

Предисловие СЮ Тюменской нефтяной компании	5
Предисловие от компании Hewlett-Packard	7
Введение	9
Глава 1. Информационные системы и их функции	12
1.1. Что такое информационная система	12
1.2. Моделирование реальности в информационных системах	22
1.3. Функции информационных систем	28
1.4. Разновидности информационных систем	40
1.5. Общие тенденции развития информационных систем	48
Литература к главе 1	59
Глава 2. Технологии баз данных	62
2.1. Общая оценка состояния	62
2.2. Архитектурные подходы	72
2.3. Объектные базы данных	82
2.4. От реляционных баз данных к объектно-реляционным	94
2.5. Интеграция средств текстового поиска в СУБД	105
2.6. XML-ориентированные базы данных	106
2.7. Системы поддержки принятия решений	108
2.8. Очень большие базы данных	117
2.9. Вызовы времени и перспективы	119
Литература к главе 2	124
Глава 3. Технологии текстового поиска	130
3.1. Основные понятия и круг проблем	130
3.2. Краткая история	133
3.3. Принципы текстового поиска	139
3.4. Модели текстового поиска	150
3.5. Дополнительные возможности	156
3.6. Нетрадиционные направления технологий поиска	160
3.7. Текстовый поиск и базы данных	162
3.8. Текстовый поиск в Web	168

3.9. Перспективы систем текстового поиска	175
Литература к главе 3	178
Глава 4. Технологии Web нового поколения	183
4.1. Предпосылки создания платформы XML	184
4.2. Радикальные перемены в Web	189
4.3. Организация и функции платформы XML	193
4.4. Расширяемость языка и платформы XML	201
4.5. Преемственность с технологиями HTML	204
4.6. Моделирование данных XML	206
4.7. Метаданные и семантика XML-документов	210
4.8. Семантический Web	212
4.9. Сферы применения стандартов XML	216
4.10. XML-ориентированные базы данных	221
4.11. Перспективы платформы XML	231
Литература к главе 4	232
Заключение	241
Глоссарий	242
Предметный указатель	274

Предисловие СЮ Тюменской Нефтяной Компании

Современный мир переживает бум развития и применения корпоративных информационных систем (КИС) и ERP-систем. Подготовка и принятие решений в среде ERP-систем стало естественным фактором в конкурентной борьбе компаний: крупные холдинги, функционирующие в соответствии с мировыми стандартами, обязаны опираться на самые современные информационные технологии, получившие мировое признание. Разработкой крупных ERP-систем в России не занимаются, а вот их внедрение осуществляется все чаще.

Надо признаться, что сегодня в России нет достаточного количества квалифицированных кадров, которые обладают необходимым опытом внедрения и эксплуатации крупных ERP-систем. При этом, как мне кажется, отсутствует и продуманная национальная система подготовки таких специалистов в российских вузах. Есть прекрасные школы, выпускающие программистов и специалистов по компьютерной аппаратуре. Но практически нигде не учат внедрению современных бизнес-приложений в рамках корпоративных информационных систем и работе с ними. Мне не приходилось встречать серьезных работ российских авторов по экономике информационных систем.

В нефтяном бизнесе конкуренция может быть самая жестокая — мы обязаны работать на очень высоком уровне подготовки и принятия решений с использованием последних достижений в области информационных технологий. Высокая конкурентоспособность компаний, в том числе и нашей ТНК, не может быть обеспечена только иностранными специалистами. Как воздух нужны молодые российские таланты, блестяще сочетающие глубокие знания и широкий кругозор в экономике, информационных системах и математике. Совместный проект МГУ и АйТи, названный «ИТ-экономика», служит как раз этой благородной цели — созданию методического фундамента для подготовки специалистов нового комплексного профиля.

Ежегодно в ТНК устраиваются на работу выпускники МГУ, проходят практику, участвуют во внедрении современных информационных систем бакалавры и магистры экономического факультета МГУ. Таким образом, мы участвуем в подготовке нужных нам специалистов. Жизнь научила нас больше рассчитывать на собственные силы даже в деле выращивания квалифицированных кадров. Для нас поддержка совместного проекта «ИТ-экономика» — это и честь, и насущная жизненная необходимость.

Желаю успеха этому проекту, и пусть изданные в его рамках учебные курсы и книги помогут появиться в России новому поколению талантливых специалистов в области экономики и применения корпоративных информационных систем.

*Александр Блох,
корпоративный управляющий
информационными технологиями (CIO) ОАО ТНК*

ПРЕДИСЛОВИЕ ОТ КОМПАНИИ HEWLETT-PACKARD

Развитие корпоративных информационных систем в России в последние годы становится частью общей стратегии бизнеса предприятия. Однако такая роль информационных систем и информационных технологий требует не только управленческой воли (хотя она, безусловно, необходима), но и решения целого ряда сопутствующих проблем.

Прежде всего это методологические проблемы. Информатизация становится в ряд равноправных бизнес-проектов, а значит, на всех стадиях планирования и учета рассчитывается и контролируется финансовый результат. Для этого как минимум необходима адекватная методика расчета. Существующие методики, увы, отличаются фрагментарностью, затрагивая лишь отдельные виды информационных систем либо отдельные стадии их жизненного цикла. Отдельную и также нерешенную проблему представляет организация сбора исходных данных для этих целей.

Не менее важна кадровая проблема. Необходимы топ-менеджеры, не только придающие должное значение информатизации предприятия, но и имеющие широкий кругозор в области перспективных технологий информационных систем, способные принимать стратегические решения в процессе создания и развития корпоративных информационных систем. Также нужны менеджеры среднего звена, реализующие новые методики оценки финансового результата непосредственно в бизнес-процессах учета и планирования. Однако на сегодняшний день налицо серьезный дефицит специалистов такого уровня.

Все эти факторы придают особую важность совместному проекту Московского государственного университета и компании «АйТи», названному «ИТ-экономика». В рамках проекта создаются руководства, решающие целый ряд методических проблем, а также база для подготовки квалифицированных кадров в области управления проектами корпоративных информационных систем и их экономической оценки.

Для компании Hewlett-Packard этот проект имеет принципиальное значение. Во-первых, осознание потребителями реальных затрат и результатов, связанных с ИТ, делает рынок в целом более надежным и предсказуемым. Во-вторых, более рациональное и прозрачное принятие решений потребителями снижает риски компаний-поставщиков. Проект «ИТ-экономика» направлен на обеспечение подготовки специалистов, способных преодолеть понятийный барьер между предпринимателями, мыслящими в экономических категориях, и инженерами в области ИТ, ориентированными прежде всего на технические категории.

В заключение хочу пожелать проекту удачи. Надеюсь, что создаваемые в его рамках курсы и книги будут способствовать подготовке нового поколения специалистов в области экономики ИТ, которые, в свою очередь, придадут новое качество рынку ИТ в России.

*Горюнов Павел,
директор Хьюлетт-Паккард Консалтинг Россия и СНГ*

Эта книга подготовлена на основе конспектов лекционных курсов, которые читались автором на протяжении последних 15 лет для студентов бакалавриата и магистратуры экономического факультета МГУ. В ней используются также материалы отечественной и зарубежной периодики, ряда монографических изданий, результаты исследований и публикаций автора, труды крупных научно-технических конференций, посвященных проблематике информационных систем, спецификации стандартов информационных технологий.

Цель книги — способствовать формированию у читателя общего кругозора в области современных информационных систем, представлений о важнейших информационных технологиях, на которых они базируются, о главных достижениях и современном состоянии этой сферы, о перспективах дальнейшего развития. Предполагается, что читатель владеет начальными знаниями в рассматриваемой области.

В настоящее время наибольшее распространение получили информационные системы, основанные на технологиях баз данных, текстового поиска и Web, которые с полным правом можно назвать базовыми технологиями информационных систем. Выбор какой-либо из них для реализации конкретной системы в существенной степени определяет ее организацию и функциональные возможности.

Нужно заметить, что наряду с «чистыми» системами указанных классов создаются многочисленные крупные системы, совместно использующие различные базовые технологии. Интеграция информационных технологий стала характерной чертой разработок современных информационных систем. Поэтому для создания эффективной системы важно не только владеть отдельными направлениями базовых технологий, но и иметь представление о возможностях каждого из них, о целесообразных сферах их использования и «точках их соприкосновения».

Базовые технологии информационных систем в процессе своего развития стали довольно разветвленными, затрагивают разнообразные стороны создания и функционирования систем. Поэтому их изучение — довольно сложная задача.

Положение осложняется тем, что, к сожалению, отсутствуют какие-либо доступные издания, в которых эти технологии рассматривались бы совместно и комплексно с единых позиций. Мало выпускается литературы фундаментального характера.

В последнее время издан ряд переводов глубоких монографий по технологиям баз данных. Выпускаются книги по различным аспектам

приложений технологий Web. Однако практически отсутствуют публикации, которые бы давали читателю концептуальное представление о ключевых идеях технологий Web, их достоинствах и ограничениях, о предпосылках создания технологической платформы Web нового поколения, которая в настоящее время разрабатывается усилиями большого отряда специалистов, о ее возможностях и перспективах. Синтезировать целостную картину положения дел в технологиях Web на основе документации многих десятков стандартов, разрабатываемых консорциумом W3C, под силу не каждому.

Что касается систем текстового поиска, то основательных изданий, посвященных их проблематике, выпускалось на русском языке крайне мало. В отечественную классику, безусловно, вошла широко известная монография Михайлова А.И., Черного А.И. и Гиляровского Р.С. «Основы информатики» (1968, издательство «Наука»). Главным систематическим источником информации о технологиях текстового поиска, пожалуй, остаются переводы всемирно известных монографий Дж. Сэлтона, выпущенные еще в 70–80-е годы. В них обсуждаются результаты новаторских исследовательских проектов, выполненных под его руководством в 60-е годы, которые надолго определили пути развития систем текстового поиска. Однако издания такого уровня, отражающие текущее положение дел в этой области, к сожалению, отсутствуют. Источником информации об этом являются только труды нескольких крупных международных конференций и многочисленные публикации в периодике.

В этой сравнительно небольшой по объему работе, состоящей из четырех глав, сделана попытка обрисовать целостную картину сегодняшнего состояния и тенденции развития современных информационных систем, основные достижения и перспективы их базовых информационных технологий.

В первой главе книги представлен общий взгляд на информационные системы различных типов. Прежде всего вводится понятие об информационной системе. В контексте данной работы наиболее продуктивным представляется рассмотрение информационной системы как средства поддержки *динамической информационной модели реальности*. Моделирование реальности средствами системы и предоставление пользователям возможности избежать наблюдений и измерений состояния самой интересующей их реальности, а также происходящих в ней процессов благодаря взаимодействию с моделью — это то главное, ради чего создаются информационные системы. Хотя в системах разных типов используются различные подходы к моделированию, именно поддержка информационной модели реальности составляет главное предназначение любой

информационной системы. В этой главе достаточно подробно обсуждаются проблемы моделирования реальности в информационных системах, различные типы используемых моделей и их особенности. Описываются основные функции и причины разнообразия информационных систем, а также общие тенденции их развития, знание которых позволит читателю получить представление о передовых достижениях в рассматриваемой области и о системах будущего.

Три остальные главы книги посвящены обсуждению состояния, новых достижений, перспектив и некоторых важных областей применения базовых технологий информационных систем и свойств систем, на них основанных. При этом уделяется внимание архитектурным аспектам систем, моделированию данных, роли и месту объектных технологий, проблемам интеграции информационных ресурсов и интеграции различных технологий, а также возможностям объектных баз данных, текстового поиска в базах данных и в Web, XML-ориентированных баз данных и др. Кроме того, рассматриваются важнейшие стандарты технологий информационных систем.

Все разделы книги снабжены вопросами для самопроверки, а каждая ее глава завершается списком использованной литературы. Приложение содержит детальный глоссарий используемых в книге терминов.

Автор выражает искреннюю благодарность Лугачеву М.И., заведующему кафедрой экономической информатики экономического факультета МГУ, за его инициативу по созданию серии «ИТ-экономика», за внимание к этой работе и ее поддержку. Хотелось бы поблагодарить также компании ТНК, Hewlett-Packard и Price Waterhouse Coopers за поддержку проекта «ИТ-экономика» и издания этой книги.

ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ИХ ФУНКЦИИ

Создание современных электронных вычислительных машин позволило автоматизировать обработку данных во многих сферах человеческой деятельности. Без современных систем обработки данных трудно представить сегодня передовые производственные технологии, управление экономикой на всех ее уровнях, научные исследования, образование, издательское дело, функционирование средств массовой информации, проведение крупных спортивных состязаний. Значительно расширило сферу применения систем обработки данных появление персональных компьютеров.

Одним из наиболее распространенных классов систем обработки данных являются *информационные системы*. Хотя на уровне здравого смысла назначение таких систем понятно каждому, для серьезного обсуждения технологий современных информационных систем необходимо более четко определить, в чем заключаются их специфические особенности, чем они отличаются от других систем обработки данных, какие функции они могут выполнять, какими ресурсами они располагают.

Обсуждению указанных вопросов посвящена эта глава книги.

1.1. ЧТО ТАКОЕ ИНФОРМАЦИОННАЯ СИСТЕМА

Для обсуждения возможностей современных информационных систем, состояния и перспектив используемых в них информационных технологий необходимо прежде всего понять, что такое информационная система.

Зачем нужны информационные системы. Любой разумный вид деятельности основывается на информации о свойствах состояния и поведения той части реального мира, с которой связана эта деятельность. Для получения такой информации во многих случаях необходимо регулярно через некоторые интервалы времени проводить натурные измерения (или наблюдения), позволяющие определять характеристики состояния

сущностей реального мира и протекающих процессов, соответствующие моментам времени, когда эти измерения производятся.

В других случаях удается воспользоваться «материализованной» информацией, содержащейся в различного рода бумажных документах, отчетах или публикациях, которые также выступают как часть реальности. Требуемую информацию можно извлечь из них путем своего рода «наблюдения».

Однако некоторые натурные измерения или наблюдения могут оказаться неосуществимыми в отведенное для них время в связи с большой трудоемкостью, высокой стоимостью, недоступностью объекта измерения (наблюдения) и по другим причинам.

Значительно сократить объем необходимых натурных измерений позволяет *компьютерное моделирование реальности*. Если компьютерная модель адекватно (относительно информационных потребностей пользователей) отражает состояние и динамику реальности, то многие необходимые сведения можно получать с помощью такой модели, избегая тем самым натурных измерений, с существенно меньшими затратами времени, а возможно, и при более низкой стоимости. Именно для поддержки таких моделей служит специальный класс систем обработки данных — *автоматизированные информационные системы*. Заметим, что в ряде публикаций их называют более привычным для современного читателя термином — *компьютерные информационные системы*.

Определение понятия «информационная система». *Автоматизированной информационной системой* называется комплекс, включающий вычислительное и коммуникационное оборудование, программное обеспечение, лингвистические средства и информационные ресурсы, а также системный персонал и обеспечивающий поддержку *динамической информационной модели* некоторой части реального мира для удовлетворения информационных потребностей пользователей.

Часть реального мира, которая моделируется информационной системой, называется ее *предметной областью*.

Под *динамической моделью* здесь понимается изменяемость модели во времени. Это «живая», *действующая модель*, в которой отображаются изменения, происходящие в предметной области. Такая система должна обладать памятью, позволяющей ей сохранять не только сведения о текущем состоянии предметной области, но и в некоторых случаях предысторию.

Поскольку модель предметной области, поддерживаемая информационной системой, материализуется в форме организованных необходимым образом информационных ресурсов, она называется *информационной моделью*.

Автоматизированная информационная система не всегда функционирует самостоятельно. Она может входить в качестве компонента (подсистемы) в более сложную систему, такую, например, как система управления торговой компанией, САПР или система управления производством.

Информационные системы уже многие десятки и даже сотни лет существуют и используются на практике в форме различного рода картотек и/или коллекций бумажных документов. Однако в таких системах отсутствует какая-либо автоматизация обработки данных. Они позволяют лишь регистрировать и поддерживать в систематизированной форме на бумажных носителях результаты произведенных натуральных измерений.

Поскольку в данной работе обсуждаются только *автоматизированные* информационные системы, то есть системы, основанные на использовании средств вычислительной техники и программного обеспечения, будем далее опускать для краткости прилагательное «автоматизированная».

Приведенное выше определение охватывает информационные системы всех видов, в частности фактографические системы, которые основаны на технологиях баз данных и оперируют структурированными данными, системы текстового поиска, оперирующие документами на естественных языках, глобальную гипермедийную информационную систему Web и др. По этой причине в определении используется обобщенный термин *информационные ресурсы*. Частными его случаями являются *данные* для систем баз данных, *документы* для систем текстового поиска, *HTML-страницы* или *XML-документы* [12, 10, 17] для Web и т.д.

Нужно, однако, заметить, что на более низких уровнях представления (в памяти компьютеров, при передаче по каналам связи и т.д.) информационные ресурсы независимо от их природы и формы представления рассматриваются как хранимые или передаваемые данные. Термин «данные» часто используется по отношению к информационным ресурсам любого рода.

Отсутствие общепринятого определения. Важный факт состоит в том, что единого устоявшегося и общепринятого определения понятия «информационная система» в настоящее время не существует, да и вряд ли оно может существовать. Дело в том, что в зависимости от необходимости в разных случаях используются разные точки зрения на такой сложный продукт высоких технологий, каким являются современные информационные системы. Так, специалисты по системному проектированию трактуют понятие *информационная система* более широко, чем комплекс, о котором идет речь в нашем определении. При этом в состав информационной системы включаются, например, организационно-методические и технологические документы.

Проблемы, связанные с нечеткостью определения понятия «информационная система», вовсе не являются настолько безобидными, как это могло бы показаться. Например, в области системного проектирования и стандартов, касающихся этого вида деятельности [5, 6], вопрос о четком определении понятия информационной системы является особенно злободневным. От ответа на него зависит, в частности, что же следует считать результатом проектирования.

Приведем определение информационной системы, заимствованное в одном из наиболее авторитетных международных научных журналов в рассматриваемой области — «Information Systems», выпускаемом с 1975 года крупным английским издательством Pergamon Press. Редакционная коллегия журнала определяет информационные системы как «аппаратно-программные системы, которые поддерживают приложения с интенсивной обработкой данных (Data-Intensive Applications)». В этом определении акцентируется внимание на весьма важном, но лишь единственном аспекте информационных систем. Заметим, что *приложение информационной системы* понимается здесь как *надстройка* над информационной системой, обеспечивающая решение некоторого комплекса задач в интересах какой-либо сферы деятельности.

Большинство опубликованных определений информационной системы (см., например, [7]) трактует это понятие с функциональной точки зрения, а именно как «систему, предназначенную для сбора, передачи, обработки, хранения и выдачи информации потребителям и состоящую из следующих основных компонентов: программное обеспечение; информационное обеспечение; технические средства; обслуживающий персонал». При этом остается в стороне направленность этих функций, цель, для достижения которой они осуществляются.

В отличие от многих других публикаций, в приведенном определении делается акцент на главном назначении информационных систем, а не на их функциях и ресурсах, которые они не используют. Поддержка *динамической информационной модели предметной области* — это то общее, что свойственно любой информационной системе независимо от характера информационных ресурсов, которыми она оперирует, и, следовательно, от информационных технологий, на которых она основана.

Именно такой подход является наиболее продуктивным в данной работе, поскольку хотелось бы с единых позиций рассмотреть здесь базовые направления технологий современных информационных систем — технологии баз данных, систем текстового поиска, технологии Web.

Следствия общности определения. В силу того, что широко распространенные определения информационной системы формулируются

излишне общо или недостаточно полно, к категории информационных систем часто относят многие системы обработки данных, которые не только поддерживают информационную модель предметной области, но и позволяют решать на ее основе некоторые классы задач управленческого, исследовательского, конструкторского или иного характера. По сути дела, такая система представляет собой уже не информационную систему, а информационную систему вместе с приложением. В эту категорию попадают, например, так называемые корпоративные информационные системы, которые более естественно было бы называть системами управления корпорациями, или системы планирования ресурсов предприятия ERP (Enterprise Resources Planning Systems) [9]. Четкую границу между такими системами и информационной системой в определенном здесь смысле провести практически невозможно.

Ситуация усугубляется еще и тем, что 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.

Граница между системой базы данных и приложением. Частным случаем указанной выше терминологической проблемы является вопрос о границе между системой базы данных и ее приложением.

Традиционно *система базы данных* понимается как СУБД с управляемой ею базой данных, возможно, уже наполненной. В некоторых не очень частых случаях система базы данных бывает *самодостаточной*. Функциональные возможности пользовательских интерфейсов СУБД способны полностью удовлетворять информационные потребности пользователей.

Однако во многих случаях дело обстоит совсем не так, и необходимо создавать приложение. *Приложение системы базы данных*, в соответствии с приведенным выше определением приложения информационной системы, это надстройка над системой базы данных, представляющая собой комплекс средств прикладного программного обеспечения, который служит для решения каких-либо задач на основе этой системы. Приложение с помощью интерфейсов прикладного программирования (Application Programming Interface, API) СУБД получает доступ к базе данных и использует содержащиеся в ней данные для решения необходимых пользователям задач.

Таким образом, прикладная система, основанная на технологиях баз данных, представляет собой совокупность системы базы данных и приложения. Граница между ними четко определена — это *интерфейсы прикладного программирования СУБД*.

Но ситуация изменилась во второй половине 90-х годов, когда SQL-серверы баз данных стали обеспечивать некоторые возможности *интеграции приложения и системы базы данных* с помощью триггеров, хранимых процедур и внешних программ. Появилась, таким образом, возможность встраивать различные процедурные элементы приложения в систему базы данных. Соответствующие дополнения были приняты к стандарту языка SQL. Но в ситуации, когда приложение включает такие интегрированные компоненты, четкой границы между ним и системой базы данных уже не существует.

Ресурсы информационных систем. Информационные системы используют ресурсы нескольких категорий — средства вычислительной техники, системное и прикладное программное обеспечение, информационные, лингвистические и человеческие ресурсы. Кроме того, хотя об этом не говорится в известных определениях автоматизированных информационных систем, но подразумевается как само собой разумеющееся, для функционирования системы необходимы и другие ресурсы — помещения, их техническое оснащение, всевозможная оргтехника, электроснабжение и т.д. В этой книге они не рассматриваются, поскольку не имеют непосредственного отношения к информационным технологиям.

Информационные системы могут базироваться на различных *аппаратных платформах* — персональных компьютерах, мейнфреймах, суперкомпьютерах и других вычислительных системах. Они могут использовать отдельные компьютеры или вычислительные системы либо вычислительные сети различного масштаба — от локальной до глобальной сети. В информационных системах могут использоваться наряду с универсальными также и специализированные компьютеры, например так называемые машины баз данных, аппаратным путем реализующие некоторые функции реляционной алгебры.

Коммуникационное оборудование в информационных системах обеспечивает взаимодействие компонентов распределенных систем, например обмен данными между компьютерами сети, а также удаленный доступ пользователей к ресурсам системы. К числу коммуникационных ресурсов относятся выделенные или коммутируемые проводные и беспроводные каналы связи, различное сетевое оборудование, а также устройства приема-передачи информации, например телефонные или радиомодемы, антенные устройства.

Системное программное обеспечение включает операционные системы для используемых аппаратных платформ, различные операционные оболочки, повышающие уровень пользовательского интерфейса, системы программирования, разнообразные системные тесты, служебные программы для поддержки деятельности системного администратора и для других целей, сетевое программное обеспечение.

Информационные системы используют также разнообразное *прикладное программное обеспечение*, типовое и специализированное.

Типовое прикладное программное обеспечение ориентировано на классы задач. Оно может настраиваться на конкретный случай использования. Чаще всего в качестве средств используются коммерческие программные продукты: СУБД общего назначения, Web-серверы, системы текстового поиска (их по традиции часто называют информационно-поисковыми системами), системы управления документами, текстовые процессоры, конверторы данных, программы распознавания текста и речи, системы электронных таблиц, генераторы отчетов для систем баз данных и др.

Специализированное прикладное программное обеспечение создается для конкретной информационной системы или для класса систем, имеющих некоторое узкое назначение. Например, в корпоративной информационной системе это могут быть программы, предназначенные для поддержки каких-либо конкретных бизнес-процессов.

Прикладное программное обеспечение информационных систем может относиться к стадии разработки или к стадии исполнения. Оно может быть общего назначения или ориентированным на конкретную предметную область. Наконец, программное обеспечение может быть ориентированным на конкретную аппаратную платформу или мобильным.

Лингвистические ресурсы информационных систем служат для:

- ◆ представления информационных ресурсов в системе;
- ◆ описания их свойств и свойств окружающей среды, позволяющего системе адекватно интерпретировать поддерживаемые информационные ресурсы;
- ◆ обеспечения взаимодействия пользователей с системой.

В общем случае к числу лингвистических ресурсов относятся те или иные естественные или искусственные языки, а также *средства их лингвистической поддержки* — словари лексики естественных языков, тезаурусы предметной области (см. раздел 3.3), переводные словари и др.

Следует отметить, что тезаурусы играют в информационных системах двоякую роль. С одной стороны, это средство лингвистической поддержки используемого в системе естественного языка. Поэтому он должен

быть отнесен к категории лингвистических ресурсов. Вместе с тем тезаурус используется как контекст для интерпретации семантики поддерживаемых в системе документов, представленных на естественном языке. В связи с этим правомерно также считать тезаурус информационным ресурсом системы.

Используемый в конкретных случаях набор лингвистических ресурсов системы зависит от требований, предъявляемых к ней.

Информационные ресурсы системы составляют главный компонент модели предметной области, которую система поддерживает. Они являются вместе с тем «сырьем» и «конечным продуктом» работы информационной системы. Конкретный вид информационных ресурсов зависит от характера системы.

Важно заметить, что в любой информационной системе поддерживается две категории информационных ресурсов. Ресурсы первой категории непосредственно используются конечными пользователями системы. Ресурсы второй категории можно было бы назвать *метаресурсами*. Описывая свойства ресурсов первой категории, они позволяют системе корректно оперировать ими. Как уже отмечалось, ресурсы первой категории часто называют *данными* независимо от среды их представления (изображения, текстовые документы, аудиозаписи и т.д.), а метаресурсы — *метаданными*.

Используя эту терминологию, можно сказать, что метаданные — это *данные о данных*. Однако фактически метаданные могут описывать свойства не только собственно данных, но и информационной системы в целом, отдельных ее механизмов и их функций, других ее ресурсов, поддерживаемых технологий, пользователей и т.д. Конкретные функции метаданных и их состав в значительной мере зависят от специфики рассматриваемой системы и характера конкретных информационных ресурсов.

Уместно вспомнить, что данные в информационной системе представляют собой некоторую *абстрактную модель реальности*. Рассматривая соотношение между данными и метаданными, можно сказать, что метаданные — это данные более высокого уровня абстракции по отношению к описываемым ими данным.

В некоторых случаях метаданные сами являются предметом интереса пользователей, разработчиков системы или ее исследователей, становясь тем самым частью моделируемой реальности. Описывающие их метаданные естественно назвать *метаметаданными*. Подобного рода иерархия абстракций может иметь любое число уровней. В таких случаях может идти речь не только о моделях реальности, но и о метамоделях, метаметамоделях и т.д. Одним из примеров ситуации, где приходится иметь

Конец ознакомительного фрагмента.
Приобрести книгу можно
в интернет-магазине
«Электронный универс»
e-Univers.ru