



БА́ЗА ДА́ННЫХ

Авторы: А. Б. Антопольский

БА́ЗА ДА́ННЫХ, объективная форма представления и организации совокупности данных (статей и др.), систематизированных таким образом, чтобы эти данные могли быть найдены и обработаны ЭВМ. Центральным понятием теории Б. д. является модель данных (МД), под которой понимается совокупность правил структурирования данных в Б. д., допустимых операций над ними и ограничений целостности, которым они должны удовлетворять. Целостность Б. д. – свойство Б. д., означающее, что она содержит полную и непротиворечивую информацию, необходимую для корректного функционирования приложений.

Различают осн. типы МД: иерархические, сетевые, реляционные, объектные.

Иерархическая МД представляет собой древовидный граф (см. *Графов теория*), вершины которого состоят из записей определённых типов и связей между ними, причём один тип записи определяется как корневой, а остальные связаны с ним или друг с другом отношением «один-ко-многим». Иерархич. Б. д. – совокупность таких древовидных графов. Сетевая МД представляет собой граф общего вида, вершинами которого являются данные разных типов, простые (атомарные) или составные записи, а дугами – связи между этими данными. Записи – участники связей – соединяются в список, который называется набором. Наиболее известная версия – МД CODASYL (Conference on Data System Languages – Конференция по языкам систем обработки данных). Сетевая Б. д. состоит из набора записей и набора связей между этими записями. Реляционная МД представляет собой набор таблиц, называемых отношениями. Т. к. в реляционной МД можно организовать очень большое число связей, значит. часть которых избыточна, то такая МД устанавливает четыре формы (варианта) нормализации отношений. Отношения включают атрибуты и кортежи, составляющие соответственно столбцы и строки таблицы. Множество значений атрибутов называется доменом. Реляционная МД послужила основой стандарта языка реляционной Б. д. SQL (Structured Query Language – Язык структурированных запросов). Объектная МД основана на понятии объекта, т. е. сущности, обладающей состоянием и поведением. Состояние объекта определяется совокупностью его атрибутов, а поведение – совокупностью операций, допустимых для объекта. Между объектами устанавливаются связи. Объекты типизируются, причём обычно предусматривается возможность связи между типами объектов. Объектная МД представлена в стандарте ODMG (Object Database Management Group – Группа управления объектно-ориентированными базами данных).

По применяемой МД различают Б. д.: иерархические, сетевые, реляционные, объектные, а также смешанные (объектно-реляционные и т. п.). В кон. 1990-х гг. появился новый вид Б. д., предназначенный для хранения и обработки XML-документов (XML-документ представляет собой обычный текстовый файл, в котором при помощи спец. маркеров создаются элементы данных, последовательность и вложенность которых определяют структуру документа и его содержание). По принципу организации в телекоммуникац. среде различают Б. д. локальные, с сетевым доступом, распределённые, фрагментированные, централизованные, тиражированные и мобильные; по способу отображения предметной области выделяют Б. д. предметные, интегрир., концептуальные, внешние и внутр.; по характеру использования – встроенные (служебные) и общедоступные, предназначенные для

внешнего использования; по режиму доступа – открытые и конфиденциальные. Б. д. используются в рамках [банков данных](#).

Для создания и использования Б. д. важное значение имеет понятие «метаданные». Это сведения о данных, хранящихся в базе, описывающие их структуру, состав, формат представления, методы доступа, полномочия пользователя или администратора и др. Метаданные подразделяют на системные и пользовательские. Первые предназначены для поддержки системных функций, среди них осн. роль играют схемы Б. д. Вторые описывают свойства данных, представляющие интерес для конечных пользователей, прежде всего для поиска в Б. д. В состав метаданных входят информационно-поисковые языки, определяющие возможности и качество поиска в базе данных.

Осн. параметрами Б. д. принято считать её физич. или логич. объёмы. Физич. объём Б. д. выражается в байтах и производных величинах (килобайтах, мегабайтах и др.), логич. объём определяется числом записей (объектов, документов), представленных в Б. д. Однако оба параметра являются нестрогими, поэтому сравнивать по объёму можно только Б. д. с одинаковой структурой. Средний объём общедоступных Б. д. составляет ок. 1 млн. записей, объём одной записи для текстовых Б. д. варьируется в пределах 200–2000 слов. Объёмы крупнейших Б. д. составляют десятки терабайт, или десятки млрд. записей. Важный показатель качества Б. д. и эффективности вложенных в них средств – число пользователей или запросов к Б. д., однако этот показатель применим только к общедоступным базам данных.

Термин «Б. д.» введён в 1963 на первом симпозиуме, посвящённом проблеме организации данных, состоявшемся в г. Санта-Моника (Калифорния, США). Значит. влияние на развитие технологий Б. д. оказали достижения в смежных областях – операционных системах, языках и технологиях программирования. Фундам. разработками первого периода стали сетевая МД Комитета CODASYL (фактически первый стандарт в области Б. д.), иерархич. МД с языком DL-1 (Data Language 1), разработанная компанией IBM, и реляционная МД Э. Кодда (он же ввёл само понятие «модель данных»). В этот же период появились первые коммерч. системы управления базами данных (СУБД). В 1970-х гг. сформировалась наука о Б. д., была разработана технология Б. д., началось индустриальное произ-во СУБД. Создана теория реляционных Б. д. и основанные на ней технологии реляционных СУБД, которые с 1980-х гг. доминируют на мировом рынке. Коммерч. СУБД этого класса разработаны для всех аппаратных средств вычислит. техники, включая персональные компьютеры. Пик разработок языков программирования Б. д. пришёлся на кон. 1980-х гг., среди которых наибольшее распространение получил объектный язык C++. С сер. 1990-х гг. активизировалось развитие стандартов и технологий объектных Б. д., это связано в т. ч. с появлением языков программирования Java и UML (Unified Modeling Language – Унифицированный язык моделирования). В 1999 принят Стандарт SQL. В кон. 1990-х гг. появились базовые технологии управления данными для систем поддержки принятия решений: интерактивная аналитич. обработка данных (OLAP – On-line Analytical Processing), технологии хранилищ данных (Data Warehousing), технологии глубинного анализа данных (Data Mining). Созданы спец. технологии для очень больших Б. д., ориентированных на терабайты и даже петабайты информации. В 1990-е гг. в связи с созданием открытой распределённой неоднородной гипермедийной информац. системы World Wide Web (Всемирная паутина, или WWW, W3, Web), использующей коммуникац. среду Интернет, происходит активное взаимодействие технологий Б. д. и технологий WWW, среди которых наибольшее значение для Б. д. имеет расширенный язык разметки XML (Extensible Markup Language – Расширяемый язык разметки). С кон. 1990-х гг.

на базе технологий WWW ведутся многочисл. исследования, связанные с интеграцией информац. ресурсов, в т. ч. неоднородных.

Б. д. – информац. продукт, который можно приобрести либо непосредственно на разл. носителях (чаще всего на оптич. дисках типа CD-ROM), либо через Интернет или по др. протоколам, оплатив услуги по доступу к Б. д. По содержанию выделяют Б. д. библиографические, реферативные, новостные, правовые, словарные, биографические, адресно-справочные, Б. д. о пром. продукции, химич. соединениях и минералах, динамич. ряды статистич. и демографич. данных и др. До сер. 1980-х гг. доминировали Б. д. научно-технич. информации, затем резко возросло значение Б. д. коммерч. и деловой информации, в т. ч. биржевой и финансовой.

Б. д. – важнейший вид [информационных ресурсов](#), количество, качество и характер использования которых во многом определяют уровень развития страны как информационного общества. Поэтому создание и использование Б. д. в развитых странах, включая Россию, является объектом правового регулирования (см. [Информационное право](#)).

Литература

Лит.: Directory of online databases. Detroit; L., 1979–1992. [Vol. 1–13]; Gale directory of databases. Detroit; L., 1993–2004– . [Vol. 1–12–]; Когаловский М. Р. Энциклопедия технологий баз данных. М., 2002; Родионов И. И. и др. Рынок информационных услуг и продуктов. М., 2002.