читаемой форме, размещается в ло-

<sup>4</sup> Михайлов А. И., Болотин И. А., Кулебякин А. З. Сеть автоматизированных центров научно-технической информации.— М.: ВИНТИ, 1979.— 56 с.

кальных автоматизированных информационных центрах.

Под базой документов/данных в данном случае понимается записанная на магнитных лентах или других машиночитаемых носителях совокупность сведений о научных и технических документах, позволяющих идентифицировать документ и оценить его релевантность в соответствии с запросом пользователя.

Совокупность баз документов/данных и комплекс языковых и программных средств, обеспечивающий создание, накопление, хранение, ведение и обработку этих баз, образуют автоматизированные банки документов/данных. Система локальных автоматизированных банков документов/данных АИЦ, которые взаимодействуют в рамках общих принципов формирования массивов и функционируют под единым управлением, образует распределенный автоматизированный банк данных (РАБД) сети.

РАБД должен развиваться как по пути создания локальных АБД в каждом АИЦ, так и по пути организации связи между локальными АБД для обеспечения их взаимодействия в процессе информационного обслуживания абонентов сети.

РАБД сети создается с целью упорядочения хранения информации, интеграции и реорганизации массивов без изменения обслуживающих программ, направленного доступа к информации с минимальными затратами времени, однократного ввода информации в систему и многократного ее использования в нужных для абонентов сочетаниях.

Таким образом, распределенный автоматизированный банк документов/данных в составе сети выступает как новая функционально-технологическая компонента, реализующая

справочно-информационную функцию и обеспечивающая эффективное накопление, хранение и использование информационных фондов сети.

Основными принципами формирования РАБД являются:

— централизованная обработка мирового потока НТИ на уровне АИЦ всесоюзного статуса для накопления документов по видам;

— распределение усилий в подготовке и формировании массивов документов/данных на машиночитаемых носителях между всесоюзными и отраслевыми АИЦ;

— децентрализованное хранение массивов.

Наряду с рассмотренной информацией имеется и другая, которая возникает в определенном регионе. Это отчеты по НИР и ОКР, отчеты о заграникомандировках, переводы, обзоры, материалы проведенных в регионе конференций и симпозиумов, депонированные рукописи, информационные листки, нормативно-технические и инструктивные материалы, информация о зарубежных книгах и периодических изданиях, различные фактографические данные. Вся эта информация также может быть использована в процессе создания проблемно-ориентированных баз данных.

**Программное обеспечение создания базы данных** — это совокупность математических методов, алгоритмических языков и программного обеспечения (в дальнейшем — математическое обеспечение). Проблемно-ориентированные базы данных создаются на основе рассмотренных выше информационных ресурсов и в дальнейшем будут составлять распределенную базу документов/данных в определенном регионе или отрасли. Поэтому математическое обеспечение не должно принципиально отличаться от обеспечения, применяемого при создании

баз данных САЦНТИ или любых других ретроспективных баз. Такое тематическое обеспечение разработано в Республканской автоматизированной системе научно-технической информации (РАСНТИ)<sup>5</sup>. Оно позволяет осуществлять:

— хранение вторичных документов и адресный указатель к первичным документам;

— хранение поисковых образов документов и баз данных НТИ как в исходном виде, так и в виде, подготовленном для проведения поисковых операций;

— хранение адресных шифров первичных и вторичных документов;

— смысловой поиск информации по семантическим, логическим, позиционным, весовым и фактографическим критериям;

— информационно-логическую переработку информации;

— адресный поиск и выдачу абонентам вторичных документов;

— адресный поиск и выдачу абонентам первичных документов или их копий.

Поступающая информация разделяется на:

— документальную информацию на МЛ (вторичные документы);

— запросы на информацию (запросы ИРИ, разовые тематические запросы, запросы на копии первичных документов);

— лингвистические средства для формирования поисковых предписаний.

Входная информация регистрируется и, в зависимости от ее вида, проходит различную предварительную обработку.

Обработка вторичных документов включает создание машиночитаемого фонда данных, на базе которого осуществляется избирательное распространение и ретроспективный поиск информации. Основу технологии обработки вторичных документов представляет машинная обработка информации на ЭВМ типа ЕС с использованием пакета прикладных программ. Весь процесс машинной обработки разбит на технологические цепочки, каждая из которых предоставляет некоторый законченный частичный технологический процесс (например, загрузку описаний документов). Для каждой такой технологической цепочки разработан соответствующий пакет заданий и инструкция для его обработки на ЭВМ. Технология обработки баз данных на ЭВМ представлена на рис. 3<sup>6</sup>.

Для функционирования Подсистемы формируются и ведутся основные системные наборы данных и информационные массивы.

Системные наборы данных многократно используются монитормыми (управляющими) программами. Доступ к этим наборам осуществляется с помощью интерфейсных программ ввода/вывода, входящих в состав служебных модулей или базисных системных модулей библиотеки программ.

Математическое обеспечение РАС НТИ также включает программные средства для информационного фактографического обслуживания потребителей по разовым и постоянным запросам. С этой целью разработана базовая система данных ФАКТ. Она представляет собой комплекс про-

<sup>5</sup> Автоматизированная система научно-технической информации Литовской ССР.: (АСНТИ ЛитССР) / Сост. Бронюкайтис Р. и др.— Вильнюс: ЛитНИИНТИ, 1980.— 48 с.

<sup>6</sup> Определение информационных потребностей специалистов народного хозяйства в республике: Аналит. обзор / Аткачюнене З., Бронюкайтис Р.— Вильнюс: ЛитНИИНТИ, 1985.— С. 22.

граммных и языковых средств, позволяющих создать и вести базу фактографических структурированных данных и осуществлять поиск и выдачу данных по запросам пользователей.

Функциональные, программные, языковые, и технические средства системы иллюстрируются на рис. 4 и н 5<sup>7</sup>.

**Технические средства обработки и хранения данных** включают средства вычислительной техники, связи и передачи данных, микрофильмирования, полиграфии, полноформатного копирования и размножения документов, оргтехники. Предполагается, что для создания и эксплуатации проблемно-ориентированных баз данных будут использоваться технические средства РАСНТИ, состав и типы которых обоснованы в техническом проекте РАСНТИ<sup>8</sup>.

**Программное обеспечение (ПО)** для поиска требуемой информации не отличается от программных средств РАСНТИ и состоит из следующих компонентов:

- 1) общесистемного или машинного ПО,
- 2) системного ПО,
- 3) специального ПО.

Общесистемное ПО является составной частью ПО ЕС ЭВМ и СМ ЭВМ, предназначенного для повышения эффективности использования аппаратных средств и повышения производительности труда программистов и обслуживающего персонала. Оно представляет собой набор основных операционных систем, прикладных программ и программ технического обслуживания. В настоящее время для ЕС ЭВМ

основными являются две операционные системы: дисковая операционная система (ДОС/ЕС) и операционная система (ОС/ЕС), а для СМ ЭВМ — операционная система с разделением времени (ОС/РВ). Для АСНТИ ЛитССР в качестве базовых намечено использовать операционные системы ОС/ЕС и ОС/РВ.

Системное ПО предназначено для реализации основных функций и задач АСНТИ и состоит из совокупности ППП (Пакет прикладных программ) различного функционального назначения. В состав системного ПО ЕС ЭВМ входят следующие ППП: ДИАЛОГ, МУЛЬТИПЛЕКС, ФАКТ-ДИАЛОГ и ПОИСК-4.

Специальное программное обеспечение состоит из совокупности программ, предназначенных для наращивания возможностей, а также комплексирования средств системного ПО в соответствии с требованиями и спецификой задач, решаемых в системе. Оно образует «микрোকружение» системного ПО и создается для вполне конкретных случаев применения. В состав специального ПО АСНТИ ЛитССР входят в основном программы, преобразующие форматы представления данных (программы-конверторы), программы, реализующие библиотечные функции учета и контроля абонентов, управление СИФ-ом и учет деятельности республиканской сети библиотек, программы, обеспечивающие выборку и вывод данных обработки из ЭВМ, а также комплекс программ автоматического формирования информационных потребностей.

Проведенный выше внутренний ана-

<sup>7</sup> **Определение информационных потребностей специалистов народного хозяйства в республике:** Аналит. обзор / Аткиюнене З., Бронюкайтис Р.— Вильнюс: ЛитНИИНТИ, 1985.— С. 24—25.

<sup>8</sup> **Развитие автоматизированной системы научно-технической информации Литовской ССР:** Техн. проект.— Вильнюс: ЛитНИИНТИ, 1982.— 201 с.

лиз создания ПОБД показывает, что в настоящее время имеются все предпосылки для создания БД на основе поступающих и генерируемых на ме-

сте информационных ресурсов, а также информационно-технологического, лингвистического, программного и технического обеспечения РАСНТИ.

Литовский научно-исследовательский институт научно-технической информации и технико-экономических исследований (ЛитНИИИТИ)

Подготовлено к печати в 1985 г.

## CREATION OF PROBLEM-ORIENTED DATA BASES

ZENONA ATKOŠCIŪNIENĖ, ROMUALDAS BRONIUKAITIS

### S u m m a r y

Program objective method is currently employed for planning the directions of overall industrial intensification and introduction of scientific and technical achievements into practice.

Objective complex programs demand respective reorganization of information activities.

The fundamentally new approach to data organization lies in the creation of data banks (DB) representing problem-oriented systems for collective use.

The accomplished analysis of practice in creation of oriented DB shows that at present there are possibilities to create DB resting upon information resources either provided or generated on the spot as well as on information and technological, linguistic, program and technical supply arranged by Republican automatic system and technical information.