ТЮМЕНСКАЯ НЕФТЯНАЯ КОМПАНИЯ

Когаловский М. Р.

Перспективные технологии информационных систем

Научный редактор серии Лугачев М. И.



Серия «ИТ-Экономика»

Москва

УДК 004.6
ББК 32.81
К57

Когаловский М. Р.
К57 Перспективные технологии информационных систем. — М.: ДМК Пресс. — 288 с.

ISBN 5-94074-200-9

Эта книга призвана способствовать формированию у читателя общего кругозора в области современных информационных систем, представлений о важнейших информационных технологиях, на которых они базируются, об их главных достижениях и современном состоянии, о перспективах дальнейшего развития.

В книге рассматриваются общие свойства информационных систем, обсуждаются состояние и перспективы развития технологий баз данных, текстового поиска, а также Web-технологий. Особое внимание уделяется новым архитектурным подходам, вопросам моделирования предметной области, проблемам интеграции информационных ресурсов, возможностям технологической платформы Web нового поколения, основанной на языке XML.

Для студентов, аспирантов и преподавателей вузов, слушателей учебных курсов по информационным системам, а также специалистов в области разработки и исследования информационных систем.

Все права защищены. Любая часть этой книги не может быть воспроизведена в какой бы то ни было форме и какими бы то ни было средствами без письменного разрешения владельцев авторских прав.

Материал, изложенный в данной книге, многократно проверен. Но поскольку вероятность технических ошибок все равно существует, издательство не может гарантировать абсолютную точность и правильность приводимых сведений. В связи с этим издательство не несет ответственности за возможные ошибки, связанные с использованием книги.

ISBN 5-94074-200-9

© Когаловский М. Р.
© Компания АйТи
© Оформление, издание ДМК Пресс

Содержание

Предисловие CIO 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 CIO 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

Введение

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