

Дорогие читатели!

Журнал “Проблемы информатики” открывает новую рубрику “Дискуссионный клуб”. Задача публикуемых в ней статей — поиск истины в научном споре. Не на все вопросы можно дать точные ответы, но наука не должна отказываться от их поиска.

Научная деятельность состоит из постановки задачи, выбора понятий и аксиом и решения задачи. Каждый этап по-своему сложен и интересен. Истории науки известны случаи, когда возвращение к предыдущему этапу, уточнение формулировок и понятий приводили к новым решениям или к постановке более глобальных задач. Всегда полезно вести работу над уточнением понятий, используя более широкий взгляд на исследуемый объект. Первая публикация в новой рубрике посвящена обсуждению общеизвестного понятия “база данных” с разных точек зрения.

Если у Вас есть тема для дискуссии в профессиональной сфере или желание высказаться в ответ на текущую публикацию, пишите на электронный адрес редакции problem-info@sscc.ru с пометкой “Дискуссионный клуб”.

Главный редактор журнала “Проблемы информатики” проф. В. К. Попков

К ФОРМАЛЬНОМУ ОПРЕДЕЛЕНИЮ ПОНЯТИЯ “БАЗА ДАННЫХ”

Е. А. Мирошниченко

Институт кибернетики Национального исследовательского
Томского политехнического университета, 634034, Томск, Россия

УДК 541.64:547.759.32

Показано, что понятие “база данных” не имеет общепризнанного определения, несмотря на длительную историю существования и наличие стандартов. Классифицированы основные расхождения в определениях. Рассмотрены недостатки ряда формулировок. Выполнен анализ существенных признаков понятия “база данных”.

Ключевые слова: база данных, информационная система, система управления базами данных.

The paper shows the lack of universally recognized definition of database concept even with a glance of long-term history and presence of standards. Main differences in existing definitions are classified. Drawbacks of some wordings are discussed. Essential attributes of database concept are analyzed.

Key words: database, information system, database management system.

Введение. Точное определение понятия имеет большое значение в науке, так как является решающим фактором взаимопонимания исследователей. Проиллюстрируем эту мысль примерами из личного опыта автора данной работы. Сотрудник промышленного предприятия может называть базой данных отдела сейф с документами. Другие специалисты под проектированием базы данных, содержащей геолого-геофизическую информацию, понимают разработку иерархии каталогов файловой системы для размещения файлов первичных данных, т. е. разработку примитивного файлового архива.

Базы данных (БД) используются начиная с 1955 г., когда появилось программируемое оборудование для обработки записей. За это время опубликовано большое количество работ по этой теме (см., например, [1–5]). Определение понятия “база данных” можно найти в законодательстве РФ [6], глоссарии IEEE [7], стандарте ISO [8]. Приведенные выше примеры показывают, что распространенность понятия не является гарантией его единообразного понимания и трактовки: «Специалисты в разных областях, не являясь профессионалами в области информационных систем, часто полагаются на кажущийся интуитивно ясным смысл понятия “информационная система” и в результате весьма вольно с ним обращаются, как и с другими “заезженными” терминами. Так обстоит дело, например, с термином *база данных*. Часто базой данных называют любую совокупность данных, независимо от того, идет ли разговор в контексте технологий баз данных» [9]. Проблема состоит в существенности различий, иными словами, в неэквивалентности определений. Следует отметить, что ряд специалистов в данной области не пытаются дать корректное определение этого понятия [3].

Известные определения. Приведем ряд определений БД из источников, заслуживающих доверия: отечественных и зарубежных научных изданий по теории БД, международных стандартов, законодательных актов.

База данных — некоторый набор перманентных (постоянно хранимых) данных, используемых прикладными программными системами какого-либо предприятия [1].

База данных — организованная в соответствии с определенными правилами и поддерживаемая в памяти компьютера совокупность данных, характеризующая актуальное состояние некоторой предметной области и используемая для удовлетворения информационных потребностей пользователей [2].

База данных — совместно используемый набор логически связанных данных (и описание этих данных), предназначенный для удовлетворения информационных потребностей организации [4].

Термином “база данных” в соответствии с традицией принято обозначать набор данных, находящихся под контролем СУБД [5].

Базой данных является представленная в объективной форме совокупность самостоятельных материалов (статей, расчетов, нормативных актов, судебных решений и иных подобных материалов), систематизированных таким образом, чтобы они могли быть найдены и обработаны с помощью электронной вычислительной машины [6].

База данных — совокупность взаимосвязанных данных, совместно хранимых в одном или нескольких компьютерных файлах [7].

База данных — совокупность данных, хранимых в соответствии со схемой данных, манипулирование которыми выполняется в соответствии с правилами средств моделирования данных [8].

Приведенные выше определения содержат такие нечеткие понятия, как “перманентность”, “пользователь”, “актуальность”, “предприятие”, “организация”, “объективная форма”, “самостоятельность”, “совместность хранения”. Определение, приведенное в работе [5], тавтологично. Определение, приведенное в [7], позволяет назвать базой данных любой структурированный компьютерный файл, например офисный документ или исполняемую программу. В определении ISO [8] использованы технические термины “схема данных”, “манипулирование данными” и “средство моделирования данных”, введенные позднее, чем термин “база данных”.

Основные разногласия и причины их возникновения. Поиск общей части приведенных определений позволяет сделать следующий вывод: единственное понятие, использованное во всех формулировках, — это “набор (совокупность) данных”.

Проанализируем основные положения, по которым расходятся определения.

Первое разногласие заключается в том, считать ли базой данных исключительно электронный набор данных (т. е. набор данных, хранящихся и обрабатываемых вычислительной системой) или любой набор данных, в том числе на бумажных носителях (архив, картотеку, библиотеку и т. п.). Определения Коннолли — Бегг и ISO не указывают на связь БД с вычислительной системой, несмотря на то что оба они подразумевают компьютерную природу БД. Определение Гражданского кодекса РФ указывает на существование косвенной связи БД с ЭВМ. Остальные определения явно предполагают компьютерную природу БД.

Второе разногласие состоит в том, относить ли к определяющим признакам цель создания БД. В определениях Дейта, Гарсия-Молины — Ульмана — Уидом, IEEE и ISO цель создания БД не упоминается, в отличие от определений Когаловского и Коннолли — Бегг, предлагающих считать целью создания БД удовлетворение информационных потребностей. Определение в Гражданском кодексе также указывает на цель (поиск и обработку в ЭВМ).

Третье разногласие заключается в определении потребителя информации, хранящейся в БД. Согласно определению Когаловского потребителем называется любой пользователь. В определениях Дейта и Коннолли — Бегг потребителем считается предприятие (организация). В остальных определениях понятие “пользователь” отсутствует.

Четвертое противоречие состоит в признании существенности выделенных метаданных (схем, описаний) (см. определения Коннолли — Бегг и ISO).

Пятое, наиболее существенное расхождение касается определения характеристически важного набора данных. В частности, в определениях Гарсия-Молины — Ульмана — Уидом и ISO не указаны свойства набора данных. К. Дж. Дейт использует неясный термин “перманентность”. В комментариях поясняется, что имеется в виду не способ или срок хранения, а существенность сохранения данных, в отличие от временных или промежуточных данных приложений. М. Р. Когаловский упоминает такие свойства набора данных, как организованность и связь с предметной областью. Т. Коннолли и К. Бегг выделяют логическую взаимосвязанность данных в наборе. Определение Гражданского кодекса содержит список свойств: объективность, самостоятельность, систематизированность и приспособленность к компьютерной обработке. Определение IEEE указывает на взаимосвязанность и совместность хранения данных в наборе.

Отсутствие строгого формального определения понятия, имеющего длительную историю использования, свидетельствует о невозможности выведения его из системы базовых понятий предметной области. Это означает, что оно является базовым, следовательно, ему нельзя дать формальное определение, можно только находить объяснение. Итак, причина противоречивости определений понятия “база данных” — его аксиоматичность.

Анализ признаков БД. Поиск идеального определения понятия “база данных” невозможно вести формально путем объединения (добавления всех упоминаемых авторами признаков) или выделения общей части (включения исключительно признаков, упоминаемых всеми авторами) существующих понятий. Более эффективен путь анализа сходств и различий существующих формулировок, однако недостижимость текста идеального определения очевидна.

Компьютерная природа понятия БД не вызывает сомнения, что подтверждается всеми работами, из которых приведены цитаты. Само слово “database” изобретено специали-

стами по разработке программных систем обработки данных и введено в употребление на симпозиумах в 1963–1965 гг., организованных компанией System Development Corp. [2, 10]. Предлагается считать, что внекомпьютерные хранилища информации не являются базами данных.

Очевидно, что БД создаются для накопления, поиска и обработки данных в предметной области с конкретной целью. Однако то же относится к книгам, файлам и архивам. Следовательно, включение в определение признака “наличие цели удовлетворить информационные потребности” не отражает специфику объекта.

Аналогично признак “пользователь” (“потребитель”) может обозначать человека, организацию или компьютерную программу, для функционирования которой необходима БД.

Признак метаданных является очень важным, несмотря на его отсутствие в некоторых определениях. Исследователи истории БД отмечают, что кристаллизация концепции базы данных, отделение ее от файловых систем, зарождение универсальных СУБД произошли на фоне явного выделения схемы БД [10]. В настоящее время не существует систем баз данных, в которых отсутствуют метаданные. Например, известные 12 правил Э. Кодда предписывают наличие динамического каталога (схемы) реляционной БД. Работа с информационной схемой унифицирована в стандарте SQL. Таким образом, признак наличия метаданных является ключевым.

Существует еще одно свойство, упоминаемое в большинстве источников. Его называют организованностью [2], взаимосвязанностью [4, 7], систематизированностью [6]. Во всех случаях речь идет об упорядоченности БД, включенности в систему. По сути, БД как модель предметной области должна строиться как система, т. е. обладать структурой, целенаправленностью, сильной связностью и т. д. В данном случае структурированность БД понимается не на уровне средств хранения байтов информации, а на уровне логической модели данных.

Выводы. Проведенное исследование позволяет сделать следующие выводы. В настоящее время отсутствует единое общепризнанное определение базы данных, а существующие определения значительно различаются. Основной причиной различий является аксиоматичность понятия БД. Однако можно утверждать следующее: 1) любая БД хранится и обрабатывается в вычислительной системе; 2) БД является сильно структурированной системой; 3) БД включает схему данных.

Список литературы

1. Дейт К. Дж. Введение в системы баз данных. 8-е изд. / Пер. с англ. М.: Издат. дом “Вильямс”, 2005. 1328 с.
2. КОГАЛОВСКИЙ М. Р. Энциклопедия технологий баз данных. М.: Финансы и статистика, 2002. 800 с.
3. КУЗНЕЦОВ С. Д. Основы баз данных: Учеб. пособие. 2-е изд., испр. М.: Интернет-университет информационных технологий: БИНОМ. Лаборатория знаний, 2007. 484 с.
4. Коннолли Т., Бегг К. Базы данных. Проектирование, реализация и сопровождение. Теория и практика. 3-е изд. / Пер. с англ. М.: Издат. дом “Вильямс”, 2003. 1440 с.
5. ГАРСИА-МОЛИНА Г., УЛЬМАН ДЖ., УИДОМ ДЖ. Системы баз данных. Полный курс / Пер. с англ. М.: Издат. дом “Вильямс”, 2003. 1088 с.
6. Гражданский кодекс РФ. Ч. 4 от 18.12.2006 г. № 230-ФЗ. Права на результаты интеллектуальной деятельности и средства индивидуализации. М.: Б. и. С. 1260.
7. IEEE Standard Glossary of Software Engineering Terminology: IEEE Std. 610.12-1990.

8. ГОСТ Р ИСО МЭК ТО 10032-2007. Эталонная модель управления данными (идентичен ISO/IEC TR 10032:2003 Information technology Reference model of data management). Введ. 01.09.2008.

9. КОГАЛОВСКИЙ М. Р. Перспективные технологии информационных систем. М.: ДМК Пресс: Компания АйТи, 2003. 288 с.

10. HAIGH T. How data got its base: Information storage software in the 1950s and 1960s // IEEE Ann. History Comput. 2009. N 4. [Electron. resource]. URL: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5370776&isnumber=5370770>.

*Мирошниченко Евгений Александрович — канд. техн. наук,
доц. Института кибернетики ТПУ; тел. (3822) 701-609; e-mail: mir@tpu.ru*

Дата поступления — 03.02.11 г.