

Димов Э.М., Диязитдинова А.Р.,
Маслов О.Н., Новаковский В.Ф.

**НОВЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ:
ПОДГОТОВКА КАДРОВ И ОБУЧЕНИЕ ПЕРСОНАЛА**

**Часть 3. Интеллектуальные информационные
системы и управление бизнес-процессами
в инфокоммуникациях**

Издательство СамНЦ РАН
Самара
2017

УДК 681.518: 339.13
ББК 74.202

Печатается по решению Издательского совета
Самарского научного центра Российской академии наук

Димов Э.М., Диязитдинова А.Р., Маслов О.Н., Новаковский В.Ф.

Новые информационные технологии: подготовка кадров и обучение персонала. Часть 3. Интеллектуальные информационные системы и управление бизнес-процессами в инфокоммуникациях – Самара: Издательство СамНЦ РАН, 2017. – 440 с.: ил. 101.

Научное издание

ISBN

Рецензенты: Заслуженный деятель науки РФ, д.т.н., профессор Сизов А.С.
(Юго-Западный университет, г. Курск)
Заслуженный деятель науки РФ, д.т.н., профессор Червяков Н.И.
(Северо-Кавказский Федеральный университет, г. Ставрополь)

Монография посвящена вопросам разработки и применения интеллектуальных информационных систем в интересах управления сложными системами социально-экономического типа, в том числе в сфере инфокоммуникаций. Книга рассчитана на специалистов в области корпоративного управления, прикладной информатики и информационного менеджмента. Она может быть полезна в качестве учебного пособия для студентов, магистрантов и аспирантов соответствующих специальностей.

Табл. 32. Ил. 101. Библ. 118.

Научное издание

Димов Эдуард Михайлович, Диязитдинова Альфия Радмировна,
Маслов Олег Николаевич, Новаковский Владимир Феликсович.

ISBN

© Авторы 2017

СОДЕРЖАНИЕ

Стр.

СПИСОК ПРИНЯТЫХ СОКРАЩЕНИЙ	5
ВВЕДЕНИЕ	6
1. РОЛЬ ИНТЕЛЛЕКТУАЛЬНЫХ ИНФОРМАЦИОННЫХ СИСТЕМ В ОБЕСПЕЧЕНИИ ПРОЦЕССА ПРИНЯТИЯ УПРАВЛЕНЧЕСКИХ РЕШЕНИЙ	
1.1. Технологические принципы разработки и реализации управленческих решений	14
1.2. Применение интеллектуальных ИС в управлении	27
1.3. Автоматизация бизнеса в отрасли инфокоммуникаций	40
1.4. Эволюция интеллектуальных ИС	48
1.5. Отличительные свойства ИИС	55
1.6. Классификация ИИС	62
1.7. Выводы	81
2. ВВЕДЕНИЕ В ТЕОРИЮ ЭКСПЕРТНЫХ СИСТЕМ	
2.1. История развития и архитектура ЭС	83
2.2. Экспертные системы как продукт эволюции ИС	100
2.3. Представление и преобразование знаний в ЭС	125
2.4. Онтология предметной области	132
2.5. Методы извлечения знаний	137
2.6. Проблема неопределенности знаний	157
2.7. Выводы	169
3. ПРИНЦИПЫ РАЗРАБОТКИ И ПРОЕКТИРОВАНИЯ ЭКСПЕРТНЫХ СИСТЕМ	
3.1. Этапы разработки ЭС	172
3.2. Предпроектная стадия	180
3.3. Формирование требований пользователя к ЭС	196
3.4. Стадия проектирования ЭС	212
3.5. Разработка структуры БЗ и механизмов рассуждения	224
3.6. Методы и стратегии перебора	230
3.7. Методы поиска решений в пространстве состояний	235
3.8. Критерии оценки функционирования ЭС	251

3.9. Стадия разработки ЭС	254
3.10. Стадии внедрения и опытной эксплуатации	258
3.11. Выводы	262

4. ИНТЕЛЛЕКТУАЛЬНАЯ ПОДДЕРЖКА УПРАВЛЕНИЯ БИЗНЕС-ПРОЦЕССАМИ

4.1. Квазиоптимальное управление иерархическими СС с позиций классической теории управления	263
4.2. Особенности управления знаниями	278
4.3. Принципы принятия решений в условиях неопределенности знаний	286
4.4. Снижение неопределенности выбора управленческих решений с помощью СИМ по МДМ	291
4.5. Выбор и обоснование метода проектирования БЗ	303
4.6. Информационные технологии организации удаленного доступа и выбор структуры распределенной БД	313
4.7. Принципы построения расчетно-диагностических ЭС	333
4.8. Выводы	336

5. ПОСТАНОВКА И ПРИМЕРЫ РЕШЕНИЯ ЗАДАЧ РЕАЛИЗАЦИИ ЭКСПЕРТНЫХ И СОВЕТУЮЩИХ СИСТЕМ

5.1. Предметная область ЭС диагностики неисправностей оборудования ИКК	338
5.2. Предпроектная стадия создания ЭС диагностики неисправностей оборудования ИКК	342
5.3. Стадия проектирования ЭС диагностики неисправностей оборудования ИКК	345
5.4. Описание БЗ на выбранном языке представления знаний ...	365
5.5. Стадия разработки ЭС диагностики неисправностей оборудования ИКК	371
5.6. Особенности предметной области экономической ЭС расчетно-диагностического типа	375
5.7. Технологические принципы создания экономической ЭС расчетно-диагностического типа	378
5.8. Использование функций полезности при управлении выводом в ЭС расчетно-диагностического типа	391
5.9. Реализация ЭС для управления региональной КСС	394

5.10. Принципы эксплуатации и пример работы ЭС расчетно-диагностического типа	413
5.11. Выводы	418
ЗАКЛЮЧЕНИЕ	420
СПИСОК ЛИТЕРАТУРЫ	430

СПИСОК ПРИНЯТЫХ СОКРАЩЕНИЙ

АСУ – автоматизированная система управления
АТС – автоматическая телефонная станция
БД – база данных
БП – бизнес-процесс
БЗ – база знаний
ИВЗ – инновационные знания
ИИ – искусственный интеллект
ИИС – интеллектуальная информационная система
ИКК – инфокоммуникационная компания
ИКС – инфокоммуникационная система
ИЛМ – инфологическая модель
ИС – информационная система
ИТ – информационные технологии
ИУ – интеллектуальное управление
КОВ – коэффициент относительной важности
КСС – компания сотовой связи
ЛПР – лицо, принимающее решение
МАС – мультиагентная система
МДМ – метод Димова-Маслова
ММ – математическая модель
ММК – метод Монте-Карло
ОМС – онтологическая модель ситуации
ОУ – объект управления
СИМ – статистическое имитационное моделирование
СОД – система обработки данных
СППР – система поддержки принятия решений
СУ – система управления
ТЗ – техническое задание
ФЭА – финансово-экономический анализ
ЭС – экспертная система

ВВЕДЕНИЕ

Информация является ключевым ресурсом, необходимым для эффективной деятельности любой современной корпорации – предприятия, компании, учреждения фирмы. Контроль производственной деятельности, анализ состояния внешней среды, принятие управленческих решений, реинжиниринг бизнеса – все эти функции сводятся, так или иначе, к работе с информацией. От качества этой информации: оперативности, достоверности, потенциальной пользы и т.д., во многом и самым непосредственным образом зависит успех действий коллектива корпорации [51].

Наша страна до сих пор имеет огромный неиспользуемый потенциал для компьютеризации и информатизации общества, интеграции в мировое информационное пространство. В то же время Россия обладает необходимым сочетанием всех благоприятных факторов для развития услуг в области разработки, проектирования и реализации *информационных систем* (ИС). Суть рыночных изменений, охвативших все сферы деятельности современного человека, заключается в том, что материальная составляющая в структуре жизненных благ уступает место информационной – поэтому никого не надо убеждать в том, что любой успешный бизнес немислим сегодня без применения *информационных технологий* (ИТ), методы и средства реализации которых представляют уникальный инструмент для повышения уровня эффективности решения задач управления корпорацией.

На протяжении своей многовековой истории человечество накапливало знания, полученные путем наблюдений и опыта, старалось по-возможности распространить, сохранить и передать их последующим поколениям – делая как верные, так и неверные выводы из этих знаний, пытаясь применить их на практике. В этой связи телекоммуникационный рынок всегда представлял собой благодатное поле для развития и внедрения информационных систем – однако в настоящее время бизнес по созданию мощных ИС вышел из младенческого возраста и приобрел черты конкурентного рынка. Из всех секторов экономики именно в сфере инфокоммуникаций ИС представляют собой основу жизнеобеспечения отрасли, без которой немислим ее коммерческий успех. Если другие отрасли, такие как транспорт, торговля, органы государственного управ-

ления, по уровню внедрения ИТ подчас отстают у нас от мировых стандартов на годы и десятилетия, то конкурентный и динамичный рынок инфокоммуникаций просто вынудил его участников направить значительные ресурсы, как финансовые, так и интеллектуальные, на разработку и внедрение самых современных ИС.

Поскольку для *инфокоммуникационных компаний* (ИКК) информация представляет собой основу бизнеса, игнорирование или даже невнимание к необходимости активно использовать ИТ зачастую ведет к потере жизнеспособности ИКК. Отнюдь не случайно все более новые и перспективные возможности, открывающиеся для создания и использования телекоммуникационных средств, настолько привлекают к себе внимание участников рынка, что специалисты в последние годы ведут речь о глобальном явлении, именуемом *инфокоммуникационной революцией*. Передача и прием информационных сигналов в любое нужное время на любые нужные расстояния перестали быть только технологическими проблемами, поскольку сам предмет передачи – информация – приобрел новые качественные характеристики и стал непосредственным объектом любого экономического процесса.

Телекоммуникационные системы и сети, принадлежащие ИКК, являются полноценными «участниками» жизни общества, неотъемлемой частью постоянного информационного обмена – без которого с невозможно представить ни международное сотрудничество государств, ни повседневный быт современного человека. Эти ИТ формируют фундаментальную основу ИС, имеющих дело с организацией и обработкой гигантских массивов данных в компьютеризированных корпоративных системах; обеспечивающих ИТ-поддержку решений менеджеров, экспертов и других лиц, принимающих решения (ЛПР); позволяющих объективно оценить уровень развития бизнеса корпораций, выявить резервы и обеспечить успех их дальнейшей деятельности путем применения правомерных и взвешенных решений.

Поддержка принятия решений на базе использования ИС означает обеспечение ЛПР необходимыми первичными данными, а также информацией и знаниями, облегчающими принятие ими эффективных решений. В структуре поддержки решений при этом, согласно [80], выделяют три основные составляющие: *информационную* (для предоставления пользователю необходимых первич-

ных данных), *модельную* (для обеспечения сведениями о взаимосвязях подсистем и элементов в составе исследуемой системы и возможном поведении ее в будущем), а также *экспертную* (с целью обеспечения пользователя правилами и знаниями формирования дедуктивного вывода и экспертного анализа человеческих эмпирик при выборе наиболее эффективных и практически приемлемых вариантов решения поставленной задачи).

В соответствии с классификацией проблем, предложенной Г. Саймоном, задачи, описанные количественными и качественными признаками с преобладанием последних, относятся к слабоструктурированным проблемам. Научный подход к решению таких проблем базируется на методах *искусственного интеллекта* (ИИ) – при этом, прежде всего, речь идет о развитии аналитических ИС, которые в значительной мере «разгрузили» бы ЛПР на этапе аналитической обработки больших объемов данных и были бы способны давать советы и рекомендации по наблюдаемой текущей ситуации. Сегодня пользователю необходим полный пакет программного обеспечения, адаптированного под конкретную задачу и (или) конкретную отрасль – чтобы с его помощью эффективно управлять информационными потоками и оперативно принимать необходимые решения, что, на языке бизнеса, позволяет «обеспечить конкурентное преимущество» корпорации. Интеллектуальная ИС поддержки управленческих решений, включающая в себя банк бизнес-процессов корпорации, подсистему функционального и структурного формирования новых бизнес-процессов, а также подсистему генерации управленческих решений, позволяет накапливать и применять опыт лучших экспертов-менеджеров для дальнейшего эффективного управления экономическими объектами.

Хотя создать универсального электронного эксперта, способного заменить человека, до настоящего времени не удалось и вряд ли удастся создать в ближайшем будущем, существующие наработки позволяют создавать таких экспертов в узких областях, частично освобождая рядовых ЛПР – менеджеров, управленцев и экспертов – от утомительной, однообразной и рутинной аналитической работы. Основная сложность этого безусловно прогрессивного процесса заключается в том, что, по словам чемпиона мира по шахматам и выдающегося специалиста по компьютерным технологиям

М. Ботвинника, человек с точки зрения математики *алогичен и неформализуем*, поэтому досконально «просчитать» на компьютере его деятельность до сих пор невозможно. Цель настоящей книги – представить научное направление, занимающееся использованием компьютерных ИС для целей поддержки управленческих решений, которые получили название «экспертные системы» (ЭС).

В рыночной экономике XXI века, которую специалисты-футурологи именуют постиндустриальной, инновационной, сетевой, кластерной, информационной и т.д., а, в конечном счете, *экономикой знаний*, актуальным является вопрос о том, какими по виду могут быть эти знания и какую разумную (справедливую) цену ЛПР имеет смысл платить за разные виды знаний. Предваряя подробный ответ на поставленный вопрос, отметим, что наибольший интерес для бизнеса представляют *инновационные знания* (ИВЗ) – даже с учетом риска внедрения тех или иных новшеств (инновационного риска), стоимость которых для их обладателя (соответственно, цена для пользователя) является максимально высокой.

Инструментарий для проведения исследований в данной области, в том числе путем анализа и прогнозирования инновационного риска, дают постнеоклассические методы изучения организационно-технических (социально-экономических, экологических, военных и т.п.) сложных систем (СС), способных к саморазвитию, саморегулированию и самоорганизации. Неотъемлемыми элементами таких СС являются ЛПР: руководители, менеджеры и эксперты (по терминологии [18] акторы холонических СС), от правильности действий которых напрямую зависит эффективность их функционирования.

Понимая под ЛПР индивидуума, размышляющего о своем жизненном назначении, мы приходим к постановке той же задачи и в некоем обобщенном виде – не претендуя на вклад в ее решение, отметим, что любая попытка анализа и моделирования саморазвития холонических СС имеет ярко выраженный гуманистический смысл – поскольку позволяет разделить истинные и ложные ценности, сформулировать (формализовать в первом приближении) их непреходящую суть. Актуальность и значимость этой проблемы для экономики знаний, на наш взгляд, сомнений не вызывает.

Выделим два наиболее важных аспекта этой проблемы. Первый аспект обусловлен информационным обменом между ЛПР и имеет важное консолидирующее (конъюнктурное, оперативно-тактическое) значение. Стратегический процесс наследования традиций в науке также связан с накоплением, архивированием и передачей научных знаний – предметом обмена при этом могут быть как *явные верифицированные* (объективные, многократно доказанные и проверенные) знания, так и *неявные аксиологические* знания (предзнания в виде убеждений, личного опыта и недостоверных гипотез). Формы обмена могут быть самыми разными: от издания научных статей и монографий до размещения материалов в Internet.

Второй аспект проблемы продиктован развитием отечественной инновационной науки, целью которой является технический прогресс, связанный с активным внедрением ИВЗ в производственно-технологическую сферу. Точность и достоверность верифицированных знаний гарантируются надежностью научных методов и средств, использованных для их получения – их рыночная цена, однако, не может быть особо большой, поскольку определяется стоимостью обучения ЛПР, которые их приобретают. Еще меньшей стоимостью, как правило, обладают аксиологические знания в виде гипотез, суждений и предположений о свойствах и характеристиках рассматриваемых СС, обусловленных опытом предыдущих разработок, взглядами и намерениями ЛПР. Эти знания именуется также интуитивными, эвристическими, феноменологическими, иррациональными знаниями. Они базируются на субъективном понимании, а не объяснении принципов работы и моделирования, состава и структуры СС. Отметим, что, как и верифицированные знания, они способны обладать полнотой, логической связанностью и взаимной непротиворечивостью.

Граница между знаниями разного вида представляется не запретной чертой, но достаточно эластичным контуром, проходить через который наука сегодня просто обязана и постоянно это делает – поскольку питать ее в условиях рынка, кроме государственного бюджета, просто некому, *а временную переходную область* образуют как раз ИВЗ с максимальной рыночной стоимостью. Задача оперативно зафиксировать и сохранить эти ИВЗ в виде, доступном для всех нуждающихся пользователей (для ЛПР своей корпорации

бесплатно, для других ЛПР – за взаимно приемлемую плату) означает найти исключительно выгодную рыночную нишу в области практического применения ИТ, связанную с применением систем поддержки принимаемых решений (советующих ИС, экспертных систем).

Экспертные системы представляют собой достаточно хорошо известное «коммерческое» приложение систем искусственного интеллекта – о них издано сегодня такое количество научной и научно-популярной литературы, что невольно возникает ощущение, что существует целая теория создания современных ЭС. Однако это совсем не так, и при детальном изучении проблемы в учебных целях (авторы монографии, как это видно из ее названия, активно занимаются подготовкой новых кадров и обучением ЛПР) становится ясно, что большая часть исследований в области ЭС решает частные задачи без применения какой-либо теории, построенной вне области их конкретного применения. Поэтому одним из способов разобраться в указанной проблеме является попытка систематизировать и обобщить имеющиеся наработки в области ЭС, что, собственно говоря, и реализует поставленную авторами данной книги цель. Представляют интерес и практические примеры применения ЭС совместно с другими ИС, которые изучаются в высшей школе, а также анализ эволюционного развития интеллектуальных информационных систем (ИИС) в направлении теории и практики управления бизнесом.

В соответствии с поставленной целью, **первая глава** монографии посвящена роли и месту интеллектуальных информационных систем в процессе управления: здесь представлены технологии разработки и реализации управленческих решений, показаны перспективы применения ИС в практике управления – на примере телекоммуникационной отрасли, с учетом автоматизации данной сферы информационного бизнеса. Рассмотрены вопросы эволюционного развития ИИС – как с точки зрения повышения эффективности их характеристик, так и в системном плане – в рамках современной классификации иерархических СС.

Вторая глава представляет собой введение в теорию ЭС – в объеме, необходимом для первого ознакомления с ЭС, однако едва ли достаточном для дальнейшей профессиональной работы в данной области. Кратко освещена история развития ЭС; рассмотрены

их специфика и типовая архитектура; подробнее изложены вопросы, связанные с применением в интересах разработки, реализации и эффективной эксплуатации ЭС теории знаний: особенности процессов создания и преобразования знаний; классификация знаний; онтология предметной области; методы извлечения знаний и проблемы, связанные с этим важнейшим для практики создания ЭС интеллектуальным процессом.

Третья глава содержит принципы разработки и проектирования ЭС, она достаточно подробно и последовательно представляет ключевые этапы данного технологического процесса: предпроектную стадию, которая включает обследование объекта, обоснование необходимости создания и формирование требований пользователя к ЭС; стадию проектирования – разработку архитектуры, выбор метода формализации знаний в ЭС, логическую и фреймвые модели представления знаний, выбор формализма для представления знаний, разработку структуры базы знаний (БЗ) и механизмов рассуждения, управление выводом, метод и стратегия перебора, критерии оценки функционирования ЭС, оформление рабочего проекта; стадию разработки ЭС, включающую тестирование и подготовку документации; а также стадию внедрения ЭС – вплоть до опытной эксплуатации и обучения пользователей.

Четвертая глава освещает широкий круг вопросов, связанных с теорией управления бизнес-процессами при использовании ЭС: от квазиоптимального управления иерархическими СС в рамках классического подхода и введения в проблему управления знаниями до принципов построения типовых ЭС различного назначения. Здесь обоснована возможность снижения неопределенности выбора управленческих решений с помощью статистического имитационного моделирования (СИМ) по версии, условно именуемой методом Димова-Маслова (МДМ); показаны роль и место СИМ по МДМ в концепции квазиоптимального управления нерелефторными СС; рассмотрены проблема обоснованного выбора метода проектирования БЗ и особенности применения ИТ для организации удаленного доступа, включая выбор структуры распределенной базы данных (БД); изложены принципы построения ЭС расчетно-диагностического типа.

В пятой главе приведены примеры решения задач, связанных с реализацией ряда конкретных ЭС: показаны предметная область

для ЭС диагностики неисправностей оборудования ИКК и ключевые стадии ее создания: предпроектная стадия, стадии проектирования и разработки; аналогичным образом представлены предметная область и технологические особенности создания ЭС расчетно-диагностического типа, ориентированной на применение в социально-экономической сфере и отрасли инфокоммуникаций. Рассмотрены вопросы использования функций полезности при управлении выводом в ЭС расчетно-диагностического типа; подробно описан вариант реализации ЭС для управления региональной компании сотовой связи (КСС); изложены принципы эксплуатации и описание процесса функционирования расчетно-диагностической ЭС.

Каждый раздел содержит краткие выводы, которые в обобщенном виде приводятся в виде заключения. Список литературы содержит все источники, на которые имеются ссылки в тексте.

Авторы выражают глубокую признательность рецензентам: Заслуженному деятелю науки РФ, д.т.н., профессору Сизову А.С. (Юго-Западный университет, г. Курск) и Заслуженному деятелю науки РФ, д.т.н., профессору Червякову Н.И. (Северо-Кавказский Федеральный университет, г. Ставрополь). Также они считают приятным долгом поблагодарить сотрудников кафедры Прикладной информатики (до 2017 г. – кафедра Экономических информационных систем) Поволжского государственного университета телекоммуникаций и информатики, которые, как это видно из совместных публикаций, оказывали им постоянную помощь в проведении исследований.

1. РОЛЬ ИНТЕЛЛЕКТУАЛЬНЫХ ИНФОРМАЦИОННЫХ СИСТЕМ В ОБЕСПЕЧЕНИИ ПРОЦЕССА ПРИНЯТИЯ УПРАВЛЕНЧЕСКИХ РЕШЕНИЙ

1.1 Технологические принципы разработки и реализации управленческих решений

Термин *проблема* соответствует одному из ключевых и наиболее актуальных понятий теории СС – что вполне понятно, поскольку вся наша жизнь «пронизана» проблемами и непосредственным образом связана с необходимостью их правильного решения. Проблем вокруг так много и решения по ним принимаются так часто, что делается это во многих случаях подсознательно – человек четко не осознает, по какой причине он предпринимает те или иные конкретные управленческие действия. Хотя в сложных ситуациях, разумеется, наиболее важные и трудные решения всегда выделяются и становятся предметом предварительного обстоятельного анализа (индивидуального или коллективного) со стороны ЛПР.

Не обходится при этом без курьезов, символом одного из которых является «буриданов осел», помещенный на равном расстоянии от двух одинаковых связок сена, который обречен на голодную смерть, так как не в силах теоретически правильно выбрать решение, то есть предпочесть ту или иную связку. Примечательно, что в сочинениях французского философа XIV века Жана Буридана, по имени которого явление получило свое название, этот образ так и не был найден – по-видимому, данный парадокс абсолютного детерминизма в ситуации полной информационной определенности компрометировал средневековую науку настолько, что пришлось закрыть на него глаза.

Хотя, если речь идет об ЛПР, колеблющемся при выборе между двумя равнозначными возможностями, то в этом смысле «буриданов осел» модель вполне современная – достаточно напомнить, например, о региональном политикане, выбирающем столичного покровителя, или о диссертанте, решающем, к какой из взаимно враждебных научных школ ему примкнуть. Почти каждый из нас оказывался в подобном положении, не зная, какую из возможных альтернатив предпочесть. Поэтому, как ни полемично прозвучит подобное утверждение, но самая распространенная и трудная проблема, с которой сталкивается любой человек, – и в повседневной

жизни, и в своей профессиональной сфере, – это проблема выбора одного (желательно наилучшего) варианта из нескольких потенциально возможных решений.

В [32] представлен следующий, достаточно наглядный и убедительный пример: некий добросовестный ЛПР опаздывает на работу и видит, что к остановке, которая находится на противоположной стороне улицы, подходит автобус, тогда как на светофоре горит красный свет. В его распоряжении несколько вариантов действий – пересечь улицу на красный свет с риском для жизни, поехать на следующем автобусе или поймать такси. Конечно, с научной точки зрения, для принятия правильного решения ему надо точно знать, когда появится следующий автобус, хватит ли у него денег на такси, а также объективно соотнести потери от двух неприятных возможностей: опоздать на работу и неудачно перейти улицу.

Но такой информации у него нет, а время не терпит – поэтому ЛПР принимает решение на свой страх и риск, а затем также оперативно и самостоятельно претворяет это решение в жизнь. В более сложных и разноплановых ситуациях (хотя и рассмотренный случай, строго говоря, нельзя считать элементарным событием для его главного фигуранта), так называемые управляющие решения могут иметь значительно более важные научно-технологические, социально-экономические и другие последствия – особенно если выбор и реализация принятых единоличных или коллективных решений возлагаются на управляющие элементы ресурсоемкой и дорогостоящей СС.

Если ситуация, требующая принятия решения, является типовой в том смысле, что повторяется достаточно часто, решение к ЛПР приходит как бы само собой, автоматически, поскольку основано на имеющемся предыдущем опыте. Если же ситуация недостаточно хорошо знакома или ЛПР вообще не располагает необходимыми сведениями о ней, то принятие решения существенно усложняется: как правило, в таких случаях человек вынужден сравнивать между собой возможные варианты и выбирать тот, который кажется ему наиболее предпочтительным или наименее опасным.

Для подавляющего большинства принимаемых решений, однако, нельзя рассчитать и достоверно оценить все их последствия: поэтому можно лишь предполагать, что среди возможных вариантов будет тот, который приведет ЛПР к наилучшему результату. В

целом решения людей являются исключительно важным для практики и интересным для науки объектом исследования: уступая ЭВМ в скорости и точности вычислений, человек способен замечательно быстро оценивать обстановку, выделять главное в ней и отбрасывать все второстепенное, соизмерять противоречивые оценки и восполнять неопределенности знаний своими гипотезами, догадками и предположениями [32]. Проблеме принятия решений посвящены многочисленные статьи, обзоры литературы, монографии и другие публикации, однако ясности в данном вопросе и однозначных указаний по поводу универсальных или, напротив, объектно-ориентированных, объективных или субъективных, коллективных или индивидуальных научных методов, в настоящее время, к сожалению, нет.

Таблица 1.1. Дисциплинарные корни науки о принятии решений [80]

Система	Описательные дисциплины	Аналитические дисциплины
Индивидуум	Психология Психиатрия Литература	Исследование операций Экономика Логика
Группа	Социальная психология Организационное поведение Антропология	Теория игр Финансы Клиническая терапия
Организация	Теория фирмы Социология Политология	Стратегическое планирование Производственный менеджмент Организационное планирование
Государство	Социология Антропология Макроэкономика	Философия права Политология

Проблемы принятия решений и выбора интенсивно разрабатываются во многих областях знания. В таблице 1.1 приведены научные дисциплины, разрабатывающие различные аспекты проблемы принятия решений. Представленные дисциплины можно разделить на две категории: описательные и аналитические. Описательные дисциплины сообщают, как в действительности ЛПР принимают решения, аналитические дисциплины – как такие решения

принимать, руководствуясь некоторыми выбранными априори критериями.

На практике бывает достаточно сложно отнести ту или иную науку к разряду сугубо аналитических или полностью описательных дисциплин. Психология, например, первоначально изучавшая информационные способности человека и его устойчивые предупреждения – в основном путем проведения лабораторных экспериментов – долгое время могла рассматриваться как описательная дисциплина. Когда же, основываясь на понимании полученных результатов, психологи стали разрабатывать способы корректировки этих предубеждений – их методы стали относиться к аналитической сфере. Другим характерным примером является современная экономика, где большое внимание всегда уделялось и уделяется до сих пор разработке и применению моделей выбора – определяющих, каким образом ЛПР должны поступать в том или ином конкретном случае, что соответствует аналитическому аспекту проблемы. В то же время сегодня все больше внимания здесь уделяется учету неопределенности (недостаточности, неточности, неадекватности и т.п.) информации, существенно влияющей на конечный выбор индивидуума, что характеризует описательный аспект проблемы.

Таким образом, науку о принятии решений можно считать дисциплиной, интегрирующей описательный метод исследования с разработкой конкретных аналитических рекомендаций. Описательный метод дает возможность понять взаимосвязь наблюдаемых явлений и процессов, а также определить специфику типовых действий, предпринимаемых ЛПР. Под аналитическими рекомендациями при этом понимают методы, разрабатываемые для улучшения указанных типовых решений.

Содержание задачи принятия решения позволяет сформулировать некоторые принципиальные моменты, характеризующие особенности принятия управленческих решений [34].

1. Неизвестные элементы задачи (ситуации, цели, ограничения, предпочтения, решения) имеют, прежде всего, содержательный характер и только частично определяются количественными характеристиками, при этом число неизвестных элементов задачи часто существенно больше, чем известных.

2. Все элементы задачи описываются характеристиками, часть которых может быть измерена объективно, а другая – только на основе субъективных оценок (приоритеты целей, субъективные вероятности событий и т.д.).
3. Решение часто принимается в условиях неопределенности, обусловленной неполнотой информации о проблемной ситуации в СС и невозможностью достаточно точно оценить ожидаемые последствия.
4. Принимаемые решения могут непосредственно затрагивать интересы ЛПР и экспертов, поэтому мотивы их поведения (лояльность), уровень притязаний и т.п. могут оказывать существенное влияние на результат их работы.

В итоге считается, что процесс нахождения наилучшего решения не может быть полностью формализован, однако вполне реально уменьшение присущей ему неопределенности через последовательность этапов, которые описывают так называемый **нормативный процесс принятия решений**, то есть некоторую стандартную и неизменную последовательность шагов, которые следует выполнить ЛПР, чтобы принять рациональное решение. Процесс принятия управленческих решений охватывает три основных фазы: подготовка, принятие и реализация, каждая из которых представляет собой самостоятельный сложный процесс, который, в свою очередь, состоит из ряда поэтапных действий.

В настоящее время известно достаточно много различных подходов к выделению этапов принятия управленческих решений. Часть из них иллюстрирует таблица 1.2, где представлены этапы основных методик системной последовательности принятия решений, разработанные различными исследователями. Из анализа таблицы 1.2 видно, что этапы принятия решений различаются в деталях, но в целом дают одно и то же и, по-видимому, верное описание логики развития этого процесса, начиная с формулировки проблемы и постановки целей и заканчивая контролем результатов реализации принятого решения.

Системный подход к решению проблем можно представить последовательностью следующих шагов (см. рис. 1.1):

- определение проблемы и постановка цели принятия решения;
- анализ факторов, влияющих на принятие решения;
- разработка и оценивание альтернатив;

- выбор альтернативы, признанной наилучшей;
- реализация принятого решения и контроль его результатов.

Поиск решений является многошаговым процессом, в ходе которого поставленные частные цели постоянно сравниваются с шансами на их достижение, отыскиваются новые возможности и резервы, а сами цели корректируются: в сторону их усложнения.

Таблица 1.2. Этапы ряда современных методик системной последовательности реализации процесса принятия решений

Автор	Этапы
Оптнер С.Л.	<ol style="list-style-type: none"> 1. Идентификация симптомов. 2. Определение актуальности проблемы. 3. Определение целей. 4. Определение структуры системы и ее дефектов. 5. Определение возможностей. 6. Нахождение альтернатив. 7. Оценка альтернатив. 8. Выработка решения. 9. Признание решения. 10. Запуск процесса решения. 11. Управление процессом реализации решения. 12. Оценка реализации и ее последствий.
Янг С.	<ol style="list-style-type: none"> 1. Определение цели. 2. Выявление проблемы. 3. Диагноз. 4. Поиск решения. 5. Оценка и выбор альтернативы. 6. Согласование решения. 7. Утверждение решения. 8. Подготовка к вводу в действие. 9. Управление применением решения. 10. Проверка эффективности.

Таблица 1.2 (продолжение)

Автор	Этапы
Янг С.	<ol style="list-style-type: none"> 1. Определение цели. 2. Выявление проблемы. 3. Диагноз. 4. Поиск решения. 5. Оценка и выбор альтернативы. 6. Согласование решения. 7. Утверждение решения. 8. Подготовка к вводу в действие. 9. Управление применением решения. 10. Проверка эффективности.
Федоренко Н.П.	<ol style="list-style-type: none"> 1. Формулирование проблемы. 2. Определение целей. 3. Сбор информации. 4. Разработка максимального количества альтернатив. 5. Отбор альтернатив. 6. Построение модели в виде уравнений, программ или сценария. 7. Оценка затрат. 8. Испытание чувствительности решения (параметрическое исследование).
Никаноров С.П.	<ol style="list-style-type: none"> 1. Обнаружение проблемы. 2. Оценка актуальности проблемы. 3. Анализ ограничений проблемы. 4. Определение критериев. 5. Анализ существующей системы. 6. Поиск возможностей (альтернатив). 7. Выбор альтернативы. 8. Обеспечение признания. 9. Принятие решения. 10. Реализация решения. 11. Определение результатов решения.

Таблица 1.2 (окончание)

Автор	Этапы
Черняк Ю.И.	<ol style="list-style-type: none"> 1. Анализ проблемы. 2. Определение системы. 3. Анализ структуры системы. 4. Формулирование общей цели и критерия. 5. Декомпозиция цели, выявление потребности в ресурсах, композиция целей. 6. Выявление ресурсов, композиция целей. 7. Прогноз и анализ будущих условий. 8. Оценка целей и средств. 9. Отбор вариантов. 10. Диагноз существующей системы. 11. Построение комплексной программы развития. 12. Проектирование организации для достижения целей.
Максимов В.И.	<ol style="list-style-type: none"> 1. Формулирование проблемы. 2. Определение вектора целей. 3. Задание базисных факторов. 4. Выбор вектора управляющих воздействий. 5. Выбор наблюдаемых факторов. 6. Построение структурно-целевой модели системы. 7. Структурно-целевой анализ (определение непротиворечивости вектора целей, согласованности вектора управлений вектору целей, эффективности управлений). 8. Решение прямой задачи (моделирование при заданном векторе управления). 9. Решение обратной задачи (поиск векторов управлений для достижимости вектора целей).

Это происходит, если выявлены дополнительные ресурсы или другие благоприятные обстоятельства, или в сторону упрощения, если первоначально сформулированные цели оказались нереальными. Следовательно, процесс решения проблемы самым тесным образом связан с анализом условий и целей деятельности ЛПР.

Кроме того, на него во многом влияет тип решения: в теории СУ выделяют индивидуальные решения ЛПР и решения, принимаемые коллективом. Человек, приступая к нахождению и **индивидуальному решению** проблемы, обычно преследует определенную цель – связанную с желанием удовлетворить те или иные биологические, социальные и другие потребности. Надо также учитывать, что на принимаемые им решения влияют такие субъективные качества как отношение к существованию проблемы, сбору и использованию сведений о ней), а также его ограниченные информационные возможности – включая способность воспринимать, помнить и обрабатывать данные.

Многочисленные эксперименты [68] демонстрируют частое отклонение реального поведения ЛПР от наиболее выгодных им рациональных вариантов и определяют **эвристики** (от греч. *heurisko* – нахожу, в данном случае – совокупности логических приемов и методических правил), которые используются на практике при принятии решений.

К наиболее известным относятся следующие эвристики:

- **суждение по представительности** – ЛПР часто судят о вероятности того, что объект А принадлежит к классу Б только на основании внешней схожести А на объекты класса Б, игнорируя априорные вероятности, влияющие на данную принадлежность;
- **суждение по встречаемости** – люди склонны определять вероятности событий по тому, как часто они сами сталкивались с подобными событиями и насколько важными для них лично были впоследствии эти встречи;
- **суждение по точке отсчета** – если при определении вероятности того или иного события ЛПР используется имеющаяся начальная информация о нем, то она, как своеобразная точка отсчета, существенно влияет на последующий результат;
- **сверхдоверие** – экспериментальным путем доказано, что большинство людей чрезмерно доверяют своим субъективным суждениям, особенно в тех случаях, когда они были непосредственными участниками тех или иных событий в прошлом;
- **стремление к исключению риска** – люди склонны исключать альтернативы, связанные с риском и соглашаются на худшие из них лишь потому, что могут возникнуть ситуации, когда при самых малых вероятностях возможны большие потери для них.

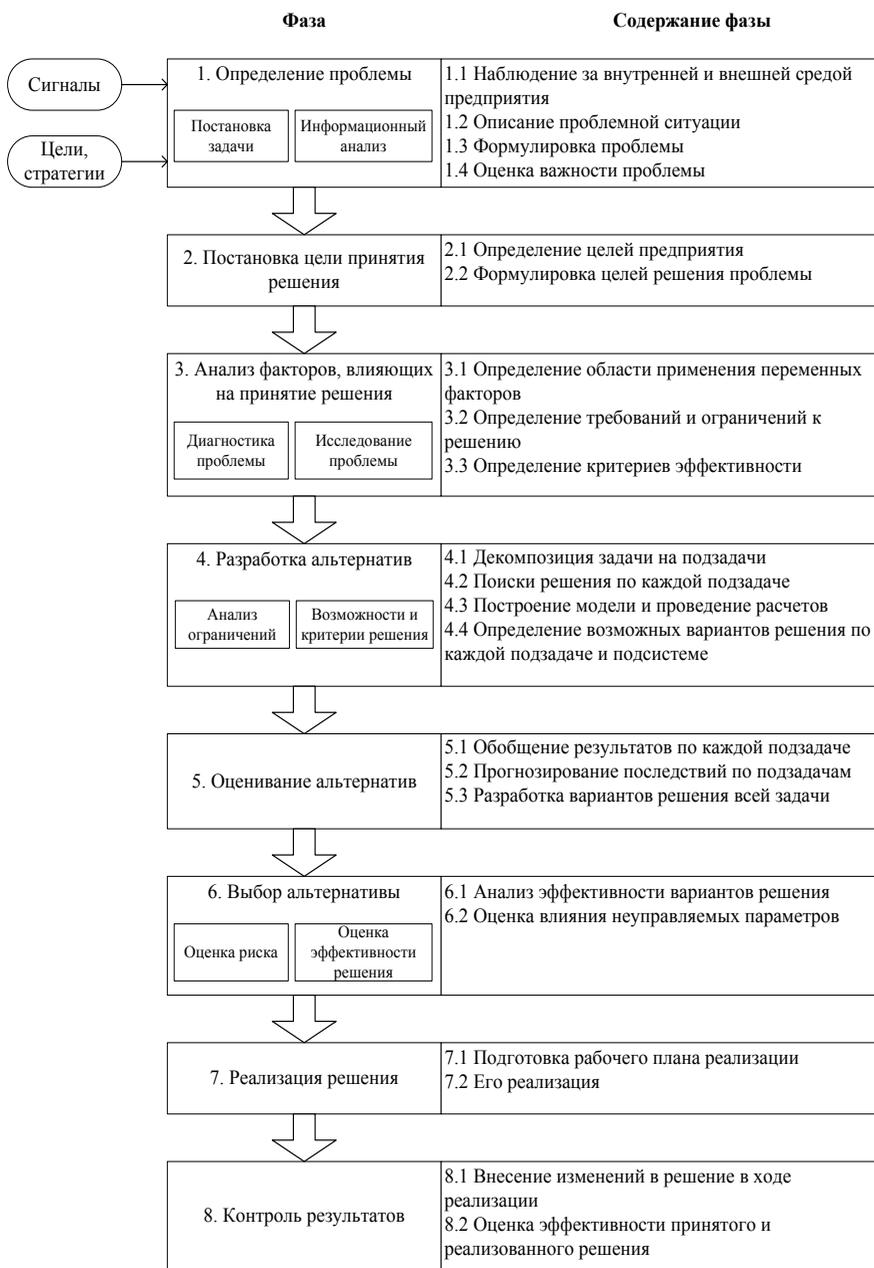


Рис. 1.1. Типовая технология принятия управленческого решения

Коллективные решения включают в себя групповые решения; решения, принимаемые организацией (корпорацией) и общественные (государственные) решения. Важность **групповых решений**, получаемых путем взаимодействия нескольких индивидуумов, общеизвестна: она объясняется, с одной стороны, профессиональной специализацией высококвалифицированных ЛПР, с другой стороны – ограниченностью физических и информационных возможностей каждого отдельного человека. Этапы группового решения проблемы мало отличаются от соответствующих этапов аналогичного индивидуального процесса – так как общими исходными моментами здесь являются ценности, представления и ограничения индивидуумов, составляющих группу. Однако группа имеет и свои особые, синергетические характеристики: например, доверие и сплоченность, выходящие за пределы простой экстраполяции индивидуальных качеств членов группы.

Результативность групповой работы ЛПР определяется как ее внешними и внутренними ресурсами, так и стратегией использования имеющихся ресурсов – неудачная выработка стратегий, составляющих основу процессов, происходящих в группе, приводит к потерям, связанным с излишними затратами, расчетами и обсуждениями. Важными аспектами, связанными с социальным влиянием на групповые решения, являются **конформизм** и **эффект поляризации**.

Одной из дилемм в групповом принятии решений является тот факт, что, хотя два человека знают больше, чем один, их взаимодействие друг с другом не всегда дает возможность реализовать это преимущество. В [80] описан следующий классический эксперимент по выявлению **группового конформизма**: в процессе эксперимента людям показывают три отрезка разной длины (А, В, С) и отрезок-эталон. У каждого участника эксперимента спрашивают – какой из трех отрезков равен отрезку-эталону. Эксперимент состоит из нескольких фаз. На первой фазе эксперимента вопрос задается каждому участнику по одиночке. На второй фазе тот же вопрос задается каждому участнику в присутствии дополнительного заранее проинструктированного человека, который, до того, как ответит участник, всегда дает неверный ответ, указывая на отрезок А. На третьей (и последующих) фазах вопрос задается, но уже в при-

сутствии двух (а затем трех, четырех и т.д.) дополнительных лиц, отвечающих первыми, всегда одинаково и всегда неверно.

Интересно сравнение процента ошибок, допущенных в ответах участников эксперимента (без учета ответов дополнительных лиц, выполняющих роль условий эксперимента). При переходе с первой фазы ко второй процент ошибки повышается с 1% до 3%. При переходе на третью фазу процент ошибки составляет уже 13%. Четвертая и все последующие фазы дают ошибку в 33%. Если в последнем случае, усилив условия эксперимента, еще сообщить его участникам, что группа получит тем большую премию, чем больше ее участников ответят на вопрос правильно, процент ошибки возрастает до 47%. Однако, если из семи дополнительных лиц шесть выберут отрезок А (неправильный ответ) и только один выберет отрезок В (правильный вариант), процент ошибки снижается до 6%. Данный эксперимент показывает наличие сильного влияния группы на ее членов даже для очень простого и очевидного случая. Легко можно представить себе степень усиления этого влияния для более сложных решений. Одним из выходов здесь может быть воздержание от обмена мнениями и попытка прийти к консенсусу простым голосованием (как, например, в методе Дельфи).

Второй проблемой, связанной с социальным влиянием на групповые решения, является так называемый *эффект поляризации*. Р. Стоунер [68] заметил, что после групповых дебатов об инвестировании в определенный проект часто принимались решения более рискованные, чем индивидуальные предложения членов группы до дебатов. Это было удивительно, поскольку считалось, что групповые решения скорее нивелируют экстремальные индивидуальные точки зрения, чем усиливают их. Однако при многих решениях была отмечена существующая тенденция поляризовать мнение большинства группы, если в ней имеется некоторая исходная установка. В случае, когда в группе имеются две сильные фракции, одна из них выступает за, а другая против некоторого предложения, прения обычно стихают, поскольку группа формирует некоторое промежуточное мнение.

Однако если в группе имеется исходная склонность к альтернативе X по сравнению с альтернативой Y, то имеется значительная вероятность того, то в результате прений мнение группы переместится еще дальше в направлении альтернативы X. Каковы бы ни

были причины эффекта поляризации, влияющего на групповой выбор, данные показывают, что группа не просто усредняет индивидуальный выбор ее членов, а имеет тенденцию изменить его. Направление такого изменения зависит от многих переменных и в каждом случае требует дополнительного анализа

Конформизм и поляризация – не единственные характеристики процесса принятия групповых решений. Главными характеристиками уровня взаимоотношений между членами группы остаются доверие, коммуникация и сотрудничество, а также показатели эффективности совместной работы. Хорошая коммуникация усиливает возможности сотрудничества при создании компромиссных решений, в то время как недостаток доверия подрывает как сотрудничество, так и коммуникацию.

При *принятии решений организациями* важную роль играют типы используемой модели, среди которых выделяют:

- унитарную модель, которая исходит из того, что только одно лицо от имени организации выбирает рациональные стратегии для достижения поставленных перед организацией целей;
- организационную модель, которая строится из предположения, что на предприятии существует несколько ЛПР (обособленных разделением труда или закрепленными за ними функциями), которые должны кооперироваться в своей деятельности;
- политическую модель, построенную таким образом, что в ней цели организации могут заменяться целями отдельных работников или подразделений;
- контекстную модель, которая делает упор на конкретную ситуацию, сопровождающую принятие решений в организации. В каждой компании существует большое количество неразрешенных конфликтов, подразделения и отдельные работники различаются своими ценностями, желаниями и отношением к риску, что создает оригинальный контекст, являющийся движущейся силой принимаемых решений, причем в большей степени, чем поставленные перед организацией цели.

Общественное принятие решений касается процессов, имеющих важные социальные последствия, то есть напрямую затрагивающих интересы множества самых разных людей. Данные решения всегда принимаются в рамках определенной социальной (общественной) политики, причем для оценки альтернатив здесь

обычно используются два трудноформализуемых критерия: эффективность и справедливость. Так, предлагаемая политика признается эффективной, если благосостояние общества при ее реализации существенно повышается, а одним из возможных количественных критериев является соотношение эффекта и затрат (при использовании данного критерия политика признается эффективной, если суммарный положительный эффект – выигрыш для всех членов общества превышает их суммарные потери).

Заметим, что в этих целях справедливая политика часто перераспределяет наиболее рациональным, с точки зрения ЛПР, способом как выигрыш, так и затраты между различными социальными группами.

В итоге можно сделать выводы о том, что:

- подэтапы реализации процесса принятия решений могут проходить не в той очередности, на которую указывает рис. 1.1: они могут срывать, перескакивать, подчиняться обратным связям, перекрытиям, параллельному движению и т.п.;
- процесс принятия решения тем более индивидуален, чем это решение по своей природе сложнее;
- ограниченный объем имеющейся информации ограничивает рациональность решения, увеличивает роль интуиции;
- предварительные установки по возможным альтернативам существенно влияют на выбор окончательного решения;
- не следует стремиться к наилучшему (оптимальному) решению, если у ЛПР имеется удовлетворяющий его вариант;
- участие нескольких лиц и другие организационные условия изменяют порядок прохождения подэтапов процесса принятий решений;
- менеджеры способны различным образом вмешиваются в структуру и процесс принятия решений, тем самым влияя (как в положительную, так и в отрицательную сторону), на их качество.

1.2. Применение интеллектуальных ИС в управлении

Составные части процесса решения проблемы представлены на рис. 1.2. Решение проблем призвано помочь корпорации в достижении плановых целей, установленных для нее стандартами, которые оценивают ее желаемое состояние. Кроме этих целей, ЛПР должен получать информацию, описывающую реальное текущее

положение дел – например, состояние производственного или коммерческого предприятия. Если текущее состояние корпорации совпадает с желаемым, то проблем нет, и ЛПР может не предпринимать никаких управленческих действий. Если указанные состояния различаются, это говорит о наличии проблемы, которая должна быть решена – тогда обязанностью ЛПР является разработка возможных альтернатив решения, призванного улучшить ситуацию.

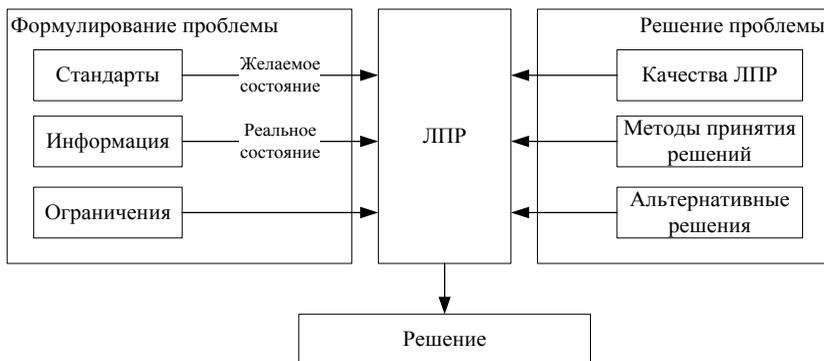


Рис. 1.2. Составные части процесса решения проблемы

На данном этапе процесса решения ЛПР обычно больше полагается на свой субъективный опыт, чем на помощь компьютера. После разработки альтернатив наступает этап их оценки, в процессе которого принимаются во внимание внутренние и внешние ограничения, связанные с реализацией разработанных альтернатив. Наибольшую сложность для ЛПР представляют два взаимосвязанных аспекта процесса решения. Первым является значительный объем информации, который перерабатывает менеджер при выработке альтернативных решений; вторым – собственно выбор лучшей из этих альтернатив. Данные аспекты непосредственным образом зависят друг от друга: поскольку чем больше объем переработанной информации, тем проще сделать выбор в пользу конкретного варианта решения.

Выбор лучшего варианта обычно происходит в условиях неопределенности – можно строить правдоподобные догадки о будущем развитии событий, но никогда нельзя точно сказать, какими будут последствия при выборе того или иного варианта решения.

Кроме того, каждый из данных вариантов имеет, как правило, и свои привлекательные стороны, и недостатки – сопоставление их между собой требует оценки вариантов решения по заданной системе критериев, что само по себе также является непростой задачей для ЛПР. С неопределенностью и многокритериальностью оценки вариантов связаны основные трудности при выборе решений. В [68; 89; 112] выделены следующие подходы к их преодолению.

Экономико-математический подход. Активное изучение задач принятия решений началось в прошлом XX веке: первыми за них взялись экономисты, поскольку задача выбора является одной из центральных в экономической науке. Два основных действующих лица торгового бизнеса: производитель и покупатель (потребитель) вовлечены в процессы постоянного выбора. Производитель решает, во что вкладывать капитал, какие изделия следует производить и продавать. Потребитель решает, что покупать и за какую цену. Одним из оснований экономической теории служит положение о рациональности выбора, которое предполагает, что решение человека является результатом его упорядоченного (разумного, научного, логически обоснованного и т.п.) мышления в объективно заданной системе тех или иных ценностных координат. Понятие упорядоченности определяется в строгой математической форме: вводится ряд формальных утверждений о поведении человека, именуемых **аксиомами рационального поведения**.

Наиболее известными и важными из этих аксиом являются:

- **аксиома транзитивности**: если для ЛПР исход A наблюдаемого (возможного) события предпочтительнее исхода B , а исход B – исхода C , то A предпочтительнее C ;
- **аксиома безразличия**: если у наблюдаемого события возможны три исхода: A ; B и C , и если ЛПР предпочитает исход A исходу B , а исход B исходу C , то с вероятностью P существует вариант, при котором для ЛПР одинаковую ценность будут иметь две альтернативы: 1) получить B наверняка; 2) принять участие в игре, где можно выиграть A с вероятностью P или C с вероятностью $(1 - P)$;
- **аксиома независимости**: если для ЛПР исходы A и B имеют одинаковую ценность, то для него безразлично, какой из них может быть реализован: A или B ;

– **аксиома рациональности**: если исход A для ЛПР предпочтительнее исхода B , он предпочтет игру с выигрышем A игре с выигрышем B .

Задачи принятия решений с оценкой полезностей и вероятностей событий были первыми из привлечших внимание исследователей в области управления экономическими СС. Постановка таких задач обычно заключается в следующем: ЛПР принимает решения, на исход (результат реализации) которых влияют случайные события, во многом неподвластные людям. Однако, зная (или утверждая, что знает) вероятности этих событий, ЛПР может аналитическим путем определить наиболее выгодный для себя вариант – по возможности избегая процедуры многокритериального выбора и пользуясь самыми простыми и «прозрачными» схемами моделирования объектов.

На основе упомянутых аксиом рациональности Дж. фон Нейманом и О. Моргенштерном была сформулирована и доказана теорема о существовании **функции полезности**, определяющей решения людей [84; 113]. Полезностью при этом была названа некая условная (воображаемая) величина, которую в процессе выбора максимизируют ЛПР с рациональным экономическим мышлением. Можно также сказать, что полезность – это воображаемая мера психологической ценности различных благ для человека, который как бы взвешивает на своих «внутренних весах» различные альтернативы и выбирает из них ту, которая сулит ему максимальную выгоду (наибольший выигрыш или наименьший ущерб).

Теория объективной ожидаемой полезности исходит из того, что указанные вероятности событий, а также наилучшие (оптимальные) варианты управленческих решений для разных СС уже существуют в природе сами по себе и ЛПР нужно лишь как можно ближе «подойти» к ним, распознать и использовать их на практике.

Теория субъективной ожидаемой полезности представляет себе те же вероятности как результат совместных эвристических представлений, согласованное мнение экспертов о них – оставляя за ЛПР возможность конструировать на их основе необходимые для эффективного управления СС варианты конкретных решений [113; 117].

Опыт практического применения обеих теорий полезности позволил, однако, установить, что многие разумные люди отнюдь не

склонны следовать их рекомендациям, а ведут себя, по меньшей мере, «нерационально». Устойчиво повторяющиеся отклонения от линии рационального поведения стали называть *парадоксами*. Характерным примером парадоксальной иррациональности действий ЛПР является выбор из четырех пар совершенно одинаковых (с точки зрения теории объективной ожидаемой полезности) тестовых альтернатив:

- гарантированный выигрыш 99 € или выигрыш 100 € с вероятностью 0,99 – большинство ЛПР не рискует и выбирает гарантированный выигрыш 99 €;
- гарантированный выигрыш 1 € или выигрыш 100 € с вероятностью 0,01 – большинство ЛПР в данной ситуации рискует, чтобы попробовать выиграть 100 €;
- гарантированный проигрыш 99 € или проигрыш 100 € с вероятностью 0,99 – большинство ЛПР рискует и выбирает проигрыш 100 €, поскольку есть призрачная возможность его избежать;
- гарантированный проигрыш 1 € или проигрыш 100 € с вероятностью 0,01 – большинство ЛПР не рискует и выбирает проигрыш 1 € во избежание большего проигрыша.

Совместные исследования психологов и экономистов, проведенные в последние годы, показали, что человеческое поведение вообще существенным образом отличается от рациональной модели – люди используют в своих суждениях эвристики, которые ведут к ошибкам и противоречиям, пренебрегают малыми значениями вероятностей, по-разному относятся к выигрышам и потерям и т.д. К попыткам разработать аксиоматическую теорию, которая учитывала бы индивидуальные особенности поведения ЛПР, можно отнести *теорию проспектов*. В этой теории вместо вероятностей используется функция от вероятностей, построенная специальным способом, что позволяет избежать ряда известных парадоксов – на смену которым, однако, приходят новые.

Важным шагом в развитии теории полезности стала *многокритериальная теория* полезности, в рамках которой существенно расширен круг аксиом, которым должна соответствовать функция полезности (проверка выполнения большинства аксиом рассматривается как непростая самостоятельная задача). В случае если условия аксиом удовлетворяются, дается математическое доказательство существования функции полезности в том или ином виде

(например, в аддитивном смысле: полезность является суммой полезностей оценок по отдельным критериям). Общим достоинством всех аксиоматических видов и разновидностей теории ожидаемой полезности, таким образом, остается их прочный научный фундамент в виде достаточно строгого и всестороннего математического обоснования.

Эвристические методы. Применение аксиоматических методов требует проверки выполнения аксиом, что является, как уже было сказано, достаточно трудной самостоятельной задачей. Кроме того, построение функции полезности требует затрат времени ЛПР, которые оправдываются лишь при наличии большого числа подлежащих анализу альтернатив. Все эти обстоятельства послужили стимулом для возникновения нормативных методов принятия решений, не имеющих строго научного теоретического обоснования, для большинства которых общим является использование метода взвешенных сумм оценок по заданным критериям. Примером такого эвристического многокритериального подхода служит метод аналитической иерархии.

Вербальный анализ решений. Большинство исследователей в области принятия решений признают наличие глубины противоречий между требованиями нормативных методов и реальными возможностями человека в качестве прикладной ИС: для преодоления этих противоречий разработан вербальный (порядковый) анализ возможных решений. В рамках вербального анализа к методам принятия решений, обычно имеющим междисциплинарное научное обоснование, предъявляются следующие требования:

- естественный язык описания проблемы, используемый ЛПР и его окружением, должен сохраняться на всех этапах ее анализа без каких-либо преобразований в числовые величины;
- способы получения информации от других людей должны соответствовать возможностям ЛПР как человеческой системы переработки информации;
- все логические операции преобразования словесных переменных (необходимые для оценки альтернатив по избранным критериям) должны быть математически корректными и четко определять тот или иной вид каждого решающего правила;
- должны быть предусмотрены средства проверки информации на непротиворечивость, необходимо применение методов поиска

противоречий в информации, получаемой от ЛПР, с целью исключения данных противоречий.

Причины указанных ошибок и противоречий заключаются, прежде всего, в ограниченной емкости кратковременной памяти ЛПР, где и осуществляются основные операции сравнения и выбора. Человеческая система переработки информации приспособлена к решению большинства задач, с которыми *homo sapiens* сталкивался в ходе своего многовекового эволюционного развития – в известной мере (в определенных природой пределах) это касается и решения многофакторных задач при относительно небольшом числе факторов. Кроме того, человек обладает набором эвристик, позволяющих решать задачи практически любой сложности, предварительно упрощая их и приспособляя к своим возможностям. Однако сегодня в сфере управления СС растет число задач, которые непосильно трудны для абсолютного большинства ЛПР.

Сложности работы ЛПР обусловлены тем, что бизнес и его внешняя среда в настоящее время чрезвычайно усложнились и имеют тенденцию к еще большему усложнению. Одной из важных проблем при этом является рост объемов корпоративной информации, которую ЛПР нужно собирать, оперативно обрабатывать, поддерживать в актуальном состоянии, передавать по назначению и принимать, надежно хранить – с тем, чтобы обеспечить соответствие корпорации ее бизнес-целям, уменьшить рыночные риски и стимулировать инновации [55].

Необходимо иметь в виду, что объем информации растет в зависимости от развития бизнеса по экспоненциальному закону: например, при росте производства вдвое объем перерабатываемой информации увеличивается примерно в четыре раза. Поэтому рост объемов информации и потребности в ней управленческого персонала базируется на объективных закономерностях развития экономики.

Данные таблицы 1.3 демонстрируют динамику факторов, влияющих на принятие управленческих решений в частности, которая показывает, что принятие решений сегодня является более сложным, чем вчера, в силу двух основных обстоятельств. Во-первых, вследствие совершенствования технологий и средств управления (коммуникационных систем) увеличивается число возможных альтернатив по каждому из принимаемых решений.

Таблица 1.3. Факторы, воздействующие на процесс принятия управленческих решений [80]

Факторы	Тенденции	Результаты
Технология	Усложнение	Увеличение числа альтернатив принимаемого решения
Коммуникация, компьютеры	Увеличение	
Структура фирмы	Усложнение	Повышение стоимости ущерба от неправильных решений
Конкуренция	Ужесточение	
Мировой рынок	Усложнение	Увеличение неопределенности в отношении будущего корпорации
Потребление	Увеличение	
Политическая стабильность	Уменьшение	
Вмешательство государства	Увеличение	

Во-вторых, в силу автоматизации производства и сложности проводимых коммерческих операций, а также ввиду механизма «цепной реакции», которую могут вызвать в элементах и подсистемах корпорации отдельные неверные решения, растет стоимость потенциального ущерба вследствие ошибочных действий ЛПР. Усложнению принятия решений способствует и увеличение неопределенности в отношении будущего корпорации (предприятия, компании, организации, фирмы) вследствие снижения политической стабильности, вмешательства государства в бизнес, усложнения ситуации на мировом рынке и динамики процессов рыночного спроса и потребления.

Характерный для современного этапа экономики рост объемов управленческой информации остро ставит проблему *информационной перегрузки* ЛПР – под которой понимают ситуацию, когда к менеджерам поступает такое большое количество информации, что они из-за недостатка времени (или по другим причинам) или не могут обработать ее в полном объеме, или рассмотреть и оценить вообще, или не могут обработать качественно, но во всех случаях неспособны принимать правильные управленческие решения. Следует также понимать, что поток информации включает в себя са-

мые разнообразные данные вербального (устного), письменного и визуального характера, несущие ЛПР сведения, необходимые для выполнения процессов управления и регулирования взаимоотношений как внутри коллектива, так и между ним и внешней средой.

Мировой объем оцифрованной информации растет по экспоненте. По данным компании IBS, к 2003 г. мир накопил 5 эксабайтов данных (1 ЭБ = 1 млрд гигабайтов). К 2008 г. этот объем вырос до 0,18 зеттабайта (1 ЗБ = 1024 эксабайта), к 2011 г. – до 1,76 зеттабайта, к 2013 году – до 4,4 зеттабайта. В мае 2015 г. глобальное количество данных превысило 6,5 зеттабайта. К 2020 г., по прогнозам, человечество сформирует 40-44 зеттабайтов информации.

Развитие человеческого общества исторически связано с преодолением *информационных барьеров* [19; 66]. Самый первый информационный барьер был достигнут и преодолен в тот давний период, когда экономические связи полностью замыкались в рамках ограниченных коллективов (род, семья, племя) и сложность управления этим коллективом стала превосходить способности отдельного человека. Это произошло многие тысячи лет назад и вызвало изменения в *технологии систем управления* (СУ), которые состояли в изобретении двух механизмов управления экономикой: первый механизм был обусловлен созданием иерархических СУ (где руководитель назначает себе помощников, а те, в свою очередь, распределяют функции между подчиненными и т.д.); второй механизм был связан с введением правил взаимоотношения между людьми и социальными коллективами: предприятиями, регионами, государствами и т.д. (эти функции первоначально выполняла религия, в последующем – законодательная система государства). Одним из действенных способов реализации данного механизма сегодня являются экономические регуляторы, основанные на введении рыночных товарно-денежных общественных отношений.

Второй информационный барьер связан с ограниченной способностью к переработке информации у населения страны – ввиду того, что сложность задач управления экономикой растет быстрее, чем число занятых в ней людей. Теоретические исследования тенденций роста численности управленческого персонала подтверждаются и многолетней статистикой: в развитых западных странах, например, в начале XX века на одного конторского работника приходилось 40 рабочих; в сороковые годы – 10 рабочих; в пятидеся-

тые годы – 6 рабочих; а к началу семидесятых – всего 1 рабочий. Отечественной статистикой аналогичным образом зафиксирован рост численности управленческого персонала до 40% и более от общей численности работников большинства производственных предприятий.

Иными словами, в ходе исторического разделения труда, которое привело к формированию самостоятельных сфер интеллектуального производства, *информационная деятельность* (неразрывно связанная с переработкой и обменом информацией, приобретением знаний) для определенных категорий людей становится важнейшей составной частью их профессиональной деятельности. К числу таких категорий в первую очередь относятся ученые, инженеры, педагоги, журналисты, врачи, юристы и другие работники умственного труда. Эффективность труда указанных специалистов во многом зависит от того, насколько успешно существующие в обществе средства коммуникации обеспечивают их информационную деятельность.

Процесс коммуникации, помимо непосредственного информационного обмена – связанного с восприятием, переработкой и выдачей пользователю нужной ему информации – включает целый ряд вспомогательных операций: таких как оформление, сбор, аналитико-синтетическую обработку, хранение, поиск и распространение информационных сообщений. На ранних этапах развития науки и техники многие из этих вспомогательных операций работники умственного труда (в том числе управленцы и другие ЛПР) выполняли самостоятельно. Однако, по мере расширения и усложнения сферы коммуникаций, удельный вес вспомогательных операций в общем балансе времени ЛПР увеличился настолько, что это стало отрицательно сказываться на производительности их труда, а, следовательно, и на эффективности общественного производства в целом.

Возникла необходимость в выделении и обособлении групп информационных работников, для которых выполнение тех или иных операций, обеспечивающих функционирование ИС, является основной профессией: так появились библиотекари, библиографы, переводчики, архивариусы, редакционно-издательские сотрудники, информаторы-аналитики и др. Формирование данных групп стало очередным шагом в разделении труда, означающим дальнейшую

специализацию работников умственного труда. Эта функциональная специализация сформировала особую область человеческой деятельности – *сферу информационного обслуживания* или *информационного сервиса*, направленную на удовлетворение информационных потребностей современного общества.

Все перечисленные операции в большинстве случаев могут быть четко отделены от других видов инженерной и творческой деятельности, поскольку наиболее квалифицированно могут выполняться именно ИТ-специалистами. Соответственно, многочисленные виды ИТ-бизнеса – такие как формирование информационных массивов, создание информационно-поисковых систем, подготовка библиографических указателей, фактографических справок, написание рефератов и обзоров, редактирование, рецензирование, перевод текстов и многое другое, традиционно рассматриваемое как разные виды информационной деятельности, сегодня представляют собой разные операции в сфере информационного обслуживания. Также следует иметь в виду, что избыточность информации может быть следствием не только объективных закономерностей, присущих техническому и технологическому прогрессу, но и результатом недостатка внимания к методам обработки и выдачи информации ЛПР.

Таким образом, в условиях возрастания сложности бизнес-процессов, реализуемых в социально-экономических и организационно-технических системах, задача выработки эффективных управленческих решений приобретает постоянно возрастающую значимость. Однако управление данными в СС сопровождается трудностями, которые связаны со следующими обстоятельствами:

- решения необходимо принимать в условиях неопределенности, нечеткости данных об объекте управления, размытости и быстрого изменения целей управляемой СС;
- внешняя среда для реальных СС также все больше приобретает черты нестабильности и неопределенности: нестабильность проявляется в том, что темпы изменения внешней среды возрастают, а неопределенность – в том, что возникающие ситуации все чаще становятся совершенно новыми;
- отсутствуют строгие и адекватные математические модели реальных СС, позволяющие определить механизмы управления ими и принципы работы СУ;

- информация, необходимая для выработки и реализации адекватных управленческих решений, как правило, рассредоточена по множеству разнообразных хранилищ, фондов, баз данных (БД), а также распределена территориально и хранится в различных форматах, она обрабатывается по-разному, недостаточно актуализирована и т.д.;
- доступ к информации для принятия управленческих решений сопряжен с целым рядом других экономических, организационных и технологических трудностей.

В результате рассмотрения текущих изменений в бизнесе и перспектив развития этих тенденций в будущем можно сделать вывод о том, что возможностей эмпирического метода проб и ошибок недостаточно для принятия управленческих решений и для эффективного управления СС необходимо использовать специальные научные методы количественного и качественного анализа. В первую очередь – связанные с применением вычислительной техники и ИС, которые сегодня нашли применение и получили развитие практически в любой области человеческой деятельности.

Естественным результатом развития информационных систем стали *интеллектуальные информационные системы (ИИС)*, сосредоточившие в себе наиболее наукоемкие технологии с высоким уровнем автоматизации не только процессов подготовки информации для принятия решений, но и самих процессов выработки их альтернативных вариантов, опирающихся на данные, полученные в информационных системах.

Возможность и целесообразность применения ИИС в процессе принятия управленческих решений могут быть определены на основании следующих соображений:

- в зависимости от структурированности решаемых задач;
- с учетом уровня иерархии управления корпорацией, на котором решение должно быть принято;
- в зависимости от принадлежности решаемой задачи к той или иной функциональной сфере бизнеса;
- в зависимости от вида используемой ИТ.

Применение вычислительной техники, как известно, привело к существенному изменению положения и роли человека в процессе производства.

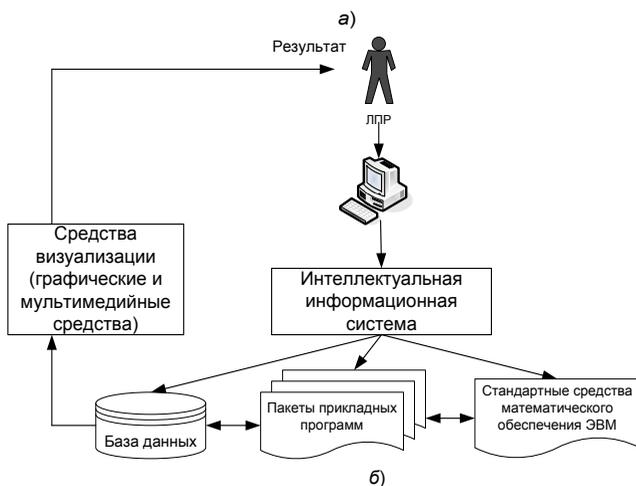


Рис. 1.3. Подходы к решению прикладных задач:
 а) традиционный, б) с помощью ИИС [102]

Одной из наиболее перспективных форм организации процесса управления экономическими объектами является интерактивный метод принятия управленческих решений на основе интеллектуально-информационной базы, предусматривающий возможность привлечения методов ИИ и позволяющий сочетать вычислительные возможности ЭВМ с интеллектом ЛПР наиболее гибким и оптимальным способом, что, в конечном счете, позволяет добиться требуемого высокого качества управления.

На рис. 1.3а представлен традиционный подход к решению прикладных задач, в рамках которого, на всех его этапах, привлекаются знания конкретных специалистов – эти знания или присутствуют в неявном виде, или «защиты» в соответствующие алгоритмы, то есть имеет место аккумуляция коллективного опыта разных поколений ЛПР по решению каждого конкретного класса задач. Схема на рис. 1.3б объединяет все виды знаний и процедурные блоки, необходимые для решения прикладных задач согласно рис. 1.3а, в единую ИИС – с целью повышения эффективности труда ЛПР, сокращения времени ожидания результатов, а также уменьшения зависимости от экспертов и других специалистов-посредников, субъективные взгляды и суждения которых могут исказить представление ЛПР о решаемой задаче и интерпретацию полученных результатов.

1.3. Автоматизация бизнеса в отрасли инфокоммуникаций

Мировая индустрия инфокоммуникаций, частью которой являются отечественные ИКК, представляет собой огромный по масштабам бизнес – общий доход которого составляет более триллиона долларов в год. В ряде стран инфокоммуникационный сектор обеспечивает значительную долю валового внутреннего продукта, при этом завершение многолетнего господства и отраслевого протекционизма ведет к трансформации рынка, резкому обострению конкуренции и буму новых ИТ. Наряду с традиционными видами услуг (электросвязь, передача данных, телевизионное и радиовещание) интенсивно развиваются такие приложения, как мобильная связь, электронный бизнес и торговля, Internet вещей, мультимедийные технологии и т.п. Этим, собственно, говоря, и обусловлен переход от традиционного термина «компания электросвязи» к термину ИКК.

В таблице 1.4 приведены основные функции инфокоммуникационной системы (ИКС), владельцем которой является ИКК. Как объект управления, ИКК характеризуется не только значительным количеством составляющих элементов, компонентов и подсистем, но и многообразием выполняемых ими функций. Список задач, которые необходимо решать при управлении в телекоммуникациях, весьма широк: к их числу можно отнести управление финансовыми потоками, бюджетами, продажами, сложной сетевой инфраструктурой, разработку и внедрение новых услуг, задачи биллинга и работы с клиентами и др.

В настоящее время операторы услуг работают в разных странах, где законодательства сильно отличаются друг от друга; их бизнес-стратегии и подходы к решению проблем конкуренции также весьма различны. Однако они имеют и общие черты:

- их конкурентоспособность все больше зависит от эффективного управления информацией и телекоммуникационными сетями;
- подход к управлению услугами они выбирают в соответствии с принципами ведения бизнеса и эксплуатации телекоммуникационных сетей;
- все шире используют сквозное управление процессами, учитывая мнение клиентов, используют автоматизацию процессов управления сетью, услугами и обслуживанием клиентов;
- им необходима интеграция новых OSS/BSS-систем и наследуемых систем;
- делают акцент на предложении новых информационных услуг, основное внимание уделяют повышению эффективности всей совокупности услуг, в частности, повышению уровня удовлетворенности клиентов;
- интегрируют существующие и новые технологии, все чаще покупают системы, а не создают их самостоятельно, осуществляют интеграцию систем от разных производителей.

Как отмечено в [94], стать эффективным оператором, с точки зрения сохранения объема и прибыли от предоставления традиционных ИТ-услуг, ИКК мешают следующие причины:

- высокие операционные затраты и недостаточная степень автоматизации технологических и бизнес-процессов;
- недостаточно эффективное владение и управление сквозными бизнес-процессами;

- использование изолированного или не отвечающего современным требованиям производственного оборудования;
- использование плохо скоординированных производственных и бизнес-систем;
- низкое качество обслуживания клиентов и неточная информация о запросах и требованиях клиентов.

Аналогичным образом развитию новых источников доходов ИКК препятствуют:

- длительные циклы запуска новых ИТ-продуктов;
- негибкие и разрозненные (изолированные друг от друга) бизнес-процессы;
- несогласованные действия функциональных подразделений;
- негибкие системы с жесткими связями;
- длительные циклы «заказ-оплата»;
- неэффективные основные и вспомогательные бизнес-процессы;
- негибкие и несистемные (разрозненные) бизнес-процессы.

Таблица 1.4. Основные функции ИКК

Функция	Содержание функции
Социальная	Обеспечивает связи между людьми, организациями, государственными учреждениями, а также доступ к информационным системам и передачу информационных сообщений в интересах всех граждан
Экономическая	Обеспечивает передачу информации для хозяйствующих субъектов для принятия управленческих решений и экономии времени (народно-хозяйственная эффективность)
Технология	Обеспечивает технический прогресс в области информатизации общества, внедрения инфокоммуникаций в высокотехнологические отрасли
Государственная	Обеспечивает органы власти необходимыми данными и средствами связи для управления страной, в том числе обеспечения национальной безопасности; позволяет осуществлять прямой контакт органов власти и граждан

В мировой практике решением вопросов, относящихся к области построения бизнес-процессов телекоммуникационных компаний, занимается международная некоммерческая организация TeleManagement Forum (TMF).

С момента своего создания TMF занимался разработкой единых методологий, рекомендаций и стандартов для индустрии телекоммуникаций, а также наиболее эффективного способа построения работы операторов связи с точки зрения себестоимости и временных затрат.

Свою главную цель TM Forum видит в том, чтобы предоставить всем операторам единый подход для сквозной автоматизации операционных бизнес-процессов оказания клиентам информационных и телекоммуникационных услуг. TM Forum с самого начала определил бизнес-процессы как инструмент достижения согласия и взаимопонимания между операторами связи. Такой подход способствовал сближению позиций и формированию в этой области широкой общей базы, которая обеспечивает:

- проведение подразделениями TM Forum целенаправленной работы по детальному определению требований бизнеса, информационных соглашений, контрактов в отношении бизнес-приложений, спецификаций совместно используемой информационной модели и модели данных и по проведению анализа полученных результатов с точки зрения их противоречивости;
- сопоставление потребностей бизнеса с имеющимися или разрабатываемыми стандартами;
- формирование общей концепции процессов для поставщиков оборудования, разработчиков приложений и интеграторов, которое позволит создавать системы управления с одновременным использованием как собственных разработок, так и разработок, закупаемых у других.

Предполагается, что с ее помощью операторам связи будет легче интегрировать в свою среду продукты, приобретаемые для управления бизнесом и оперативного управления сетями, информационными технологиями и услугами, а также снизить затраты на обеспечение сквозной автоматизации.

Кроме того, наличие общеотраслевой концепции процессов и информации облегчит взаимодействие «оператор-оператор», «оператор-клиент» и «оператор-поставщик/партнер» на уровне процессов, что будет способствовать быстрому предоставлению услуг и решению проблем в условиях глобальной конкуренции.

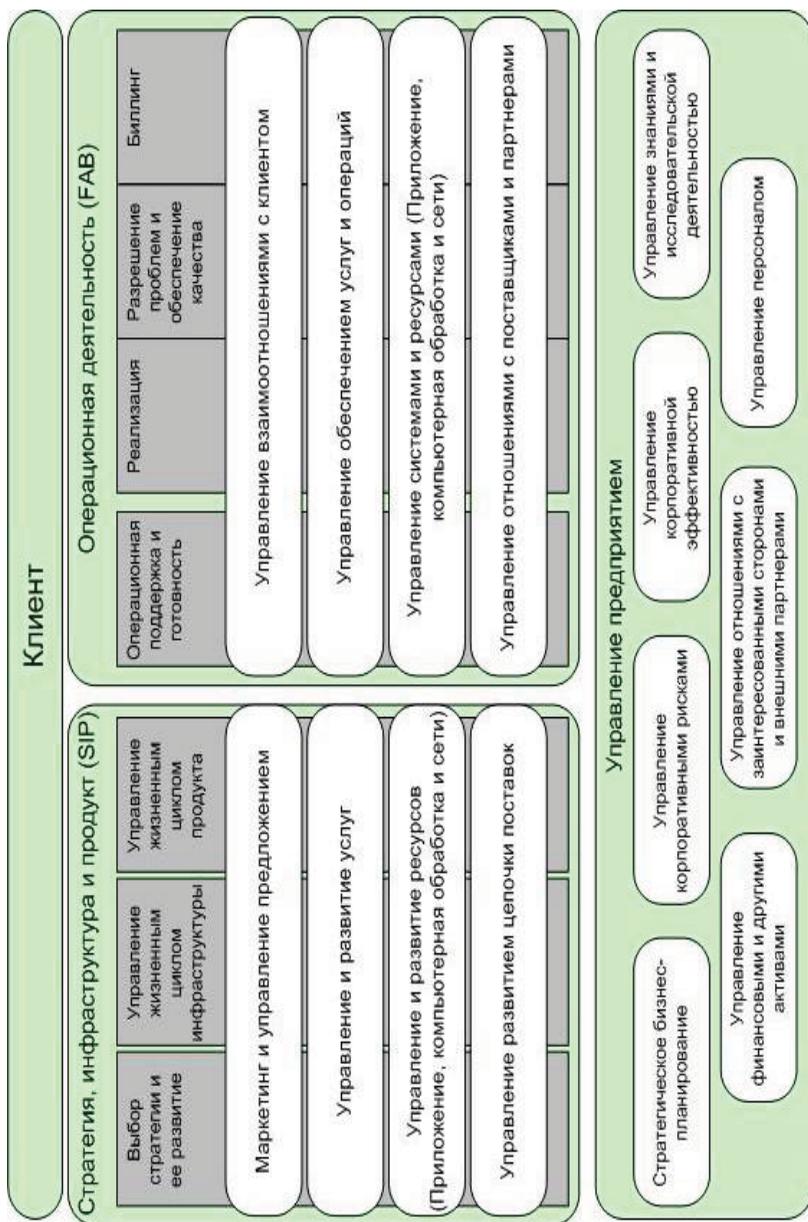


Рис. 1.4. Расширенная карта процессов eTOM

Сегодня разработки TMF ведутся на основании жизненного цикла и методологии NGOSS (New Generation Operations Systems and Software, новое поколение операционных систем и программного обеспечения), в которой определены заинтересованные стороны, этапы разработки OSS и информация (артефакты), передаваемая при смене этапов [5].

На рис. 1.4 представлена расширенная карта процессов eTOM. В карте eTOM проводится различие между стратегическими процессами развития и управления жизненным циклом услуг, с одной стороны, и операционными процессами – с другой. На рис. 1.4 показано в общей сложности семь групп процессов по вертикали. Эти окончательные (end-to-end), функционально законченные процессы, которые требуются для поддержки пользователей и управления бизнесом компании.

Центральными здесь являются процессы эксплуатационной поддержки/сетевые операции пользователей (customer operations processes), которые объединены под общей аббревиатурой FAB (Fulfillment Assurance Billing).

Процессы поддержки эксплуатации и готовности систем связи (operations support & readiness) в предлагаемой схеме функционально отделены от FAB. Это вызвано тем, что процессы, составляющие FAB, происходят в реальном масштабе времени (минуты, секунды), а процессы эксплуатации (например, ремонт оборудования и замена вышедших из строя блоков) проводятся за существенно большее время (десятки минут, часы). Процессы FAB имеют прямые интерфейсы с пользователями услуг связи и находятся в центре производственной деятельности компании.

Стратегия развития (Strategy & Commit) управления жизненным циклом инфраструктуры (Infrastructure Lifecycle Management) и управления жизненным циклом продуктов (Product Lifecycle Management) функционально разделены. Они, в отличие от сетевых операций, не связаны с непосредственной поддержкой пользователей и также функционируют в другом масштабе времени. Для создания инфраструктуры телекоммуникации, строительства зданий и сооружений связи требуются годы, в то время как для проверки состояния счета пользователя перед установлением сеанса связи требуются секунды.

Управление инфраструктурой и управление продуктами и услугами представляют собой основу стратегического развития компании. Очевидно, что без плана капитального строительства, схем развития сетей связи, плана ввода емкости ни один оператор связи не может нормально реализовать свои бизнес-планы. Сюда же относится планирование предоставления новых услуг, таких как Internet, IP-телефония и т.п. Все эти вопросы объединяются в рамках стратегии развития компании связи и совокупно обозначаются SIP. Процессы SIP необходимы для гарантии того, что сетевые процессы/операции пользователя (Customer Operation Processes) полностью отвечают требованиям пользователя, в том числе в части сроков предоставления, стоимости, уровня поддержки и доступности услуги связи. Процессы SIP не имеют прямых интерфейсов с пользователями услуг связи; они важны для внутрипроизводственной деятельности предприятия связи.

Особо следует отметить процессы, связанные с разработкой и управлением цепочками поставок (Supply Chain Management). Эти процессы важны для электронного бизнеса, а также в случае поставок продуктов и услуг другими операторами и партнерами. Например, провайдер услуг сети Internet может арендовать у оператора связи группы серийных телефонных номеров для коммутируемого доступа в Internet на определенный период времени.

Как видно из рис. 1.4, деятельность ИКК представляет собой сложный взаимосвязанный процесс, изменения в любой части которого неизбежно приводят к изменениям в других областях деятельности компании. Для эффективного менеджмента в телекоммуникациях сегодня одинаково важно рассматривать как сферу производственных процессов, так и социальную направленность бизнеса ИКК – в обоих случаях при этом речь идет об управлении многоуровневыми иерархическими СС, которые реализуют множество достаточно разных (основных и вспомогательных) бизнес-процессов, что требует проведения грамотной (осознанной, гибкой и компетентной) политики управления ими. Задачи по руководству ИКК предъявляет к ЛПР повышенные требования, удовлетворить которые без соответствующей информационной и организационной поддержки в настоящее время не удастся. Поэтому исследования в области ИТ, направленные на повышение эффективности

ИКК, вызывает интерес у ЛПР данных предприятий и являются актуальным направлением развития теории управления СС.

Высокий уровень конкуренции в отрасли заставляет операторов ИТ-сетей и поставщиков телекоммуникационных услуг одновременно решать следующие комплексные задачи [103]:

- удержание имеющихся клиентов и повышение их лояльности;
- привлечение как можно большего числа новых клиентов, в том числе путем повышения качества обслуживания и предложения новых услуг;
- снижение собственных издержек за счет оптимизации внутренних процессов и уменьшения необязательных потерь.

Таким образом, ИКК сегодня вынуждены активно искать свои конкурентные преимущества, совершенствуя и оптимизируя сложившийся бизнес, снижая себестоимость и максимально сокращая время разработки и вывода на рынок новых ИТ-услуг. Очевидно, что эта деятельность невозможна без применения специализированных ИС и СУ, которые способствуют рациональной организации деятельности как всей ИКК, так и ее отдельных филиалов и структурных подразделений, упорядочению и обновлению бизнес-процессов, эффективному использованию имеющихся ресурсов.

Управление производственными процессами ИКК представляет собой достаточно сложную комплексную задачу, поэтому в настоящее время не вызывает сомнений необходимость создания ИС, способных обеспечивать функционирование производства и бизнеса ИТ-компании. Одним из способов решения сопутствующих этому процессу проблем является внедрение ИС в практику управления ИКК – опыт лидеров мирового рынка показывает, что использование ИТ, лежащих в основе систем управления предприятием, позволяет повысить эффективность и конкурентоспособность любого бизнеса. Новые ИТ-продукты, которые появились в последние годы на рынке, подталкивают руководство ИКК к необходимости пересмотра своего подхода к управлению бизнесом, а усиливающаяся конкуренция значительным образом ускоряет этот процесс.

Единая СУ, обслуживающая все бизнес-процессы ИКК, позволяет в короткий срок добиться следующих результатов [110]:

- улучшить соотношение «цена-качество» предоставляемых пользователю услуг;

- усовершенствовать процессы управления производительностью и уровнем обслуживания пользователей услуг;
- добиться реальной экономии времени и средств;
- обеспечить более эффективное управление конфигурацией предоставляемых услуг;
- увеличить скорость предоставления услуг и обслуживания пользователей за счет автоматизации производственных процессов;
- добиться более точного учета ресурсов при предоставлении услуг ИКК.

Одним из инструментов оптимизации бизнес-процессов и повышения их эффективности может выступать использование на всех уровнях управления ИКК систем поддержки принятия решений в виде специализированных ИС и ИИС.

1.4. Эволюция интеллектуальных ИС

Область искусственного интеллекта (Artificial Intelligence, в английской и русской аббревиатурах, соответственно, AI и ИИ) направлена на создание компьютерной системы с человекоподобным интеллектом. При этом она должна обладать следующими характеристиками [101]: личностные качества, желание, внимание, креативность, интуиция и др. Данная цель труднодостижима и поэтому разработчики систем ИИ сконцентрировались в основном на создании узкоспециализированных ИИС, то есть программного обеспечения, которые проявляет «интеллект» относительно определенных вопросов, касающихся существенно ограниченных – «узких» областей знаний.

Сегодня общепризнано, что предметом интереса области ИИ являются проблемы, которые не могут быть решены с помощью традиционных ИТ. Множество концепций, технологий и архитектур, которые предлагает ИИ, предназначены для решения комплексных проблем в тех случаях, когда чисто арифметические или математические решения либо неизвестны, либо малоэффективны. Кратко остановимся на наиболее значимых этапах развития ИИС. В истории развития ИИ как отдельного научного направления сегодня принято выделять две относительно самостоятельные ветви [59; 101; 105] развития: *нейрокибернетику* и *«кибернетику черного ящика»*. Однако идея создания искусственного подобия человеческого разума для решения сложных задач и моделирования

мыслительной способности витала в воздухе с древнейших времен. Впервые ее выразил Р. Луллий, который в XIV веке пытался создать машину для решения различных задач на основе всеобщей классификации понятий. В XVIII веке Г. Лейбниц и Р. Декарт, независимо друг от друга, развили эту идею, предложив универсальные языки классификации всех известных в то время наук.

Эти идеи легли в основу первых теоретических разработок в области создания ИИ, которые приобрели практическое значение после создания ЭВМ в сороковые годы XX века. В это же время Н. Винер опубликовал свои основополагающие работы по новой науке – кибернетике, в популяризации которой немалую общественную роль сыграли литераторы: достаточно вспомнить о сказочных персонажах Пиноккио и Буратино, упомянуть «Планету обезьян» Пьера Буля и законы роботехники Айзека Азимова.

Сам термин ИИ (англоязычный вариант AI), судя по разным источникам, был предложен Дж. МакКарти и впервые получил признание в 1954-56 г.г. на семинаре с аналогичным названием в Дартмутском колледже США. Однако надо отметить, во-первых, что семинар в Дарсмуте был посвящен разработке методов решения логических, а не вычислительных задач. Во-вторых, что в английском языке термин AI не имеет той слегка фантастической антропоморфной окраски, которую он приобрел в довольно неудачном русском переводе: слово *intelligence* означает «умение рассуждать», а вовсе не «интеллект», для которого есть термин *intellect*. В-третьих, что на том же семинаре наметился и продолжился в дальнейшем раскол кибернетики на нейрокибернетику и «кибернетику черного ящика».

Основную идею *нейрокибернетики* можно сформулировать следующим образом. Единственный объект, способный мыслить, – это человеческий мозг, поэтому любое «мыслящее» устройство должно каким-то образом ему соответствовать, так или иначе воспроизводить его структуру. Таким образом, нейрокибернетика ориентирована на аппаратное моделирование структур, подобных структуре мозга. Поскольку сегодня физиологами установлено, что основой человеческого мозга является большое количество (до 10^{10}) связанных между собой и взаимодействующих нервных клеток – нейронов, усилия «нейрокибернетиков» сосредоточены на создании элементов, аналогичных нейронам, и их объединении в

функционирующие системы – которые принято называть *нейронными сетями*, или нейросетями.

В основу *кибернетики «черного ящика»* лег принцип, противоположный нейрокибернетике: не имеет значения, как устроено «мыслящее» устройство, главное, чтобы на заданные входные воздействия оно реагировало примерно так же, как человеческий мозг. Это направление ИИ было изначально ориентировано на поиски алгоритмов решения интеллектуальных задач на существующих ЭВМ. В 1956-63 г.г. велись интенсивные поиски таких моделей и алгоритмов, отражающих процесс человеческого мышления, создавались первые комплексы ИИ-программ. Оказалось, что ни одна из существующих наук: философия, психология, лингвистика и т.д. не могут предложить методологию создания таких алгоритмов, вследствие чего кибернетикам пришлось создать свои собственные методы и модели, опробовать различные подходы.

В конце пятидесятих лет XX века родилась модель *лабиринтного поиска*: данный подход представляет решаемую задачу как некий граф, отражающий пространство состояний, и в этом графе проводится поиск оптимального пути от входных данных к итоговым результатам. После первых успехов и немалых усилий работа по данной модели была остановлена, поскольку в плане решения практических задач идея большого распространения не получила.

На смену ей с начала шестидесятих лет пришла эпоха *эвристического программирования*. В разделе 1.1 было отмечено, что понятие «эвристика» соответствует совокупности логических приемов и методических правил, используемых для принятия решений – добавим, что эти приемы и правила не всегда строго обоснованы в теоретическом плане, но зато позволяют получить важный практический эффект: в данном случае сократить число переборов в пространстве поиска. В соответствии с этим, эвристическое программирование – это разработка стратегии действий на основе известных, заранее заданных эвристик. В семидесятые годы к ним стали подключать методы *математической логики*.

История развития ИИ в нашей стране также имеет свою хронологию. В 1954 г. в Московском государственном университете под руководством А. Ляпунова начал работу семинар «Автоматы и мышление», в котором участвовали крупнейшие отечественные

специалисты: физиологи, лингвисты, психологи, математики. Принято считать, что именно в это время родился ИИ в России – как и за рубежом, при этом сразу выделились сторонники и нейрокибернетики, и кибернетики «черного ящика».

Среди значимых результатов, полученных отечественными учеными в то время, следует отметить алгоритм «Кора» М. Бонгарда, моделирующий деятельность человеческого мозга при распознавании образов, а также ряд программ, предназначенных для исследования процесса поиска решений логических задач. В Ленинградском отделении математического института им. В.А. Стеклова была создана программа АЛПЕВ ЛОМИ, автоматически доказывающая теоремы – основанная на оригинальном обратном выводе С. Маслова, аналогичном методу резолюций Дж. Робинсона.

В 1965-80 г.г. получает развитие новая наука – *ситуационное управление* (соответствует *представлению знаний* в западной терминологии). Представителями данной научной школы, основоположником которой был Д. Поспелов, разработаны специальные модели представления ситуаций – представления знаний. В 1980-90 г.г. проводятся активные исследования в области представления знаний, разрабатываются необходимые специализированные языки и экспертные системы (общим числом более 300). В МГУ создается язык РЕФАЛ, в 1988 г. организуется Ассоциация ИИ.

Двум направлениям развития ИИ соответствует два основных метода их реализации: «нисходящий» и «восходящий» [2]. Нисходящий метод соответствует дедуктивному подходу, в рамках которого на этапе становления ИИ разрабатывались программы, способные решать сложные задачи на основе логической обработки содержащихся в них знаний. Примерами таких программ являются всемирно известные «Логик-Теоретик» и GPS – универсальный решатель задач, в создании которых участвовали А. Ньюэлл, А. Тьюринг, К. Шеннон, Г. Саймон, Дж. Шоу и другие всемирно известные специалисты. «Восходящий метод» развит в работах Дж. Маккалока, У. Питса, Ф. Розентблатта и их коллег, занятых созданием самоорганизующихся СС и «самообучающихся машин». Эти ученые опирались на идею Н. Винера об обратной связи, благодаря которой все живое на Земле универсальным образом приспосабливается к окружающей среде и добивается своих целей.



Рис. 1.5. Базовые парадигмы исследований по ИИ

На рис. 1.5 в схематичном виде показана эволюция основных целей разработчиков ИИС [102]. Отметим, что каждая из представленных здесь парадигм преследует собственные цели: так, например, с точки зрения *логической парадигмы*, чем шире класс решаемых задач, тем беднее выглядят возможности системы при решении конкретных проблем. Стратегия *квазилогической парадигмы* привела к созданию значительного числа языков представления знаний и трансляторов для них, но не породила резкого продвижения методов ИИ вперед.

Современная *семиотическая парадигма* позволила перейти к индустрии ИИС, поскольку была изначально ориентирована на «снабжение» ЭВМ множеством высококачественных специальных знаний о заданной проблемной области с целью интеллектуализации программного продукта. Этому в немалой мере способствовали успехи инженерии знаний, связанной с разработкой моделей и методов переноса знания из различных источников знаний в компьютерную программу, получившую впоследствии название экспертной системы (ЭС), а в более широком смысле – системы, основанной на знаниях.

В таблице 1.5 представлены ключевые события в истории ИИ и инженерии знаний, начиная с работы Дж. Маккалока и У. Питса в 1943 г. и до современных тенденций в комбинированных усилиях экспертных систем, нечеткой логики и нейронных вычислений в современных системах, основанных на знаниях, способных осуществлять вычисления при помощи слов [1].

На сегодняшний день исследования в области искусственного интеллекта ориентированы на создание смешанных человеко-машинных, или интерактивных интеллектуальных систем, на симбиоз возможностей естественного и искусственного интеллекта.

Таблица 1.5. Краткий перечень главных событий в истории ИИ и инженерии знаний

Период XX века	События
Рождение ИИ (1943-56 г.г.)	Маккалок и Питс: Логическое исчисление идей, присутствующих в нервной деятельности, 1943. Тьюринг: Вычислительная машина и интеллект, 1950. Шеннон: Программирование компьютера для шахматной игры, 1950.
Подъем ИИ (1956- конец 60-ых лет)	Маккарти: LISP – язык программирования искусственного интеллекта. Куллиан: Семантические сети для представления знаний, 1966. Ньюэлл и Саймон: Универсальный решатель задач (GPS), 1961. Минский: Структуры для представления знаний (фреймы), 1975.
Открытие и разработка ЭС (начало 70-ых – середина 80-ых лет).	Фейгенбаум, Буханан и др. (Стэнфордский университет): Экспертная система DENDRAL. Фейгенбаум, Шортлиф: Экспертная система MYCIN Стэнфордский исследовательский центр: Экспертная система PROSPECTOR Колмероз, Ковальски и др. (Франция): Язык логического программирования PROLOG.

Таблица 1.5 (окончание)

Период XX века	События
Возрождение искусственный нейронных сетей (после 1965 г.)	<p>Хопфилд: Нейронные и физические сети с эмергентными коллективными вычислительными способностями, 1982.</p> <p>Кохонен: Самоорганизующиеся топологически правильные карты, 1982.</p> <p>Румельхарт и Макклеланд: Распределенная параллельная обработка данных, 1986.</p>
Эволюционное вычисление (начало 70-ых и далее)	<p>Рехенберг: Эволюционные стратегии – оптимизация технических систем по принципам биологической информации, 1973.</p> <p>Холланд: Адаптация в естественных и искусственных системах, 1975.</p> <p>Коза: Генетическое программирование средствами естественного отбора, 1992.</p> <p>Фогель: Эволюционное вычисление – направление новой философии в машинном интеллекте, 1995.</p>
Нечеткие множества и нечеткая логика (середина 60-ых и далее)	<p>Заде: Нечеткие множества, 1965., Нечеткие алгоритмы, 1969.</p> <p>Мамдани: Применение нечеткой логики в приближенном рассуждении с использованием лингвистического синтеза, 1977.</p>
Вычисления при помощи слов (конец 80-ых и далее)	<p>Нейгоца: Экспертные системы и нечеткие системы, 1985.</p> <p>Коско: Нейронные сети и нечеткие системы, 1992; Нечеткое мышление, 1993.</p> <p>Ягер и Заде: Нечеткие множества, нейронные сети и мягкие вычисления, 1994.</p> <p>Коско: Нечеткая инженерия, 1996.</p> <p>Заде: Вычисления при помощи слов, 1996.</p>

Важнейшими проблемами в этих исследованиях является оптимальное распределение функций между естественным и искусственным интеллектом и организация диалога между человеком и машиной.

1.5. Отличительные свойства ИИС

Несмотря на представленную выше историю, до настоящего времени не существует единого определения, которое однозначно описывает научную область под названием *«искусственный интеллект»*. Среди многочисленных точек зрения и суждений по данному поводу доминируют три основных [2]. Согласно первой из них, разработки в области ИИ относятся к фундаментальным исследованиям, в процессе которых разрабатываются новые модели и методы решения задач, традиционно считавшихся интеллектуальными, и не поддававшихся ранее формализации и автоматизации. Согласно второй точке зрения, ИИ – это направление, связанное с новыми идеями решения задач на ЭВМ, с разработкой новых технологий программирования и с переходом к компьютерам «нефоннеймановской» архитектуры. Третья точка зрения, которая представляется наиболее прагматической, основана на том, что в результате исследований, проводимых в области ИИ, появляется множество прикладных систем, способных решать задачи, для которых ранее создаваемые системы были непригодны. По последней трактовке ИИ являются экспериментальной научной дисциплиной, в которой роль эксперимента заключается в проверке и уточнении ИИС, представляющих собой аппаратно-программные информационные комплексы.

Обобщением указанных трех подходов может быть мнение одного из основоположников вычислительной техники А. Тьюринга, который полагал, что компьютер или систему ИИ можно считать разумными, если они способны заставить нас поверить в то, что мы имеем дело не с машиной, а с человеком. Задачей ИИ, таким образом, является придание ИИС, а точнее их прикладным процессам, способности обучаться и «думать». При этом, естественно, предполагается, что алгоритм решения задачи, выполняемой ИИ, заранее никому не известен. По аналогии с теорией компьютерных наук, можно говорить о том, что ИИ вместо последовательного программирования с заранее создаваемыми программами, предполагает динамическое формирование программы из накопленных в памяти ЭВМ фактов, навыков и правил, которые машина умеет самостоятельно применить к сложившейся (заданной, конкретной) ситуации.

Таким образом, сегодня ИИ в большей мере связан с наукой и инновационными технологиями, чем с коммерческой сферой. Сегодня это совокупность понятий и идей, предназначенных в основном для достижения исследовательских целей – хотя одновременно являющихся научной основой ряда успешно развивающихся коммерческих технологий. Главными прикладными технологиями ИИ, помимо ЭС, являются ИИС для поддержки решений, системы обработки естественного языка и понимания речи, а также ИИС на основе нечеткой логики, робототехника и сенсорные системы, компьютерное зрение и распознавание образов.



Рис. 1.6. Дисциплины, на которых базируется ИИ и его приложения

В процессе развития ИИС затрагивались различные науки и технологии: лингвистика, психология, философия, техническое и программное обеспечение ЭВМ, механика и др. На рис. 1.6 представлены основные дисциплины и приложения ИИ: пересечение интересов психологии и ИИ сосредоточено в областях когнитологии и психолингвистики; философия и ИИ сотрудничают в логике, философии языка и разума. Пересечения между инженерией и ИИ

включают обработку изображений, распознавание образов и робототехнику. Существенный вклад в сотрудничество с ИИ вносят менеджмент и теория организации (по вопросам принятия и реализации решений), а также физика, статистика, математика, теория управления, эвристическое программирование, информационные системы менеджмента и целый ряд других дисциплин.

По мнению большинства специалистов, сегодня ИИ соотносится с двумя базовыми идеями [1]. Во-первых, ИИ вовлечен в изучение мыслительных процессов человека (для понимания, что представляет собой интеллект); во-вторых, ИИ имеет дело с представлением этих процессов через специализированные машины (компьютеры и роботы).

Исследование термина *интеллектуальное поведение* позволяет выделить способности как ЛПР, так и ЭВМ, которые можно считать признаками их интеллектуальности:

- обучение или понимание из опыта;
- выявление смысла из двусмысленности или противоположных сообщений;
- быстрый и адекватный отклик на новую ситуацию (разнообразные реакции, гибкость);
- использование рассуждений при решении проблем и эффективном направлении поведения;
- использование относительной важности различных элементов в ситуации;
- мышление и рассуждение.

Таким образом, в целом действительно можно утверждать, что целью ИИ является создание программно-аппаратного комплекса, который имитирует человеческий интеллект. Отметим также, что в обширной литературе по проблеме ИИ находят применение следующие рабочие определения [101]:

- *искусственный интеллект* – это совокупность метапроцедур – представления знаний, рассуждений, поиска релевантной информации в среде имеющихся знаний, их пополнение, корректировка и т.п. – имитирующих деятельность человека;
- *интеллектуальная информационная система* – это аппаратный и информационно-программный комплекс, действие которого аналогично действию механизмов мышления человека и неотличимо от действий, которые принимались бы человеком-

экспертом, то есть высококвалифицированным профессионалом в данной предметной области.

Любая ИИС обладает рядом особенностей, к числу которых относятся [1; 101; 106]:

- ориентация на эксплицитное (осознанное) представление знаний о предметной области, где они будут работать. Напомним, что *эксплицитные знания* базируются на вербальных формах представления, *имплицитные* – как на вербальных, так и на невербальных. Эксплицитные знания могут быть представлены в двух формах: первая – это хорошо структурированные знания, отражаемые в семантических сетях, вторая – вербальные ассоциации, существующие временно, на период коммуникативного акта. Имплицитность выражается в смутном представлении каких-либо объектов и отличается обрывочностью, хаотичностью вербальных и иных ассоциаций;
- развитые коммуникативные способности – характеризуют способ взаимодействия (интерфейса) конечного пользователя с системой, в частности, предполагают возможность формулирования произвольного запроса в диалоге с ИИС на языке, максимально приближенном к естественному;
- умение решать сложные плохоформализуемые задачи, то есть задачи, которые требуют построения оригинального алгоритма решения в зависимости от конкретной ситуации, для которой могут быть характерны неопределенность и динамичность исходных данных и знаний;
- способность к самообучению – возможность автоматического извлечения знаний для решения новой конкретной задачи из накопленного опыта предыдущих конкретных ситуаций;
- адаптивность – способность к развитию системы в соответствии с объективными изменениями модели проблемной области;
- наличие моделей механизмов мышления, то есть метапроцедур, работающих в систем знаний, представленных моделью предметной области;
- интеллектуальные систем основаны на символическом представлении и манипуляции: в ИИ символ представляет собой букву, слово или число, которые используются для представления объектов и их отношений.

Несмотря на все свои уникальные свойства, ИИС представляет собой относительно узкий класс «обычных» программных систем, обладающий, однако, рядом отличительных характеристик (см. таблицу 1.6) [1]. Центральным компонентом, который делает ИИС «особенной», является использование *базы знаний* (БЗ), которая содержит факты, понятия и отношения между ними, поскольку все свои рассуждения и заключения ИИС делает на основе использования БЗ. Содержимое БЗ реализуется в символьном виде, поэтому возникает важный вопрос о том, каким именно образом ИИС с ее помощью может осуществлять свои рассуждения и делать свои заключения.

Таблица 1.6. Сравнение ИИ с традиционным программированием

Характеристики	Искусственный интеллект	Традиционное программирование
Обработка	В основном символьная	В основном алгоритмическая
Характер входной информации	Может быть неполной	Должна быть полной
Поиск	Эвристический (в большинстве)	Алгоритмический
Объяснение	Обеспечивается	Обычно не обеспечивается
Главный интерес	Знания	Данные, информация
Структура	Управление отделено от знаний	Управление интегрировано с информацией (данными)
Характер выходной информации	Может быть неполной	Должна быть правильной
Сопровождение и модернизация	Легко осуществимы	Обычно затруднительно
Техническое обеспечение	Главным образом рабочие станции и персональные компьютеры	Все типы
Способности к рассуждению	Ограничены, но улучшаются	Нет

Ограничимся указанием на то, что базовой техникой здесь являются поиск и сопоставление образцов, а специфика алгоритма работы состоит в следующем: получив первоначальную информацию, компонент ИИС (так называемый решатель) просматривает БЗ, осуществляет поиск специфичных условий или образцов, после чего ищет подобия и соответствия, которые удовлетворили бы установленному критерию решения задачи, после чего приступает к ее решению.

Несмотря на то, что решение задач при помощи ИИ не реализуется напрямую алгоритмически, алгоритмы используются для осуществления процесса поиска. Поэтому в целом ИИС следует воспринимать как узкий класс систем, позволяющих имитировать процесс мышления человека и помогающего, помимо прочего, понять, как думает человек и каким способом можно лучше всего использовать человеческий интеллект при решении управленческих задач. Кроме того, интеллектуальные ИТ и методы ИИ могут «продвинуть» сами компьютеры – сделать их более удобными для ЛПР и сделать доступными большие объемы знаний.

На сегодняшний день ясно, что ИИС занимают важное место в теории *интеллектуального управления* (ИУ) – данный термин был впервые предложен в 1971 г. японским ученым К. Фу, в 1993 г. с целью уточнения базовых понятий и определений в области ИУ в американском институте инженеров по электротехнике и радиоэлектронике IEEE была создана специальная рабочая группа возглавляемая П. Анцаклисом. В качестве общих целей ИУ обычно называют следующие:

- полнее использовать достигнутые знания об объекте и среде для того, чтобы обеспечить эффективное управление объектом на основе заданного критерия (например, в виде желаемой траектории движения, функционала качества или некоторого целевого множества);
- управлять в присущей человеку творческой (интеллектуальной) манере, прогнозируя изменения на объекте и в его внешней среде, сохраняя работоспособность СУ даже при больших указанных изменениях (путем, например, реконфигурации СУ, координируя, а возможно, и пересматривая цели и критерии качества управления);

– улучшать с течением времени показатели качества управления путем накопления и обработки экспериментальных знаний об объекте и его внешней среде.

В [15] приведены определения *систем управления*, интеллектуальных в «малом», «большом», «целом» (см. далее), и выделены пять принципов, на основании которых сформулированы данные определения.

1. Взаимодействие с реальным внешним миром с использованием информационных каналов связи: интеллектуальная СУ имеет своей целью активное воздействие на объект управления, используя для этого информацию, поступающую с датчиков параметров объекта и окружающей среды. Для формирования управляющих воздействий при этом может использоваться модель внешнего мира, хранимая в БЗ.

2. Принципиальная открытость СУ с целью повышения интеллектуальности и совершенствования собственного поведения – другими словами, важнейшими функциями, реализуемыми в составе интеллектуальных СУ, должны быть функции самонастройки, самоорганизация и самообучения. Выполнение этого принципа требует организации в СУ процесса приобретения и пополнения знаний.

3. Наличие механизмов прогноза изменений внешней среды и поведения СУ в динамично изменяющемся внешнем мире: данный принцип определяет такое важное качество системы, как способность «экстраполировать» свое поведение и изменения среды, обеспечивать возможность «выживания» системы в динамически изменяющемся мире за счет своевременной корректировки своих действий и представлений о внешнем мире.

4. Наличие многоуровневой иерархической структуры, построенной в соответствии с правилом повышения интеллектуальности и снижения требований к точности моделей по мере повышения уровня иерархии в системе (и наоборот). Согласно данному принципу, по мере продвижения к высшим уровням иерархии СУ должен повышаться удельный вес ее интеллектуальных функций при одновременном снижении требований к точности их выполнения (и наоборот).

5. Сохраняемость функционирования (возможно, с допустимой частичной потерей качества или эффективности) при разрыве свя-

зей или потере управляющих воздействий от высших уровней иерархии управляющей структуры, то есть обеспечение отказоустойчивость СУ.

В зависимости от того, в какой мере (полностью или частично) соблюдаются данные принципы, выделяются:

- СУ, интеллектуальные «в малом» – это системы, удовлетворяющие хотя бы одному из указанных принципов и использующие в процессе своего функционирования знания (например, в виде правил) как средство преодоления непрерывности входной информации, описания управляемого объекта или его поведения;
- СУ, интеллектуальные «в большом» – системы, организованные и функционирующие в соответствии с перечисленными пятью принципами в полном их объеме;
- СУ, интеллектуальные «в целом» – системы, обеспечивающие, помимо регулирования объекта, выбор необходимых обратных связей, терминальных условий и критериальной базы, определяющих оперативную (текущую) цель управления, а также реконфигурацию способа регулирования и программы действий, исходя из заданной высшей цели функционирования СУ.

Отличительными признаками наиболее совершенной из них – интеллектуальной «в целом» СУ, являются:

- непрерывный контроль и слежение за выполнением цели управления и ее коррекция, что приближает СУ к интеллектуальному «пониманию» собственного предназначения;
- подчиненность друг другу задач регулирования планирования тактических действий и выбора стратегии поведения СУ;
- расширение информационных каналов и усложнение задач классификации по результатам контроля состояния как объекта, так и его внешней среды, влияющих на мотивацию цели и выбор программы действий СУ;
- использование при решении трудно-формализуемых задач интеллектуального типа процедур манипуляции со знаниями, учитывающих ограничения на имеющиеся ресурсы.

1.6. Классификация ИИС

Развитие методов и технологий ИИ привело к развитию значительного числа ИИС в самых разных областях промышленности, медицины, химии, банковской сферы, телекоммуникаций и т.п.

Наибольшее распространение в мире нашла классификация ИИС, представленная в [106]. Из других вариантов отметим оригинальную классификацию, предложенную в [97]. В обобщенном виде и с учетом тенденций развития исследований в области ИИ, в данном разделе предпринята попытка классифицировать ИИС с учетом различных классификационных признаков, результаты которой иллюстрирует рис. 1.6.



Рис. 1.6. Классификация интеллектуальных ИС

Рассмотрим ветвь данной классификации, соответствующую признаку *технология представления знаний*. Рубрика *нечеткая логика* соответствует термину Fuzzy Logic, который был впервые введен американским ученым персидского происхождения Л. Заде в 1965 г. В основе нечеткой логики лежит понятие лингвистической переменной, то есть такой переменной, которая выражается не числом, а словом на естественном языке. Прилагательное «fuzzy» (в переводе нечеткий, размытый, ворсистый) было введено с целью дистанцироваться от традиционной математики и аристотелевой логики, оперирующих четкими понятиями: «принадлежит – не принадлежит», «истина – ложь».

Теория нечетких множеств позволяет описывать нечеткие понятия и знания, оперировать ими и делать нечеткие выводы – такого рода ИИС основаны на правилах продукционного типа, однако в качестве посылки и заключения в них используются лингвистические переменные, что позволяет избежать ограничений, присущих классическим продукционным правилам.

Рассмотрим конкретный пример. В четкой булевой алгебре символ «1» представляет истину, а «0» – ложь. То же имеет место и в нечеткой логике, но, кроме того, здесь используются также все дроби между 0 и 1, чтобы указать на ту или иную частичную истину. Запись в отношении роста человека X : « P (высокий X) = 0,75» говорит о том, что предположение « X – высокий», в необходимом нам смысле на три четверти истинно, но на одну четверть является ложным. Слабым звеном нечеткой логики, однако, является функция принадлежности: предположим, что X имеет возраст 35 лет – насколько истинно тогда предложение, что он старый? Равна ли эта величина 0,5 – если считать, что X прожил примерно полжизни? Или оценка вроде 0,4 или 0,6 будет более реалистичной? Ответа этот вопрос дать нельзя.

Тем не менее, основанные на данной теории методы построения компьютерных нечетких систем существенно расширяют области применения ЭВМ. Больше того: в последнее время управление сложными системами становится одной из самых активных и результативных областей применения теории нечетких множеств – нечеткие системы управления. Оказывается особенно полезным, когда технологические процессы являются слишком сложными для анализа с помощью общепринятых количественных методов, или когда доступные источники информации интерпретируются качественно, неточно или неопределенно, то есть в самых трудных, но наиболее часто встречающихся на практике ситуациях.

Экспериментальным путем и на практике показано, что нечеткая логика дает лучшие результаты, по сравнению с получаемыми при общепринятых алгоритмах управления. Во многом это объясняется тем, что она, во-первых, существенно ближе к человеческому мышлению и естественным языкам, по сравнению с традиционными логическими системами. Во-вторых, нечеткая логика обеспечивает эффективные формальные средства отображения неопределенностей и неточностей реального мира – а наличие математиче-

ских средств отражения нечеткости исходной информации позволяет построить модели, адекватные реальности.

В настоящее время СУ, основанные на нечетких множествах, разработаны и успешно внедрены в таких областях, как управление технологическими процессами и транспортом, медицинская и техническая диагностика, финансовый менеджмент, биржевое прогнозирование, распознавание образов. Спектр приложений нечеткой логики очень широк – от видеокамер и бытовых стиральных машин до средств наведения ракет и управления вертолетами. В бизнесе и финансах нечеткая логика получила признание после того как в 1988 г. ЭС на основе нечетких правил для прогнозирования финансовых индикаторов единственная предсказала биржевой крах. Настоящий бум данного научно-технологического направления продолжается в Японии – в «стране восходящего солнца» успешно функционирует специализированная лаборатория LIFE (Laboratory for International Fuzzy Engineering Research), программой которой является создание близких человеку вычислительных устройств. Сегодня LIFE объединяет 48 компаний, в числе которых находятся Hitachi, Mitsubishi, NEC, Sharp, Sony, Honda, Mazda, Toyota; из иностранных участников LIFE можно выделить: IBM, Fuji Xerox, а также NASA.

Относительными недостатками нечетких СУ являются:

- отсутствие общей стандартной методики их конструирования;
- невозможность математического анализа нечетких СС существующими методами;
- применение нечеткого подхода по сравнению с вероятностными методами не приводит к повышению точности и достоверности результатов проводимых вычислений.

Эволюционные вычисления – это направление ИИ использует и воспроизводит биологическую эволюцию, его алгоритмы: эволюционное программирование, генетические алгоритмы, эволюционные стратегии и т.д. моделируют базовые положения теории биологической эволюции – процессы отбора, мутации и воспроизводства. Идею генетических алгоритмов высказал Дж. Холланд в конце шестидесятых лет XX века [100]: он заинтересовался свойствами процесса естественной эволюции (в том числе фактом, что эволюционируют хромосомы, а не сами живые существа) и пришел к идее составить и реализовать в виде компьютерной программы ал-

горитм, который будет решать сложные задачи так, как это делает природа – то есть путем эволюции, постепенного развития живых объектов. Роль хромосом в генетических алгоритмах играют последовательности двоичных цифр (единиц и нулей), по аналогии с объектами живой природы получивших название хромосом – данные алгоритмы имитировали эволюционные процессы в поколениях указанных искусственных хромосом.

В области ИИС и СУ генетические алгоритмы отражают эволюцию методов решения проблем, и, в первую очередь, задач оптимизации – это процедуры поиска, основанные на механизмах естественного отбора и наследования, где используется принцип выживания наиболее приспособленных особей. Генетические алгоритмы отличаются от традиционных методов оптимизации базовыми элементами и в частности:

- обрабатывают не значения параметров самой задачи, а их закодированную форму;
- осуществляют поиск решения, исходя не из единственной точки, фиксирующей СС в пространстве состояний, а из некоторой их популяции;
- используют только целевую функцию, а не ее производные либо иную дополнительную информацию;
- применяют вероятностные, а не детерминированные правила выбора.

Достоинство генетических алгоритмов в том, что этот метод очень гибок, и, будучи построенным в предположении, что о внешней среде нам известен лишь минимум информации (как это часто бывает в технических СС), алгоритм успешно справляется с широким кругом проблем, особенно в тех задачах, где не существует общеизвестных алгоритмов решения или высока степень априорной неопределенности знаний об СС.

Мультиагентные (многоагентные) технологии – это направление ИИ, которое для решения сложных задач или проблемы использует системы, состоящие из множества взаимодействующих агентов. Ключевым элементом мультиагентных (МА) систем является программный агент, способный воспринимать ситуацию, принимать решения и коммуницировать с другими агентами. Для таких ИИС и СУ характерна интеграция в БЗ нескольких разнородных источников знаний, обменивающихся между собой получае-

мыми данными на динамической основе. Решение задачи одним агентом на основе инженерии знаний представляет собой точку зрения классического ИИ, согласно которой агент, обладая глобальным видением проблемы, имеет все необходимые способности, знания и ресурсы для ее решения. Напротив, в распределенном ИИ и вообще в области МА-систем предполагается, что отдельный агент может иметь лишь частичное представление об общей задаче и способен решить лишь некоторую ее часть – то есть свою подзадачу. Поэтому для решения сколько-нибудь сложной проблемы, как правило, требуется взаимодействие агентов, которое неотделимо от организации МА-системы.

Главными достоинствами МА-систем являются их универсальность и гибкость, благодаря способности к самоорганизации. Каждая такая система может быть дополнена и модифицирована без переписывания значительной части программы, эти ИИС обладают способностью к самовосстановлению и обладают устойчивостью к сбоям при наличии достаточного запаса компонентов, а также указанной самоорганизации, они самостоятельно находят оптимальное решение задачи без внешнего вмешательства – при этом под оптимальным понимается такое решение, на которое тратится наименьшее количество ресурсов заданного вида при наличии одинаковых ограничений на другие ресурсы.

Нейронные сети можно рассматривать как современные вычислительные СС, которые преобразуют информацию по образу процессов, происходящих в мозгу человека. Обработка информации имеет численный характер, что позволяет использовать нейронную сеть, например, в качестве модели объекта с совершенно неизвестными характеристиками. Нейронную сеть моделирует направленный граф со взвешенными связями, где искусственные нейроны являются узлами.

В основе нейронных сетей лежит преимущественно поведенческий подход к решаемой задаче: сеть «учится на примерах» и подстраивает свои параметры при помощи так называемых алгоритмов обучения через механизм обратной связи. Реализованные на основе искусственных нейронных сетей ИИС позволяют решать проблемы распознавания образов, выполнения прогнозов, оптимизации, ассоциативной памяти и управления СС. В области ИИ данное направление стабильно держится на первом месте – се-

годня здесь активно совершенствуются алгоритмы обучения и классификации объектов в масштабе реального времени, обработки естественных языков, распознавания изображений, речи, сигналов, а также создание моделей интеллектуального интерфейса, подстраивающегося под пользователя. Среди других актуальных задач, решаемых с помощью нейронных сетей – финансовое прогнозирование, «добыча» (поиск и обнаружение) данных, диагностика СС, контроль за деятельностью сетей, шифрование данных и т.д. В последние годы идет усиленный поиск эффективных методов синхронизации работы нейронных сетей на параллельных устройствах.

«Классическая» алгоритмическая обработка – под классическим алгоритмом принято понимать описанную на некотором языке точную конечную систему правил, определяющую содержание и порядок действий над некоторыми объектами, строгое выполнение которых дает решение поставленной задачи. Поскольку большинство знаний в системах ИИ представлено продукциями (правилами вида «если ... то...»), алгоритмическая обработка представляет собой логически упорядоченную последовательность команд, необходимых для решения компьютером поставленной задачи.

Перейдем к ветви классификации **экспертные системы (ЭС)** – рубрики которой, согласно рис. 1.6, объединяют ИИС, в которые «включены» знания специалистов о заданной предметной области и которые в пределах данной области способны принимать нужные пользователю эвристические решения. Подробней о них речь пойдет далее, здесь же ограничимся кратким указанием на наиболее интересные современные решения по реализации идей в области ЭС.

Медицинские и химические ЭС – в историческом плане это самые «старые» ИИС. Назначение медицинских ЭС – консультировать врачей при постановке диагноза пациента (как в простых, так и в более сложных случаях) путем формирования гипотез диагноза с определением для каждой из них определенного «веса» - на чем процесс ее функционирования завершается, поскольку самостоятельного выбора окончательного диагноза ЭС, как правило, не делает. Система ABEL, например, разработанная специалистами Массачусетского технологического института (США) и доведенная ими до уровня исследовательского прототипа [101], помогает кли-

ницистам диагностировать нарушения кислотно-щелочного и водно-солевого баланса у пациентов, применяя знания о заболеваниях и вызываемых ими симптомов. Конкретное назначение данной ЭС – на основе причинно-следственных моделей возможных заболеваний упорядочить вопросы клинициста к пациенту для направления процесса его логических рассуждений к наиболее вероятному диагнозу.

Близкой ЭС по назначению и принципу действия является диагностическая система MYSIN, где был использован статистический подход для выбора схемы лечения на основе анализа чувствительности организма к тем или иным лекарственным препаратам, которая до настоящего времени применяется для обучения врачей. Среди химических ЭС выделим систему DENDRAL для анализа состава почв, которая использовалась при разведке месторождений полезных ископаемых – эта ЭС интересна тем, что генерация гипотез в ней была основана на алгоритме перечисления вершин плоского графа.

Промышленность – в основе ЭС, принадлежащих данной области, лежит идея создания полностью автоматизированных киберзаводов, гибких экономных производств, синхронизация цепочек снабжения и т.п. Система PTRANS [101] помогает управлять производством и распределением компьютерных систем компании DEC: она использует описание заказа клиента и информацию о работе завода, чтобы разработать план сборки и тестирования заказанной компьютерной системы, в том числе определяет сроки ее изготовления. Задача ЭС: наблюдать за тем, как персонал справляется с выполнением данного плана, диагностирует возникающие трудности, предлагает способы их преодоления и прогнозирует возможные тормозящие нехватки материалов или их излишки. Система PRTANS, разработанная компанией DEC совместно с Университетом Карнеги-Меллон (США) и доведенная до уровня исследовательского прототипа – это основанная на правилах ЭС с прямой цепочкой рассуждений, реализованная на языке OPS5.

Банки и финансовые учреждения – данная сфера является в настоящее время одним из «пионеров» широкого и эффективного использования ЭС. О масштабах исследований и предложений на рынке ЭС в области финансов можно судить как по регулярным международным конференциям, проводимым на самом высоком

уровне (например, Artificial International Applications on Wall Street, New York), так и по коммерческим каталогам, где цена конкретных приложений колеблется в пределах \$ 10³... 10⁵. Ряд ведущих корпораций США установили ЭС для решения задач в таких специфических областях, как торги на фондовой бирже, автоматический анализ новостей, управление рисками, построение портфелей кредитов и инвестиций, оценка рейтинга банков, кредитный анализ, автоматизация аудита, предсказание изменений на финансовом рынке и т.д. Компании Bear; Sterns & Company's Broker Monitoring System; Athena Group's Portfolio Advisor и Trader's Assistant и др. пользуются консультативными ЭС, которые разработаны Author D. Little Corporation и Knowledge-Based Network Corporation совместно с финансовыми институтами США.

В Японии один из крупнейших мировых банков Sanwa Bank применяет ЭС Best Mix для улучшения качества своей информации по инвестициям. Система Nikko Portfolio Consultation Management System, разработанная для внутреннего использования фирмой Nikko Securities, Ltd., помогает управляющим фондами выбрать оптимальный портфель для своих клиентов. Данная ЭС основана на БД с информацией за пять лет продаж акций и на СУ с нового поколения, которая вычисляет и оптимизирует портфель ценных бумаг для страховки от различных рисков; она освобождает управляющие фондами от рутинных вычислений и дает им возможность более быстро составить оптимальный портфель ценных бумаг.

Компания IDS Financial Services, подразделение финансового планирования корпорации American Express Company, использовала знания в области финансовые экспертизы своих лучших управляющих для создания ЭС Insight, доступной всем ее планировщикам и другим ЛПП. Одним из основных результатов применения данной ЭС в компании IDS стало то, что процент покинувших фирму клиентов упал более чем наполовину. Система логического программирования финансовой экспертизы FLiPSiDE фирмы Case Western Reserve University, которая для проведения финансового анализа и прогнозирования использует искусственные нейронные сети, решает следующие задачи: мониторинг состояния рынка ценных бумаг; мониторинг состояния текущего портфеля ценных бумаг; поддержка обзора будущих условий рынка; планирование и выполнение продаж. Фирма Promised Land Technologies предлагает

недорогой пакет, обещающий существенное улучшение эффективности инвестиций. Chase Manhattan Bank применяет гибридную систему распознавания образов с нейронной сетью для оценки риска при выдаче займов. Фирма Foster Ousley Conley использует систему, разработанную в компании Hecht-Nielsen Corporation, для оценки стоимости собственности в Калифорнии. Система Target Marketing System используется компанией Veratex Corporation для оптимизации рыночной стратегии. Фирма Spiegel Incorporation. использует программы, созданные компанией Neural-Ware Incorporation для определения потенциальных покупателей, что ведет к экономии до \$ 1 млн. в год за счет увеличения продаж и сокращения затрат на работу с малоперспективными покупателями.

Телекоммуникации – система ACE, разработчиком которой является всемирно известная компания Bell Laboratories (США), прошла опытную эксплуатацию и доведена на уровня коммерческой системы [101]. Данная ЭС определяет неисправности в телефонной сети и дает рекомендации по необходимому ремонту и восстановительным мероприятиям, она работает без вмешательства ЛПП, анализируя сводки-отчеты о состоянии оборудования ИКК, получаемые ежедневно с помощью программы CRAS, которая следит за ходом ремонтных работ в кабельной сети. После того, как ЭС обнаруживает неисправные телефонные кабели и решает, нуждаются ли они в планово-предупредительном ремонте, она выбирает, какой тип ремонтных работ, вероятнее всего, будет наиболее эффективным. Свои рекомендации ACE хранит в специальной БД, к которой у пользователя есть доступ; представление знаний в ЭС основано на общих принципах и правилах формирования БЗ, используется схема управления посредством прямой цепочки рассуждений. Система ACE реализована на OPS4 и FRANZ LISP и работает на микропроцессорах серии AT&T 3B-2, размещенных на подстанциях наблюдения состояния кабеля.

ЭС диагностики оборудования – примером является система DART [101], предназначенная для оказания ЛПП помощи при диагностировании неисправностей компьютерных систем с использованием сведений о конструкции и предполагаемом поведении каждого диагностируемого устройства. Система помогает ЛПП как при эксплуатации действующих сетей, так и при поиске проектных недочетов в заново создаваемых устройствах. Данная ЭС реализо-

вана на языке MRS, она была разработана и доведена до уровня исследовательского прототипа в Стэнфордском университете США. В результате ее применения, например, в японской сталелитейной компании Nippon Steel [91] диагностика более 500 единиц электромеханического оборудования осуществляется каждые 10 мс – причем ЭС за это время регистрирует и обрабатывает показания 13 000 датчиков и выносит решение о текущем состоянии производственного процесса в целом.

Ветвь классификации на рис. 1.6 «*универсальные приложения*» начнем с рубрики *OLAP-анализ и Data mining*. Термином OLAP (Online Analytical Processing – аналитическая обработка в реальном времени) здесь обозначается технология обработки данных, заключающаяся в подготовке суммарной (агрегированной) информации на основе больших массивов данных, структурированных по многомерному принципу. В основе OLAP-технологий лежит представление информации в виде OLAP-кубов, которые в экономике содержат бизнес-показатели, используемые для анализа и принятия управленческих решений: прибыль, рентабельность, совокупные активы, собственные средства, заемные средства и т.д. При этом бизнес-показатели хранятся в кубах не в виде простых таблиц, как в обычных системах учета или бухгалтерских программах, а в разрезах, представляющих собой основные бизнес-категории деятельности организации: товары, магазины, клиенты, время продаж и т.д. Благодаря детальному структурированию информации, OLAP-кубы позволяют оперативно осуществлять анализ данных и формировать отчеты в различных разрезах и с произвольной глубиной детализации. Процесс построения OLAP-куба, однако, не является тривиальной задачей, поскольку без продуктивной комплексной переработки потока данных просто накапливаются в архивах и не приносят никакой практической ценности.

В качестве иллюстрации приведем отрывок из книги А. Азимова «Ловушка для простаков», который достаточно удачно иллюстрирует недостатки разрозненных данных с точки зрения получения новой информации: «Существует Земля, где половина населения обслуживает правительство и только то и делает, что считает, а на всех других планетах есть вычислительные центры. И все равно многие сведения теряются. Каждая планета знает что-то такое, чего не знают другие. Даже почти каждый человек. Возьмите нашу ма-

ленькую группу. Вернадский не знает биологии, а я ничего не понимаю в химии. Ни один из нас, кроме Фоукса, не мог бы пилотировать самый простой патрульный космолет. Поэтому мы и работаем вместе: каждый приносит те познания, которых не хватает другим. Но тут есть одна зацепка. Ни один из нас не знает точно, что именно из того, что он знает, важно для других при данных обстоятельствах. Мы же не можем сидеть и рассказывать друг другу все, что знаем. Поэтому приходится гадать, и не всегда правильно. Например, есть два факта, *A* и *B*, которые очень хорошо вяжутся друг с другом. И *A*, который знает факт *A*, говорит *B*, который знает факт *B*: «Почему же ты мне это не сказал десять лет назад?». *A* *B* отвечает: «Я не знал, что это так важно», или «А я думал, об этом все знают».

Термин *Data Mining* переводится как «добыча» или «раскопка данных», что означает исследование и обнаружение компьютером (алгоритмами, средствами ИИ) в сырых данных скрытых знаний, которые ранее не были известны – нетривиальных, практически полезных, доступных для интерпретации ЛПР. Нередко *Data Mining* трактуют как «обнаружение знаний в БД» и «интеллектуальный анализ данных» – эти определения можно считать синонимами *Data Mining*.

По сути *Data Mining* – это процесс анализа БД, направленный на поиск новой полезной информации, обычно представленной в форме ранее неизвестных отношений между переменными. В основу современной версии технологии *Data Mining* (*Discovery-Driven Data Mining*) положена концепция шаблонов (паттернов), отражающих фрагменты многоаспектных взаимоотношений в анализируемых данных. Эти шаблоны представляют собой закономерности, свойственные «подвыборкам» данных, которые могут быть компактно выражены в понятной ЛПР форме. Поиск указанных шаблонов производится методами, не ограниченными рамками априорных предположений о структуре выборке и виде распределений значений анализируемых показателей.

Технологию *Data Mining* часто применяют для обработки больших неструктурированных данных, получивших название *Big Data*. Данный термин появился как альтернатива СУБД и стал одним из модных трендов ИТ-индустрии. Под *Big Data* понимают огромный (сотни терабайт) массив данных, который нельзя обработать тра-

диционными способами; реже – инструменты и методы обработки этих данных.

Признаки Big Data определяются как «три V»: Volume – объем (действительно большие); Variety – разнородность, множество; Velocity – скорость (необходимость очень быстрой обработки). Примеры источников Big Data: события RFID, сообщения в соцсетях, метеорологическая статистика, информация о местонахождении абонентов сетей мобильной сотовой связи и данные с устройств аудио- и видеорегистрации. Поэтому «большие данные» широко используются на производстве, в здравоохранении, госуправлении, интернет-бизнесе – в частности, при анализе целевой аудитории.

Приведем пример использования Big Data для бизнеса оператора связи. В общем случае здесь выделяют четыре основные направления – первые три нацелены на улучшение внутренней работы самой компании, а последнее является дополнительным рыночным продуктом для внешних клиентов [66]:

- высокоточный маркетинг (precise marketing) – адресное предложение продуктов и услуг тем потребителям, которые наиболее готовы к их приобретению (новые тарифные планы, дополнительные сервисы, платежные терминалы и пр.);
- управление качеством услуг для клиента (Customer Experience Management) для повышения его удовлетворенности с целью предотвращения оттока пользователей;
- оптимизация внутренней работы оператора и планирование развития (ROI-based Network Optimization and Planning) на основе учета всех объективных факторов и мнений потребителей с целью максимальных гарантий возврата инвестиций в кратчайшие сроки;
- монетизация информационных активов (Data Asset Monetization) – продажа в той или иной форме (в том числе в виде долевого участия в проектах) имеющихся у оператора данных партнерам, чтобы они могли с их помощью решать свои задачи.

Развернув решение больших данных Huawei Telco Big Data, один из ведущих китайских сотовых операторов China Unicom, стал собирать и анализировать существенно больше информации о поведении и интересах своих клиентов, в том числе об интенсивности использования сотовой связи и географическом местополо-

жении. Причем все эти сведения можно было увязывать с данными о работе самой сотовой сети, в том числе о ее загрузке, о возникающих сбоях и пр.

Возможности применения подобных методов видны по полученным результатам. В начале 2013 г. эффективность маркетинговых предложений (для клиентов, которые их приняли) при общей массовой рассылке составляла 0,7%. К концу года, за счет сегментации абонентов (по возрасту, полу, сроку подписки), эта величина была доведена до 4%, а в течение 2014 года – повышена сначала до 11% (учет интенсивности использования услуг и местоположение клиентов) и затем до 24% (учет предпочтительных вариантов получения предложения – голосовые звонки, SMS, E-mail, социальные сети и пр.). За год удалось сократить число нерезультативных обращений к клиентам на 11 млн., существенно снизив затраты на рекламные кампании.

На основе анализа 85 параметров поведения абонентов была выделена «группа риска», потенциально готовая к уходу от оператора. Внутри нее также была проведена сегментация, и для каждой категории клиентов выработан комплекс мероприятий по повышению уровня их лояльности (скидки, другие тарифные планы, подарки и пр.). Было проведено исследование, в рамках которого «группу риска» разделили на две подгруппы: с первой проводились специальные действия по удержанию, в другой подгруппе не делалось ничего. Анализ работы за год показал, что компания смогла существенно сократить отток своих действующих потребителей, удержав более 200 тыс. абонентов – при этом нужно учитывать, что стоимость удержания клиента всегда значительно ниже, чем привлечения нового пользователя.

До использования больших данных расширение географической сети оператора фактически выполнялось только на основе информации о плотности застройки и населения, но после этого решения China Unicom перешел к развитию своей деятельности на базе многофакторного анализа, который учитывал такие показатели, как реальная загруженность трафика и востребованность услуг (например, с учетом места работы людей), «ценность» клиентов (по уровню жизни), требования к качеству связи (расстояние между станциями приема), востребованность разных категорий услуг (от которой зависит использование различной аппаратуры) и пр.

В плане монетизации клиентских данных для внешних партнеров были приведены два примера: во-первых, оптимизация размещения наружной рекламы, причем как в географическом плане (место проживания, работа или транспортные коммуникации нужных клиентов), так и с учетом времени для динамической рекламы (в зависимости от времени суток, дней недели и сезонов года состав публики может меняться), а во-вторых, аналогичные предложения по развитию торговых сетей (с учетом местоположения и ассортимента). Кроме того, эффективной и выгодной оказалась целевая рассылка мобильной рекламы в реальном времени в соответствии с графиком занятости человека, его интересов и физического пребывания (например, рассылка информации о фильмах-боевиках, которыми клиент интересуется, именно в его свободное время и с учетом близлежащих кинотеатров). Опыт показал, что такие адресные методы позволяют повысить доходы от распространения рекламы в несколько раз.

Другой из подобластей Data Mining является технология Text Mining, которая ориентирована на обработку текстовой информации [27]. Задачей Text Mining является анализ не синтаксиса, а семантики значения текстов, выбор из него наиболее значимой для пользователя информации – по аналогии и в тесной связи с контент-анализом. В практическом плане выделяют следующие приложения Text Mining: реферирование текстов на естественном языке; классификацию (тематическое индексирование) текстовых документов; кластеризацию текстовых документов и их фрагментов; построение онтологии текстового документа (основных терминов и связей между ними в виде семантической сети); визуализация полученных знаний.

Автоматизированный анализ естественных языков – основными направлениями данного анализа являются лексический, морфологический и терминологический анализ, выявление незнакомых слов, распознавание национальных языков, перевод, коррекция ошибок, эффективное использование словарей. Сфера применения ИИС подобного рода – системы перевода, справочные и поисковые системы, программы автоматического реферирования и аннотирования текстов. Естественно-языковой интерфейс ЭВМ предполагает трансляцию естественно-языковых конструкций на внутримашинный уровень представления знаний.

Чтобы из области фантастики (где в книге П. Буля «Идеальный робот», например, описан процесс сочинения рассказа носителем ИИ), эти ИИС перенести в реальный мир, необходимо решить достаточно сложные задачи морфологического, синтаксического и семантического анализа и синтеза высказываний на естественном языке. **Морфологический анализ** здесь «отвечает» за распознавание и проверку правильности написания слов, в ходе данного анализа осуществляется разбор каждого слова в отдельности (включая определение части речи, рода, числа, падежа, времени и т.д.). **Синтаксический анализ** реализует разложение входных сообщений на отдельные компоненты, проверку на соответствие грамматическим правилам внутреннего представления знаний и выявление недостающих частей. **Семантический анализ** обеспечивает установление смысловой правильности синтаксических конструкций и т.д.

Робототехника – это активно развивающаяся прикладная наука, специалисты которой заняты разработкой автоматизированных «человекоподобных» по виду и принципам функционирования технических СС – что как мечта привлекала людей с древнейших времен, тогда как сам термин появилось в двадцатые годы XX века благодаря чешскому писателю К. Чапеку. На сегодняшний день известны три поколения таких систем: на смену роботам с жесткой схемой управления (это практически все современные промышленные роботы – программируемые манипуляторы), приходят адаптивные роботы с сенсорными устройствами – сегодня уже есть их лабораторные образцы, которые в промышленности пока не используются. Цель робототехники завтрашнего дня – самоорганизующиеся (интеллектуальные) роботы, основным препятствием на пути создания которых остается проблема машинного зрения.

Когнитивная графика – относящиеся к данной рубрике ЭС позволяют осуществлять интерфейс пользователя с ИИС при помощи графических образов, которые генерируются в соответствии с происходящими событиями. Графические образы в наглядном интегрированном виде описывают множество параметров изучаемой ситуации, освобождая ЛПР от анализа тривиальных ситуаций, способствует быстрому освоению программных средств и повышению конкурентоспособности разрабатываемых ИИС. Системы когнитивной графики широко используются также в обучающих и

тренажерных системах на основе использования принципов виртуальной реальности, когда графические образы моделируют ситуации, в которых обучаемому необходимо принимать решения и выполнять определенные действия.

Обработка изображений – основными задачами в области представления и анализа изображений являются сжатие, кодирование при передаче с использованием различных протоколов, обработка биометрических образов, снимков с искусственных спутников Земли и других мобильных (летательных) аппаратов. Дальнейшие развитие в этих ИИС получают средства поиска, индексирования и анализа смысла изображений, согласования содержимого справочных каталогов при автоматической каталогизации, организация защиты от копирования, а также машинное зрение, алгоритмы распознавания и классификации образов.

Самоорганизующиеся СУБД – данные СУ предназначены для управления БД и призваны гибко подстраиваться под профиль решаемой конкретной задачи без какого-либо внешнего администрирования.

Ветвь классификации ИИС по признаку *оперативность* является наиболее естественной и традиционной: в нее входят статические, квазидинамические и динамические ИИС. **Статические системы** разрабатываются в предметных областях, где БЗ и интерпретируемые данные не меняются во времени.

Квазидинамические системы интерпретируют ситуацию, которая меняется с некоторым фиксированным интервалом времени. **Динамические системы** работают в сопряжении с датчиками объектов в режиме реального времени с непрерывной интерпретацией поступающих данных.

Ветвь **модель представления знаний** на рис. 1.6 открывает рубрика **логические модели**, где фигурируют ИИС, основанные на системе исчисления предикатов первого порядка. Данные методы являются системами дедуктивного типа, поскольку в них используется модель получения вывода из заданной системы посылок с помощью фиксированной системы правил вывода. Дальнейшим развитием предикатных систем являются системы индуктивного типа, в которых правила вывода порождаются системой на основе обработки конечного числа обучающих примеров.

Продукционная модель в силу своей простоты получила наиболее широкое распространение. В этой модели знания представляются в виде совокупности правил, а само продукционное правило представляет собой упорядоченную пару цепочек символов, которая содержит левую и правую части – посылку и следствие. На множестве правил продукции для представления знаний задано отношение порядка, парадигма использования данных правил состоит в том, что левой части ставится в соответствие некоторое условие, тогда как правой части – действие в соответствии со схемой: ЕСЛИ <перечень условий>, ТО <перечень действий>. В такой интерпретации левая часть правил сравнивается по отношению к БД системы, и, если эта оценка в определенном смысле соответствует логическому значению ИСТИНА, то выполняется действие, заданное в правой части продукции. Важное место в продукционных моделях занимают стратегии вывода, то есть перехода от одного правила к другому.

Модель семантической сети – это модель, основой нотации которой является формализация знаний в виде ориентированного графа с размеченными вершинами и дугами. Вершинам соответствуют объекты или ситуации, дугам – отношения между ними. Вершины могут соответствовать общим понятиям, константам, типовым переменным, событийным фреймам, фреймам-характеристикам, логическим функциям и предикатам. Дуги представляют теоретико-множественные, логические, квантифицированные и лингвистические отношения. Термин «семантическая» при этом эквивалентен термину «смысловая», поскольку семантика – это наука, устанавливающая отношения между символами и объектами, которые они обозначают, то есть наука, определяющая смысл знаков.

Систематизация отношений конкретной семантической сети зависит от специфики знаний предметной области и является достаточно сложной задачей. В семантических сетях декларативные и процедурные знания не разделены – поэтому БЗ не отделена от механизма вывода, а сама процедура логического вывода обычно представляет совокупность процедур обработки сети. Семантические сети получили широкое применение в системах распознавания речи и ЭС.

Фреймовая модель представления знаний основана на теории фреймов М. Минского, которая представляет собой систематизированную психологическую модель памяти человека и его сознания. Эта теория имеет весьма абстрактный характер, поэтому только на ее основе создание конкретных языков представления знаний невозможно. Отметим, что на практике редко удается обойтись рамками одной модели и, кроме комбинированного представления знаний при помощи разных моделей, разработчики ИИС обычно используют специальные средства, позволяющие отразить особенности конкретных знаний о предметной области, а также различные способы устранения и учета неточности и неполноты знаний.

В заключение отметим еще одно важное обстоятельство: непреходящий самостоятельный интерес представляют исследования в области ИИ, ориентированные на интересы военных и других силовых ведомств. Значительные интеллектуальные усилия и материальные ресурсы сегодня вкладываются в исследования по распознаванию речи, создаются экспертные и консультационные системы, призванные автоматизировать рутинные работы и снизить нагрузку на экипажи боевых машин (самолетов, надводных и подводных кораблей, танков и т.п.).

Нейронные сети применяются для обработки сигналов сонаров при различении подводных камней и мин, генетические алгоритмы используют для эвристического поиска решения уравнений, определяющих работу военных устройств (систем ориентации и навигации), а также в задачах распознавания образов – для разделения искусственных и естественных объектов, типов военных машин, анализа изображений, получаемых от телекамер с низким разрешением или инфракрасных датчиков. Примером исследований в области нейронных сетей, позволяющих получить хорошие (хотя и приближенные) результаты при решении сложных задач управления войсками, является проект Smart Sensor Web, финансируемый военным научным агентством DARPA. Данный проект предусматривает организацию распределенной сети датчиков разного типа, синхронно работающих на поле боя – каждый такой объект (стоимостью не более \$ 300) в составе сети представляет собой источник данных – визуальных, электромагнитных, цифровых, инфракрасных, химических и т.п. Поскольку проект потребовал новых математических методов решения многомерных задач оптимизации, в

его рамках ведутся работы по автоматическому распознаванию целей, анализу и предсказанию сбоев военной и вспомогательной техники по отклонениям от типовых параметров ее работы (например, по звуку) и т.п. Операции, проводимые США на Ближнем востоке («Буря в пустыне» и др.), стали стимулом к развитию ЭС с продвинутым ИИ в области снабжения войск. На разработках, связанных с технологиями машинного зрения, основано современное высокоточное оружие.

1.7. Выводы

Задача принятия решений – это наиболее распространенная задача, решаемая в процессе управления любым объектом, в том числе СС. Данная задача обладает рядом типовых особенностей:

- наличие неопределенных (недостаточно известных) элементов задачи: цели, ограничения, критерии, приоритеты и др. обычно имеют содержательный характер и лишь частично определяются количественными характеристиками; причем число известных элементов и взаимосвязей между ними обычно существенно меньше, чем число неизвестных;
- неизвестные элементы задачи, способные повлиять на итоговое решение проблемы, как правило, до конца неформализовано;
- элементы задачи могут быть описаны характеристиками, часть которых измерена объективно, а часть – выбрана субъективно;
- условия неопределенности, при которых необходимо принимать решения, обусловлены неполным знанием проблемной области и недостаточно точной оценкой ожидаемых последствий.

Поиск решений – это многошаговый процесс, тесным образом связанный с анализом условий и целей деятельности ЛПР. С информационной точки зрения процесс принятия решений связан с уменьшением неопределенности знаний – при этом его представление (структурирование) в виде совокупности логически упорядоченных процедур позволяет определить модель процесса, на основе которой можно рационально организовать сбор, обработку и хранение необходимой информации.

Указанные процедуры могут выполняться путем мышления ЛПР, включая экспертов, то есть творчески (неформальным способом), а также с применением математических методов и соответствующих программных средств обеспечения, то есть формальны-

ми методами. Проблема информационной перегрузки ЛПР сегодня выводит на первый план целесообразность применения ИИС в повседневную деятельность по управлению СС. Область инфокоммуникаций не является здесь исключением, поскольку современные ИКК – владельцы ИКС – для поддержания рыночной конкурентоспособности идут по пути активного сближения ИТ и бизнеса. Сегодня деятельность ИКК представляет собой совокупность тесно переплетенных между собой сфер управления: оборудованием и взаимоотношениями с клиентами; биллингом и расчетами; системами и ресурсами; развитием и расширением номенклатуры услуг; управление персоналом и т.д. Это многообразие ведет к необходимости оптимизировать бизнес-процессы и непрерывно повышать их эффективность, одним из возможных направлений на этом пути является внедрение в управленческую практику средств ИИ.

История развития систем ИИ включает периоды как взлета, так и падения интереса со стороны специалистов конкретных областей знаний. Развитие методов и технологий ИИ идет по пути решения частных задач из самых разных областей промышленности, финансов, медицины, телекоммуникаций и т.д., что затрудняет классификацию ИИС. Приведенная в разделе 1.6 классификация систем ИИ иллюстрирует многогранность затрагиваемой области. В то же время деление ИИС по разным признакам во многом является условным, поскольку имеют место взаимное проникновение и смешение направлений ИИ, технологий и терминов, используемых при создании ИИС – что, в свою очередь, затрудняет понимание сути наблюдаемых явлений и процессов.

Несмотря на существование значительного числа ИИС в самых разных областях, их массового внедрения в практику управления СС пока что не происходит. Это обусловлено и трудностью создания ИИС как самостоятельного программного продукта, и неясностью содержания конкретных этапов по созданию ИИ, и сложностью формирования БЗ (путем «извлечения» знаний из экспертов как основных носителей истины). В последующих разделах монографии предпринята попытка осветить все эти трудности и проблемы более подробно.

2. ВВЕДЕНИЕ В ТЕОРИЮ ЭКСПЕРТНЫХ СИСТЕМ

2.1. История развития и архитектура ЭС

К настоящему моменту времени *экспертные системы* (ЭС) являются собой пример наиболее успешного с коммерческой точки зрения практического приложения ИИ. Именно они позволили создателям ИИ перейти от игр и головоломок к массовому использованию выдвинутых научных идей для решения практически значимых задач. Однако путь первоначального развития ЭС был достаточно тернист, что подтверждает перечень его ключевых этапов: интерес и увлечение новой наукой (1960 г.), период именованной лженаукой (1960-65 г.г.), успехи при решении головоломок и игр (1965-75 г.г.), разочарование при попытках решения практических задач (1970-85 г.г.), первые успехи на этом направлении (1992 г.), начало коммерческого использования при решении практических задач (1993-95 г.г.).

Своеобразной точкой отчета для работ по созданию ЭС принято считать 1965 г. – когда исследователь в области ИИ Э. Фейгенбаум задумал сконструировать «мыслящую» СС, которая в идеале сможет думать и рассуждать подобно человеку. Перед ученым сразу возник вопрос, какое научное направление для экспериментов следует выбрать – ответить на него помог нобелевский лауреат биохимик Д. Ледерберг, предложивший создать компьютерного помощника, способного определять путем расчета молекулярную структуру химических соединений. Такой помощник, по словам Ледерберга, был просто необходим в органической химии. В 1965 г. проект построения ЭС в Стэнфордском университете, где Э. Фейгенбауму и Д. Ледербергу примкнул Б. Бученен, получил название DENDRAL.

Одной из главных проблем, которую создателям ЭС надлежало решить, было построение гибкой программы, оперирующей с многочисленными знаниями и работающей по правилам логики («если – то»). Однако вскоре выяснилось, что сложнее всего создать базу данных(БД), включающую знания многих специалистов в органической химии. Для этого разработчикам DENDRAL пришлось опросить большое число экспертов, поскольку оказалось, что приобретение знаний – совсем не такой легкий процесс, как это представлялось на первый взгляд. Одно дело собрать разные факты,

совсем другое – аккумуляровать и формализовать познания конкретного человека. Поэтому опрашиваемым специалистам Б. Бученен предоставил специальную программу, которая производила некоторые «умозаключения», правдивость или ложность которых нужно было установить и объяснить – таким образом ему удалось отделить механизм логического вывода от *базы знаний* (БЗ) и предложить практический инструмент для создания ЭС.

Одной из самых первых таких программ была META-DENDRAL, ставшая эффективным вспомогательным средством для построения других ЭС. Среди многочисленных «потомков» DENDRAL следует упомянуть MYCIN – предназначенную для диагностики бактериальных инфекций и помощи при выборе антибиотиков; PUFF – для диагностики легочных болезней; ONCOCIN – для назначения лечения при онкологических заболеваниях, а также MOLGEN и GENESIS, оказывающих помощь специалистам в области молекулярной биологии при планировании экспериментов по изучению ДНК и клонированию.

Начиная с семидесятых годов XX века, разработка ЭС становится ведущим направлением для исследований, проводимых в области ИИ. В этот период было создано множество разнообразных ЭС и диагностических систем, большая часть которых действует и сегодня. Рассмотрим более подробно особенности и возможности двух наиболее известных из них – упомянутой MYCIN и PROSPECTOR, предназначенной для геологической разведки месторождений полезных ископаемых.

Создателем первой версии MYCIN почти полвека назад в Стэнфордском университете был врач и специалист в области вычислительной техники Э. Шортлайф – с тех пор медицина является одной из самых популярных областей применения ЭС. Объяснить это достаточно просто: для успешного выздоровления пациента диагностика заболеваний должна производиться оперативно, а кроме того, каждый человек, идя на прием к врачу, надеется, что его примет профессионал, который внятно объяснит причину недомогания и предложит одну или несколько эффективных методик лечения. Любая медицинская ЭС, содержащая знания и логику рассуждения лучших специалистов в мире, может это реализовать.

Процесс диагностики с помощью MYSIN осуществляется следующим образом: доктор или сам больной вводит в MYCIN симп-

томы болезни, а ЭС задает уточняющие вопросы и, в конце концов, ставит возможный диагноз и предлагает методы лечения. Кроме того, ЭС на любом шаге может «объяснить» свои доводы и соображения. Механизм логического вывода в MYCIN включает первоначальный опрос пациента, прямой вывод с использованием правил продукции и нечеткой логики, а также обратный вывод.

По эффективности работы, как показали сравнительные исследования в области диагностики бактериальных инфекций, ЭС MYCIN уступает группе профессиональных врачей не более чем на 20%. При этом ее БЗ постоянно расширяется, благодаря чему ЭС «осваивает» все новые области медицины. В настоящее время MYCIN используется преимущественно для обучения врачей, а ее механизм логического вывода E-MYCIN был успешно применен для создания других ЭС: таких, как NEOMYCIN и PUFF для исследования легочных заболеваний.

Система PROSPECTOR, разработанная SRI International в 1974-83 г.г. для геологических изысканий, относится к интерпретирующему типу ЭС, которые выводят некоторые «умозаключения» по результатам наблюдений. Данная программа располагает динамическим количеством геологических моделей, каждая из которых содержит знания об определенных видах полезных ископаемых. Так же, как и MYCIN, PROSPECTOR вовлекает геолога в диалог с тем, чтобы опираясь на его наблюдения в качестве источника исходных данных, выбрать модель и дать ответ на вопросы: «Где бурить?» и «Где искать?» В 1984 г. система достаточно точно предсказала существование месторождения молибдена, оцененного в многомиллионную сумму.

При разработке последующих ЭС были учтены особенности и недостатки PROSPECTOR и MYCIN. Благодаря этому такие диагностические ЭС, как INTERNIST и CASNET, основанные на ассоциативном и казуальном (от casual – случайный) подходах, приобрели более мощные механизмы вывода. На основе системы CASNET для диагностики и выдачи рекомендаций по лечению глазных заболеваний был разработан язык инженерии знаний EXPERT, с помощью которого впоследствии создавался целый ряд других медицинских диагностических систем.

Немалый интерес представляют также ЭС «серии» HEARSAY – HEARSAY-2 – HEARSAY-3 – AGE. Первые две системы из этого

ряда являются развитием интеллектуальных информационных систем (ИИС) для распознавания слитной человеческой речи, слова которой берутся из заданного словаря. Эти ЭС отличаются оригинальной структурой, основанной на использовании доски объявлений. В дальнейшем на основе данных ЭС были созданы инструментальные системы HEARSAY-3 и AGE (Attempt to Generalize – попытка обобщения) для построения ЭС.

Системы AM (Artificial Mathematician – искусственный математик) и EURISCO были разработаны в Стэнфордском университете Д. Ленатом для исследовательских и учебных целей. По мнению разработчика данной ЭС, во-первых, чтобы система была способна к обучению, в нее должно быть введено около миллиона сведений общего характера. Это примерно соответствует объему информации, каким располагает четырехлетний ребенок со средними способностями. Во-вторых, поскольку ценность и эффективность любой ЭС определяется закладываемыми в нее знаниями, путь создания узкоспециализированных ЭС с уменьшенным объемом знаний является тупиковым.

В систему AM первоначально было заложено около 100 правил вывода и более 200 эвристических алгоритмов обучения, позволяющих строить произвольные математические теории и представления. Первые результаты работы ЭС были весьма многообещающими: система могла сформулировать понятия натурального ряда и простых чисел, она синтезировала вариант гипотезы Гольдбаха о том, что каждое четное число, большее 2, можно представить в виде суммы двух простых чисел (заметим, что до сих пор математикам не удалось ни найти доказательства данной гипотезы, ни опровергнуть ее). Однако в дальнейшем развитие ЭС замедлилось и было отмечено, что, несмотря на проявленные на первых порах «математические способности», она неспособна синтезировать новые эвристические правила, поскольку ее возможности определяются только теми эвристиками, которые были заложены в нее изначально.

При разработке ЭС EURISCO была предпринята попытка преодолеть указанные недостатки системы AM. Как и при эксплуатации AM, первые результаты, полученные с помощью EURISCO, были эффективными: сообщалось, что ЭС может успешно участвовать в очень сложных играх. С ее помощью в штабной игре, прово-

димой ВМФ США, была разработана стратегия, содержащая ряд оригинальных тактических ходов. Согласно одному из них, например, предлагалось взрывать свои корабли, получившие повреждения – на том основании, что корабли, оставшиеся неповрежденными, получают необходимое пространство для выполнения боевого маневра. Вскоре обнаружилось, что ЭС вообще не всегда корректно переопределяет первоначально заложенные в нее правила: так, например, она стала нарушать строгое предписание обращаться к программистам с вопросами только в определенное время суток. В итоге EURISCO, также как и ее предшественница AM, остановилась в своем развитии, достигнув предела, предопределенного в конечном счете ее разработчиком.

В восьмидесятые годы XX века существовало стойкое мнение, что ЭС позволят создать новое направление ИТ – машинную экспертизу, которая заменит труд специалиста. Однако число успешных практических применений ЭС было столь незначительно, что многие ЛПП пришли к выводу о том, что существующая на тот момент времени технология создания ЭС является тупиковым направлением в развитии ИТ. Компьютерные системы, которые могут лишь воспроизводить логические рассуждения и выводы своего создателя-эксперта, сегодня принято относить к ЭС первого поколения. Причинами их неудач сегодня считаются, во-первых, недооценка разработчиками ЭС объемов и роли неявных знаний, а, во-вторых, неудачный выбор модели данных и знаний, на которых были основаны первые экземпляры данных ЭС [71, 91].

В девяностые годы ЭС возродились в виде систем с БЗ (или систем, основанных на знаниях). «Овеществление» знаний специалистов, распространение знаний наиболее квалифицированных экспертов среди остальных коллег через взаимодействие с ЭС, использование ИИС такого рода в подготовке специалистов и, в частности, в составе различных тренажеров, позволили улучшить эффективность и качество их работы во многих практически важных областях.

Сегодня число ЭС в мире исчисляется десятками тысяч. В развитых зарубежных странах сотни фирм занимаются их созданием и внедрением в различные сферы жизни, активно ведутся разработки самостоятельно обучаемых ЭС. Кроме того, в проблеме ИИ обозначилось такое направление, как инженерия знаний, отвечающая

за поиски передовых методов в сборе, представлении, хранении и преумножении информации. В качестве современных ЭС можно назвать систему OMEGAMON компании Candle (с 2004 г. корпорация IBM) для отслеживания состояния корпоративной информационной сети, а также G2 фирмы Gensym – коммерческую ЭС для работы с динамическими объектами. Обе эти ЭС отличаются быстродействием и служат для принятия решений за считанные секунды с момента наступления внештатных или критических ситуаций.

В современном обществе неструктурированные и слабоструктурированные задачи управления и контроля сложных производственных процессов и объектов часто встречаются в таких областях, как авиация, энергетика, машиностроение, медицина, микроэлектроника и др. Поэтому многие отрасли промышленности сейчас не могут обходиться без применения ЭС в управлении технологическими процессами и процессами подготовки производства.

Приведем ряд определений, необходимых для понимания того, чем обусловлена уникальность и успешность ЭС.

Определение 1. ЭС – это сложные программные комплексы, аккумулирующие знания специалистов в конкретных предметных областях и тиражирующие этот эмпирический опыт для консультаций менее квалифицированных пользователей.

Определение 2. ЭС – это система ИИ, использующая знания из сравнительно узкой предметной области для решения возникающих в ней задач, причем так, как это делал бы эксперт-человек, то есть в процессе диалога с заинтересованным ЛПР, поставляющим необходимые сведения по конкретному вопросу.

Определение 3. ЭС – это машинная программа, которая пытается моделировать знания качественного характера и опыт человека-эксперта.

Анализируя приведенные определения, можно утверждать, что основная задача ЭС – тиражирование опыта и знаний ведущих специалистов практически во всех областях человеческой деятельности в интересах бизнеса конкретных ЛПР.

Согласно [59], знания традиционно существуют в двух видах – коллективный опыт и личный опыт (см. рис. 2.1). Если большая часть знаний о предметной области представлена в виде коллективного опыта (это относится, например, к таким дисциплинам, как высшая математика, начертательная геометрия и др.), то такая

предметная область не нуждается в ЭС (по крайней мере, создание их для такой предметной области является экономически нецелесообразным). Если же в предметной области большая часть знаний является «личной собственностью» специалистов высокого уровня (экспертов, профессионалов и других опытных ЛПР) и если эти знания слабо структурированы и неформализованы (например, для установления диагноза пациенту следует учитывать множество самых разных нюансов; при ведении биржевой торговли следует знать обстановку на рынках в мире и т.п.), то разработка ЭС здесь скорее всего объективно необходима и будет экономически оправдана.



Рис. 2.1. Предметные области знаний ЛПР
 а) непригодная для создания ЭС; б) пригодная для создания ЭС

Проведем некоторое (безусловно субъективное и поверхностное) сравнение ЭС с экспертами-людьми. **Преимущества** ЭС заключаются в следующем:

- **постоянство** – чтобы сохранить свой профессиональный уровень в некоторой предметной области, эксперт должен постоянно учиться, практиковаться и упражняться. Знания в ЭС сохраняются навсегда (если только не произойдет непредвиденная авария с памятью компьютера);
- **относительная легкость**, с которой можно передавать и воспроизводить знания с помощью ЭС – передача знаний от одного человека другому является трудоемким, долгим и дорогим процессом, который, как известно, называется обучением и подготовкой кадров, передача искусственной компетенции ЭС – это простой процесс копирования данных;

- **устойчивость** – эксперт-человек может принимать различные решения в однотипных ситуациях под влиянием различного рода эмоциональных факторов, чего не может быть в ЭС;
- **относительно невысокая стоимость** – эксперты, особенно высококвалифицированные, очень ценятся и, следовательно, обходятся очень дорого ЛПР, тогда как ЭС могут быть дороги в разработке, но существенно дешевле в эксплуатации.

К **слабым сторонам** ЭС могут быть отнесены следующие обстоятельства:

- люди обладают значительно большей **способностью к творчеству** и изобретательностью, чем даже самая умная программа – поэтому ЭС тяготеют к рутинному, лишённому творчества поведению, а эксперты лучше справляются с неожиданными поворотами событий с помощью воображения и новых подходов к решению задач;
- другой областью, где человеческая компетентность существенно превосходит искусственную, является **обучение**: опытные эксперты легко адаптируются к изменяющимся условиям, тогда как ЭС, напротив, мало приспособлены к обучению новым концепциям и правилам;
- эксперты способны **непосредственно воспринимать** весь комплекс входной сенсорной информации, то есть визуальную, звуковую, осязательную или обонятельную информацию – поскольку в распоряжении ЭС есть только символы, сенсорную информацию необходимо преобразовать в символьную форму, которая будет понятна системе, при этом теряется некоторая (возможно существенная) часть информации;
- эксперты-люди **могут исследовать** все аспекты проблемы и понять, как они относятся к основной задаче, ЭС, напротив, стремится сосредоточить все внимание на самой задаче, игнорируя те аспекты, которые, хотя и связаны с основной задачей, но не входят в нее явно.

В целом ЭС стоят обособленно как в классе прикладных ИС, так и в классе ИИС. От «обычных» прикладных программ ЭС отличает наличие **следующих особенностей** [90]:

- ЭС моделируют механизм мышления человека применительно к решению задач в конкретной проблемной области (а не физическую или иную природу данной проблемной области), при этом

основные усилия создателей ЭС направлены на то, чтобы понять и в программном виде смоделировать методiku решения проблем, применяемую экспертами – хотя, разумеется, нельзя сказать, что ЭС воспроизводит психологический портрет человека-эксперта (в настоящее время это невозможно, так как механизм мышления человека в нужной степени не изучен);

- ЭС, помимо выполнения вычислительных операций, формирует определенные соображения и выводы, основываясь на тех знаниях, которыми она располагает – как правило, эти знания в ЭС представлены на специальном языке и хранятся отдельно от собственно программного кода, поэтому мощность ЭС напрямую зависит от мощности БЗ, а также от используемого механизма обработки знаний и формирования заключений для конечного пользователя;
- при решении поставленных задач ЭС в большей степени оперирует эвристическими и приближенными методами, которые, в отличие от алгоритмических, не всегда гарантируют успешное (точное и достоверное) решение – эти эвристические методы приобретаются человеком по мере накопления практического опыта решения аналогичных проблем, поэтому всегда существует лишь определенная степень уверенности (или неуверенности) в том, что предлагаемое решение является верным;
- одна из отличительных черт ЭС – это выдача качественных, а не количественных выходных результатов: с одной стороны, это облегчает процесс взаимодействия пользователя с системой, а с другой стороны – достаточно сильно затрудняет процесс тестирования БЗ в ЭС.

По сравнению с другими видами программ из области ИИ современные ЭС имеют *следующие отличия*:

- ЭС ориентированы на решение конкретных практических задач, преследующих коммерческую выгоду, поэтому объектами манипуляции в них являются предметы реального мира, операции с которыми обычно требуют наличия значительного опыта, накопленного ЛПР – множество программ из области ИИ являются сугубо исследовательскими и основное внимание в них уделяется абстрактным математическим проблемам или упрощенным вариантам реальных проблем, а целью решения каждой такой проблемы является повышение «уровня интуиции» или

отработка конкретной методики (по аналогии можно указать на то, что хотя вряд ли кто-то будет оспаривать важность задач робототехники, следует иметь в виду, что имеющиеся на сегодня образцы сильно упрощены);

- требования к производительности, то есть скорости получения конечного результата и его достоверности (надежности) у ЭС существенно выше: эта их особенность вытекает из предыдущей, поскольку исследовательские программы ИИ могут не быть слишком быстрыми, в них можно мириться с отказами и сбоями в реальных ситуациях, тогда как сфера применения ЭС не допускает ничего подобного – здесь необходимо за приемлемо короткое время найти решение «не хуже» того, которое может предложить квалифицированный ЛПР в данной области;
- ЭС выгодно отличает «умение объяснить» предложенное решение и доказать его обоснованность: пользователь может получить всю необходимую информацию, чтобы быть уверенным в логичности предлагаемых действий, тогда как исследовательские системы зачастую «общаются» только со своим создателем, который заранее знает, на чем основывается тот или иной получаемый результат. В отличие от них, ЭС проектируются в расчете на взаимодействие с различными категориями конечных пользователей, для которых работа ЭС по возможности должна быть прозрачной и понятной, полученный ответ и ход решения должны быть объяснены доступным пользователю способом (характерный пример – дедуктивный метод Шерлока Холмса, демонстрирующий логическую цепочку умозаключения, чтобы ответ на поставленный вопрос стал очевидным).

Суммируя изложенное, можно сделать важный промежуточный вывод: ЭС содержит знания в определенной узкой проблемной области, накопленные в результате практической деятельности людей (ЛПР, экспертов, менеджеров) и использует их для решений проблем, специфичных для рассматриваемой области.

Типовая архитектура ЭС. На архитектуру современных ЭС свой отпечаток наложили их отличительные особенности, такие как:

- способность проводить экспертизу только в конкретной проблемной (предметной) области;
- БЗ и механизм вывода являются различными компонентами ЭС;

- неформализованность решаемых ЭС задач;
- решение задач, как правило, дедуктивным методом;
- умение объяснить ход решения задачи и мотивировать выводы понятным пользователю способом;
- ЭС строится по модульному принципу, что позволяет постепенно наращивать мощность БЗ.

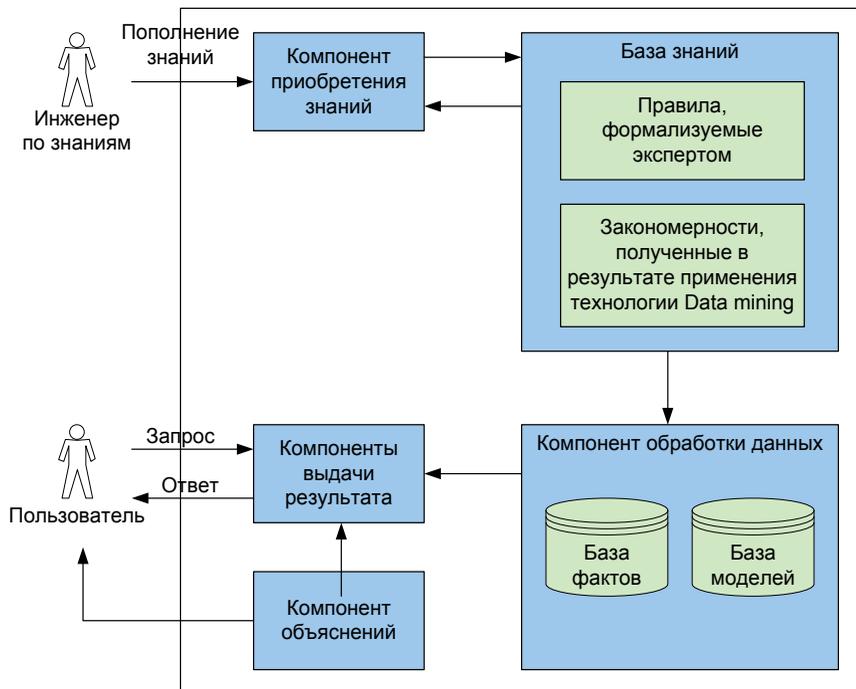


Рис. 2.2. Типовая архитектура ЭС

Типовую архитектуру ЭС демонстрирует рис. 2.2. В целом процесс функционирования ЭС можно представить себе следующим образом: пользователь, желающий получить необходимую информацию, через пользовательский интерфейс посылает запрос к ЭС; решатель, пользуясь БЗ и компонентом обработки данных, генерирует и выдает пользователю подходящую рекомендацию, объясняя ход своих рассуждений при помощи компонента объяснений.

База знаний представляет собой ядро ЭС – фактически это совокупность знаний в заданной предметной области, записанная на

машинный носитель в форме, понятной эксперту и пользователю. Знания, хранимые в базе знаний, можно разделить на правила, формализуемые экспертом, и закономерности, полученные в результате применения технологии Data Mining. Большая часть **правил, формализуемые экспертом**, реализуется при помощи выбранного метода формализации знаний в процессе устного общения инженера по знаниям с экспертом, данная часть БЗ соответствует процессу извлечения знаний. **Закономерности, полученные при помощи Data Mining**, являются результатом процесса приобретения знаний. Как было отмечено в разделе 1.6, Data Mining (в русском переводе: добыча данных, интеллектуальный анализ данных, глубинный анализ данных) представляет собой собирательное название, используемое для обозначения совокупности методов обнаружения в данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых ЛПР для принятия решений в различных сферах человеческой деятельности. Отметим, что в некоторых ЭС в составе БЗ выделяют так называемую **базу прецедентов**, под которой понимают результаты обследования экспертов-людей и информацию о последующей успешности (неуспешности) их деятельности.

Компонент приобретения знаний – данный компонент в составе БЗ, которую часто называют сердцем ЭС, отвечает за ее объем и наращивание мощности. При этом под приобретением знаний понимается передача потенциального опыта решения проблем от некоторого источника знаний и преобразование его в вид, позволяющий использовать эти знания в программном коде. Основным источником знаний, как правило, выступает человек-эксперт (опытный ЛПР, наставник, менеджер). Передача данных осуществляется в результате длительных и пространственных собеседований между специалистом по проектированию ЭС (так называемым **инженером по знанию**) и **экспертом**. При этом обязательным условием является способность эксперта четко и логично сформулировать его опыт в виде логической цепочки. По имеющимся оценкам, таким образом за день можно сформировать от двух до пяти «единиц знаний» – причина столь низкой производительности может заключаться как в эксперте (ввиду его неумения вербально выразить те или иные знания), так и в инженере по знаниям (из-за неумения задавать наводящие вопросы и кратко формулировать

получаемые знания). Функцию приобретения знаний считают «узким местом» в технологии разработки ЭС по следующим причинам:

- эксперты, являющиеся профессионалами в своей довольно узкой предметной области, часто используют «жаргонизмы» – профессиональные обозначения или термины, которые довольно трудно понять и перевести на обычный язык. Поскольку смысл жаргона не всегда очевиден, требуется достаточно много дополнительных вопросов для уточнения его смыслового (логического или математического) содержания – специалисты по военной стратегии, например, охотно говорят об «агрессивной демонстрации» военной мощи противника, но не могут объяснить, чем такая «агрессивная» демонстрация отличается от демонстрации, не несущей угрозы;
- факты и принципы, лежащие в основе большинства конкретных областей знаний, не могут быть четко сформулированы в терминах математической теории или детерминированных моделей, свойства которых ясны ЛПР – так, например, эксперту в финансовой области может быть известно, что определенные события способны стать причиной роста или снижения котировок на фондовой бирже, но он ничего не скажет точно о механизмах, которые приводят к такому эффекту, или о количественной оценке влияния данных факторов – тогда как статистические модели могут помочь сделать общий долговременный прогноз, но, как правило, не работают в отношении курсов конкретных акций на коротких временных интервалах;
- для того, чтобы решить проблему в заданной предметной области, эксперту недостаточно обладать суммой знаний о фактах и принципах, «принадлежащих» этой области – например, опытный ЛПР знает, какого рода информацией нужно располагать для формулировки того или иного суждения, насколько надежны различные источники информации и как можно расчленить (декомпозировать) сложную проблему на более простые, которые можно решать независимо друг от друга – но выявить в процессе собеседования такого рода знания, основанные на личном опыте и плохо поддающиеся формализации, значительно сложнее, чем получить простой перечень каких-то фактов или общих принципов;

- экспертный анализ в узкой области, выполняемый человеком, очень часто как бы помещен в довольно обширный контекст, который включает атрибуты, кажущиеся эксперту само собой разумеющимися, но для постороннего ЛПР отнюдь таковыми не являющимися.

Компонент выдачи результата – в литературе данный компонент ЭС также называют «решатель», «интерпретатор» или «механизм вывода». Его основное назначение – моделировать ход рассуждений эксперта на основании знаний, имеющихся в БЗ. Фактически механизм вывода запускает ЭС в работу, определяя, какие правила надо вызвать и организовав к ним доступ в БЗ. Поскольку ЭС предназначены для решения слабоформализованных (или неформализованных) задач, с точки зрения организации управления это приводит к тому, что процесс решения нельзя представить в виде детерминированной последовательности правил, то есть невозможно выявить множество всех мыслимых ситуаций, которые возникнут в ходе работы ЭС при различных входных данных.

Основные отличия управляющей компоненты от других элементов ИИС состоят в следующем:

- отдельные правила вызываются не по имени, а по описанию ситуации;
- способ взаимосвязи правил формируется в процессе решения задачи, так как выбор очередного правила зависит от текущей ситуации и не может быть сформирован заранее.

Иллюстрацией здесь может быть вопрос о том, чем отличается мужская логика (формализованная задача) от женской (неформализованная задача) – если допустить, что предугадать решение мужчины сложно, но все-таки можно, то предугадать действия дамы невозможно в принципе (ибо она сама не всегда знает, чем объяснить свои действия).

Механизм работы компонента выдачи результата напрямую зависит от используемой модели представления знаний. Наибольшее распространение получила продукционная модель, при которой компонента выдачи результата, как правило, выполняет две функции:

- просмотр существующих данных (фактов) из компонента обработки данных и правил (или продукций) из БЗ и добавление по мере возможности новых фактов в компонент обработки данных;

– определение порядка просмотра и применения правил (разрешение конфликтного множества).

В подавляющем большинстве ИИС, основанных на знаниях, механизм вывода представляет собой некую программу и содержит два компонента, один из которых реализует собственно вывод, а другой управляет этим процессом. Действие **компонента вывода** основано на применении правила *modus ponens* – «Если существует правило вида «Если А, то В» и известно, что истинно утверждение А, то тогда утверждение В также истинно» Таким образом, правила срабатывают, когда находятся факты, удовлетворяющие их левой части: если истинная посылка, то должно быть истинным и соответствующее ей заключение.

Управляющий компонент определяет порядок применения правил и выполняет четыре функции:

- сопоставление – образец правила сопоставляется с имеющимися фактами;
- выбор – если в конкретной ситуации могут быть применены сразу несколько правил, то из них выбирается одно, наиболее подходящего по заданному критерию (разрешение конфликта);
- срабатывание – если образец правила при сопоставлении совпал с какими-либо фактами из рабочей памяти, это правило срабатывает;
- действие – рабочая память подвергается изменению путем добавления в нее заключения сработавшего правила, при этом если в правой части содержится указание на какое-либо действие, то оно выполняется.

Компонент выдачи результата работает циклически – как это показано на рис. 2.3. В каждом цикле он просматривает все правила, чтобы выявить те, посылки которых совпадают с известными на данный момент фактами из компонента обработки данных. После выбора правило срабатывает, его заключение заносится в компонент обработки данных, затем цикл повторяется снова. В одном цикле может сработать только одно правило. Если несколько правил успешно сопоставлены с фактами, то интерпретатор производит выбор по определенному критерию единственного правила, которое срабатывает в данном цикле.

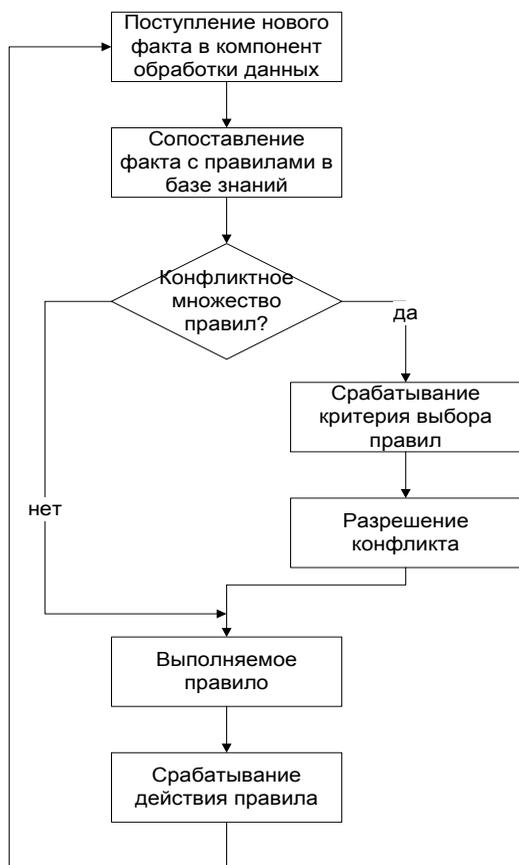


Рис. 2.3. Цикл работы компонента ЭС выдачи результата

Информация из компонента обработки данных последовательно сопоставляется с посылками правил для выявления успешного сопоставления. Совокупность отобранных правил составляет так называемое конфликтное множество. Для разрешения конфликта компонент обработки данных имеет критерий, с помощью которого он выбирает единственное правило, которое срабатывает – это выражается в занесении фактов, образующих заключение правила, в компонент обработки данных или в изменении критерия выбора конфликтующих правил. Если в заключение правила входит название какого-либо действия, то выполняется это действие.

Компонент обработки данных – основным назначением данного компонента является хранение данных, соответствующих объектам, связь между которыми задается правилами в БЗ. Распространенными синонимами данного компонента являются также «рабочая память» и **база данных**. Термин БД по названию, но не по смыслу, совпадает с термином, используемым в информационно-поисковых системах и СУ для работы с этими БД (их принято именовать СУБД) с целью обозначения всех данных (в первую очередь долгосрочных), хранимых в ИС [91]. Во избежание недоразумений и путаницы, использование термина БД в рамках перспективных ЭС некоторыми специалистами не рекомендуется.

Компромиссным решением, на взгляд авторов монографии, может быть выделение в составе данного компонента двух составляющих: базы фактов и базы моделей. **База фактов** предназначена для хранения исходных (фактов) и промежуточных данных, решаемой в текущий момент задачи. На основе базы фактов запускается работа интерпретатора.

Наряду с обеспечением доступа к данным, ЭС должна обеспечивать доступ пользователя к **базе моделей**, применение которых обеспечивает способность ЭС к проведению анализа, необходимого для принятия решений. Модели используют математическую интерпретацию проблем и при помощи определенных алгоритмов способствуют нахождению информации, полезной для принятия корректных решений.

Компонент объяснений – его основное назначение заключается в разъяснении принятого системой решения. Способность ЭС объяснить методику принятия решений иногда называют прозрачностью системы, под которой понимается, насколько просто пользователю выяснить, что делает ЭС и почему. Представление информации о поведении ЭС важно по следующим причинам:

- пользователи, работающие с системой, нуждаются в подтверждении того, что в каждом конкретном случае заключение ЭС является корректным;
- инженеры по знаниям должны убедиться, что сформулированные ими знания применены правильно (то есть поставленная задача решена корректно и верно);

- экспертам желательно проследить ход рассуждений ЭС и конкретный способ использования тех сведений, которые с их слов были введены в БЗ;
- разработчикам необходимо иметь инструмент, позволяющий видеть «внутренний мир» ЭС для дальнейшей модернизации и наращивания ее функциональности.

2.2. Экспертные системы как продукт эволюции ИС

В начальный период применения ЭВМ их основной функцией в сфере экономики и бизнеса была обработка данных, включавшая расчеты заработной платы, учета товарно-материальных ценностей и другие бухгалтерские операции. Такого рода ИС называют сегодня **системами обработки данных** (СОД) – при их создании ставится задача автоматизировать трудоемкие, регулярно повторяющиеся и рутинные операции над большими массивами данных. Основные функции СОД – это сбор, хранение, поиск и обработка данных, необходимых для выполнения экономических расчетов, с наименьшими производственными затратами.

По мере развития компьютерных технологий их применение все более перемещалось в сферу управления, для создания и обработки информации, необходимой менеджменту для принятия решений. Появились **автоматизированные системы управления** (АСУ), создающие управленческие отчеты, предназначенные для повышения эффективности бизнеса. В функции АСУ входили расчеты, связанные с решением задач управления, выбором оптимальных вариантов планов на основе экономико-математических методов, моделей и т.п. Дальнейшее совершенствование АСУ в направлении их специализации на решение плохо структурированных задач привело к появлению **систем поддержки принятия решений** (СППР).

Особенностями современных СППР являются [80]:

- ориентация на решение плохо структурированных и слабоформализованных задач, которые характерны, главным образом, для высоких уровней управления;
- возможность сочетания традиционных методов доступа и обработки компьютерных данных с возможностями математических моделей и методов решения задач на их основе;

- направленность на непрофессионального пользователя посредством использования диалогового режима работы;
- высокая адаптивность, обеспечивающая возможность приспосабливаться к особенностям имеющегося технического и программного обеспечения, а также к требованиям пользователя.

Дальнейшее развитие СППР привело к появлению интеллектуальных ИИС в целом и ЭС в частности, моделирующих в определенной мере процессы человеческого мышления. Современные ИИС и ЭС сосредоточили в себе наиболее наукоемкие технологии с высоким уровнем автоматизации не только процессов подготовки информации для принятия решений, но и самих процессов выработки вариантов решений, опирающихся на полученных ИС данные. Созданы ИИС и ЭС, способные диагностировать состояние предприятия, оказывать помощь в антикризисном управлении, обеспечивать выбор оптимальных решений по стратегии развития предприятия и его инвестиционной деятельности.

Благодаря наличию средств естественно-языкового интерфейса появляется возможность их применения в качестве средств поддержки процессов анализа, оценки и принятия экономических решений бизнес-пользователями, не владеющими языками программирования. Сферой успешного применения ИИС и ЭС, помимо текущего анализа деятельности предприятия, являются стратегическое планирование и инвестиционный анализ, оценка рисков и формирование портфеля ценных бумаг, углубленный финансово-экономический анализ, маркетинг и т.п. [97]. Перечисленные компьютерные системы часто выступают в качестве подсистем единой корпоративной ИИС: при этом выходные данные СОД и АСУ являются входными по отношению к ЭС и т.п. Однако принципиальным и наиболее существенным отличием СОД и АСУ от ИИС и ЭС принято считать то обстоятельство, что в СОД и АСУ обработке и использованию подлежат **данные**, а в ИИС и ЭС – **знания** ЛПР, с учетом того, что эти знания обычно рождаются в результате переработки значительных по объему данных. Таким образом, поскольку изучение вопросов формирования знаний без понимания основ обработки данных невозможно, цель данного раздела – краткий обзор вопросов организации и обработки данных в рамках «традиционных» ИС обработки данных, а также проблем организации знаний в ЭС.

Автоматизированные системы управления. В соответствии с названием, СОД служат для решения задач, содержащих большое число операций по обработке данных. К таким операциям относятся сбор и хранение данных, манипулирование ими, подготовка документов. Поэтому СОД в большей степени ориентированы на данные, имеющие временной аспект, а не на информацию, которая содержится в них. Опыт создания и эксплуатации СОД позволил определить следующие принципы их построения и методы разработки:

- **принцип интеграции:** обрабатываемые первичные данные вводятся в СОД один раз; решаемые задачи взаимно увязываются таким образом, чтобы первичные данные и данные, являющиеся результатом решения одних задач, использовались как исходные для возможно большего числа других задач – что устраняет дублирование операций сбора, подготовки и контроля данных, обеспечивается их комплексное использование;
- **принцип централизации** обработки данных: большая часть информации аккумулируется в едином репозитории СОД;
- **принцип системной организации** технологического процесса: при создании СОД необходимы комплексная механизация и автоматизация операций на всех этапах сбора и обработки данных, сопряжение применяемых технических средств по пропускной способности и другим техническим параметрам.

Разработке СОД предшествует детальное обследование и анализ объекта управления, состава и структуры СУ, потоков и содержания информации. На основе материалов обследования разрабатывается информационная модель СОД, фиксирующая связь между задачами обработки данных, выявляются новые потоки информации. По определяемым на основе информационной модели СОД объемам перерабатываемых, хранимых и передаваемых данных производится выбор технических средств и разрабатывается технология реализации СОД.

По сравнению с СОД, АСУ используется при худшей структурированности решаемых задач – например, когда система осуществляет поиск и обработку входных данных, поступающих из СОД корпорации и ее окружения (внешней среды). В АСУ имеется возможность работы в запросно-ответном режиме за счет появления в ее составе **системы управления базами данных** (в принятых

обозначениях – СУБД); АСУ выдает выходную информацию в виде управленческих отчетов, осуществляющих фильтрацию и агрегирование данных, представляемых в удобном для принятия решения виде. В то же время АСУ не предназначены для использования в чисто автоматическом режиме, все решения, на любом уровне управления объектом, здесь принимает ЛПР.

В общем случае АСУ можно рассматривать как совокупность взаимосвязанных процессов и объектов, главной целью управления которыми является повышение эффективности использования потенциальных возможностей объекта.

Также выделяют ряд других целей:

- предоставление ЛПР релевантных данных для принятия управленческих решений;
- ускорение выполнения операций по сбору и обработке данных;
- снижение числа решений, которые должен принимать ЛПР;
- повышение уровня контроля и исполнительской дисциплины;
- повышение оперативности процесса управления объектом;
- снижение затрат на выполнение вспомогательных процессов;
- повышение степени обоснованности принимаемых решений.

К основным характеристикам АСУ относят:

- подготовку информации, полезной на уровне среднего управленческого персонала;
- структурирование (упорядочение) информационных потоков;
- интеграцию (объединение) данных, получаемых от СОД по функциональным сферам бизнеса (АСУ производства, АСУ технологических процессов и т.п.);
- создание запросно-ответной системы и составление отчетов руководству с использованием БД.

Таким образом, СОД и АСУ дают возможность удовлетворить информационные потребности пользователя посредством оперативного и своевременного доступа к любым необходимым данным для получения отчетов (построенных с различной степенью обработки данных), облегчающих процесс принятия решений. В случае применения СППР и ИИС, напротив, правильнее говорить о способности системы создавать новую информацию (часто в виде готовых альтернатив) для принятия решений. Заметим также, что термин *автоматизированные СУ* в широком смысле включает в себя любые виды рассмотренных компьютерных систем обработки

информации (СОД, АСУ, СППР и пр.), используемых в интересах ЛПР. В узком смысле он означает вид информационной системы, производящих управленческие отчеты.

Процесс создания АСУ, являющейся стратегическим инструментом управления, обладает приоритетом перед другими технологиями, но подчиняется основным стратегиям и целям предприятия, среди которых важно правильно определить его место. Существует ряд методов и вариантов разработки АСУ, использование которых зависит от различных факторов: таких как размеры компании, ее структура, уже внедренные ИС, цели создания новых ИС, имеющиеся ресурсы и др. Опыт разработки и эксплуатации разных классов современных ИС демонстрирует высокую эффективность (в том числе экономическую) их применения, особенно на крупных предприятиях – которая отражается в хорошей организации труда и производства, повышении точности планирования и реализации поставленных задач, в обеспечении ритмичности работы компании, эффективном (в том числе оперативном) информационном обеспечении различных категорий пользователей и т.д. В то же время возможны варианты, зависящие, например, от начальных условий, в которых ведется разработка: создается ли новая система по результатам обследования предприятия при существующей модели его бизнес-процессов, или в компании уже есть ИС, которая может быть использована в качестве прототипа для новых бизнес-процессов, или разрабатываемой ИС должна быть интегрирована с действующей и т.п.

Детализированная разработка проекта АСУ предполагает наличие полного комплекта организационной, конструкторской, технологической и эксплуатационной документации. Напомним, что проектирование любого нового объекта предполагает:

- определение его функционального назначения (зачем нужен, что и как будет делать);
- выявление его логических связей (как осуществляет свое функциональное назначение, какая информация и в какой последовательности будет обрабатываться);
- выбор материальных средств реализации – функционально-технологический и технический аспекты (носители, средства обработки данных и др.);

- определение пространственного (территориального) размещения материальных средств реализации на выделенных или возможных для использования площадях;
- формирование организационно-управленческой структуры объекта (состав подразделений, полномочия и функциональные обязанности работников).

Таблица 2.1. Этапы проектирования АСУ

Номер	Наименование этапа	Характеристика	Результат
1	2	3	4
1	Формирование требований к АСУ	Обследование объекта и обоснование необходимости создания АС, определяются основные задачи АСУ, проводится декомпозиция задач по модулям и определяются функции, с помощью которых решаются эти задачи	Отчет о выполненной работе, в котором описание функций осуществляется на языке производственных (описание процессов предметной области), функциональных (описание форм обрабатываемых документов) и технических требований (аппаратное, программное, лингвистическое обеспечение)
2	Разработка концепции АСУ	Формализация бизнес-модели, разработка логической модели бизнес-процессов	Отчет о выполненной работе, включающий: – концептуальную модель АСУ, состоящую из описания предметной области, ресурсов и потоков данных, перечень требований и ограничений к технической реализации АСУ

Таблица 2.1. Этапы проектирования АСУ (продолжение)

1	2	3	4
3	Техническое задание	Разработанная концептуальная модель формализуется, то есть воплощается в виде логической модели АСУ; разрабатывается диаграмма «сущность-связь» (ER (Entity-Relationship) – CASE-диаграммы)	Техническое задание, включающее в том числе разработанное информационное обеспечение АСУ: схемы и структуры данных для всех уровней модульности АСУ, документацию по логической структуре АСУ
4	Эскизный проект	Разработка предварительных проектных решений по системе и ее частям	Эскизный проект; сгенерированные скрипты для создания объектов БД
5	Технический проект	Разработка АСУ: выбирается лингвистическое обеспечение (среда разработки – инструментарий), проводится разработка программного и методического обеспечения. Разработанная на предыдущих этапах логическая схема воплощается в реальные объекты, при этом логические схемы реализуются в виде объектов БД, а функциональные схемы – в пользовательские формы и приложения	Технический проект; разработка программного кода с использованием выбранного инструментария

Для автоматизации различных видов деятельности (управление, проектирование, исследование и т.п.), включая их сочетания, используются положения ГОСТ 34.601-90.

Таблица 2.1. Этапы проектирования АСУ (окончание)

1	2	3	4
6	Рабочая документация	Осуществляется корректировка информационного, аппаратного, программного обеспечения, проводится разработка методического обеспечения (документации разработчика, пользователя) и т.п.	Комплект документации: разработчика, администратора, пользователя
7	Ввод в действие	Проведение предварительных испытаний; проведение опытной эксплуатации; проведение приемочных испытаний	Работоспособная АСУ; оптимальный состав и эффективное функционирование АСУ
8	Сопровождение АСУ	Наращиваемость и масштабируемость АСУ; послегарантийное обслуживание	Действующая АСУ

В таблице 2.1 согласно ГОСТ 34.601-90 приведены типовые этапы разработки АСУ, характеристика каждого этапа и его результаты. Центральным вопросом при создании АСУ, как и любой другой ИС, является адекватное отображение бизнес-модели (схем бизнес-процессов и существующих информационных потоков) в объекты БД. При этом должна соблюдаться двусторонняя зависимость: структура БД не должна модифицироваться до тех пор, пока изменения в реальном мире не потребуют соответствующей корректировки в ней для того, чтобы эта модель продолжала соответствовать предметной области БД. Кратко рассмотрим основные особенности процесса проектирования БД.

Выбор методологии построения базы данных. Разрабатываемая АСУ, как и любая другая ИС, содержит информацию об определенной предметной области. Поэтому процесс ее создания начинается определения *инфологической модели* (ИЛМ) предметной

области. Данная ИЛМ не зависит от используемых при построении ИС технических и программных средств – в ней должны быть отражены информационные потребности пользователей ИС и описаны важнейшие информационные процессы: например, формирование различных показателей в ИС.

Основными задачами инфологического проектирования являются определение предметной области системы и формирование взгляда на нее с позиций сообщества будущих пользователей БД. В этой связи ИЛМ представляет собой описание структуры и динамики предметной области, характера информационных потребностей пользователей в терминах, понятных пользователю и не зависящих от способа реализации БД. Это описание выражается в терминах не отдельных объектов, принадлежащих предметной области и связей между ними, а их типов, связанных с ними ограничений целостности и тех процессов, которые приводят к переходу предметной области из одного состояния в другое.

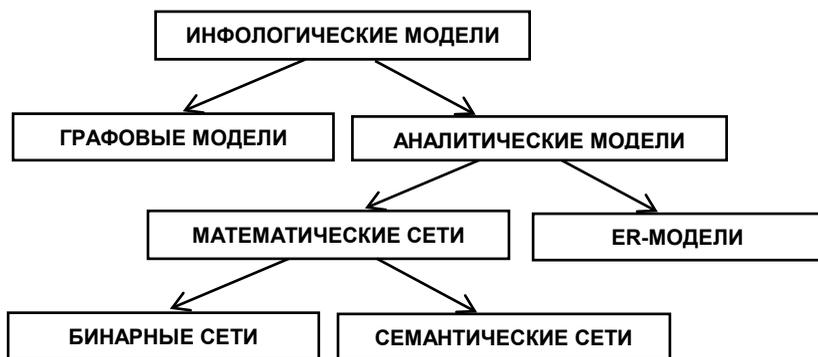


Рис. 2.4. Классификация инфологических моделей данных

Начальным этапом является выбор методологии инфологического проектирования БД. Среди множества известных способов представления реальности в виде моделей инфологическое моделирование отличают:

- **селективность**, поскольку не все реальные объекты и процессы можно отразить с помощью ИЛМ и разные методологии их создания предназначены для разных объектов, на них накладываются ограничения (хотя обычно нестрогие и «размытые»), что

связано с тем, что сфера ИЛМ ограничена предметной областью: частью реального мира, информация о которой является объектом рассматриваемой ИС;

- *используемый язык*, то есть синтаксические и семантические правила построения: по языку ИЛМ можно разделить на аналитические и графовые, как это показано на рис. 2.4.

Аналитические модели используют различные математические и другие аналитические языки, представление предметной области в виде формул и зависимостей. Наиболее часто в ИЛМ применяют аппарат реляционной алгебры и теории множеств, но возможны и другие аналитические представления предметной области. Алфавитом в графовых ИЛМ являются различные графические символы. Наиболее известной методикой проектирования БД, использующей аналитическую ИЛМ, является методика проектирования К. Дейта [33]

Методологии проектирования с применением аналитических моделей используют строгий математический аппарат, что позволяет их автоматизировать. На сегодняшний день программные средства автоматизации проектирования на основе данных моделей не получили широкого распространения – возможно потому, что они рассчитаны на специалистов, имеющих хорошую математическую подготовку и значительный опыт. При проектировании БД аналитические модели практически невозможно использовать (ввиду их сложности, громоздкости, неадаптивности и трудоемкости) при отображении большого числа разнотипных процессов и объектов, свойств и ограничений. Такие модели подходят для систем с числом объектов и процессов не более пяти, имеющих ограниченное количество свойств. Кроме того, моделируемая предметная область должна быть относительно статична, что нехарактерно для социально-экономических СС и других систем организационно-технического типа.

В графических ИЛМ может быть использован различный математический аппарат (см. рис. 2.4). Распространение моделей на основе разного вида математических сетей (бинарные, семантические) сдерживают объективные сложности в создании и реализации алгоритмов перехода от них к физическим моделям, описанным на широко применяемых языках описания данных.

Отдельным классом графовых ИЛМ являются ER-модели. В настоящее время они наиболее широко распространены, вследствие своей эффективности, возможности использования для отображения многих областей реального мира, относительной простоты построения и наглядности. Широкое применение ER-моделей привело к появлению сначала алгоритмов перехода от них к иерархическим, сетевым и реляционным физическим моделям, а затем средств автоматизации, как процесса построения ER-модели, так и процесса получения по ней физической модели.

Таблица 2.2. Сравнительный анализ разных типов ИЛМ

Наименования характеристик	Аналитические модели	Сетевые модели	ER-модели
Используемый математический аппарат	Теория множеств, аналитические формулы, теория графов	Бинарные и семантические сети	Начала теории графов, алгебра теории множеств
Языки	Аналитический	Графический	Графический
Наличие преобразования в физическую модель	+	+	+
Средства автоматизации:			
а) построения	-	-	+
б) ведения словарей, пулов	+	-	+
в) преобразования в физическую модель	+	-	+
г) разработки программных средств администрирования БД	-	-	+
Наглядность	Низкая	Высокая	Очень высокая
Требуемая квалификация проектировщика	Очень высокая	Высокая	Средняя

Результаты сравнительного анализа разных типов ИЛМ представлены в таблице 2.2. Наиболее предпочтительным для построения ИЛМ предметной области является использование ER-

моделей. В этой связи имеет смысл проанализировать основные виды методик проектирования БД с использованием ER-моделирования более подробно.

Проектирование БД с применением ER-моделей. Цель моделирования состоит в обеспечении разработчика ИС концептуальной схемой БД в форме одной или нескольких локальных моделей, которые могут быть отображены в любую систему БД. Наиболее распространенным средством моделирования данных являются диаграммы «сущность-связь» (ERD), с помощью которых определяются все важные для предметной области объекты (сущности), их свойства (атрибуты) и отношения друг с другом (связи). В этой связи ER-модель – это отражение реального мира в виде сущностей (Entity) и связей между ними (Relationship).

В настоящее время ER-моделирование является самым распространенным методом построения ИЛМ предметной области, в том числе и социально-экономических СС. Многие из методик ER-моделирования легли в основу современных автоматизированных систем проектирования БД, используются они и при их неавтоматизированной разработке.

Методики построения ER-моделей анализируются по следующим основным аспектам:

- терминологический аппарат, лежащий в основе методики;
- семантические возможности модели для отображения различных ситуаций реального мира;
- наличие алгоритма перехода от ER-модели к различным физическим моделям и конкретно к реляционной физической модели;
- эффективность алгоритма перехода;
- технология построения модели, степень сложности процесса моделирования.

В связи с тем, что построение ER-модели производится на этапе ИЛМ, а преобразование ее в структуры данных – на этапе физического моделирования, анализ моделей далее производится отдельно от анализа алгоритмов перехода. Кроме того, ER-модель может быть построена с использованием одной методики, а преобразована в структуры таблиц – при помощи другой. Рассмотрим наиболее известные виды ER-моделей.

ER-модель П. Чена – идеология построения данной модели [115] является своеобразным стандартом ER-моделирования. Дру-

гие модели могут отличаться набором графических символов, некоторыми особенностями моделирования, но основные правила остались теми же. Под **сущностью** понимается «нечто», что можно идентифицировать. Сущности могут попадать в различные типы сущностей, которые на ER-диаграммах изображаются в виде прямоугольников. Реальный мир содержит множество объектов, которые можно четко идентифицировать. Некоторые из них представляют интерес при моделировании, другие – нет. В задачу проектировщика входит отобрать именно те типы сущностей, которые нужны и будущей ИС (см. рис. 2.5).

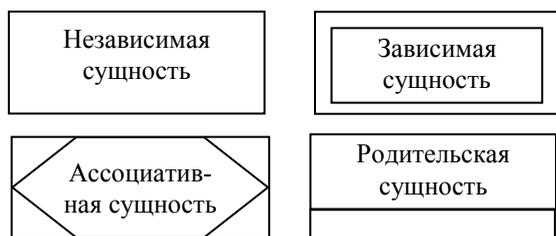


Рис. 2.5. Символы сущности в нотации Чена

Между сущностями могут существовать **связи (отношения)**, которые разделяются по типам. В самом общем виде отношения представляют собой связи между двумя и более сущностями. Именованное отношение определяется с помощью грамматического оборота глагола («имеет», «определяет», «может владеть» и т.п.). На диаграммах ER-моделей связи изображаются в виде ромбов, соединенных линиями с типами сущностей (см. рис. 2.6).

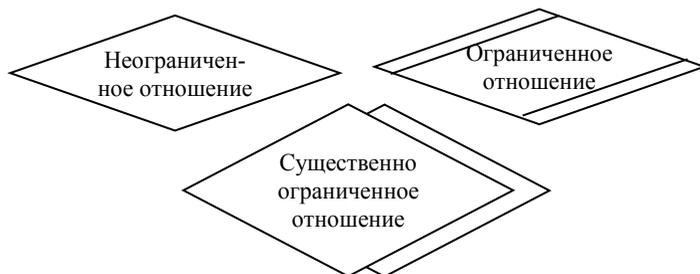


Рис. 2.6. Символы отношений в нотации Чена

Значение связи характеризует ее тип и, как правило, выбирается из следующего множества:

{«0 или 1», «0 или более», «1», «1 или более», «r:q» (диапазон)}.

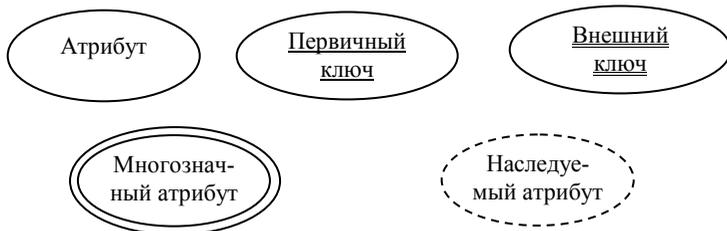


Рис. 2.7. Символы атрибутов в нотации Чена

Пара значений связей, принадлежащих одному и тому же отношению, определяет тип этого отношения. Практика показывает, что для большинства приложений достаточно использовать следующие типы отношений:

- 1*1 (один-к-одному). Отношения данного типа используются, как правило, на верхних уровнях иерархии модели данных, а на нижних уровнях встречаются сравнительно редко;
- 1*n (один-к-многим). Отношения данного типа являются наиболее часто используемыми;



Рис. 2.8. Фрагмент ER-модели в нотации Чена

- $n*m$ (многие-к-многим). Отношения данного типа обычно используются на ранних этапах проектирования с целью прояснения ситуации. В дальнейшем каждое из таких отношений должно быть преобразовано в комбинацию отношений первого и второго указанных типов (возможно, с добавлением вспомогательных сущностей и введением новых отношений).

Каждая сущность обладает одним или несколькими **атрибутами**, которые однозначно идентифицируют каждый экземпляр сущности. При этом любой атрибут может быть определен как ключевой (см. рис. 2.7). На рис. 2.8 приведен пример ER-модели в нотации Чена.

ER-модель Р. Баркера. Дальнейшее развитие ER-подход получил в работах Р. Баркера, предложившего оригинальную нотацию, которая позволила на верхнем уровне интегрировать предложенные П. Ченом средства описания моделей. В нотации Баркера используется только один тип диаграмм – ERD. Сущность на ERD представляется прямоугольником любого размера, содержащим внутри себя имя сущности, список имен атрибутов (возможно, неполный) и указатели ключевых атрибутов (знак "#" перед именем атрибута).

Каждая сущность должна обладать уникальным идентификатором. Каждый экземпляр сущности должен однозначно идентифицироваться и отличаться от других экземпляров данного типа сущности. Сущность должна обладать следующими свойствами:

- иметь уникальное имя, причем к одному и тому же имени должна всегда применяться одна и та же интерпретация. Одна и та же интерпретация не может применяться к различным именам, если только они не являются псевдонимами;
- обладать одним или несколькими атрибутами, которые либо принадлежат сущности, либо наследуются через связь;
- обладать одним или несколькими атрибутами, которые однозначно идентифицируют каждый экземпляр сущности;
- может обладать любым числом связей с другими сущностями модели.

Все связи являются бинарными и представляются линиями с двумя концами (соединяющими сущности), для которых должно быть определено имя, степень множественности (один или много

объектов участвуют в связи) и степень обязательности (то есть обязательная или необязательная связь между сущностями).

Обозначение	Обязательность	Кардинальность
-----	Необязательная	0:1
—————	Обязательная	1:1
----->	Необязательная	0:N
—————>	Обязательная	1:N

Рис. 2.9. Связи в нотации Баркера

Для множественной связи линия присоединяется к прямоугольнику сущности в трех точках, а для одиночной связи – в одной точке. При обязательной связи рисуется непрерывная линия до середины связи, при необязательной – пунктирная линия (см. рис. 2.9).

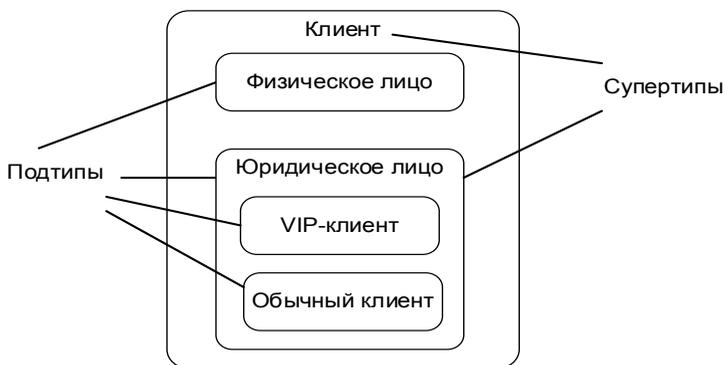


Рис. 2.10. Подтипы и супертипы

Связи может даваться имя, выражаемое грамматическим оборотом глагола и помещаемое возле линии связи. Имя каждой связи между двумя данными сущностями должно быть уникальным, но имена связей в модели не обязаны быть уникальными. Имя связи всегда формируется с точки зрения родителя, так что предложение может быть образовано соединением имени сущности-родителя, имени связи, выражения степени и имени сущности-потомка. По-

нтия категория и общая сущность заменяются Баркером на эквивалентные понятия подтипа и супертипа, соответственно (см. рис. 2.10). На рис. 2.11 приведен фрагмент ERD в нотации Баркера.

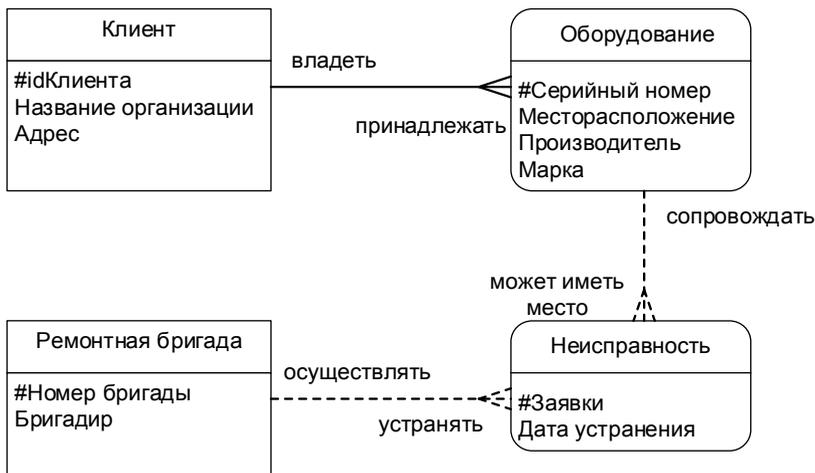


Рис. 2.11. Фрагмент ER-модели в нотации Баркера

ER-модель Дж. Мартина. Для построения диаграмм можно также использовать нотацию Мартина, основные элементы которой иллюстрирует рис. 2.12.



Рис. 2.12. Основные элементы нотации Мартина

Список атрибутов приводится внутри прямоугольника, обозначающего сущность. Ключевые атрибуты подчеркиваются. Связи изображаются линиями, соединяющими сущности, вид линии в месте соединения с сущностью определяет кардинальность связи (см. рис. 2.13).

Обозначение	Кардинальность
	нет
	1:1
	0:1
	M:N
	0:N
	1:N

Рис. 2.13. Кардинальность связи в нотации Мартина

Имя связи указывается на линии ее обозначающей. Пример диаграммы по нотации Мартина приведен на рис. 2.14.

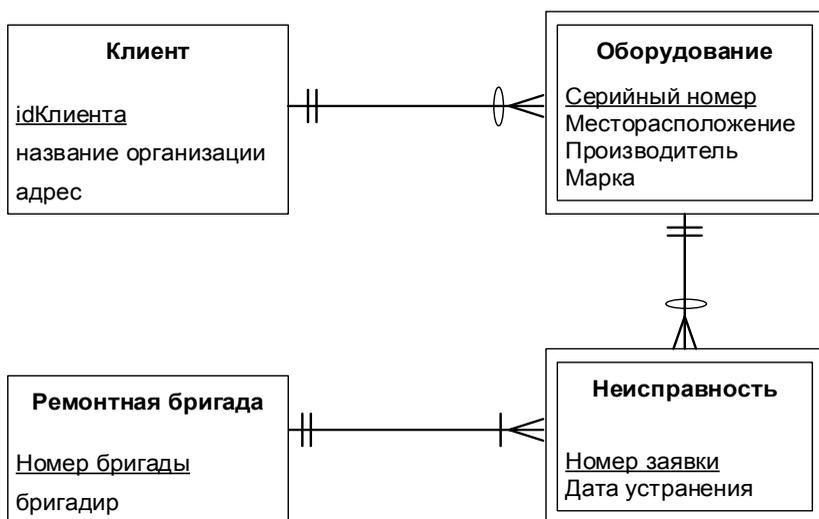


Рис. 2.14. Фрагмент ER-модели в нотации Мартина

ER-модель Г. Джексона – в работе [35] изложены принципы построения ER-моделей по методологии Г. Джексона. **Сущность** определяется как некоторый объект, представляющий интерес для

организации, этот объект должен иметь экземпляры, отличающиеся друг от друга и допускающие однозначную идентификацию. **Связь** представляет собой «соединение между двумя или более сущностями». **Атрибут** есть свойство сущности, атрибут или набор атрибутов, используемых для идентификации экземпляра сущности, называется **ключом сущности**.

На диаграммах Джексона сущности изображаются в виде прямоугольников, связи – в виде ромбов, соединенных линиями с сущностями. Под сущностью могут быть указаны ее атрибуты, при этом ключевые атрибуты подчеркиваются. Связи классифицируются по степени связи и по классу принадлежности. Степени связи бывают «1: 1», «1: n» и «m: n». На диаграммах степень связи обозначается символами «1», «n» или «m» вблизи соответствующей сущности.

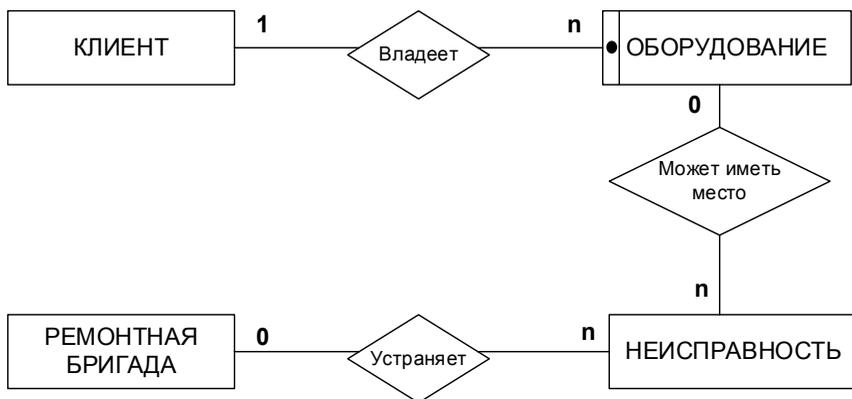


Рис. 2.15. Пример диаграммы Джексона со связью 1 : n, необязательной с одной стороны и обязательной с другой

Класс принадлежности может быть либо обязательным, либо необязательным. Обязательность связи изображается на ER-диаграммах следующим образом: в месте соприкосновения сущности и связи изображается небольшой (гораздо меньше сущности) прямоугольник и внутри него ставится точка.

IDEF1-методология. Метод IDEF1, разработанный Т. Рэймеем (Т. Ramey), также основан на подходе П. Чена и позволяет построить модель данных, эквивалентную реляционной модели в третьей

нормальной форме. В настоящее время на основе совершенствования методологии IDEF1 создана ее новая версия – методология IDEF1X. IDEF1X разработана с учетом таких требований, как простота изучения и возможность автоматизации.

IDEF1-методология инфологического проектирования структур данных имеет в виду процесс проектирования БД, состоящий из пяти стадий:

- стадия 0 – начальные действия, которые включают определение проекта, план сбора данных, авторские соглашения, стандарты и т.п.;
- стадия 1 – этап идентификации и определения сущностей;
- стадия 2 – идентификация и определение отношений;
- стадия 3 – идентификация и определение ключей;
- стадия 4 – идентификация и определение «неключевых» атрибутов.

Остановимся подробнее на стадиях IDEF1-методологии. IDEF1-модель данных должна быть описана и определена в терминах как ограничений, так и целей. На стадии 0 перед разработчиками ставятся следующие **цели**: определение проекта (общая формулировка того, что должно быть сделано и как это будет достигнуто); исходный материал (план сбора исходной информации, ее регистрация и структуризация); авторские соглашения (выбор основополагающих соглашений и методов, используемых при построении модели).

На стадии 1 определяются **сущности**, которые в IDEF1-модели представляют собой множество **предметов**, обладающих связанными с ними данными. Элементы представляемого каждой сущностью множества обладают общим набором атрибутов (характеристик). Большинство сущностей могут быть определены на основе исходного материала, собранного на стадии 0. Если при моделировании расширяется или детализируется предшествующая модель данных, то соответствующие сущности должны быть выделены из прежней модели. В настоящее время, к сожалению, IDEF-методология предлагает пользователю отдельные продукты, не имеющие общего хранилища данных, поэтому сущности каждый раз приходится выбирать «вручную». В процессе IDEF1-проектирования ведется глоссарий сущностей, который имеет локальное («местное») значение.

На стадии 1 глоссарий представляет собой набор определений сущностей, включающих имя, определение и синонимы каждой сущности. Целью стадии 2 являются выявление и определение основных отношений между сущностями. Отношение в IDEF-методологии – это ассоциация или связь между двумя сущностями, причем IDEF1 ограничивается бинарными отношениями. Больше того: в IDEF1 принято изображать ИЛМ в виде отношений *родитель-потомок*. Такое отношение – это ассоциация между типом родительской сущности и типом сущности-потомка, при которой каждый экземпляр родительской сущности ассоциирован с произвольным (в том числе и нулевым) числом экземпляров сущности-потомка, а каждый экземпляр сущности-потомка ассоциирован в точности с одним экземпляром родительской сущности.

Если сущность-родитель и сущность-потомок представляют один и тот же объект реального мира, то родительская сущность является общей сущностью, а сущность-потомок является *сущностью-категорией*. Ассоциация между этими двумя сущностями называется *отношением категоризации*.

IDEF1-модель допускает только отношения «родитель-потомок» и категоризации. Эти отношения называются *специфическими*. Отношения типа «многие-ко-многим» и «n-арные» отношения моделью не поддерживаются, они подлежат разбиению на несколько специфических отношений. Изображается ER-модель в виде диаграммы уровней сущностей: на ней сущности представляются прямоугольниками, а отношения – линиями, связывающими сущности с точками на концах. На стадии определения ключей прежде всего проводится детализация всех неспецифических отношений. В этом процессе возникают новые сущности – *сущности-пересечения*, которые должны быть добавлены в глоссарий сущностей.

Следующим шагом является определение первичных, вероятных и внешних ключей. На диаграмме модели в прямоугольнике сущности проводится горизонтальная линия, и внутри блока, выше этой линии указывается первичный ключ. Все атрибуты, не принадлежащие первичному ключу указываются ниже горизонтальной линии. Внешний ключ появляется в сущности-потомке (категории), он является первичным ключом сущности-родителя (общей сущности). После каждого атрибута, принадлежащего внешнему ключу следует (FK).

Процесс передачи ключа от сущности к сущности называется *миграцией*. Определение атрибутов – это завершающая стадия разработки ER-модели, которая начинается с построения пула (гlossария) атрибутов, где каждый атрибут должен быть приписан к одной сущности-владельцу. В *пуле атрибутов* указываются идентификатор атрибута, имя сущности-владельца, описание атрибута, указания на то, является ли данный атрибут первичным или вероятным ключом.

Сущность в методологии IDEF1X является независимой от идентификаторов или просто независимой, если каждый экземпляр сущности может быть однозначно идентифицирован без определения его отношений с другими сущностями.



Рис. 2.16. Сущности в нотации IDEF1X

Сущность называется зависимой от идентификаторов или просто зависимой, если однозначная идентификация экземпляра сущности зависит от его отношения к другой сущности (см. рис. 2.16). Каждой сущности присваивается уникальное имя и номер, разделяемые косой чертой «/» и помещаемые над блоком. Обозначения идентифицирующей и неидентифицирующей связи иллюстрирует рис. 2.17.

Обозначение	Обозначение
-----	Неидентифицирующая связь
_____	Идентифицирующая связь

Рис. 2.17. Обозначение связи в нотации IDEF1X

Связь может дополнительно определяться с помощью указания степени или мощности (числа экземпляров сущности-потомка, которое может существовать для каждого экземпляра сущности-

родителя). В IDEF1X могут быть выражены следующие мощности связей:

- каждый экземпляр сущности-родителя может иметь ноль, один или более связанных с ним экземпляров сущности-потомка;
- каждый экземпляр сущности-родителя должен иметь не менее одного связанного с ним экземпляра сущности-потомка;
- каждый экземпляр сущности-родителя должен иметь не более одного связанного с ним экземпляра сущности-потомка;
- каждый экземпляр сущности-родителя связан с некоторым фиксированным числом экземпляров сущности-потомка.

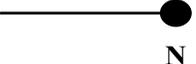
Обозначение	Обозначение
	1:1
	0: M
	0:1
	1: M
	Точно N (N – произвольное число)

Рис. 2.18. Мощность связи в нотации IDEF1X

Связь изображается линией, проводимой между сущностью-родителем и сущностью-потомком с точкой на конце линии у сущности-потомка. Мощность связи обозначается как показано на рис. 2.18 (мощность по умолчанию – N). Если экземпляр сущности-потомка однозначно определяется своей связью с сущностью-родителем, то связь называется *идентифицирующей*, в противном случае – *неидентифицирующей* (см. обозначения на рис. 2.17).

Атрибуты изображаются в виде списка имен внутри блока сущности. Атрибуты, определяющие первичный ключ, размещаются наверху списка и отделяются от других атрибутов горизонтальной чертой. Сущности могут иметь также внешние ключи (Foreign Key), которые могут использоваться в качестве части или целого

первичного ключа или неключевого атрибута. Внешний ключ изображается с помощью помещения внутрь блока сущности имен атрибутов, после которых следуют буквы FK в скобках.

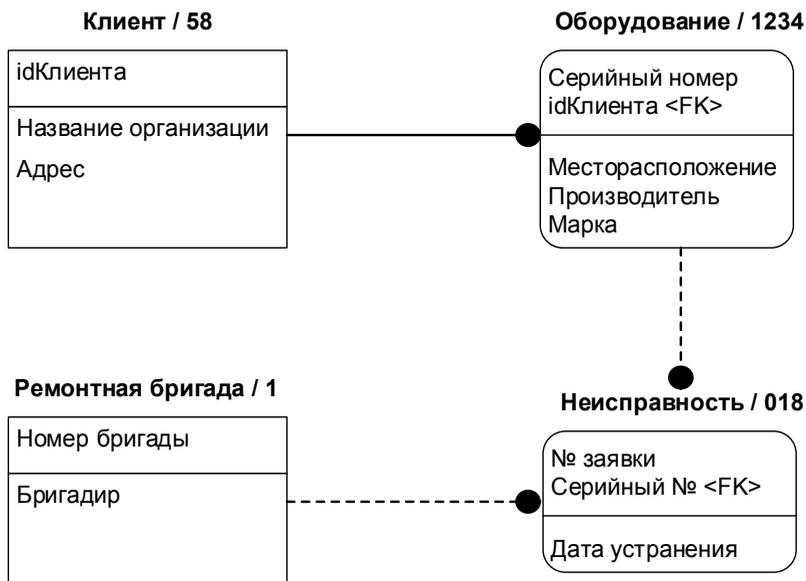


Рис. 2.19. Фрагмент ER-модели в нотации IDEF1X

Характеристики рассмотренных методик ER-моделирования, существенные для их применения в реальных проектах, представлены в таблице 2.3. Отметим, что недостатком ряда методологий моделирования является отсутствие стройного понятийного аппарата – вследствие чего специалисты вынуждены использовать неопределенные термины, что затрудняет построение моделей реальных систем.

Для проектирования БД современных экономических ИС с использованием ER-моделирования необходимы язык с высокой семантической мощностью, позволяющий отражать все необходимые виды сущностей и отношений, а также наличие алгоритма преобразования ER-модели в физическую, средств автоматизации этого алгоритма и ведения словарей данных. Создаваемая модель при этом должна оставаться максимально наглядной и простой для обеспечения участия пользователей в обсуждении проектных ре-

шений – причем лучше всего, во-первых, чтобы пользователь мог самостоятельно построить первоначальный вариант модели, а затем совместно с проектировщиком они доработали бы его, исключили аномалии и ошибки, связанные с особенностями ER-модели, уточнили содержание объектов, их свойств и отношений.

Таблица 2.3. Характеристики основных методологий ER-моделирования

Наименования характеристик	Чен	Баркер	Мартин	Джексон	IDEF1
Количество видов сущностей	1	1	3	1	2
Возможность указания отношений	–	+		+	+
Наличие n-арных отношений	+	–	–	+	–
Возможность отображения «род-вид»	–	+	–	–	–
Вид физической модели, для которой предложены алгоритмы преобразования	Сетевая	Реляционная	Реляционная	Реляционная	Реляционная
Наличие средств автоматизации	–	+	–	–	+
Отражение на диаграмме атрибутов	+	+	+	–	+
Выделение на диаграмме ключей	+	+	+	–	+

Во-вторых, чтобы указанные аномалии и ошибки помогало обнаруживать специализированное программно-техническое средство типа CASE: поскольку использование неавтоматизированных моделей при создании систем обработки данных для больших ИС достаточно затруднено и предпочтительно использовать средства автоматизации проектирования. Отметим, что ER-модели, использующие средства автоматизации, как правило, не дают выигрыша по времени при построении БД. Преимущества использования

CASE-средств выражаются на стадии физического проектирования и при разработке программных средств.

Анализ существующих ИЛМ показывает, что ни одна из них не предлагает сегодня полного набора средств для отображения разных классов объектов предметной области. При проектировании больших и сложных ИС этот недостаток является весьма существенным, так как объекты и отношения реального мира многочисленны и разнообразны, а потому отсутствие средств их адекватного отображения может привести к ошибкам восприятия и реализации модели.

2.3. Представление и преобразование знаний в ЭС

Анализ особенностей знаний, используемых в ЭС, начнем с цитаты, смысл которой остается актуальным до настоящего времени: «Огромное и все возрастающее богатство знаний разбросано сегодня по всему миру. Этих знаний, вероятно, было бы достаточно для решения всего громадного количества трудностей наших дней – но они рассеяны и неорганизованы. Нам необходима очистка мышления в своеобразной мастерской, где можно получать, сортировать, суммировать, усваивать, разъяснять и сравнивать знания и идеи» (Г. Уэллс, 1940 г.).

Отличительной особенностью ИИС сегодня называют их способность содержать, обрабатывать и применять знания в процессе решения задач в той или иной предметной области. Прежде чем рассматривать методы представления и обработки знаний, необходимо показать различие понятий «данные» и «знания». С точки зрения целей глобального информационного общества, где доминируют ИТ, знания – это интеллект, используемый в работе [58]. В своем абсолютном большинстве знания, полученные путем фактического опыта, продуктивны лишь тогда, когда они используются или интегрируются в процесс практической деятельности. Сегодня ЛПП вынуждены работать с огромными объемами информации, которой так много, что они зачастую просто неспособны ее воспринять и продуктивно использовать. При этом знания приобретают разные формы, поэтому управлять ими становится все сложнее и ответственнее – знания являются чем-то большим, чем просто информацией и данными о событиях, продуктах или процедурах.

По нашему мнению, **данные** – это отдельные факты, характеризующие объекты, процессы и явления предметной области, а также их свойства. **Знания** – это закономерности предметной области (принципы, связи, законы), полученные в результате теоретической и практической деятельности, а также профессионального опыта, которые позволяют специалистам ставить и решать задачи в этой области.

Можно также сказать, что если данные – это относительно недолго живущие новости, временные записи и т.п., непредназначенные для длительного использования, то информация представляет собой полуструктурированные (или агрегированные) данные, служащие, например, опорой для периодического принятия каких-либо решений. В свою очередь знания, являющиеся результатом переработки информации, имеют весьма длительный цикл жизни, несут определенную идею и снабжены контекстом, определяющим область их эффективного применения в каждом данном месте и в данное время. Другими словами, информация – это то, что может иметь отношение к решаемой задаче, а знания – это то, что необходимо для решения данной задачи.

Знания могут быть охарактеризованы следующими признаками [101; 108]:

- **внутренняя интерпретируемость** – существование возможности разъяснения смысла знаний;
- **структурированность** – возможность декомпозиции сложных объектов и установления связей между составляющими объектами с использованием отношения типа «часть – целое», «род – вид» или «элемент – класс»;
- **связность** – описание закономерностей относительно фактов, явлений, процессов и причинно-следственных отношений между ними, позволяя специалистам ставить и решать практические задачи в данной предметной области;
- **активность** – возможность программной генерации новых знаний (в отличие от данных, которые пассивно хранятся в памяти ЭВМ), то есть знания предполагают целенаправленное использование информации, способность управлять информационными процессами для решения определенных задач.

Рассмотрим данные свойства подробнее. **Внутренняя интерпретируемость** (интерпретация) предполагает, что в компьютере

хранятся не только «собственно (сами) данные», но и «данные о данных» (метаданные), что позволяет их интерпретировать (см. рис. 2.20). Имея такую информацию, можно ответить на вопрос типа «В каком городе находится Шоколадная фабрика «Россия?» или «Какие предприятия выпускают кондитерские изделия?». При этом сверху в таблице находятся метаданные, а в строках ниже – сами данные.

Предприятие	Местонахождение	Продукция
ОАО «Красный Октябрь»	119172, Москва, наб. Берсеневская, 6	Кондитерские изделия
ОАО «Кондитерский концерн Бабаевский»	107140, Москва, Малая Красносельская ул., 7	Кондитерские изделия
Самарский булочно-кондитерский комбинат	443125, Самара, Московское шоссе, 234	Хлебобулочные изделия
«Шоколадная фабрика ООО Россия»	443087, Самара, пр. Кирова, 257	Кондитерские изделия

Рис. 2.20. Иллюстрация внутренней интерпретации знаний

Структурированность знаний включает в себя наличие внутренней структуры связи, наличие внешней структуры связей и шкалирование. Наличие **внутренней структуры связи** – предполагает, что в качестве информационных единиц используются не отдельные данные, а их упорядоченные определенными отношениями (родо-видовыми, причинно-следственными и др.) структуры. Например, «век – год – месяц – неделя – день». Внутренняя структура связей позволяет описывать отдельный объект (понятие). Однако объекты (понятия) способны находиться и в других отношениях (вступать в ситуативную связь). Например: объекты «автомеханик Петров» и «двигатель №56» могут находиться в связи «выполнил осмотр».

Шкалирование подразумевает введение соотношений между различными информационными единицами (то есть их измерение посредством какой-либо шкалы – порядковой, классификационной,

метрической и т.п.), а также их упорядочение путем измерения интенсивности отношений и свойств. Например, объект «учебная группа ПИЭ-52» и понятие «успеваемость» связаны отношением «занимает первое место», которое имеет численную характеристику (шкалу).

Разумной деятельности человека свойственно стремление к упорядоченности: поэтому опытные ЛПР всегда пытаются систематизировать, «разложить по полочкам» все те явления, события, факты (а точнее, информационные единицы), с которыми имеют дело. Для этой цели используются различного рода шкалы: это могут быть строгие метрические шкалы, такие как шкала упорядочивания людей по возрасту, шкала воинских званий или весовых категорий в боксе. Но это могут быть и «размытые» шкалы типа: «никогда»; «очень редко»; «редко»; «редко, но не очень»; «часто»; «очень часто»; «почти всегда»; «всегда».

Наличие *семантической метрики* означает установление смысловой связи между понятиями и объектами, не имеющие «количественного» толкования. На практике ЛПР довольно часто встречаются с понятиями, к которым неприменимы количественные шкалы, но существует потребность в установлении их близости (например, понятия «искусственный интеллект» и «искусственный разум»). *Семантики* классифицируются следующим образом:

- значение, то есть объективное содержание;
- контекстуальный смысл, определяемый связями данного понятия с другими, соседствующими в данной ситуации;
- личностный смысл, то есть объективное значение, отраженное через систему взглядов эксперта;
- прагматический смысл, определяемый текущим знанием о конкретной ситуации (например, фраза «информация получена» может иметь как негативную, так и позитивную оценку – в зависимости от того, нужно это было ЛПР или нет).

Семантика отношений может носить декларативный или процедурный характер. Например, два или более объекта могут быть связаны отношением «одновременно», что может трактоваться или отношением «причина – связь», или отношением «быть рядом». Приведенные отношения характеризуют декларативные знания. Если между двумя информационными единицами установлено от-

ношение «аргумент – функция», то оно характеризует процедурное знание, связанное с вычислением определенных функций. Также необходимо различать отношения структуризации и функциональные отношения, каузальные и семантические отношения. С помощью первых задаются иерархии информационных единиц, вторые несут процедурную информацию, позволяющую находить (вычислять) одни информационные единицы через другие, третьи задают причинно-следственные связи, четвертые соответствуют всем остальным отношениям.

Активность – данное свойство принципиально отличает понятие «знание» от понятия «данные». Например, знания человека, как правило, активны, поскольку ему свойственна познавательная активность (обнаружение противоречий в знаниях становится побудительной причиной их преодоления и появления новых знаний, стимулом активности является также неполнота знаний, которая ведет к необходимости их пополнения). В отличие от данных, знания позволяют выводить (получать) новые знания. Будучи активными, знания позволяют ЛПР решать не только типовые, но и принципиально новые, нетрадиционные задачи.

Рассмотрим процесс работы ЭВМ, по заданной программе обрабатывающей те или иные данные. В программном компьютерном коде содержится процедурное знание, которое хранит информацию о том, как надо действовать, чтобы получить желаемый результат. Эти данные представляют собой декларативные знания, то есть они хранят информацию о том, над чем надо выполнять действия, заложенные в программном коде. Таким образом, программа играет роль активатора данных. У человека же очень часто в процессе мышления декларативные знания являются активатором процедурных.

Существует ряд **классификаций знаний**, систематизирующих знания разного вида и разных предметных областей. На абстрактном уровне их рассмотрения можно говорить о признаках, по которым подразделяются знания – как это видно на схеме рис. 2.21. По своей природе знания можно разделить на декларативные и процедурные [2]. **Декларативные знания** представляют собой описание фактов и явлений, фиксируют наличие или отсутствие таких фактов, а также включают описания основных связей и закономерностей, в которые эти явления и факты входят.

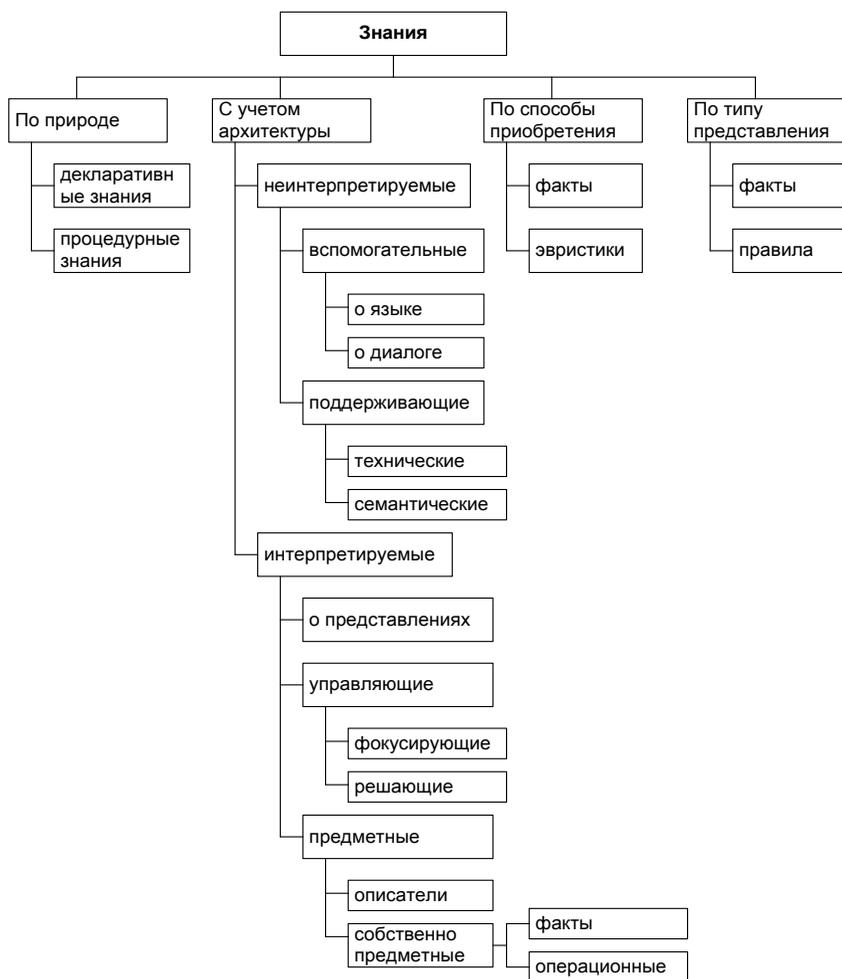


Рис. 2.21. Классификация знаний по различным признакам

Процедурные знания – это описания действий, которые возможны при манипулировании фактами и явлениями для достижения намеченных целей.

С учетом иерархической архитектуры ИИС [91] знания целесообразно подразделить следующим образом:

- **интерпретируемые** – знания, которые способен интерпретировать решатель ЭС, в состав которых входят

- **предметные знания** – содержат данные о предметной области и способах преобразования этих данных при решении поставленных задач. В предметных знаниях можно выделить:
 - **описатели**, содержащие определенную информацию о предметных знаниях: такую, как коэффициент определенности правил и данных, меры важности и сложности;
 - **собственно предметные знания** – включают **факты** (они определяют возможные значения сущностей и характеристик предметной области);
 - **исполняемые утверждения** (они содержат информацию о том, как можно изменять описание предметной области в ходе решения задач, то есть исполняемые утверждения – это знания, задающие процедуры обработки, которые могут быть заданы не только в процедурной, но и декларативной форме;
- **управляющие** знания, которые можно разделить на следующие виды:
 - **фокусирующие** – они описывают, какие знания следует использовать в той или иной ситуации, обычно эти знания содержат сведения о наиболее перспективных объектах или правилах, которые ЭС целесообразно использовать при проверке соответствующих гипотез: в первом случае внимание фокусируется на элементах рабочей памяти, во втором – на правилах БЗ;
 - **решающие** – они содержат информацию, используемую для выбора способа интерпретации знаний, подходящего к текущей ситуации, эти знания применяются для выбора стратегий или эвристик, наиболее эффективных для решения данной задачи;
 - **знания о представлении**, которые содержат информацию о том, каким образом (в каких структурах) в системе представлены интерпретируемые знания;
- **неинтерпретируемые** – к числу которых относятся все знания, относительно которых решателю неизвестны ни их структура, ни их содержание, если эти знания используются каким-либо компонентом ЭС, то он не «осознает» их. Данные знания подразделяются на следующие виды:

- **вспомогательные** знания, которые содержат информацию о лексике и грамматике языка общения ЭС, структуре диалога с пользователем – эти знания обрабатываются естественно-языковой компонентой, но ход данной обработки решатель не осознает, так как этот тип обработки входных сообщений является вспомогательным для проведения экспертизы;
- **поддерживающие** знания, которые используются при создании ЭС и выполнении объяснений – они имеют описательный характер и необходимы для описания (обоснования) как интерпретируемых знаний, так и действий системы, их можно разделить на следующие знания:
 - **технологические** – они содержат сведения о времени создания описываемых ими знаний, об авторе (владельце) знаний и т.п.;
 - **семантические** – они содержат смысловое описание приводимых значений, то есть дают информацию о причинах ввода и назначении знаний, описывают способ использования знаний и получаемый эффект.

По способу приобретения [2] знания можно разделить на две категории: **факты** и **эвристики**. Первая категория обычно указывает на хорошо известные в данной предметной области обстоятельства, вторая категория основана на субъективном опыте ЛППР – эксперта, работающего в данной предметной области, который он приобрел в результате многолетней практики. По типу представления знания делятся на **факты** и **правила**, причем здесь факты – это знания типа «*A* – это *A*», что характерно для БД и сетевых моделей, а правила, или продукции – это знания типа «Если *A*, то *B*».

2.4. Онтология предметной области

В последние годы в области ИИ появилось и активно используется термин «онтология», пришедший из философии, где обозначал часть метафизики. **Онтология** (ontologia от греч. «онтос» – сущее, то, что существует и «логос» – учение, наука) – по традиционной терминологии учение о сущем; о бытии как таковом; это раздел философии, изучающий фундаментальные принципы бытия, его наиболее общие сущности и категории, структуру и закономерности. В инженерии знаний под онтологией понимают детальное описание проблемной (предметной) области, которое ис-

пользуется для формального и декларативного определения ее концептуализации. Онтология включает упорядоченные понятия предметной области и моделирует основные функциональные связи или отношения между ними – онтологией часто называют БЗ специального вида, которую можно разделять, отчуждать и самостоятельно использовать в рамках рассматриваемой предметной области [27].

Как и многие другие термины ИИ, онтология имеет многозначную интерпретацию – сегодня, в частности, известны следующие ее рабочие определения [20-21; 23]:

- онтология как философская дисциплина представляет собой систему категорий, являющихся следствием определенного взгляда на мир;
- онтология как неформальная система концептуализации знаний предполагает построение описания множества выделенных объектов, понятий, связей и отношений в заданной области знаний – при этом могут использоваться как формальные, так и неформальные средства. В простейшем случае онтология может описывать только иерархию понятий, связанных отношениями «элемент-класс», а также содержать набор аксиом и правил вывода, позволяющих выразить другие отношения между понятиями и ограничить область интерпретации понятий;
- онтология как представление концептуальной системы в виде логической теории означает использование определенного синтаксиса для представления знаний.

Таким образом, под онтологией сегодня понимается нечто большее, чем детализированный набор понятий и отношений. В онтологии также включаются ограничения, накладываемые на отношения в рамках конкретной предметной области.

Применительно к сфере ИИ термин «онтология» был впервые использован рядом исследовательских сообществ вначале в области инженерии знаний и при обработке естественных языков, а затем в представлении знаний. На рубеже XX-XXI столетий это понятие стало широко использоваться в таких областях, как интеллектуальная интеграция информации, поиск информации в Internet и управление знаниями. Позже онтологии стали рассматриваться в качестве ключевого элемента в проекте Семантической Сети – нового этапа развития Web-сети WWW (Word Wide Web). Если су-

шествующая Web-сеть – это преимущественно множество документов, которые связаны перекрестными ссылками, то создаваемая Семантическая Сеть должна добавить к ним множество онтологий и метаописаний знаний, содержащихся в документах Web-сети (включая стандарты и программные инструменты).

С успехом понятие онтологии применяется сегодня и при моделировании бизнес-процессов – поскольку позволяет использовать бизнес-знания корпорации (предприятия, организации, компании, фирмы) в формализованном и потому стандартизированном и легко модифицируемом виде [56-57]. Это означает, что традиционные корпоративные знания, которые раньше хранились лишь «на бумаге», в электронном формате текстовых документов или «в головах» ЛПР в виде, непригодном для чтения и анализа ЭВМ, теперь становятся предметом компьютерной обработки и входной информацией для программных ИС.

Содержание понятий онтологии отражается с помощью концептов. Концептом может быть описание задачи, функции, действия, стратегии и т.п. Формально *концепт* отождествляется с объектом (классом), имеющим связи (отношения) с другими классами СС. *Класс* определяется как множество экземпляров СС с общими свойствами и содержит описания собственных экземпляров систем и их свойств.

Под формальной моделью онтологии O согласно [27] принято понимать упорядоченную тройку вида $O = \langle C, R, F \rangle$, где C – конечное множество концептов (понятий) предметной области, которую определяет онтология O , а естественными ограничениями, накладываемыми на множество C , являются конечность и непустота; R – конечное множество отношений между концептами (понятиями) предметной области; F – конечное множество функций интерпретации (аксиоматизация), заданных на концептах и/или отношениях онтологии O .

Для проведения онтологического анализа может быть использована методология IDEF5, которая обеспечивает наглядное представление данных, полученных в результате обработки онтологических запросов в простой и естественной графической форме. Для поддержания процесса построения онтологии в IDEF5 разработаны специальные онтологические языки: схематический язык (Schematic Language-SL) и язык доработок и уточнений (Elaboration

Language-EL). Онтологические системы строятся на основе следующих общих принципов [26]:

- формализации – описание объективных элементов действительности в единых, строго определенных образцах (терминах, моделях и др.);
- использования ограниченного количества базовых терминов (сущностей), на основе которых конструируются все остальные понятия;
- внутренней полноты и логической непротиворечивости.

В отличие от обычного словаря для онтологической системы характерно внутреннее единство, логическая взаимосвязь и непротиворечивость используемых понятий. При разработке онтологии следует помнить следующие правила:

- не существует единственного правильного способа моделирования предметной области;
- разработка онтологии – это обязательно итеративный процесс;
- понятия в онтологии должны быть близки к объектам (физическим или логическим) и отношениям в рассматриваемой предметной области: скорее всего, это существительные (объекты) или глаголы (отношения) в предложениях, которые описывают предметную область.

Практическая методика разработки онтологии включает следующие шаги.

Шаг 1. Определение области и масштаба онтологии – здесь необходимо определить цель разработки онтологии, будущих ее пользователей, а также составить примерный список вопросов для проверки компетентности БЗ.

Шаг 2. Рассмотрение вариантов повторного использования существующих онтологий (анализ существующих или частично подходящих аналогов).

Шаг 3. Перечисление терминов, существенных (наиболее важных) в рамках онтологии. Один из способов – выписать в алфавитном порядке все существительные, встречающиеся при описании предметной области.

Шаг 4. Определение классов и иерархии классов. Известны три подхода при разработке иерархии классов: процесс нисходящей разработки, процесс восходящей разработки и процесс комбинированной разработки. Выбор подхода в большей степени зависит от

личного взгляда аналитика на предметную область. Как правило, базисом для формирования классов выступает список терминов, сформированный на третьем шаге. Затем полученные классы организуют в иерархическую таксономию.

Шаг 5. Определение свойств классов – слотов. Как правило, кандидатами на слоты выступают неиспользованные термины и списка, составленного на третьем шаге. Слотами могут стать несколько типов свойств объектов: «внутренние», «внешние», «части», отношения и т.п.

Шаг 6. Определение facets слотов, которые описывают мощность слота (единичная, множественная), тип значения слота (строка, число, логический, нумерованный), домен слота и диапазон значений слота и др.

Шаг 7. Создание экземпляров классов в иерархии, то есть заполнение значений слотов экземпляров

Применение онтологий ограничено рядом аспектов, игнорирование которых может привести к неактуальности (нецелесообразности) онтологического подхода.

Во-первых, значительные объемы постоянной пополняемой информации о предметной области требуют разумного компромисса в комбинированном применении онтологических и лингвистических методов, используемых для развития и поддержки web-порталов [67]. Тезаурус портала UMLS, например, содержит до миллиона концептов, которым соответствуют примерно два миллиона строк на английском языке. При таком количестве концептов поддержка отношений в актуальном состоянии становится трудно выполнимой задачей. Для того чтобы поддерживать соответствие столь внушительного количества концептов и новых терминов, приходится применять NLP (Natural Language Processing) – лингвистические методы выявления новых терминов из актуальных текстовых материалов.

Во-вторых, особым обстоятельством являются фундаментальные отличия и множественность национальных языков народов, населяющих планету Земля. По данным ЮНЕСКО сегодня существует более 2700 языков разных народов и народностей – таким образом, для конкретной предметной области очень сложно создать одну «универсальную» онтологию, доступную и понятную абсолютно для всех участников. Отметим, что пользователи онто-

логии необязательно могут говорить на разных языках: даже носители одного языка могут не понять друг друга – если запутаются в его омонимах и синонимах (использование слов «поребрик» и «бордюр», например, позволяет почти безошибочно отличить в России петербуржца от москвича).

В-третьих, актуальной проблемой являются вопросы создания средств для совместного использования и поддержки онтологий, разработанных независимо друг от друга. Наиболее проблематичным пунктом здесь является использование словарей, соответствующих разным онтологиям, в которых сходная информация представлена разными терминами и понятиями.

Кратко проиллюстрируем вышесказанное: пусть дана предметная область «Такси». Являются ли равнозамменяемыми следующие понятия: «клиент», «пассажир», «заказчик»? На первый взгляд, это три равноценных синонима. Однако можно дать подробное определение каждого понятия:

- заказчик – это человек, вызвавший такси;
- клиент – это человек, оплативший поездку;
- пассажир – это человек, совершивший поездку.

Если все указанные действия совершил один человек, то все три слова можно использовать как синонимы. Если же имеет место другая ситуация – секретарша директора (заказчик) вызвала такси для встречи гостя в аэропорту (пассажир), поездку которого оплатит директор (клиент) – то подобное использование слов как синонимов неправильно и недопустимо.

2.5. Методы извлечения знаний

Извлечением знаний называют процесс получения знаний от экспертов [22]. Многообразие задач, ситуаций и источников знаний обусловило появление большого количества методов извлечения, приобретения и формирования знаний. Одна из возможных классификаций методов извлечения знаний приведена на рис. 2.22, на первом уровне которой выделены два больших класса. Первый класс образуют **коммуникативные методы**, которые ориентированы на непосредственный контакт инженера по знаниям с экспертом (источником знаний), второй класс – **текстологические методы**, основанные на приобретении знаний из документов и специальной литературы. Начнем с обширной группы коммуникатив-

ных методов которые разделяются на пассивные и активные (при решении конкретных задач обычно используемые совместно), различающиеся тем, что в пассивных методах ведущую роль играет эксперт, в активных – инженер по знаниям.



Рис. 2.22. Классификация методов извлечения знаний

Активные методы делятся на индивидуальные и групповые: в групповых методах знания получают от множества экспертов, в индивидуальных – от единственного эксперта. Индивидуальные методы получили более широкое применение на практике по сравнению с групповыми. Подробное описание указанных методов можно найти в работах [2; 22-23; 87; 93], ниже приведены краткие характеристики по каждому способу.

Пассивные» методы, несмотря на свое название, зачастую требуют от инженера по знаниям больших усилий, чем активные: в рамках общения с экспертом инженер по знаниям должен незаметно для опрашиваемого направлять нить разговора в нужную сторону и не давать эксперту углубляться на отвлеченные темы. С другой стороны, иногда такое отступление необходимо, поскольку именно оно может являться звеном в логической цепочке, приво-

димой к решению задачи. К этой группе относят такие методы, как наблюдения, анализ протоколов «мыслей вслух», лекции.

В процессе *наблюдений* инженер по знаниям находится непосредственно рядом с экспертом во время его профессиональной деятельности или имитации этой деятельности. До начала сеанса извлечения знаний, эксперту необходимо объяснить цель наблюдений и попросить максимально комментировать свои действия. Во время сеанса инженер по знаниям должен фиксировать все действия эксперта, его реплики и объяснения. Обязательным условием этого метода является невмешательство аналитика в работу эксперта.

Сеансы наблюдений могут потребовать от аналитика:

- овладения техникой стенографии для фиксации действий эксперта в реальном масштабе времени;
- ознакомления с методиками хронометрирования для четкого структурирования производственного процесса по времени;
- развития навыков «чтения по глазам», наблюдательности к жестам, мимике и другим невербальным компонентам общения;
- серьезного предварительного знакомства с предметной областью, так как из-за отсутствия «обратной связи» иногда многое непонятно в действиях экспертов.

Протоколы наблюдений после сеансов тщательно расшифровываются, а затем обсуждаются с экспертом. Наблюдение – один из распространенных методов, но обычно он применяется в совокупности с другими методами.

Протоколирование «мыслей вслух» отличается от наблюдений тем, что эксперта просят не просто прокомментировать свои действия и решения, но и объяснить, как это решение было найдено, то есть продемонстрировать всю цепочку своих рассуждений. Во время рассуждения эксперта все его слова протоколируются инженером по знаниям: при этом полезно отмечать даже паузы и междометия. Вопрос об использовании для этой цели диктофонов является дискуссионным, поскольку сам факт записи иногда действует на эксперта парализующе, разрушая атмосферу доверительности, которая может и должна возникать при непосредственном общении.

Основной трудностью при протоколировании «мыслей вслух» является сложность для любого человека объяснить, как он думает.

Психологами в рамках проведения экспериментов неоднократно было доказано, что люди очень часто не в состоянии достоверно описать собственные мыслительные процессы. Кроме того, эксперты практически никогда не озвучивают свои ошибочные рассуждения, хотя такие логические цепочки и вызывают наибольший интерес. Кроме того, часть знаний, хранящихся в невербальной форме (например, различные процедурные знания типа «как сложить бумажный самолет»), вообще слабо коррелируют с их словесным описанием.

Расшифровка полученных протоколов производится инженером по знаниям самостоятельно с коррекциями на следующих сеансах извлечения знаний. Удачно проведенное протоколирование «мыслей вслух» является одним из наиболее эффективных методов извлечения, поскольку для большинства экспертов это самый приятный и лестный способ извлечения знаний.

Самым известным способом передачи знаний является лекция. Лекторское искусство издавна ценилось во всех областях науки, но в процессе извлечения знаний более значимым является способность инженера по знаниям слушать и усваивать материал лекции. Чаще всего выбор эксперта predetermined, сменить его сложно (поскольку эксперт – это штучный экземпляр), поэтому тут важно умение инженера подстраиваться под стиль изложения лекционного материала экспертом. В случае наличия у эксперта опыта преподавания (например, если он профессор медицины или опытный руководитель производства), то использование лекции – это приемлемый способ получения концентрированного знания.

В лекции эксперту предоставлена свобода для самовыражения, хотя для него необходимо сформулировать тему и задачу лекции, чтобы он мог заранее структурировать свои знания и ход рассуждений. От инженера по знаниям в этой ситуации требуется лишь грамотно законспектировать лекцию и в конце задать необходимые вопросы. Продолжительность лекции стандартная – два академических часа (90 мин, с перерывом 5-10 мин в середине периода), курс обычно включает от двух до пяти лекций.

Метод извлечения знаний в форме лекций, как и все пассивные методы, используют в начале разработки для быстрого и эффективного «погружения» инженера по знаниям в предметную об-

ласть. При прослушивании лекций инженеру по знаниям рекомендуется следовать следующим принципам:

- нужна предварительная подготовка к теме лекции (следует заранее ознакомиться с предметной областью, чтобы не тратить время эксперта на разъяснение азбучных истин);
- необходимо слушать излагаемый материал с максимальным вниманием, но при этом следует уметь отдыхать во время слушания (например, когда эксперт приводит цифры, которые достаточно легко найти в справочниках или открытых публикациях);
- нужно уметь слушать одновременно и лектора, и самого себя, если параллельно рассуждениям лектора по ассоциации возникают и собственные мысли;
- необходимо умение одновременно слушать и записывать излагаемый текст, однако записывать его следует с использованием условных сокращений, чтобы не отвлекать лектора. Вопрос использования диктофона необходимо заранее согласовать с лектором;
- текст лекция подлежит расшифровке обязательно в тот же день;
- не следует спорить с лектором во время слушания (во-первых, это отвлекает его от изложения материала; во-вторых, у вас будет время задать вопросы после окончания лекции).

Активные индивидуальные методы извлечения знаний являются наиболее распространенными, поскольку к ней прибегают при разработке практически любой экспертной системы. К основным активным методам принято относить: анкетирование; интервью, свободный диалог; игры с экспертом.

Анкетирование – наиболее жесткий метод, поскольку наиболее стандартизованный. Инженер по знаниям заранее составляет вопросник или анкету, размножает ее и использует для опроса нескольких экспертов. Это основное преимущество анкетирования. Сама процедура может проводиться двумя способами:

- аналитик вслух задает вопросы и сам заполняет анкету по ответам эксперта;
- эксперт самостоятельно заполняет анкету после предварительного инструктирования.

Выбор способа зависит от конкретных условий.

Отдельно стоит рассмотреть процесс составления вопросника (анкеты), от которой напрямую зависит успех анкетирования:

- анкета не должна быть монотонной и однообразной;
- анкета должна быть приспособлена к языку эксперта (содержать терминологию предметной области, не быть достаточно тривиальной и в то же время язык анкеты должен быть ясен, понятен и предельно вежлив);
- следует учитывать, что вопросы влияют друг на друга, и поэтому их последовательность должна быть строго продумана;
- желательно стремиться к оптимальной избыточности. Известно, что в анкете всегда много лишних вопросов, часть из них необходима – это так называемые контрольные вопросы, а другую часть нужно минимизировать.

Интервью – это специфичная форма общения инженера по знаниям и эксперта, в которой инженер по знаниям задает эксперту серию заранее подготовленных вопросов с целью извлечения знаний о предметной области. Интервью очень близко к анкетированию, когда аналитик сам заполняет анкету, заносит туда ответы эксперта. Основное отличие в том, что интервью позволяет аналитику опускать ряд вопросов в зависимости от ситуации, вставлять новые вопросы в анкету, изменять темп, разнообразить ситуацию общения. Кроме того, у аналитика появляется возможность очаровать эксперта своей харизмой, заинтересовать его самой процедурой и тем самым увеличить эффективность сеанса извлечения.

На успех интервью немаловажное влияние оказывает умение инженера по знаниям задавать вопросы. Существует даже специальная ветвь математической логики – логика вопросов. Вопрос в интервью является не просто средством общения, но и способом передачи мыслей и позиции аналитика. Отсюда необходимость фиксировать в протоколах не только ответы, но и вопросы, предварительно отработав их форму и содержание. Смысл вопроса напрямую зависит от контекста, кроме того, сама постановка вопроса позволит эксперту понять, насколько «в теме» находится инженер по знаниям – несколько откровенно «глупых» вопросов могут разочаровать эксперта и отбить охоту к дальнейшему сотрудничеству.

Свободный диалог – это метод извлечения знаний в форме беседы инженера по знаниям и эксперта, в которой нет жестко регламентированного плана и вопросника. Это определение не означает, что к свободному диалогу не надо готовиться. Напротив,

внешне свободная и легкая форма этого метода требует высочайшей профессиональной и психологической подготовки.

Рекомендуется строить свободный диалог с соблюдением стадий:

- начало беседы (знакомство, объяснение целей и задач работы);
- диалог по извлечению знаний;
- заключительная стадия (обязательно благодарность эксперту, подведение итогов, договоренность о последующих встречах).

Квалифицированная подготовка к диалогу помогает аналитику запланировать гладкое течение процедуры извлечения – от приятного впечатления в начале беседы к профессиональному контакту, – пробудив интерес и завоевав доверие эксперта. Для обеспечения желания эксперта продолжить беседу необходимо «подбадривать» эксперта и подтверждать всячески его уверенность в собственной компетентности. Так, в одном из исследований по технике журналистских диалогов экспериментально доказано, что одобрительное и поощрительное «хмыканье» интервьюера увеличивает длину ответов респондента.

В свободном диалоге важно выбрать правильный темп или ритм беседы: без больших пауз, так как эксперт может отвлечься, но и «без гонки», иначе быстро утомляются оба участника и нарастает напряженность; кроме того, некоторые люди говорят и думают очень медленно. Умение чередовать разные темпы, – напряжение и разрядку существенно влияет на результат беседы.

К *групповым активным методам* извлечения знаний относятся: ролевые игры; дискуссии за круглым столом с участием нескольких экспертов; «мозговой штурм». Основное достоинство групповых методов – это возможность одновременного плучения знаний от нескольких экспертов, взаимодействие которых вносит элемент принципиальной новизны от наложения разных взглядов и позиций.

Метод круглого стола представляет собой современную форму публичного обсуждения или освещения каких-либо вопросов, когда участники, имеющие равные права, высказываются по очереди или в определенном порядке.

Согласно старой английской легенде, король Артур впервые усадил своих соратников за круглый стол с целью сделать их как бы равными друг другу и равноправными в дискуссии – благодаря

чему круглый стол дошел до сегодняшних дней как символ плодотворной совместной работы и принятия компромиссных решений.

«Круглые столы» характеризуются тем, что:

- проводятся в форме обсуждения одного или нескольких определенных вопросов или проблем;
- обсуждаемый вопрос допускает разные мнения и толкования, а также взаимные возражения участников;
- в результате обсуждения должны быть выявлены точки зрения всех участников на данный вопрос;
- участники имеют равные права и высказываются в определенном порядке.

Обычно вначале участники высказываются в определенном порядке, а затем переходят к живой свободной дискуссии. Число участников дискуссии колеблется от трех до пяти-семи. Существует и специфика, связанная с поведением человека в группе. Во-первых, от инженера по знаниям подготовка круглого стола требует дополнительных усилий, как организационных (место, время, обстановка, минеральная вода, чай, кворум и т.д.), так и психологических (умение вставлять уместные реплики, чувство юмора, память на имена и отчества, способность гасить конфликтные ситуации и т.д.), Во-вторых, большинство участников под воздействием желания произвести впечатление на других экспертов будут говорить совсем не то, что они сказали бы в другой обстановке. Кроме того, по ходу дискуссии необходимо проследить, чтобы слишком эмоциональные и разговорчивые эксперты не подменили тему и критика позиций друг друга была обоснованной

Мозговой штурм или мозговая атака – один из наиболее распространенных методов раскрепощения и активизации творческого мышления. Впервые этот метод был использован в США как способ получения новых идей в условиях запрещения критики. Замечено, что боязнь критики мешает творческому мышлению, поэтому основная идея штурма – это отделение процедуры генерирования идей в замкнутой группе специалистов от процесса анализа и оценки высказанных идей.

Как правило, штурм длится около 40 мин. Участникам (до 10 человек) предлагается высказывать любые идеи (шутливые, фантастические, ошибочные) на заданную тему (критика запрещена). Обычно высказывается более полусотни идей. Регламент до 2 мин.

на выступление. Самый интересный момент штурма – это наступление пика (ажиотажа), когда идеи начинают «фонтанировать», то есть происходит непровольная (бессознательная) генерация гипотез участниками. При последующем анализе всего лишь 10-15% идей оказываются разумными, но среди них бывают весьма оригинальные. Оценивает результаты обычно группа экспертов, не участвовавшая в генерации.

Ведущий мозгового штурма – инженер по знаниям – должен свободно владеть аудиторией, подобрать активную группу экспертов-генераторов, не зажимать плохие идеи – они могут служить катализатором хороших. Основной девиз штурма – «Чем больше идей, тем лучше». Обязательно фиксирование сеанса мозгового штурма в виде протокола, стенографии или видеозаписи.

Игры называют вид деятельности людей, который отражает (воссоздает) другие ее виды. Под **деловой игрой** понимают эксперимент, где участникам предлагается производственная ситуация, а они на основе своего жизненного опыта, общих и специальных знаний и представлений принимают решения. Решения анализируются и вскрываются закономерности мышления участников эксперимента. Именно эта анализирующая часть деловой игры полезна для получения знаний. И если участниками такой игры становятся эксперты, то игра из деловой превращается в экспертную.

В случае **индивидуальной игры с экспертом** инженер по знаниям «примеряет» на себя ту или иную роль в моделируемой ситуации. Например, игра «учитель и ученик», в которой инженер по знаниям – это ученик, который на глазах эксперта выполняет его работу, а эксперт поправляет ошибки ученика.

В другой игре инженер по знаниям берет на себя роль врача, который хорошо знает больного, а эксперт – роль консультанта. Консультант задает вопросы, делает прогноз о целесообразности применения того или иного вида лечения. Такая игра «двух врачей» позволила, например, выявить, что эксперту понадобилось всего 30 вопросов для успешного прогноза, в то время как первоначальный вопросник, составленный медиками, содержал 170 пунктов.

Групповые экспертные игры предусматривают участие в игре нескольких экспертов. К такой игре обычно заранее составляется сценарий, распределяются роли, к каждой роли готовится портрет-описание и разрабатывается система оценивания игроков.

Таблица 2.4. Сравнительные характеристики пассивных методов извлечения знаний

Методы	Достоинства	Недостатки	Требования к эксперту	Требования к аналитику	Предметная область
Наблюдения	Отсутствие влияния аналитика и его субъективной позиции. Максимальное приближение аналитика к предметной области	Отсутствие обратной связи. Фрагментарность полученных комментариев	Собеседник или мыслитель (способность к вербализации мысли)	Мыслитель (наблюдательность, полнота, независимость)	Слабо- и средне-структурированные, слабо- и среднедокументированные
«Мысли вслух»	Свобода самовыражения для эксперта. Вербализация рассуждений. Отсутствие влияния субъективной позиции аналитика	Отсутствие обратной связи. Возможность ухода «в сторону» в рассуждениях эксперта	Мыслитель или собеседник (полнота, склад ума, открытость, рефлексивность)	Мыслитель или собеседник (полнота, независимость, способность к общению)	Слабо- и средне-структурированные, слабо- и среднедокументированные
Лекции	Свобода самовыражения для эксперта. Структурированное изложение. Отсутствие влияния субъективной позиции аналитика	«Зашумленность» деталями. Слабая обратная связь. Отсутствие влияния аналитика и его субъективной позиции	Мыслитель (лекторские способности)	Мыслитель (полнота, независимость, способность к общению)	Слабо- и средне-структурированные, слабо- и среднедокументированные

Игры с тренажерами в значительной степени ближе не к играм, а к имитационному моделированию реальной ситуации. Наличие тренажера позволяет воссоздать почти производственную ситуацию и понаблюдать за экспертом. Тренажеры широко применяются для обучения (например, летчиков или операторов атомных станций). Считается, что применение тренажеров для извлечения знаний позволяет зафиксировать фрагменты «летучих» знаний, возникающих во время и на месте реальных ситуаций, которые нередко выпадают из памяти при выходе за пределы той или иной конкретной ситуации.

Сравнительные характеристики пассивных методов извлечения знаний представлены в таблице 2.4. Сравнительный анализ активных методов приведен в таблице 2.5.

Входящие в состав второй группы *текстологические методы* включают способы извлечения знаний, основанные на изучении текстов учебников, специальной литературы и документов. Напомним, что текстология как наука в практическом плане занимается прочтением текстов, изучением и интерпретацией литературных источников, рассматривает семиотические, психолингвистические и другие аспекты извлечения знаний из текстов. В соответствии с этим, простейший алгоритм извлечения знаний из текстов [93] содержит следующие шаги.

1. Составить «базовый» список литературы для ознакомления с предметной областью.
2. Выбрать текст для извлечения знаний.
3. Беглое знакомство с текстом. Провести консультации со специалистами для определения значений незнакомых слов.
4. Сформировать первую гипотезу о макроструктуре текста.
5. Внимательно прочитать текст и выписать ключевые слова и выражения, определив тем самым «смысловые вехи».
6. Определить связи между ключевыми словами, разработать макроструктуры текста в форме графа или реферата.
7. Сформировать новое представление знаний на основании макроструктуры текста.

Таблица 2.5. Сравнительные характеристики активных методов извлечения знаний

Методы	Достоинства	Недостатки	Требования к эксперту	Требования к аналитику	Предметная область
Анкетирование	Возможность стандартного опроса экспертов. Не требует напряжения от аналитика во время процедуры анкетирования	Требует умения и опыта составления анкет. Отсутствие контекста и обратной связи. Вопросы анкеты могут быть неправильно поняты экспертом	Практик и мыслитель	Мыслитель (педантизм в составлении анкет, внимательность)	Слабоструктурированные, слабо- и среднедокументированные
Интервьюирование	Наличие обратной связи (возможность уточнения контекста и разрешения противоречий)	Требует значительного времени на подготовку вопросов интервью	Собеседник и мыслитель	Собеседник (журналистские навыки, умение слушать)	
Свободный диалог	Гибкость. Обратная связь. Возможность изменения сценария и формы сеанса	Требует от аналитика напряжения. Отсутствие формальных методик проведения. Трудность протоколирования результатов		Собеседник (наблюдательность, умение слушать, обаяние)	

Таблицы 2.5. (продолжение)

Методы	Достоинства	Недостатки	Требования к эксперти	Требования к аналитику	Предметная область
«Мозговой штурм»	Позволяет выявлять глубинные пласты знаний (на уровне бессознательного). Активизирует экспертов. Позволяет получать новые знания	Возможен только для новых интересных исследовательских проблем. Не всегда эффективен (довольно низкий процент продуктивных идей)	Мыслитель (способность к творчеству)	Собеседник или мыслитель (быстрая реакция и чувство юмора)	Слабоструктурированные и слабодокументированные с наличием перспективных «белых пятен»
Дискуссия за круглым столом	Позволяет получить более объективные фрагменты знаний. Оживляет процедуру извлечения знаний. Позволяет участникам обмениваться знаниями	Требует больших организационных затрат. Отличается сложностью проведения	Собеседник или мыслитель (искусство полемики)	Собеседник (ди-пломатические способности.)	Слабоструктурированные и слабодокументированные с наличием спорных проблем

Таблицы 2.5. (окончание)

Методы	Достоинства	Недостатки	Требования к эксперту	Требования к аналитику	Предметная область
Индивидуальные экспертные игры	Дают возможность быстро получить качественную картину принятия решения. Позволяют выяснить, какую информацию и как использует эксперт	Отсутствие методик и стандартного набора игр. Высокие профессиональные требования к аналитику	Собеседник или практик (раскованность и актерское мастерство)	Собеседник (режиссерские способности, умение создавать сценарии, актерское мастерство)	Средне- и слабоструктурированные и слабодокументированные
Групповые экспертные игры	Реалистично воссоздают атмосферу конкретной задачи. Раскрепощают экспертов. Более объективны. Выявляют логику и аргументацию экспертов	Требуют от аналитика знания основ игротехники. Сложность создания игр для конкретных предметных областей		Собеседник (способность к ведению конференса, режиссерские способности, актерское мастерство)	
Компьютерные экспертные игры	Вызывают интерес у эксперта. Привлекают дизайном и динамикой	Сложность и высокая стоимость создания специализированных игр в конкретной предметной области	Практик (без психологического барьера к ЭВМ)	Мыслитель (контакт с программистом)	

Схема извлечения знаний из специальных текстов приведена на рис. 2.23, где M_1 – смысл, заложенный автором и основанный на его собственной модели мира; M_2 – смысл, который постигает инженер по знаниям; I – интерпретация текста, изложенного словесно; T – словесное изложение знаний; V – результат вербализации. Автор книги (эксперт) на основании понимания реальности формирует собственную модель мира M_1 . Эксперт пишет текст книги и формализует лишь некоторую часть M_1 . Затем инженер по знаниям, на основании информации, полученной из текста книги T , и через призму собственных знаний и опыта, интерпретирует знания эксперта. При этом модель мира инженера по знаниям M_2 может заметно отличаться от модели мира эксперта M_1 (которая, на самом деле, также может быть далека от реальности).



Рис. 2.23. Схема извлечения знаний из текста

Научный текст T , который формируется автором на основе модели M_1 включает в себя пять компонентов, погруженных в языковую среду $L: T = (\alpha, \beta, \gamma, \delta, \lambda)$, где α – первичный материал наблюдений; β – система научных понятий; γ – субъективный взгляд автора; δ – дополнительная информация, не касающаяся научного знания (бытовые знания); λ – заимствования. На процесс понимания (интерпретации) I и на модель M_2 влияют следующие компоненты: $M_2 = [(\alpha_T, \beta_T, \gamma_T, \lambda_T); (\varpi, \varepsilon, \varphi)]$, где $(\alpha_T, \beta_T, \gamma_T, \lambda_T)$ –

экстракт компонентов, почерпнутый из текста T ; ϖ – предварительные знания аналитика о предметной области; ε – общенаучная эрудиция инженера по знаниям; φ – личный опыт аналитика. Особую сложность представляет извлечение знаний из специальной литературы и методик, поскольку в них очень высока степень концентрации специальных знаний.

Трудности в разработке ЭС, связанные с проблемой извлечения и структурирования знаний, исследует и решает **инженерия знаний** (Knowledge Engineering) – под этим научным направлением понимают разработки в области ИИС, ставящее целью создание моделей, методов и систем для получения, структурирования и формализации знаний ЛПР с целью проектирования БЗ.

Одним из центральных в инженерии знаний является понятие **поле знаний** – согласно [23; 111], это условное неформальное описание основных понятий и взаимосвязей между понятиями предметной области, выявленных из системы знаний эксперта в виде графа, диаграммы, таблицы или текста. Поле знаний как первый шаг от структурирования к формализации представляет собой **модель знаний** о предметной области в том виде, в каком ее сумел выразить аналитик (инженер по знаниям) на некотором «собственном» языке.

При формировании поля знаний главным является **процесс получения знаний**, когда происходит перенос компетентности экспертов на инженеров по знаниям. Для характеристики этого процесса среди специалистов по ЭС распространены также другие варианты его названия: приобретение, добыча, извлечение, выявление, формирование знаний. Ключевая процедура **извлечения знаний** имеет в виду взаимодействие эксперта с источником знаний, в результате которого становятся явным процесс рассуждений специалистов при принятии решения и структура их представления о предметной области.

Процесс извлечения знаний является наиболее длительным и трудоемким этапом при разработке ЭС. Обычно эксперт не прибегает здесь к процедурным или количественным методам, его основные средства – аналогия, интуиция, абстрагирование. Если эксперт не может связно и доходчиво объяснить, каким именно образом им было найдено то или иное решение, на инженера по знаниям возлагается сложная работа по преобразованию этих невразум-

мительных описаний в строгую, полную и непротиворечивую систему, которая позволила бы решать прикладные задачи не хуже, чем это делал бы сам эксперт.

Основная сложность проблемы извлечения знаний «из человека» заключается в том, что все ментальные, когнитивные процессы, а также мышление, память и работа сознания субъекта доступны для непосредственного восприятия только самому субъекту. Для всех остальных людей сознание, знания, память, мысли другого человек всегда остаются «черным ящиком». Поэтому об уровне интеллектуального развития и способах мышления каждого человека можно судить только каким-либо внешним проявлением этих характеристик, которые для остальных людей выступают своеобразными «индикаторами» при формировании суждений. Естественно, это, прежде всего, деятельность человека, его речь, какие-либо предметы, созданные человеком и имеющие определенное смысловое наполнение. Здесь необходимо также учитывать, что механизмы работы сознания и памяти человека, процессы формирования его знаний являются в настоящее время во многом непознанными, парадоксальными и неподдающимися формализации.

Следует отметить ряд проблем, сопровождающих процесс формирования поля знаний. **Во-первых**, это эффект «испорченного телефона»: Internet демонстрирует схему, в шуточной форме отображающую отношения между реальной действительностью и полем знаний (см. рис. 2.24), которая иллюстрирует фразу из популярной песенки «Сделать хотел грозу, а получил козу». Как видно из рис. 2.24, конечный результат получен после третьей трансляции: первоначальный замысел изменяется до неузнаваемости, проходя цепочку из «грозы», «дождя», «облака-барашка» и превращаясь, наконец, в «розовую козу». Данная схема действительно похожа на детскую игру в «испорченный телефон», в ходе которой специалист по извлечению знаний пытается добиться максимального соответствия модели эксперта и реальности. Кроме того, приходится учитывать, что на трактовку полученных сведений оказывает влияние (всегда не полностью предсказуемое в тех или иных деталях, которые могут оказаться важными) и личность самого эксперта. Последующие трансляции способны еще больше исказить и видоизменить модель, но теперь уже с учетом субъективного восприятия инженера по знаниям.

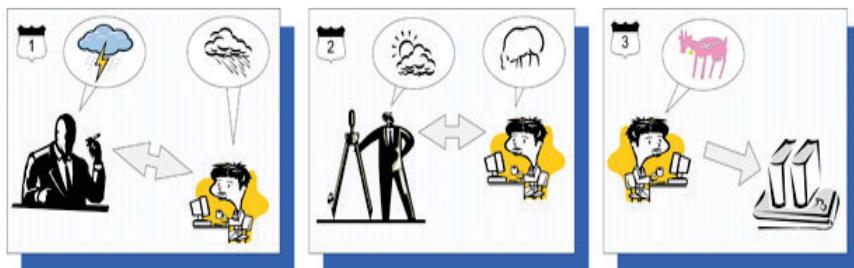


Рис. 2.24. Эффект «испорченного телефона»

Рассмотрим влияние субъективных взглядов эксперта на примере анализа следующей ситуации: два ЛПР прибегают на вокзал за 2 мин. до отхода пригородного поезда. У обычных кассы – очереди, в автоматических кассах свободно, но у обоих нет мелких денег, а разменять негде и некогда: оба опаздывают на важные встречи, тогда как следующий поезд через 40 мин.

Интерпретация ситуации первым экспертом: нельзя приходить на вокзал менее чем за 10 мин. до отхода поезда; интерпретация второго эксперта: надо всегда иметь мелочь в кармане. Вербализация первого эксперта: я опоздал к нужному поезду, так как неправильно рассчитал время; вербализация второго эксперта: я опоздал, так как на вокзале неразбериха, в кассах толпа и нигде у нас в стране порядка нет.

Во вторых, важное значение имеет проблема организации эффективного взаимодействия участников процесса передачи знаний. Чтобы сформировать суждения о знаниях, необходимо организовать такое взаимодействие с экспертом, в процессе которого он тем или иным способом проявит свои знания. Информация, в которой содержатся полученные знания, может быть представлена в числовой, текстовой, графической, звуковой или иной форме. При формировании суждений о знаниях обязательно должна учитываться специфика используемой формы взаимодействия обладателя и получателя знаний: свободный диалог, анкетирование, интервьюирование, тестирование. При разработке ЭС, основанной на правилах, до 60-70% времени тратится на извлечение знаний – что обусловлено структурной сложностью большинства правил. Эксперт, например, обдумывает альтернативные сценарии и говорит: «Я ду-

маю, что при обстоятельствах X наиболее вероятное заключение – Y , но если есть дополнительный фактор F , то более вероятное заключение могло бы быть P » и т.д.

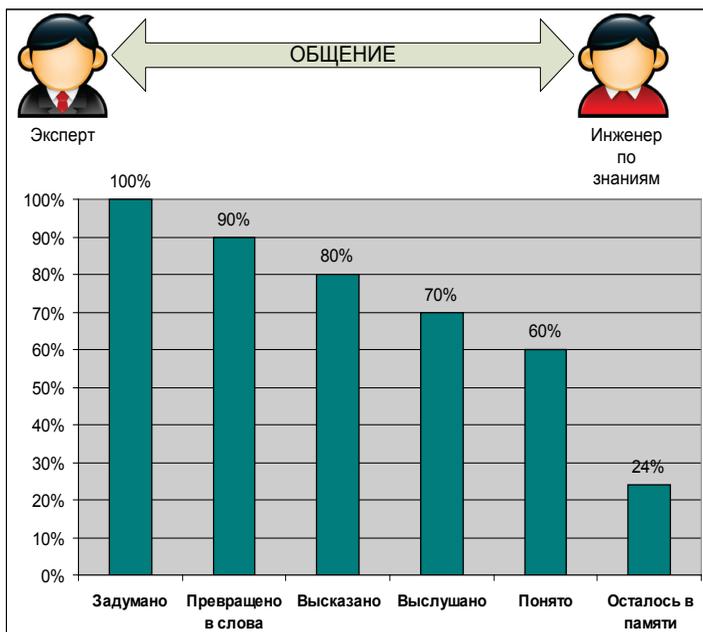


Рис. 2.25. Потери информации при разговорном общении

Причин сложности налаживания коммуникаций немало, поскольку почти всегда члены команды будущей ЭС имеют перво-классное техническое образование, но не владеют хотя бы начатками профессиональных приемов общения. С другой стороны, нередки случаи, когда эксперт не обладает навыками вербализации собственных знаний – в таких случаях процесс их извлечения также существенно затруднен. Кроме того, необходимо помнить и об известной статистике информационного содержания при разговорном общении людей – см. схему на рис. 2.25.

В-третьих, проблема интерпретации полученной информации о знаниях, вследствие которой определенный процент информации при каждом разговорном общении людей неизбежно теряется. Ведущим в процессе извлечения знаний является психологический

аспект, поскольку именно он определяет успешность и эффективность взаимодействия инженера по знаниям (аналитика) с основным источником знаний – экспертом-профессионалом. Психологический аспект выделяется и потому, что извлечение знаний происходит чаще всего в процессе непосредственного общения инженера по знаниям и разработчиков ЭС с экспертами, ввиду чего особенности их бытовой психологии могут свести на нет самый высокий интеллектуальный потенциал создаваемой ЭС. *Общение* (коммуникация) – это междисциплинарное понятие, обозначающее все формы непосредственных контактов между людьми – от дружеских отношений до деловых. Существуют десятки теорий общения, и единственное, в чем сходятся их авторы – это сложность и многоплановость процедуры общения [111].

В-четвертых, необходимо учитывать разницу в менталитете инженера по знаниям и эксперта. Пожалуй, данная проблема не так значительна, однако она также может оказать существенное влияние на процесс извлечения и приобретения знаний. Для иллюстрации приведем отрывок из романа А. и Б. Стругацких «Хромая судьба», где переводчик просит писателя-япониста помочь с переводом современных русских сказок на японский язык: «Я вспомнил о втором листке из конверта и, развернув, бегло просмотрел его. Сначала недоумения Рютаками позабавили меня, но не прошло и несколько минут, как я понял, что ничего особенно забавного мне не предстоит.

А предстоит мне объяснить, да еще в письменном виде, да еще японцу, что означают такие, например, выражения: «хватить шилом патоки», «цвести как майская роза», «иметь попсовый вид», «полные штаны удовольствия», «начистить ряшку» и «залить зенки»... Но все это было еще полбеды, и не так уж, в конце концов, трудно было объяснить японцу, что «банан» на жаргоне школьников означает «двойку как отметку, в скобках, оценку», а «забойный» означает всего-навсего «сногшибательный» в смысле «великолепный». А вот как быть с выражением «фиг тебе»? Во-первых, фигу, она же дуля, она же кукиш, надлежало самым решительным образом отмежевать от плодов фигового дерева, дабы не подумал таками, что слова «фиг тебе» означают «подношу тебе в подарок спелую, сладкую фигу». А во-вторых, фига, она же дуля, она же кукиш, означает для японца нечто иное, нежели для европейца или,

по крайней мере, для русского. Этой несложной фигурой из трех пальцев в Японии когда-то пользовались уличные дамы, выражая готовность обслужить клиента...»

В-пятых, проблема достоверности формируемой модели знаний. Для того чтобы инженер по знаниям мог оценить адекватность суждений эксперта, ему необходимо сравнить полученную модель знаний с «реальностью» – но реальные знания эксперта недоступны для его непосредственного восприятия, поэтому такое контрольное сравнение провести невозможно. Одним из методов обеспечения (повышения) достоверности и адекватности построенной конкретным субъектом модели знаний является ее сравнение с другими моделями – желательно созданными разными субъектами и разными способами – например, на основе анализа информации, полученной от других экспертов, или на основе других форм взаимодействия с ними. Тогда можно считать, что те знания, которые присутствуют одновременно в нескольких моделях, имеют требуемый высокий уровень достоверности.

В-шестых, проблема неопределенности (неполноты, неточности, нечеткости) формируемой модели знаний. Естественно, что любая построенная модель будет неполной в силу недостаточного времени взаимодействия и других ограничений, накладываемых выбранной формой взаимодействия, а также значительного числа случайных факторов, влияющих на процесс взаимодействия. На способность выражения знаний эксперта, например, существенное влияние может оказать его физиологическое и эмоциональное состояние, способность и умение ясно и кратко формулировать свои знания вербальной форме.

2.6. Проблема неопределенности знаний

Некорректно сформулированные постановки задач существуют во всех предметных областях. Одним из существенных достоинств ЭС, отличающих их от других информационных технологий, является обладание методами, позволяющими быть точными в отношении неточностей. Будем называть высказывание неточным, если его истинность (или ложность) не может быть установлена с определенностью.

При решении проблем эксперт (или пользователь) часто сталкивается со множеством источников неопределенности используемой

информации, но в большинстве случаев их можно разделить на две категории [90]:

- недостаточно полное знание предметной области;
- недостаточная информация о конкретной ситуации.

Знания эксперта о предметной области могут быть неполными или не до конца ясными: могут использоваться недостаточно четко сформулированные концепции или недостаточно изученные явления. Неопределенность знаний приводит к искажению результата. Так, располагая неполным знанием, нельзя уверенно предсказать, какой эффект даст то или иное действие. Следовательно, правила, сформулированные на базе неполных знаний, будут давать заведомо искаженный результат.

Другим случаем может служить ситуация, когда эксперт владеет достаточно полной информацией о предметной области, но по каким-то причинам (например, экономя время) может решить, что эффективнее использовать не точные, а эвристические методы. Так, методика устранения неисправности в электронном блоке путем замены подозрительных узлов оказывается значительно более эффективной, чем скрупулезный анализ цепей в поиске детали, вышедшей из строя.

Но, помимо неточных знаний, неопределенность может быть внесена неточными или ненадежными данными о конкретной ситуации. Например, любой датчик имеет ограниченную разрешающую способность и далеко не стопроцентную надежность; при составлении отчетов могут быть допущены ошибки и опечатки или в них могут попасть недостоверные сведения. Помимо всего прочего, необходимо учитывать и фактор времени: например, недостаток времени или, напротив, устаревшие данные.

Таким образом, в итоге можно сказать, что эксперты вынуждены использовать неточные методы по двум причинам: точные методы не существуют либо точные методы существуют, но не могут быть применены на практике из-за отсутствия необходимого объема данных или невозможности их накопления по соображениям стоимости, технических или программных возможностей, риска или из-за отсутствия времени на сбор необходимой информации.

Теоретической основой обработки неточной информации является теория вероятностей (теорема Байеса) и нечеткая логика Заде. Оба этих математических аппарата имеют как своих сторонников,

так и противников; каждый из методов имеет как недостатки, так и преимущества. И решение о том, каком из методов следует применять в том или ином проекте, принимается индивидуально и зависит от множества факторов: начиная от наличия достаточной статистических данных и заканчивая квалификацией эксперта по применению того или иного математического аппарата.

Подход на основе теоремы Байеса. Одним из вычислений неопределенностей, имеющим сравнительно прочные теоретические основы, является теория вероятностей. Во многих ЭС теорема Байеса используется в качестве нити, помогающей связать вместе информацию, поступающую из различных источников.

Теории Байеса опираются на предположение, что практически для любого положения имеется, какая бы малая она не была, априорная вероятность того, что данное положение истинно. Байесовский подход предполагает начальное априорное задание предполагаемых гипотез (значений достигаемых целей), которые последовательно уточняются с учетом вероятностей свидетельств в пользу или против гипотез, в результате чего формируются апостериорные (условные) вероятности.

$$P(H : E) = \frac{P(E : H) \cdot P(H)}{P(E)}, \quad (2.1)$$

$$P(\text{не}H : E) = \frac{P(E : \text{не}H) \cdot P(\text{не}H)}{P(E)}, \quad (2.2)$$

$$P(E) = P(E : H) \cdot P(H) + P(E : \text{не}H) \cdot P(\text{не}H), \quad (2.3)$$

где $P(H)$ – априорная вероятность гипотезы H ; $P(\text{не}H) = 1 - P(H)$ – априорная вероятность отрицания гипотезы H ; $P(E)$ – априорная вероятность свидетельства E ; $P(H:E)$ – апостериорная (условная) вероятность гипотезы H при условии, что имеет место свидетельство E ; $P(\text{не}H:E)$ – апостериорная (условная) вероятность отрицания гипотезы H при условии, что имеет место свидетельство E ; $P(E:H)$ – вероятность свидетельства E при подтверждении гипотезы H ; $P(E : \text{не}H)$ – вероятность свидетельства E при отрицании H .

Теперь, если имеется априорная вероятность некоторой гипотезы (например, болен ли пациент гриппом), то должны иметься некоторые данные, которые можно было бы привлечь для подтвер-

ждения вывода эксперта по данному вопросу. Если бы таких дополнительных данных не было, то указанная априорная вероятность оставалась бы навсегда неизменной. Однако при наличии относящихся к делу сведений можно модифицировать априорную вероятность той же самой гипотезы с учетом поступивших новых данных.

Вернемся к нашему примеру с пациентом, больном (или не больном) гриппом – поскольку определить данный факт и есть поставленная задача. Эксперт может ориентировочно (по имеющейся статистической выборке) определить, какова априорная вероятность того, что пациенты болеют гриппом, то есть знает $P(H)$. Но в рассматриваемом случае необходимо знать вероятность гриппа у данного, конкретного пациента.

Предположим, что врач обнаруживает у пациента жар и хочет узнать $P(H:E)$ – вероятность того, что у пациента грипп, при условии, что у него есть температура. Можно строить догадки, но это не очень простой вопрос. Намного легче ответить на вопрос противоположного свойства: если у больного грипп, какова вероятность, что у него высокая температура? Ответ получить легко – эта вероятность равна 1, а дальше по формуле Байеса можно вычислить искомую вероятность.

Предварительно необходимо прояснить еще один момент, а именно выяснить значение знаменателя в уравнении, то есть $P(E)$ – вероятность самого свидетельства, в нашем случае вероятность, что у человека имеется высокая температура. Ведь если у каждого всегда повышенная температура, то тот факт, что у нашего пациента жар, ничего нам не объяснит. У нашего пациента жар, но является ли такое состояние в общем случае редким событием?

Для вычисления $P(E)$ используем формулу (2.3). Другими словами, вероятность того, что у произвольного человека может иметься высокая температура, равна вероятности температуры у больного гриппом, умноженной на вероятность гриппа у произвольно взятого человека, плюс вероятность высокой температуры у пациента не больного гриппом, умноженная на вероятность того, что данное лицо не более гриппом.

Здесь начинают проявляться достоинства байесовского метода. Первоначальное значение $P(H)$ хранится в базе знаний ЭС. Но, вычислив новую величину $P(H:E)$ (то есть задав, допустим вопрос о

наличии у больного высокой температуры), можно отбросить первоначальную величину $P(H)$ и воспользоваться значением $P(H:E)$ как обновленным значением $P(H)$. Таким образом весь процесс можно повторять теоретически бесконечно, каждый раз несколько увеличивая или уменьшая вероятность заболевания, но каждый раз обращаясь к одной и той же формуле Байеса, подставляя в нее каждый раз новую априорную вероятность, получаемую из апостериорной вероятности, имевшейся на предыдущем шаге. В конечном итоге, собрав все сведения, касающиеся всех гипотез, ЭС может прийти к окончательному заключению, что верная гипотеза является истинной.

Байесовский подход встречает ряд критических замечаний. Первое из них относится к вычислению $P(E)$. Рассмотренное вычисление является точным и довольно простым, если известна $P(E:neH)$. Как правило, эксперт знает очень точное значение для $P(E:H)$, но значение $P(E:neH)$ не всегда может быть точно известно.

Одна из возможностей обойти это затруднение заключается в использовании общей формулы для вероятности некоторого свидетельства с учетом того, что использованная выше формула (2.3) является частным случаем выражения (2.4).

$$P(E) = \sum_{i=1}^n P(E : H_i) \cdot P(H_i), \quad (2.4)$$

Формула означает, что вероятность свидетельства E есть вероятность E при учете каждой ситуации, которая может повлечь за собой появление E , умноженная на вероятность такого события.

Если $P(H)$ постоянно уточняется в ходе работы, то $P(E)$ теоретически также может уточняться с течением времени. Трудность и основная опасность состоит в неполноте базы знаний (точнее, в отсутствии в базе знаний абсолютно всех данных, касающихся какой-либо ситуации), поскольку отсутствие какого-либо «симптома» приводит к соответствующей ошибке при вычислении $P(E)$, точно так же как и при формировании окончательного результата.

С практической точки зрения использование грубых оценок для $P(E:H)$ и для $P(E:neH)$ вполне допустимо, так как, используя более простой метод, можно экономить объем необходимых вычислений и быстрее получить ответ.

Второе замечание касается того, что в этом методе используется предположение о независимости переменных. С точки зрения теории подобное замечание весьма серьезно. Допустим, в БЗ содержатся два симптома одного и того же состояния, которые полностью коррелируют друг с другом. Таким образом, эксперт включил одно свидетельство дважды и, следовательно, увеличил или уменьшил вероятность какой-либо гипотезы по сравнению с тем, что должно было быть, так что апостериорные вероятности будут ошибочными.

В каком-то небольшом масштабе подобная трудность не может не возникнуть, поскольку большинство симптомов всегда в какой-то степени коррелируют друг с другом. И из этого положения нет никакого теоретически приемлемого способа. Эта проблема встает особенно остро, если пользователю необходимо иметь точные вероятности для каждой из рассматриваемых гипотез. Если же пользователю достаточно знать порядок отношения вероятностей, то проблема исчезает. До тех пор пока информация в базе знаний одинаково ошибочна в отношении каждого ее элемента и числа свидетельств, имеющих для каждой гипотезы, примерно одно и то же, то и относительный порядок ошибок в целом примерно одинаков.

Таким образом, при использовании байесовского подхода необходимо придерживаться одной из стратегий:

- либо априори предполагать, что все данные независимы, и использовать менее трудоемкие методы вычислений, за что придется расплачиваться снижением достоверности результатов;
- либо необходимо организовать отслеживание и анализ зависимости между используемыми данными, количественно оценить эту зависимость, реализовать оперативное обновление соответствующей нормативной информации, то есть усложнить вычисления, но получить более достоверные результаты.

Вычисления (2.1) могут быть реализованы с помощью понятий о шансах и отношении правдоподобия. Априорные шансы $O(H)$ и апостериорные шансы $O(H : E)$ могут быть выражены как

$$O(H) = \frac{P(H)}{1 - P(H)}; \quad (2.5)$$

$$O(H : E) = \frac{P(H : E)}{1 - P(H : E)}. \quad (2.6)$$

Уравнение (2.6) может быть переписано в форме

$$O(H : E) = LR(H : E) \cdot O(H); \quad (2.7)$$

где отношение правдоподобия есть

$$LR(H : E) = \frac{P(E : H)}{P(E : неH)}. \quad (2.8)$$

Согласно (2.8) апостериорные шансы могут быть вычислены как произведение $O(H)$ и $LR(H : E)$.

Таблица 2.6. Статистические данные о доходности ценных бумаг

Отрасль промышленности	Доходность ценных бумаг		Всего
	Высокая ($\geq 30\%$)	Низкая ($< 30\%$)	
Электроника	40	10	50
Другие	50	100	150
Всего	90	110	200

Рассмотрим пример [98]. Дана таблица частот, которая отражает данные о доходности 200 ценных бумаг для различных предприятий (см. таблицу 2.6). В качестве гипотезы H принято попадание в группу ценных бумаг с высоким уровнем доходности, $не H$ – попадание в группу предприятий с низким доходом, свидетельство E – ценные бумаги принадлежат электронной индустрии, $не E$ – ценные бумаги принадлежат неэлектронной индустрии.

Если нет другой информации о ценных бумагах, априорную вероятность доходности можно определить как

$$P(H) = P(\text{высокая}) = \frac{90}{200} = 0,450;$$

$$P(\text{не}H) = P(\text{низкая}) = 1 - 0,450 = 0,550.$$

В случае, если отрасль, к которой принадлежит ценная бумага, известна, то можно применить формулу (2.7). Априорные шансы равны

$$O(H) = O(\text{высокая}) = \frac{P(\text{высокая})}{1 - P(\text{высокая})} = \frac{0,450}{1 - 0,450} = 0,818.$$

Отношение правдоподобия равняется

$$\begin{aligned} LR(\text{Высокая} : \text{Электроника}) &= \frac{P(\text{Электроника} : \text{Высокая})}{P(\text{Электроника} : \text{Низкая})} = \\ &= \frac{40/90}{10/110} = 4,888. \end{aligned}$$

Следовательно, из (2.7) следует, что

$$\begin{aligned} O(\text{Высокая} : \text{Электроника}) &= \\ &= LR(\text{Высокая} : \text{Электроника}) \cdot O(\text{Высокая}) = 0,818 \cdot 4,888 = 3,996. \end{aligned}$$

Преобразование в апостериорную вероятность дает

$$P(\text{Высокая} : \text{Электроника}) = \frac{3,996}{1 + 3,996} = 0,801.$$

Таким образом, при наличии свидетельства о том, что ценные бумаги принадлежат электронной отрасли, вероятность получения дохода >30% увеличивается с 0,450 до 0,801.

Таблица 2.7. Статистические данные о рынке

Рынок	Доходность ценных бумаг		Всего
	Высокая ($\geq 30\%$)	Низкая ($< 30\%$)	
Экспорт	40	30	80
Внутренний рынок	50	80	120
Всего	90	110	200

Предполагая, что существуют дополнительные свидетельства относительно того, ориентирована ли продукция на экспорт или внутренний рынок, дополнительная таблица частот для рынка может иметь вид таблицы 2.7.

Тогда апостериорное отношение правдоподобия

$$LR(\text{Высокая} : \text{Экспорт}) = \frac{P(\text{Экспорт} : \text{Высокая})}{P(\text{Экспорт} : \text{Низкая})} = \frac{50/90}{30/110} = 2,037.$$

Следовательно,

$$O(\text{Высокая:Электроника,Экспорт}) = LR(\text{Высокая:Экспорт})$$

$$LR(\text{Высокая:Электроника}) \cdot O(\text{Высокая}) = LR(\text{Высокая:Экспорт})$$

$$O(\text{Высокая:Электроника}) = 2,037 \cdot 3,996 = 8,14.$$

$$P(\text{Высокая : Электроника, Экспорт}) = 0,8907.$$

Таким образом, располагая этими двумя элементами свидетельств – отрасль промышленности и экспортная ориентация продукции – получаем увеличение вероятности получения более высокого дохода с 0,45 до 0,8907.

Подход на основе нечеткой логики. Более простым, но менее точным методом оценки достоверности используемых знаний является применение нечеткой логики (Fuzzy Logic). Впервые идея нечеткой логики была предложена Л. Заде в 1965 г., который распространил булеву алгебру на действительные числа. Основной идеей нечеткой алгебры является использование нечетких чисел, то есть любая величина может принимать не конкретное значение, а принадлежать какому-либо множеству, причем для каждого элемента этого множества определяется вероятность того, что измеряемая величина примет именно это значение. При использовании нечеткой логики вероятности заменяются на экспертные оценки определенности фактов и применения правил (или, как их еще называют, факторы уверенности, коэффициенты уверенности, коэффициенты определенности, меры доверия). Коэффициенты уверенности, введенные Е. Шортлиффом [97] для измерения степени доверия к любому заключению, могут рассматриваться и как весовые коэффициенты, отражающие степень важности аргументов в процессе вывода заключений. Итоговые факторы уверенности получаемых решений главным образом отражают порядок достоверности результата, а не его точность, что вполне приемлемо во многих задачах.

Коэффициенты уверенности измеряются по некоторой относительной шкале, например, от –100 до 100 или от –1 до 1. В отличие от теории вероятностей сумма факторов уверенности некоторых альтернативных значений необязательно составляет 100 (или 1).

Коэффициенты уверенности не являются вероятностными мерами, они подчиняются некоторым аксиомам теории вероятностей, но не являются выборками из какой-нибудь популяции, и, следова-

тельно, им нельзя дать статистическую интерпретацию. Они просто позволяют упорядочить гипотезы в соответствии с той степенью обоснованности, которая у них есть.

Предполагается, что оценка факторов уверенностей исходных данных задается пользователем при описании конкретной ситуации, а факторы уверенности применения правил определяются инженерам знаний совместно с экспертами при наполнении БЗ.

Рассмотрим правило вида ЕСЛИ X , ТО C . Будем считать, что условие, правило и заключение не точны и заданы с некоторым коэффициентом определенности. Тогда, следуя идеологии ЭС МУСИН, коэффициент уверенности для заключения будет равен

$$cf(C) = cf(X) \cdot cf(ПР); \quad (2.9)$$

где $cf(C)$ – коэффициент уверенности заключения C ; $cf(X)$ – коэффициент уверенности условия X ; $cf(ПР)$ – коэффициент уверенности правила ПР.

В нечеткой логике определены эквиваленты логических операций вида:

$$X1 \text{ И } X2 \Rightarrow \min(X1; X2); \quad (2.10)$$

$$X1 \text{ ИЛИ } X2 \Rightarrow \max(X1; X2); \quad (2.11)$$

$$\text{НЕ } X \Rightarrow -X. \quad (2.12)$$

Таким образом

$$cf(X1 \text{ И } X2) = \min sf(X1; X2); \quad (2.13)$$

$$cf(X1 \text{ ИЛИ } X2) = \max sf(X1; X2). \quad (2.14)$$

Довольно часто одно и то же заключение поддерживается несколькими независимыми правилами. Одним из многих вкладов, привнесенных системой МУСИН, является некоторое решение проблемы комбинирования свидетельств. То, что было добавлено Е. Шортлиффом [97], – это формула уточнения, по которой новую информацию можно непосредственно сочетать со старыми результатами.

$$\begin{aligned}
 cf(C) &= cf(C_{PP1}) + cf(C_{PP2}) \cdot [1 - cf(C_{PP1})] = \\
 &= cf(C_{PP1}) + cf(C_{PP2}) - cf(C_{PP1}) \cdot cf(C_{PP2}),
 \end{aligned}
 \tag{2.15}$$

где $cf(C_{PPi})$ – коэффициент уверенности для i -го правила.

Смысл (2.15) состоит в том, что эффект второго правила ($PP2$) на заключение C при заданном первом правиле ($PP1$) сказывается в смещении cf в сторону полной определенности на расстояние, зависящее от второго правила, то есть происходит усиление заключения C . Эта формула имеет два важных свойства:

- она симметрична – порядок следования $PP1$ и $PP2$ является несущественным;
- по мере накопления подкрепляющих свидетельств $cf(C_{PPi})$ общий результат движется к определенности.

Существуют заключения, подтверждаемые не только двумя, но и большим количеством правил. Рассмотрим случай, когда количество правил равно трем. Из предыдущей формулы (2.15) коэффициент уверенности для заключения исходя из первых двух правил равен:

$$cf_{1,2}(C) = cf(C_{PP1}) + cf(C_{PP2}) - cf(C_{PP1}) \cdot cf(C_{PP2}).$$

Рассматривая это выражение в качестве единого целого, можно записать:

$$\begin{aligned}
 cf(C) &= cf_{1,2}(C) + cf(C_{PP3}) - cf_{1,2}(C) \cdot cf(C_{PP3}) = \\
 &= cf(C_{PP1}) + cf(C_{PP2}) + cf(C_{PP3}) - cf(C_{PP1}) \cdot cf(C_{PP2}) - \\
 &+ cf(C_{PP3}) \cdot [cf(C_{PP1}) + cf(C_{PP2}) - cf(C_{PP1}) \cdot cf(C_{PP2})] = \\
 &= cf(C_{PP1}) + cf(C_{PP2}) + cf(C_{PP3}) - cf(C_{PP1}) \cdot cf(C_{PP2}) - \\
 &- cf(C_{PP1}) \cdot cf(C_{PP3}) - cf(C_{PP2}) \cdot cf(C_{PP3}) + \\
 &+ cf(C_{PP1}) \cdot cf(C_{PP2}) \cdot cf(C_{PP3}).
 \end{aligned}$$

Аналогично можно получить правило для расчета коэффициента уверенности для любого количества заключений, выводимых на основании различных правил. Общий коэффициент уверенности заключения в случае, если это заключение выводится из двух правил, коэффициенты уверенности каждого из которых имеют различные знаки, рассчитывается следующим образом:

$$cf = \frac{cf(C_{PP1}) + cf(C_{PP2})}{1 - \min[abs(cf(C_{PP1})), abs(cf(C_{PP2}))]} . \quad (2.16)$$

Если оба коэффициента отрицательны, формула расчета имеет вид:

$$cf(C) = cf(C_{PP1}) + cf(C_{PP2}) + cf(C_{PP1}) \cdot cf(C_{PP2}) . \quad (2.17)$$

Рассмотрим применение аппарата нечеткой логики на примере оценки надежности поставщика, в котором кроме фактора финансового состояния учитывается и фактор формы собственности [106]. Пусть государственное предприятие не имеет задолженность с уверенностью 0,6 и предполагается, что его рентабельность удовлетворительна с уверенностью 0,8. Фрагмент множества правил имеет следующий вид:

Правило 1:

Если Задолженность = «нет» и Рентабельность = «удовл.»

То Финансовое_состояние = «удовл.» cf 1

Правило 2:

Если Финансовое_состояние = «удовл.»

То Надежность+ = «есть» cf 0,9.

Правило 3:

Если Предприятие = «государств.»

То Надежность+ = «есть» cf 0,5.

Результат выполнения первого правила:

$$cf(\text{посылки}) = \min(0,6; 0,8) = 0,6$$

$$cf(\text{Фин_сост.} = \text{«удовл.»}) = 0,6 \cdot 1 = 0,6 .$$

Результат выполнения второго правила:

$$cf(\text{Надежность} = \text{«есть»}) = 0,6 \cdot 0,9 = 0,54 .$$

Результат выполнения третьего правила:

$$cf(\text{Надежность} = \text{«есть»}) = 0,54 + 0,5 - 0,54 \cdot 0,5 = 0,77 .$$

Применение того или иного метода неточного рассуждения во многом зависит от выбранной модели представления знаний. Например, для байесовского подхода к построению продукционной базы знаний характерна большая трудоемкость статистического оценивания априорных шансов и факторов достаточности и необходимости. С другой стороны, вероятностный метод позволяет

получить рациональное решение даже при отсутствии достаточно-го объема информации.

Нечеткие числа, получаемые в результате «не вполне точных измерений», во многом аналогичны распределениям теории вероятностей, но свободны от присущих последним недостатков: малое количество пригодных к анализу функций распределения, необходимость их принудительной нормализации, соблюдение требований аддитивности, трудность обоснования адекватности математической абстракции для описания поведения фактических величин. В пределе, при возрастании точности, нечеткая логика приходит к стандартной, булевой. По сравнению с вероятностным методом, нечеткий метод позволяет резко сократить объем производимых вычислений, что, в свою очередь, приводит к увеличению быстродействия нечетких систем.

Несмотря на присущие нечетким системам недостатки – отсутствие стандартной методики конструирования нечетких систем, невозможность математического анализа нечетких систем существующими методами, применение нечеткого подхода по сравнению с вероятностным не приводит к повышению точности вычислений – использование коэффициентов определенности при продукционной модели представления знаний представляется более рациональным.

2.7. Выводы

Общей характеристикой ИИС является способность приобретать, воспроизводить и использовать знания. Подобные системы обладают следующими особенностями [65]:

- способны на основе сформированной или заданной цели определять методы, пути и средства достижения конечного результата;
- «умеют» проводить рассуждения в условиях неполноты информации с использованием правил как достоверного, так и правдоподобного выводов и, таким образом, порождать новые знания;
- обладают способностью (механизмами, методами, алгоритмами, программами и т.п.) к аппроксимации, обобщению и концептуализации знаний;
- в рамках своей компетенции должны уметь «понимать», ставить и решать задачи;

- аналогичным образом должны узнавать и распознавать ситуации, образы, процессы и явления окружающего их мира;
- являются информационно открытыми системами, расширяющими объем и содержание модели мира об окружающей среде и сфере деятельности;
- в процессе функционирования не только используют уже известную им информацию, но и генерируют новую (знания, данные), то есть выступают в качестве производителя и источника интеллектуальных информационных ресурсов.

Одной из коммерчески оправданных ветвей развития ИИС являются ЭС. Исторически значимыми являются две системы ЭС – это системы MYCIN и DENDRAL, которые продемонстрировали, что эффективность ЭС в значительно большей степени зависит от способа представления знаний и способа вывода, чем от конкретных числовых особенностей реализаций правил вывода.

Анализ существующих ЭС, а также особенности их функционирования, позволяет сделать вывод, что типовая архитектура подобных систем должна включать следующие компоненты: базу знаний; компонент приобретения знаний; компонент выдачи результатов; управляющий компонент; компонент обработки данных; компонент объяснений. Особенности проблемной области, в рамках которой разрабатывается конкретная ЭС, могут приводить к тому, что архитектура системы может дополняться новыми компонентами и модулями.

ЭС имеют и другое распространенное название – **системы, основанные на знаниях**. Подобное название подчеркивает важность понятия «знание», которое в качестве объекта исследования интересует сегодня многих исследователей – как теоретиков, так и практиков. Несмотря на то, что существует множество публикаций на эту тему, точной и однозначной трактовки данного понятия до сих пор не существует.

Главная проблема, с которой сталкиваются разработчики ЭС – это «извлечение» знаний из опыта людей-экспертов для формирования БЗ. В настоящее время вопросами «извлечение» знаний занимается инженерия знаний. Определение данной дисциплине было дано Фейгенбаумом в 1983 г.: «инженерия знаний – это раздел инженерии, направленный на внедрение знаний в компьютерные

системы для решения сложных задач, обычно требующих богатого человеческого опыта».

Еще одной сложностью, которую необходимо учитывать, является проблема неопределенности знаний ЛПР: либо недостаточно полное знание экспертом предметной области, либо недостаточная информация о конкретной ситуации. Оба эти фактора при разработке ЭС могут привести к получению некорректного результата. Математическими инструментами «уменьшения» неопределенности (точнее, снижения влияния неопределенности на точность принимаемых решений) являются теория вероятностей (в частности, теорема Байеса) и нечеткая логика Л. Заде. Решение о целесообразности применения конкретного технологического способа принимается руководителем проекта по разработке ЭС на основании совокупности принимаемых во внимание факторов.

3. ПРИНЦИПЫ РАЗРАБОТКИ, ПРОЕКТИРОВАНИЯ И РЕАЛИЗАЦИИ ЭКСПЕРТНЫХ СИСТЕМ

3.1. Этапы разработки ЭС

При проектировании ЭС необходимо соблюдать ряд основополагающих принципов:

- мощность ЭС в первую очередь зависит от мощности БЗ и возможности ее пополнения новыми знаниями, и лишь во вторую очередь – от используемых методов и процедур обработки имеющейся (в том числе вновь поступающей) информации;
- знания, позволяющие эксперту (или самостоятельной в этом отношении ЭС) получить качественные и эффективные решения задач, являются в основном эвристическими, эмпирическими, неопределенными, правдоподобными и т.п.;
- неформальный характер принимаемых решений большинства конкретных задач и используемых при этом знаний делает необходимым обеспечение активного диалога пользователя с ЭС в процессе их совместной работы.

Начиная с девяностых годов XX века, расширение сферы применения интеллектуальных информационных систем (ИИС) для решения практических задач выдвинуло на передний план проблему технологической поддержки разработки таких систем. Рассмотрим более подробно отработанные на сегодня элементы технологии создания ЭС.

Устоявшаяся «стандартная» технология проектирования ЭС, изложенная в работах Э. Попова [60; 91], Д. Уотермена [92; 107] и ряда других авторов [69, 71, 73], включает следующие этапы, показанные на рис. 3.1:

- **идентификация** – где определяются набор задач, цели разработки, ресурсы, наличие экспертов необходимой квалификации, требования будущих пользователей;
- **концептуализация** – здесь проводится содержательный анализ предметной области, включающий определение методов решения задач с выделением используемых понятий и взаимосвязей;
- **формализация** – на данном этапе определяются способы представления полученных знаний, специфицируются выделенные понятия, фиксируются способы интерпретации знаний, моделируется работа ЭС;

- **реализация** – данный этап включает программирование всех компонентов ЭС и наполнение БЗ;
- **тестирование** – в рамках этого этапа проводится комплексная проверка работоспособности и компетентности ЭС, которая осуществляется инженером по знаниям совместно с экспертом и пользователем;
- **опытная эксплуатация** – в ходе которой проверяется пригодность ЭС для решения задач ее конечных пользователей.

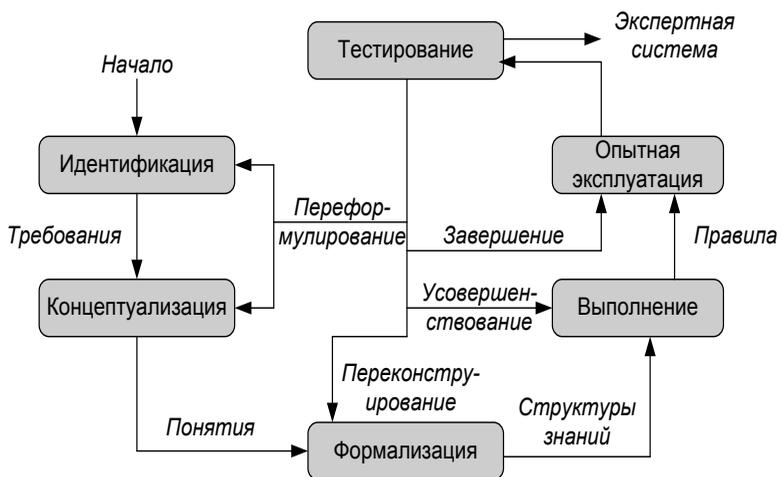


Рис. 3.1. Стандартная технология разработки ЭС

Характерной особенностью стандартной технологии являются многочисленные возвраты к любому ее этапу и пересмотр принятых там проектных решений – что удобно для быстрого создания небольших автономно функционирующих ЭС. Однако модель жизненного цикла, соответствующая данной технологии, не имеет шансов на промышленное использование из-за низкой эффективности разработки конкретной системы. Поэтому Л. Микуличем (см. [102] и др.) была предложена иная технология создания ЭС, включающая три фазы – проектирование, реализация и внедрение, жизненный цикл которой состоял из шести этапов:

- исследование выполнимости проекта;
- разработка общей концепции ЭС;
- разработка и тестирование общей серии прототипов;

- разработка и испытание головного образца;
- разработка и проверка расширенных версий системы;
- привязка системы к реальной рабочей среде.

Существенным недостатком обеих указанных технологий, однако, является тот факт, что они представляют собой декларативные методологические рекомендации, не подкрепленные общей моделью жизненного цикла. В [102] приводится также следующий вариант технологии, основанный на материалах Л. Микулича и обобщенном опыте отечественных разработчиков ЭС.

1. Фаза проектирования.

1.1. Инициирование проекта ЭС:

- составление плана-графика создания ЭС;
- определение потенциальной доступности аппаратных и программных средств для разработки ЭС;
- определение наличия и доступности экспертов;
- планирование способов управления ходом разработки ЭС и способов взаимодействия между разработчиками;
- планирование работы будущей ЭС совместно с имеющимися у заказчика другими ИС и ИИС;
- определение возможных трудностей и ограничений при разработке и установке системы.

1.2. Формирование группы разработки:

- определение состава группы в количестве 4-6 человек (руководитель проекта, ассистент руководителя, 1-2 инженера по знаниям, 1-2 программиста);
- изучение возможности привлечения специалистов по конкретным носителям машинных источников информации (если знания будут также извлекаться, например, из БД).

1.3. Определение требований к системе:

- определение четких представлений о среде, в которой будет работать ЭС;
- выяснение пожеланий (или требований) к ЭС со стороны руководства, экспертов, пользователей и разработчиков;
- формулирование требований к средствам разработки, экспертам, операционной среде, интерфейсу с пользователями, дисциплине поддержки и т.д.

1.4. Проведение исследований по выполнимости проекта:

- технические предпосылки:
 - пригодность задачи для ее решения с помощью ЭС (характеристики уместности, оправданности и возможности разработки ЭС);
 - характеристики знаний (примерный размер будущей БЗ, структура знаний, время, необходимое для получения знаний, характер знаний, соотношение между временем, затраченным на рассуждения при решении задачи, и временем, необходимым для сбора информации);
 - интерфейсы будущей системы (формы взаимодействия ЭС с различными категориями пользователей, интерфейсы между разными ЭС (если они работают в комплексе), интерфейсы между ЭС и внешними БД или оборудованием);
- экономические предпосылки:
 - оценка расходов на создание ЭС (доход больше расходов);
 - оценка расходов на поддержку функционирования ЭС;
 - культурные предпосылки (культура производства и уровень культуры среды, в которой будут работать пользователи).

1.5. Разработка общей концепции системы:

- разработка структуры БЗ и механизмов рассуждения (структура БЗ определяется разбиением всей задачи на отдельные, по возможности слабо связанные между собой подзадачи);
- выбор программных и аппаратных средств (то есть инструментальной среды) для разработки ЭС;
- разработка способа переноса ЭС на реальное оборудование и в реальную рабочую среду;
- определение критериев оценки функционирования системы.

2. Фаза разработки (по методу прототипирования):

- начальная стадия (разработка первого прототипа на основе концепции, выработанной на фазе проектирования);
- промежуточная стадия (создание на базе первого прототипа серии прототипов с последовательно улучшающимися характеристиками);
- финальная стадия (разрабатывается система, служащая основой для перевода в головной образец).

2.1. Общие соображения по прототипированию:

- выбор ширины и глубины проработки задачи, достаточной для данного прототипа;
- привлечение конечных пользователей к работе над прототипом;
- организация эффективного взаимодействия в группе разработки;
- проведение оценки функционирования очередного прототипа.

2.2. Виды стратегий прототипирования:

- стратегия I (для небольших задач): создание начального прототипа для всей задачи в целом, его тестирования и отладка на серии последующих прототипов;
- стратегия II (для крупных задач с интеграцией): создание прототипа, обладающего на поверхностном уровне всеми функциональными возможностями для всех подзадач (скелетный прототип), его тестирование, расширение последующих прототипов за счет углубленной проработки каждой из подзадач;
- стратегия III (для слабо пересекающихся задач): создание прототипов для каждой из подзадач, поочередное тестирование каждого из прототипов, а затем объединение всех составных прототипов в единый финальный прототип, решающий всю задачу в целом.

2.3. Документирование разработки.

2.4. Прототипирование с использованием стратегии I (по аналогии осуществляется прототипирование с использованием стратегий II и III):

- разработка серии начальных прототипов в интерактивном режиме;
- подведение итогов по каждому промежуточному прототипу по следующим критериям: общее впечатление от прототипа, качество разработки пользовательских интерфейсов, наличие пробелов или неадекватностей в знаниях и в форме их представления, естественность рассуждения системы;
- составление списка изменений;
- разработка финального прототипа (последний из промежуточных, служащий основой для решения вопроса о продолжении работ по созданию головного образца).

3. Фаза внедрения.

3.1. Создание на базе финального прототипа головного образца, опирающегося на реальную рабочую среду.

3.2. Перенос системы из среды разработки в среду функционирования у заказчика.

Авторы предлагают рассматривать ЭС в этом отношении как «обычную» ИС, которая по жизненному циклу и стадиям проектирования мало чем отличается от других ИС – варианты жизненного цикла и стадии которых определяются соответствующими нормативными документами (ГОСТ 34.601-90, ISO/IEC 12207:1995, Rational Unified Process (RUP), Microsoft Solution Framework (MSF) и др.). В этой связи таблица 3.1 иллюстрирует этапы жизненного цикла ЭС с учетом вышеуказанных методик ее создания [54]. Отметим, что при разработке ЭС, как правило, используется «спиральная» модель, суть которой заключается в том, что разработчики не пытаются построить конечный продукт сразу.

На начальном этапе создается прототип ЭС, удовлетворяющий двум противоречивым требованиям: умение решать задачи конкретного приложения и незначительные время и трудоемкость разработки ЭС. При выполнении заданных первоначальных условий становится возможным параллельный процесс накопления и отладки знаний, осуществляемый экспертом, и процесс выбора (разработки) программных средств, выполняемый инженером по знаниям и программистами. Для удовлетворения указанных требований при создании прототипа используются инструментальные средства, ускоряющие процесс проектирования.

Учитывая, что вопросы разработки ЭС в целом мало исследованы, перечень стадий в таблице 3.1 не представляется завершенным и исчерпывающим. Отметим, что вопросы управления проектной деятельностью по разработке «обычных» ИС также во многом открыты для дискуссий – о чем свидетельствует значительный объем публикаций, а также существующие разные методологии создания ИИС и ЭС.

В последующих разделах дано более подробное описание предлагаемых стадий работ по созданию ЭС.

Таблица 3.1. Этапы жизненного цикла ЭС

Этап	Стадии	Подэтапы	Выполняемые работы
Идентификация	Предпроектная стадия	Обследование объекта и обоснование необходимости создания ЭС	<ul style="list-style-type: none"> – сформировать команду проекта; – провести исследования по выполнимости проекта; – определить наличие и доступность экспертов; – определить потенциальную доступность аппаратных и программных средств для разработки; – определить способы управления ходом разработки и способы взаимодействия между разработчиками; – определить возможные трудности и ограничения при разработке ЭС (идентификация и оценка рисков); – провести предварительную экономическую оценку (технико-экономическое обоснование); – построить план-график выполнения работ.
Концептуализация		Формирование требований пользователя к ЭС	<ul style="list-style-type: none"> – определить требования к ЭС со стороны конечных пользователей и заказчика; – определить требования к среде функционирования ЭС (вопросы интеграции с уже имеющимися у заказчика системами); – сформулировать требования к средствам разработки, экспертам, операционной среде, пользовательскому интерфейсу (GUI), поддержке и т.п.; – разработать варианты концепции ЭС, удовлетворяющей требованиям пользователя; – разработать техническое задание.

Таблица 3.1. Этапы жизненного цикла ЭС (продолжение)

Этап	Стадии	Подэтапы	Выполняемые работы
Формализация	Стадия проектирования	Разработка общей концепции и архитектуры ЭС	<ul style="list-style-type: none"> – разработка архитектуры ЭС; – обосновать выбор метода формализации знаний; – разработать структуру базы знаний и механизмов рассуждения; – разработать способ переноса базы знаний на реальное оборудование и в реальную рабочую среду; – определить критерии оценки функционирования ЭС; – оформить рабочий проект.
Выполнение	Стадия разработки (с прототипами ЭС)	Создать очередной прототип (релиз) ЭС	<ul style="list-style-type: none"> – обосновать выбор достаточной глубины и ширины проработки задачи для данного прототипа; – заполнить БЗ.
Тестирование	Тестирование	Тестирование ЭС	<ul style="list-style-type: none"> – осуществить тестирование ЭС при помощи инженера по знаниям и эксперта; – привлечь конечных пользователей к работе над прототипом ЭС; – выявить и исправить ошибки; – провести оценку функционирования очередного прототипа; – принять решение о необходимости разработки очередного прототипа.
		Подготовка документации	<ul style="list-style-type: none"> – задокументировать процедуру инсталляции ЭС; – задокументировать процедуру наполнения базы знаний; – оформить руководство пользователя.

Таблица 3.1. Этапы жизненного цикла ЭС (окончание)

Этап	Стадии	Подэтапы	Выполняемые работы
Опытная эксплуатация	Стадия внедрения	Опытная эксплуатация	<ul style="list-style-type: none"> – установка всех модулей системы у конечного пользователя; – осуществить наладку и запуск ЭС у конечного пользователя; – внести изменения согласно предложениям, внесенным пользователями; – проведение приемочных испытаний.
		Обучение пользователей	<ul style="list-style-type: none"> – установить ЭС на рабочие места конечных пользователей; – осуществить настройку прав и уровней доступа пользователей; – провести обучение пользователей; – сформировать у пользователей знания и навыки по работе с ЭС.

3.2. Предпроектная стадия

Существует две основные точки зрения на проблему построения экспертных систем [101]. Первая точка зрения гласит, что важнее всего получить *конечный результат*, то есть приоритетным является совпадение поведения искусственно созданного интеллекта с реально существующим. При этом не учитывается внутренний механизм формирования поведения искусственной системы – то есть, таким образом, экспертная система представляется в виде «черного ящика».

Согласно второй точке зрения, изучение механизмов естественного мышления и анализ данных о способах разумного поведения людей могут создать основу для построения экспертной системы, то есть главным является *моделирование воспроизведения* техническими средствами принципов и конкретных особенностей деятельности ЛПР.

Соответственно этим точкам зрения выделяют два направления проведения научных исследований в области искусственного интеллекта:

- все данные рассматриваются на уровне их физиологических, нейропсихических и психических механизмов для последующей

реализации в виде экспертной системы – при этом исследователи стараются воспроизвести указанные механизмы при помощи соответствующих технических средств в составе как программной, так и аппаратной части ЭС. Перспективы развития данного направления предполагает, что ИИ будет базироваться на математической интерпретации работы нервной системы во главе с мозгом человека и имеет ввиду реализацию ИИ в виде нейронных сетей на базе нейрокомпьютеров;

- изучаются продукты интеллектуальной деятельности человека с целью воспроизвести их средствами вычислительной техники – при этом моделирование систем достигается за счет использования законов формальной логики, теории четких и нечетких множеств, графов, семантических сетей и других ИИ-моделей. Основные результаты в данной области заключаются в создании нечетких ЭС и систем разбора естественного языка.

Таким образом, интеллектуальные информационные системы, включая ЭС, строятся по следующему алгоритму:

- анализируются механизмы (способы) разумного поведения, которые записываются в математическом виде;
- указанные математические модели реализуются доступными разработчикам ЭС программно-техническими средствами;
- полученные результаты воплощаются в виде реальных ЭС.

Предпроектная стадия начинается с *обследования объекта и обоснования необходимости создания ЭС*. Исследование выполнимости проекта начинается с вопроса о том, следует ли разрабатывать ЭС для решения данного круга задач, насколько эта разработка будет оправдана и принесет ли она ощутимый эффект на практике. Краткий ответ на него в обобщенном виде может быть таким: начинать проект стоит, если использование будущей ЭС целесообразно, а разработка ЭС возможна, оправдана и методы инженерии знаний соответствуют решаемой задаче [60; 62; 91; 102].

Чтобы разработка ЭС *была возможной* необходимо, как минимум, одновременное выполнение следующих условий:

- разработчикам ЭС должны быть известны и доступны все нужные эксперты в данной области;
- эксперты должны сходиться в оценке предлагаемого решения, иначе нельзя будет оценить качество разработанной ЭС;

- эксперты должны уметь вербализировать и объяснить используемые ими методы, в противном случае трудно рассчитывать на то, что их знания будут «извлечены» и вложены в ЭС;
- задача не должна быть чересчур трудной для квалифицированного эксперта: на решение у него должно уходить несколько часов, а не дней или недель;
- эффективно могут быть решены задачи, требующие интеллектуальных навыков – то есть применения только рассуждений, а не действий;
- решаемые задачи, даже не будучи выраженными в явном формальном виде, все же должны относиться к достаточно «понятной» экспертам и структурированной предметной области – то есть должны быть выделены все основные понятия, отношения и известные (хотя бы эксперту) типовые способы получения решения аналогичных или близких по сути задач;
- решение задач не должно опираться на так называемый «здоровый смысл» – под которым понимается широкий спектр представлений современного человека об окружающем его мире – поскольку подобные знания пока невозможно в достаточном количестве заложить в системы ИИ.

Создание ЭС может быть возможно, но не оправдано (например, экономически нецелесообразно). Чтобы возможное использование ЭС *было оправдано* необходимо одного или нескольких из числа следующих факторов:

- решение задачи обещает принести значительный эффект;
- показано, что существует опасность постепенного утрачивания опыта решения задач в данной области;
- использование человека-эксперта невозможно либо из-за недостаточного штата экспертов, либо ввиду выполнения экспертизы одновременно в разных местах;
- условия, в которых решается задача, опасны для человека (например, наблюдается повышенная радиация).

Приложение *соответствует методам* и возможностям ЭС, если решаемая задача обладает совокупностью следующих характеристик:

- решение задачи опирается на использование операций с символами, а не с числами (то есть задача связана не с расчетом, а с

логическими рассуждениями, анализом, перебором вариантов, комбинаторикой элементов СС и т.п.);

- задача должна иметь эвристическую (не алгоритмическую) природу, чтобы ее решение можно было свести к применению эвристических правил – задачи, которые могут быть решены (при соблюдении заданных ограничений) с помощью некоторых формальных процедур, не подходят для применения ЭС;
- задача должна быть достаточно сложной и трудоемкой для ЛПР, чтобы оправдать затраты на разработку ЭС;
- задача должна быть достаточно узкой по тематике, чтобы ее можно было решить методами инженерии знаний, она также должна быть практически значимой (см. условия оправданности применения ЭС).

Команда проекта – после того, как решен вопрос о применимости технологии ЭС, необходимо сформировать команду проекта. Успешный выбор и подготовка команды во многом определяют эффективность и продолжительность процесса создания ЭС – проектирование и реализации которой представляет собой специфический вид проектной деятельности. Поскольку разработка любого программного комплекса существенно отличается от таких областей инжиниринга как, например, гражданское строительство или машиностроение, которые являются традиционными образцами проектной деятельности, прямое заимствование их известных принципов в данном случае недопустимо [4; 99].

Кроме того, любой проект по созданию ЭС является уникальным, что затрудняет выдачу универсальных рекомендаций о составе проектной команды. Тем не менее, проект должен иметь организационную структуру, которая четко определяет распределение ответственности и полномочий среди его участников, а также обязанностей и отношений отчетности (отметим, что чем меньше по объему работ и ресурсам проект, тем больше ролей приходится совмещать его исполнителям).

Роли и ответственности участников типового проекта разработки ЭС можно условно разделить следующим образом:

- **руководитель проекта** – отвечает за достижение целей проекта при заданных ограничениях (по срокам, бюджету и содержанию), осуществляет операционное управление проектом и выделенными ресурсами;

- **инженер по знаниям** – специалист по разработке ЭС, с точки зрения разделения труда в небольшом проекте, его роль может выполнять системный аналитик. Однако в целом задача инженера по знаниям достаточно широка и включает в себя взаимодействие с заказчиком по получению первоначальных требований и их уточнению в процессе работы, совместную работу с экспертом по «извлечению» знаний, наполнение БЗ, тестирование работоспособности БЗ и ЭС в целом. Также инженер по знаниям помогает эксперту выявить и структурировать знания, необходимые для работы ЭС, осуществляет выбор инструментальных средств, которые наиболее подходят для данной проблемной области, и определяет способ представления знаний – обычно на проект требуется 1-2 инженера по знаниям;
- **эксперт в проблемной области**, которой принадлежат задачи, решаемые будущей ЭС – этот специалист определяет знания (данные и правила), характеризующие указанную проблемную (предметную) область, обеспечивает полноту и правильность введенных в ЭС знаний. Число таких экспертов на проект зависит, во-первых, от объема решаемой задачи, во-вторых, от наличия доступных разработчикам ЭС экспертов в природе – но не менее одного эксперта;
- **программист** – это разработчик инструментальных средств, в зависимости от объема решаемых задач в реализации проекта принимают участие от 2 до 4 программистов-разработчиков;
- **тестировщик** – отвечает за определение целей и стратегии тестирования ЭС, а также проведение тестирования, анализ и документирование результатов; на проекте необходим как минимум один тестировщик.

Таким образом, минимальный состав команды включает 5 специалистов, в реальности он возрастает до 8-10 штатных единиц, без учета вспомогательного персонала. Численное увеличение команды происходит по следующим причинам:

- необходимость учета мнения нескольких потенциальных пользователей ЭС;
- привлечение к процессу экспертизы нескольких экспертов одновременно;
- необходимость в совместной групповой работе программистов с разной квалификации и специализации.

В целом уровень и численность команды зависят от характеристики поставленной задачи и масштаба проекта, поскольку одинаковые роли могут исполнять несколько ЛПР, а единственный ЛПР – несколько ролей. Возможны следующие варианты такого совмещения:

- руководитель проекта + инженер по знаниям;
- инженер по знаниям + специалист по тестированию;
- инженер по знаниям + проектировщик пользовательского интерфейса.

В то же время ЛПР крайне нежелательно совмещать следующие роли:

- разработчик + руководитель проекта;
- разработчик + инженер по знаниям;
- разработчик + проектировщик интерфейсов пользователя;
- разработчик + специалист по тестированию.

Подбор экспертов – от того, насколько успешно будут подобраны пары «эксперт – инженер по знаниям», во многом зависит успех как создания БЗ, так и проектирования ЭС в целом. Для организации эффективной совместной работы каждой такой пары необходимо учитывать целый ряд их субъективных личностных характеристик. В инженерии знаний большое внимание уделяется **психологическому аспекту** процесса извлечения знаний, поскольку именно он определяет результативность взаимодействия инженера по знаниям с источником знаний – экспертом. Для выявления личностных характеристик в настоящее время существует большое число психологических тестов, опросов, анкет. В [111], например, приводится следующий перечень желательных качеств эксперта:

- доброжелательность;
- готовность поделиться своим опытом и знаниями;
- умение доходчиво объяснить, о чем идет речь (педагогические навыки);
- заинтересованность (моральная и материальная) в успешности разработки.

Опытные эксперты обычно являются людьми зрелого возраста, что необходимо учитывать членам команды. Вопрос о необходимом количестве экспертов остается открытым, поскольку связан с проблемой совмещения противоречивых знаний – для ее решения с

каждым экспертом в команде при необходимости работают индивидуально, иногда даже создавая альтернативные БЗ.

Для формирования психологического портрета личности «идеального» эксперта можно также воспользоваться [102], где в табличном виде (см. таблицу 3.2) сформирован набор из наиболее интересных критериев для отбора специалистов в конкретной предметной области, претендующих на роль экспертов (указаны типовые вопросы и варианты ответов).

Таблица 3.2. Критерии для отбора специалистов в предметной области, претендующих на роль экспертов

№ п/п	Требования	Варианты ответов
1	<i>Можете ли Вы ...</i> Эксперт четко осознает границы своих познаний	Да Нет Не уверен
2	<i>Всегда ли Вы...</i> Эксперт может ответить на поставленный вопрос, если вопрос поставлен в этих границах	Всегда Не всегда Иногда
3	<i>Как часто ошибаетесь?</i> Эксперт обычно не ошибается	Часто Редко Иногда
4	<i>Можете ли Вы ...</i> Эксперта можно попросить оценить значение некоторого параметра, не поддающегося прямому измерению, и доверять этой оценке	Могу Нет Затрудняюсь ответить
5	<i>Обладаете ли Вы...</i> Эксперт обладает некоторой (хотя бы и неявной) моделью предметной области, поэтому его ответы на различные вопросы всегда согласованы между собой	Да Нет Очень приблизительно
6	<i>Всегда ли Вы можете ...</i> Эксперт может объяснить причины и/или мотивы своих решений, ответов, рекомендаций	Всегда Не всегда Иногда
7	<i>Важна ли для Вас степень детализации?</i> Чем подробнее задаваемые эксперту вопросы, тем больше он выдает информации	Важна Нет Не очень важна

Таблица 3.2 (окончание)

№ п/п	Требования	Варианты ответов
8	<i>Трудно ли для Вас ...</i> Эксперт может сравнивать несколько ситуаций (вариантов, альтернатив и т.д.), находить в них различия и отличать, принципиальны ли они и к чему приводят	Трудно Нетрудно По-разному
9	<i>Способны ли Вы ...</i> Эксперт способен учесть одновременно воздействие нескольких независимых или взаимозависимых факторов, параметров, критериев и т.п.	Да Нет Затрудняюсь ответить
10	<i>Можете ли Вы о себе сказать, что Вы ...</i> Эксперт рационален и последователен в своих предпочтениях, поэтому принимаемые им решения разумны и поддаются автоматизации	Да Нет Затрудняюсь ответить
11	<i>Важен ли для Вас ...</i> Ответ эксперта не зависит от формы и последовательности задаваемых вопросов	Да Нет В отдельных случаях
12	<i>Вы всегда искренни в своих ответах?</i> Искренность ответов эксперта не позволяет вводить в заблуждение тех, кто задает вопросы	Да Нет Не всегда
13	<i>Вы беспристрастный человек?</i> Эксперт должен быть беспристрастным	Да Нет Не всегда
14	<i>Вам удобно работать в группе?</i> Эксперт не возражает против работы в группе экспертов, что позволяет обобщать, согласовывать и интегрировать несколько мнений. Это приводит к повышению качества экспертизы	Да Нет Терплю

В заключение следует добавить, что тестирование с использованием приведенных 14 критериев можно значительно усложнить, если оценивать еще и влияние реальных условий на «идеального» эксперта – так называемых «шумовых» личностных особенностей эксперта, от которых зависит искренность ответов тестируемого (профессиональная гордость, зависимость ответа от контекста и формы ответа, умолчание, недопонимание, страх «не угадать»),

собственные интересы эксперта, зависимость ответа от состояния эксперта, усталость, торопливость и др.).

Выбор инструментальных средств – универсальной методики оценки качества инструментального средства для определения его применимости в решении конкретной задачи в настоящее время нет. Тем не менее, существует значительное количество статистического материала для оценки современных инструментальных средств, а также профессионалы из числа ЛПП, обладающие знаниями по выбору подходящего инструментального средства.

В печати периодически появляются обзоры, где содержится коллективное экспертное мнение специалистов относительно инструментальных средств для разработки ЭС. Приведем перечень критериев выбора инструментального средства по данным [102] – наиболее важные из них (имеющие экспертную оценку больше «трех» по пятибалльной шкале) будут более подробно рассмотрены в дальнейшем.

I Формализмы представления знаний

1.1 Модели представления знаний:

- правила;
- структурированные правила;
- метаправила;
- таблицы решений;
- фреймы (объекты);
- сценарии (схемы);
- формальные логики;
- другие.

1.2 Наследование свойств.

1.3 Множественные экземпляры.

1.4 Наличие подсистемы описания знаний.

1.5 Индукция.

1.6 Процедурные шаблоны и «демоны».

1.7 Активные объекты.

1.8 Библиотека случаев (прецедентов).

II Машина вывода

2.1 Машина рассуждений:

- прямой вывод;

- обратный вывод;
- смешанный вывод;
- немонотонные рассуждения;
- система поддержки истинности.

2.2 Стратегия поиска решений:

- метод «в ширину»;
- метод «в глубину»;
- метод «наилучшего»;
- метод «ветвей-и-границ»;
- метод «генерация-и-проверка»;
- метод «подъем-в-гору»;
- RETE-алгоритм;
- альфа-бета-алгоритм;
- анализ «цель-средство»;
- другие.

2.3 Разрешение конфликтов:

- присвоение правилам приоритетов;
- специфичность правил (типы правил);
- общность (частность) правил;
- давность информации;
- другие.

2.4 Обработка НЕ-факторов знаний:

- теория Байеса;
- теория Демпстера-Шеффера;
- теория нечетких множеств;
- фактор уверенности;
- порог достоверности;
- другие.

2.5 Поиска всех ответов.

2.6 «Доска объявлений».

2.7 Рекурсия.

2.8 Итерация

III Интерфейс разработчика

3.1 Быстрая разработка прототипа:

- открытая архитектура;
- взаимодействие с инструментальной системой:

- командный язык;
- пакетная обработка;
- набор меню.
- средства создания базы знаний:
 - текстовый процессор;
 - графическое представление базы знаний;
 - компиляция базы знаний;
 - инкрементный компилятор;
 - другие возможности.
- средства отладки:
 - проверка целостности базы знаний;
 - трассировка правил;
 - хранение протокола работ (история);
 - возможность просмотра правил (памяти);
 - список перекрестных ссылок;
 - средства верификации базы знаний.
- средства конструирования интерфейса ЭС:
 - графические утилиты;
 - формирование экрана (меню, поля ввода / вывода);
 - анимация;
 - конструирование подсказки.

3.2 Конструирование объяснения.

3.3 Управление логическим выводом:

- метаконтроль;
- средства управления поиском;
- система приоритетов;
- контроль доступа к БЗ.

3.4 Взаимодействие на естественном языке.

3.5 Библиотека случаев (прецедентов).

3.6 Документация, учебник.

3.7 Средства помощи.

3.8 Средства гипертекста / гипермедиа.

3.9 Другие возможности.

IV Интерфейс пользователя

4.1 Диалоговые возможности:

- окна;
 - цветовые возможности;
 - pull-down / pop-up меню;
 - командный язык;
 - взаимодействие на естественном языке;
 - речевой ввод / вывод;
 - средства настройки (макросы, переопределение клавиш);
 - другие.
- 4.2 Графически возможности:
- графическое представление результатов;
 - графическое представление дерева решений;
 - средства гипертекста / гипермедиа.
- 4.3 Множественные и неопределенные ответы на вопросы системы.
- 4.4 Объяснительные способности:
- ПОЧЕМУ (почему задан вопрос?);
 - КАК (как получен результат?);
 - ЧТО (переформулировать вопрос);
 - ЧТО ЕСЛИ (изменить информацию).
- 4.4 Несколько решений.
- 4.5 Библиотека случаев (прецедентов).
- 4.6 Документация, учебник.
- 4.7 Средства помощи.

V Обстановка и окружение

- 5.1 Аппаратное обеспечение:
- переносимость;
 - поддержка на микрокомпьютере;
 - совместимость;
 - многопроцессорная поддержка;
 - многопользовательская среда;
 - доступ к специальному оборудованию.
- 5.2 Языки реализации:
- переносимость;
 - встраиваемость;
 - совместимость.
- 5.3 Защита от копирования.
- 5.4 Пакетная обработка.
- 5.5 Работа в режиме реального времени.

5.6 Сетевая поддержка.

5.7 Безопасность:

- кодирование информации;
- пароли доступа;
- привилегии доступа.

5.8 Клиент-серверная архитектура.

VI Доступ к внешним данным

6.1 Взаимодействие с системами программирования.

6.2 Доступ к БД.

6.3 Доступ к возможностям языка реализации.

6.4 Доступ к специальному программному обеспечению.

Дальнейшие работы по инициации проекта по созданию ЭС связаны с определением способов управления ходом разработки и способов взаимодействия между разработчиками; определением и оценкой возможных рисков, разработкой технико-экономического обоснования, а также разработкой плана-графика выполнения работ – они мало чем отличаются от типового проекта по созданию ИС. Данные вопросы детально освещены в [4; 52; 96; 99; и др.].

Отдельно остановимся на аспектах распределения полномочий и ответственности между участниками команды проекта. Одним, но не единственным, инструментом является построение матрицы ответственности RACI, используемой для определения ролей и обязанностей разработчиков.

Термин RACI (или ARCI) является аббревиатурой:

- R – Responsible (исполняет);
- A – Accountable (несет ответственность);
- C – Consult before doing (консультирует до исполнения);
- I – Inform after doing (оповещается после исполнения).

Иногда можно встретить вариант аббревиатуры – RACIS, где S – supported (оказывает поддержку). Такое кодирование используется для формирования таблицы, которая характеризует участие той или иной роли при выполнении задач в процессе.

Пример матрицы RACI приведен в таблице 3.3.

Таблица 3.3. Примерная матрица RACI

Этапы работ	Руководитель проекта	Системный архитектор	Эксперт	Инженер по знаниям (аналитик)	Специалист по моделям данных	Разработчик	Администратор приложений	Специалист по тестированию
1	2	3	4	5	6	7	8	9
Организация проекта								
Согласование процедур управления	A	C						
Сбор команды проекта	A	C						
Формирование требований и разработка ТЗ								
Обследование объекта	I	I	A				C	
Анализ и обработка полученной информации	I	C		A	C			
Разработка концептуальной модели данных	I	C		R	A			C
Разработка ТЗ	C	A		R	C	C		C
Эскизный проект								
Определение общей функциональной и технической архитектур	I	A		R	C	R	R	I
Разработка логической модели данных	I	C		R	A	C		
Разработка модели БЗ	A		C	R				
Обоснование выбора ПО и технической инфраструктуры	C	A		C		C	C	I
Разработка предварительного регламента взаимодействия ИС	C	A		C	R	C		
Оформление эскизного проекта	C	A		R	R	R	R	

Таблица 3.3 (окончание)

1	2	3	4	5	6	7	8	9
Технический проект								
Определение функциональной и технической архитектур	C	A			C	R	R	
Формирование плана развёртывания	I	A				R	R	
Разработка физической модели данных	I	C			R	A		
Разработка и согласование регламентов взаимодействия ИС	I	A						I
Проектирование интерфейсов пользователя	I	C		C		A	C	
Оформление технического проекта	A	R		R	R	R	R	
Разработка программного обеспечения								
Разработка рабочей документации на систему	I	A		C		R	R	
Разработка ПО	I	A	I	C	C	R	R	
Тестирование ПО	C	C	C			A		R
Ввод в действие								
Подготовка объекта автоматизации к вводу системы в действие	I	C		I		C		
Подготовка персонала	C		C	R				
Проведение предварительных испытаний	C	A				R	R	I
Проведение опытной эксплуатации	R	R		C		C	C	
Проведение приёмочных испытаний	R	R		R		R	R	R
Завершение работ	A	R						

Вертикальный анализ (по функциональным ролям) позволяет выявить соответствующие проблемы. Если получилось:

– много R, нужно задать себе вопрос, может ли определенный человек быть ответственным за такое количество действий;

- нет пустых ячеек – нужно ли втягивать людей в такое количество операций?
- нет R или A – можно ли исключить эту функциональную роль?
- много A – правильно ли распределяются обязанности? Могут ли другие люди быть подотчетными в этих процессах?

При горизонтальном анализе рассматривают варианты:

- нет R, то тогда никто не несет ответственности за процесс, и он не будет выполнен;
- много A – будет путаница, так как любой Ответственный имеет свое видение, как должно быть осуществлено действие;
- много C – надо понять, нужно ли в реальности консультироваться с таким количеством различными функциональными ролями;
- много I – может быть ситуация, где неудачно определены роли.

Основные преимущества использования матрицы ответственности заключаются в том, чтобы прояснить разграничение ролей и ответственности как индивидуальных, так и в команде. Зачастую внутри группы возникает недопонимание, отсутствие четкой специализации и расплывчатое представление о собственных полномочиях, что приводит к ослаблению командного духа и, как следствие, к низкой производительности. Поэтому определение сфер ответственности и полномочий повышает результативность деятельности как каждого работника, так и группы в среднем.

В результате выполнения предпроектной стадии должен быть сформирован документ (технико-экономическое обоснование проекта). Ориентировочное содержание этого документа:

- ограничения, риски, критические факторы, которые могут повлиять на успешность проекта;
- совокупность условий, при которых предполагается эксплуатировать будущую ЭС: примерная концепция архитектуры системы, аппаратные и программные ресурсы, условия функционирования, обслуживающий персонал и пользователи системы;
- сроки завершения отдельных этапов, форма приемки/сдачи работ, привлекаемые ресурсы, меры по защите информации;
- описание выполняемых системой функций;
- возможности развития системы;
- информационные объекты системы;
- интерфейсы и распределение функций между человеком и системой;

- требования к программным и информационным компонентам программного обеспечения, требования к системе управления базами данных;
- границы проекта (что не будет реализовано в рамках проекта.)

3.3. Формирование требований пользователя к ЭС

Этапы данной стадии проектирования ЭС включают в себя следующие работы:

- определение требований к ЭС со стороны конечных пользователей;
- определение требований к среде функционирования ЭС;
- формулирование требований к средствам разработки, экспертам, операционной среде, пользовательскому интерфейсу и т.п.;
- разработка вариантов концепции ЭС;
- разработка технического задания.

Указанные этапы достаточно подробно освещены в [29-30], которые представляют собой совокупность взаимосвязанных отечественных государственных стандартов (ГОСТ) – все эти стандарты появились в конце восьмидесятых – начале девяностых годов XX века, заменив более ранние аналоги [7].

Несмотря на то, что эти стандарты были разработаны четверть века назад и сегодня их критикуют на том основании, что за истекшее время в сфере ИТ произошли огромные изменения, их практическая ценность этих ГОСТ не утрачена. На наш взгляд, это объясняется следующими причинами:

- использование ГОСТ не требует специальных знаний в области ИТ, они «прозрачны» и понятен рядовым ЛПР;
- все эти ГОСТ практичны в том смысле, что сразу понятно, как применять каждый стандарт и как контролировать его выполнение;
- система ГОСТ может быть эффективно использована для выработки общего языка с экспертами – неспециалистами в области ИТ и инженерии знаний.

По данным причинам при подготовке данного раздела монографии авторы пользовались двумя ГОСТ, которые представляют интерес при разработке ЭС:

- ГОСТ 34.601-90 «Автоматизированные системы. Стадии создания»;

– ГОСТ 34.602-89 «Техническое задание на создание автоматизированной системы».

Определение требований к ЭС и к среде функционирования.

Основное содержание стадии формирования требований – это сбор данных и анализ объекта, для поддержки функционирования которого предполагается создание ЭС, а также анализ существующей или планируемой среды функционирования ЭС (включая изучение информационных потоков и вопросы интеграции с имеющимися информационными системами). Обоснованное и тщательное формирование требований к ЭС – необходимое условие успешного выполнения работ по созданию данной системы. Требованием при этом именуется любое условие, которому должна соответствовать разрабатываемая система: требованием может быть как возможность, которой система должна обладать, так и ограничением, которому система должна удовлетворять.

Согласно глоссарию терминов программной инженерии IEEE, требование предусматривает:

- условия или возможности, необходимые пользователю для решения проблем или достижения целей с помощью ЭС;
- условия или возможности, которыми должна обладать ЭС или ее системные компоненты, чтобы выполнить контракт или удовлетворять действующим стандартам, спецификациям или другим формальным документам;
- документированное представление условий или возможностей ЭС, соответствующих обоим вышеуказанным пунктам.

В соответствии со стандартом разработки требований ISO/IEC 29148, под требованием понимают утверждение, которое идентифицирует эксплуатационные, функциональные параметры, характеристики или ограничения проектирования продукта или процесса, которое однозначно проверяемо и измеримо.

Обычно считается, что в практическом (прагматическом) плане требования к ЭС сводятся к тому, что она должна быть предназначена для удобной, квалифицированной поддержки информационной деятельности человека в заданной предметной области. Подразумевается наличие развитого диалога для понимания входных сообщений, выполнение всех промежуточных процедур и выдача «разумных» сообщений в удобной форме (сообщения считаются разумными, если они заслуживают доверия и «прозрачны» для

пользователей). Однако эта расплывчатая формулировка несет в себе мало практической пользы для ЛПР. Поэтому необходимо добавить, что требования должны соответствовать следующим характеристикам:

- **недвуусмысленность** – в каждое требование должна быть заложена возможность однозначной интерпретации и ясного, недвусмысленного толкования;
- **тестируемость** – специалисты по тестированию должны иметь возможность проверить, насколько корректно было реализовано требование.
- **формулировка требований** – по возможности не должна содержать определенные вербальные элементы (слова): прилагательные «устойчивый», «безопасный», «точный», «дружелюбный», «соответствующий» и т.п.; наречия и фразы с ними: «быстро», «своевременно», «точно» и т.п.; а также неспецифические слова и акронимы: «и т.д.», «и т.п.», «прочее», «и/или» и их аналоги;
- **ясность** (точность и простота) – требования не должны содержать выражений, которые не нужны для понимания их сути или вообще не относятся к данному контексту;
- **корректность** – требование должно отражать факты, которые объективны и достоверны;
- **понятность**, лингвистическая корректность – требования должны быть лексически и синтаксически выверенными, сформулированными в корректном стиле на основе соглашений, принятых ЛПР для рассматриваемой предметной области: эта характеристика предполагает замену слов типа «можно» и «возможно» на слова типа «должен» и «обязан»;
- **правдоподобность** (выполняемость) – требование должно быть выполнено в рамках существующих ограничений на процесс функционирования ЭС (например, по количеству ресурсов: денег, времени, ЛПР и другого персонала);
- **независимость** – возможность понять то или иное конкретное требование не должна быть обусловлена необходимостью знать или понимать какое-либо другое требование;
- **элементарность** (атомарность) – требование должно содержать отдельный элемент, для которого возможно отслеживание связей в иерархической пирамиде требований, что связано с его проверкой;

- **необходимость** – требование должно быть необходимым для рассмотрения ЭС в целом. Проверить необходимость можно путем определения ЛПР, которому необходимо данное требование – если такового лица не нашлось, требование является лишним; а также путем изъятия (удаления) требования из рассмотрения: если это не влияет на функциональность ЭС, требование также является лишним;
- **независимость от реализации** (абстрактность): требование не должно содержать избыточной информации о дизайне, аппаратной платформе ЭС, если от этого не зависит работа системы;
- **постоянство** – это означает, что между требованиями не должно быть конфликтов, которые возникают, когда ожидается различное поведение ЭС в заданной конкретной ситуации;
- **отсутствие повторов** – каждое требование должно быть обозначено только один раз и не должно перекрываться другими требованиями;
- **завершенность** – требование должно быть описано для всех возможных условий функционирования ЭС.

Если рассматривать процесс разработки требований к ИИС в широком смысле, то он представляет собой достаточно сложную научно-технологическую задачу и выполняется в рамках процесса **управления требованиями**, которые шире только этапа предпроектного обследования и распространяется на другие этапы, охватывая весь жизненный цикл ЭС. Изменчивость и неполнота требований могут привести к тому, что пользователь ЭС в итоге получит совсем не тот программный продукт, который представлял себе и на который рассчитывал. Поэтому необходимо не только формулировать и анализировать требования, но и эффективно управлять ими в процессе работы ЭС.

Под **анализом требований** понимают процесс сбора требований к программному обеспечению, их систематизацию, документирование, анализ, выявление противоречий, неполноты, разрешенных конфликтов – с учетом возможных противоречий требований к ЭС со стороны различных ЛПР. Анализ требований включает три типа деятельности [17]:

- сбор требований: общение с клиентами и пользователями, чтобы определить, каковы их требования;

- аналитическое исследование собранных требований: определение, являются ли они (и в какой мере) неясными, неполными, неоднозначными или противоречащими друг другу, а затем решение выявленных проблем;
- документирование требований – в таких формах, как простое описание, сценарии использования, пользовательские истории, спецификации процессов и т.п.

Анализ требований является трудоемким и длительным процессом, в рамках которого от инженера по знаниям может потребоваться наличие психологических навыков – если, например, со стороны эксперта проявляется нежелание делиться знаниями, поскольку, отдавая их в ЭС, эксперт как бы своими руками снижает свою рыночную стоимость. Парадоксально, но пользователи ЭС тоже способны препятствовать сбору требований: если они не имеют ясного понимания собственных запросов и потребностей (не знают, что конкретно им нужно, чего они от ЭС хотят), и, будучи технически неподготовленными, боятся брать на себя ответственность за некорректно сформулированные требования. Инженеры по знаниям и разработчики (программисты) ЭС могут добавить сюда проблемы за счет:

- разных трактовок и мнений по рассматриваемым вопросам (это так называемая проблема интерпретации данных);
- желания подкорректировать требования, чтобы они максимально соответствовали какой-нибудь готовой действующей модели – вместо того, чтобы разработать ЭС, полностью отвечающую потребностям заказчика.

По аналогии с другими ИС требования к ЭС можно разделить на три группы.

1. Требования к системе в целом, которые содержат:

1.1. *Требования к структурным характеристикам и режимам функционирования системы:*

- состав основных функций ЭС (состав функциональных подсистем);
- объектная структура ЭС (число уровней иерархии, основные объектные подсистемы на каждом уровне);
- требования к средствам и способам обмена информацией между объектными подсистемами в случае их территориальной разобщенности;

- требования к интегрируемости (совместимости) со смежными системами или уже реализованными элементами создаваемой ЭС, с которыми должна быть обеспечена возможность взаимодействия;
- требования к режимам функционирования системы (пакетный, интерактивный и т.д.).

1.2. **Требования к показателям назначения**, то есть к важнейшим характеристикам ЭС, определяющим степень соответствия системы ее основному назначению. Для систем продажи и резервирования железнодорожных билетов, например, показатели назначения – это пропускная способность (среднее время приобретения билета), число подключаемых терминалов кассира, обслуживаемые регионы; для информационно-справочной системы вокзала – это среднее время реакции, число терминалов пользователей, показатели достоверности выдаваемой информации (показатели степени соответствия данных, хранящихся в системе, истинной ситуации).

1.3. **Требования к надежности**, которые имеют в виду

- перечень отказов (указание на то, что понимается под отказом) системы или ее частей, по которым следует предъявлять требования по надежности;
- состав и количественные значения (нормы) показателей надежности по типам отказов для системы или ее элементов;
- требования к методам оценки и контроля надежности на разных этапах создания системы (жизненного цикла системы).

1.4. **Требования к качеству данных**, которые содержат:

- показатели достоверности данных (вводимых, хранящихся, выдаваемых ЭС) и их количественные значения; ситуации (события), при которых должна быть обеспечена сохранность данных;
- возможные способы несанкционированного доступа к данным, от которых ЭС должны быть защищена.

1.5. **Требования по стандартизации и унификации**: используемые при создании ЭС стандарты, классификаторы, требования по применению типовых программных и технических средств;

1.6. **Требования к развитию системы**: возможности модификации ЭС, включения в нее новых функций, открытости (возможности взаимодействия с другими системами), масштабируемости

(увеличения числа пользователей, числа подключаемых рабочих станций и пр.).

2. Требования к функциям (задачам), выполняемым ЭС, которые включают в себя:

- перечни задач по каждой функциональной подсистеме (комплексу ИТ) с их распределением по уровням ЭС;
- требования к качеству реализации каждой функции (задачи, комплекса задач);
- формы представления входной и выходной информации;
- временной регламент (требования к динамическим характеристикам); требования к качеству результатов (достоверности выдаваемой информации, точности расчетов и т.д.).

3. Требования к видам обеспечения (информационному, техническому, программному и т.д.) – состав этих требований непосредственно зависит от типа и назначения ЭС. Требования **к информационному обеспечению** могут включать требования к качеству, составу и способу организации данных, по условиям их совместимости со смежными системами, использованию классификаторов и унифицированных документов, методам контроля, хранения, обновления и восстановления данных. Требования **к программному обеспечению** содержат требования к качеству программных средств, к интерфейсам, используемым языкам программирования, операционной системе и т.д.

В состав требований **к техническому обеспечению** могут входить требования к функциональным, конструктивным, эксплуатационным характеристикам отдельных видов аппаратных средств: например, к быстродействию средств передачи данных, производительности, объемам запоминающих устройств, надежности подсистем ЭС, отдельных устройств и блоков или их комплексов и т.д.

Перечисленные требования могут быть представлены в следующем виде:

- список необходимых элементов (список задач, перечень способов несанкционированного доступа к данным, против ЭС должна быть защищена и т.п.);
- перечень возможных элементов: с указанием, например, на то, что в качестве линии связи могут быть использованы оптоволоконные линии или медная витая пара и т.д.;

- требований качественного типа (открытости, масштабируемости и пр.);
- количественных показателей (норм значений соответствующих показателей), регламентирующих требования к надежности, достоверности информации, временным характеристикам и т.п. рассматриваемой ЭС.

Разработка вариантов концепции ЭС. Стадия разработки вариантов реализации ЭС предполагает поиск путей удовлетворения требований пользователя на уровне концепции создаваемой ЭС – на уровне требований к структуре, функциям, программно-технической платформе и режимам работы ЭС. В ходе исполнения данной стадии рассматриваются альтернативные варианты концепции, производится их анализ и выбирается концепция ЭС, признанная наилучшей.

Разработка концепции позволит учесть требования всех заинтересованных участников, дав, таким образом, на ранних стадиях проработки вопросов создания ЭС правильным образом выстроить ожидания, закрепить концептуальные и частные требования участников, реализация которых в последующем призвана обеспечить успех и эффективность внедрения ЭС. Данный этап включает всесторонний и тщательный анализ объекта, интервьюирование будущих пользователей ЭС, проектирование бизнес-процессов и обязательное оформление подробного отчета о проделанной работе. Это выполняется инженером по знаниям с привлечением экспертов и конечных пользователей, его цель – спроектировать схемы бизнес-процесса объекта исследования «Как есть» и «Как будет» после внедрения ЭС, поэтому привлечение конечных пользователей и экспертов является обязательным условием. На основе полученных сведений, собранных в виде отчета, выбирается концепция создания ЭС, положенная в основу технического задания на проектирование ЭС.

Разработка концепции ЭС основана прежде всего на сборе материалов обследования объекта. Обследование – это изучение и диагностический анализ структуры объекта исследования с целью выявления всех характеристик объекта, внутренних и внешних потоков информационных связей, состава задач и специалистов, которые будут работать с разрабатываемой ЭС, уровень их подготовки как будущих пользователей системы и т.д. Поскольку каждая

ЭС является уникальной, то план обследования может затрагивать различные, на первый взгляд, не связанные аспекты деятельности.

Основной задачей обследования является, прежде всего, оценка реального объема проекта, а также его целей и задач, чтобы получить полное и однозначное понимание требований со стороны всех заинтересованных участников процесса создания ЭС. Кроме этого, в рамках обследования должны быть выявлены, если таковые имеются, существующие нормативно-методические и директивные материалы, на основании которых может быть уточнен перечень задач, для решения которых создается ЭС, а также возможности применения новых методов решения задач.

Инженер по знаниям (или аналитик) собирает и фиксирует информацию в двух взаимосвязанных формах:

- выполняемые экспертом функции, то есть сведения о событиях и процессах, а также мыслительных и умозаключающих выводах, выполняемых экспертом при решении задачи;
- сущности – важная информация о вещах, имеющих значение для исследуемой проблемной области и о которых что-то известно.

При изучении каждой функциональной задачи по мере возможности определяются:

- наименование задачи, сроки и периодичность ее решения;
- степень формализации задачи;
- источники информации, необходимые для решения задачи;
- показатели и их количественные характеристики;
- порядок корректировки информации;
- действующие алгоритмы расчета показателей и возможные методы контроля;
- действующие средства сбора, передачи и обработки информации;
- действующие средства связи;
- принятая точность решения задачи;
- трудоемкость решения задачи;
- действующие формы представления исходных данных и результатов их обработки в виде документов;
- потребители результатной информации по задаче.

На этапе обследования следует классифицировать планируемые к разработке функции ЭС по степени важности. Одним из возможных форматов представления такой классификации может служить метод MoSCoW – это метод расстановки приоритетов, используе-

мый в бизнес-анализе и разработке программного обеспечения, для достижения заинтересованными сторонами взаимно понимания важности, которую они придают каждому требованию.

Заглавные буквы в акрониме MoSCoW означают:

- М (англ. Must have this – должен сделать это) – необходимые функции, которые обеспечивают критичные для успешной работы системы возможности;
- S (англ. Should have this if at all possible – должен сделать это, если это вообще возможно) – желательные функции;
- C (англ. Could have this if it does not affect anything else – мог бы сделать это, если это не повлияет отрицательно на что-то другое) – возможные функции;

Реализация функций второй и третьей категорий ограничивается временными и финансовыми рамками: разрабатывается то, что необходимо, а также максимально возможное в порядке приоритета число функций второй и третьей категории.

- W (англ. Won't have this time, but Would like in the future – не будет достаточно времени на это, но в будущем хотелось бы) – отсутствующие функции. Данная категория функций особенно важна, поскольку необходимо четко представлять границы проекта и набор функций, которые будут отсутствовать в системе.

Буквы «о» добавлены в акроним MoSCoW для удобства произношения. Часто их оставляют строчными, чтобы показать, что они ничего не означают. Некоторые специалисты считают, что акроним должен выглядеть как «MuSCoW», чтобы точнее отображать слова, из которых он состоит. Но MoSCoW предпочтительнее, так как легче запоминается из-за созвучности со столицей России.

На основании полученной информации инженер по знаниям (аналитик) создает модель деятельности в двух видах:

- модель «Как есть» («As-Is»), отражающая существующие бизнес-процессы или алгоритмы решения задачи;
- модель «Как будет» («To-Be»), отражающая необходимые изменения бизнес-процессов с учетом внедрения ЭС.

Одной из рекомендаций является привлечение на ранних этапах специалистов группы тестирования для решения следующих задач:

- получение сравнительных характеристик предполагаемых к использованию аппаратных платформ, операционных систем и иного окружения;

– разработки плана работ по обеспечению надежности ЭС и ее тестирования.

Привлечение специалистов по тестированию на ранних этапах разработки является целесообразным для любых проектов. Если проектное решение оказалось неудачным и это обнаружено слишком поздно (на этапе разработки или на этапе внедрения в эксплуатацию), то исправление ошибки проектирования обходится дорого. Чем раньше специалисты группы тестирования выявляют ошибки в ЭС, тем меньше будет итоговая стоимость разработки. Время на тестирования системы и на исправление обнаруженных ошибок следует предусматривать не только на этапе разработки, но и на этапе проектирования. Для автоматизации тестирования рекомендуется использовать системы отслеживания ошибок (Bug Tracking), что позволит иметь единое хранилище ошибок, отслеживать их повторное появление, контролировать скорость и эффективность исправления ошибок, видеть наиболее нестабильные компоненты ИС а также поддерживать связь между группой разработчиков и группой тестирования.

Разработка технического задания. Результаты обследования являются основой для формирования технического задания (ТЗ), которое представляет собой перечень требований, целей и задач, поставленных заказчиком, учитываемых разработке проекта ИИС. Решение любой задачи, связанной с проектированием СС, начинается с осмысления ее сути и уточнения исходных данных. При этом требования, которые выдаются заказчиком и формулируются им на языке конечного пользователя (зачастую неспециалиста в данной предметной области), являются неопределенными и нечеткими, допускают разное толкование и не позволяют в конечном итоге объективно оценить качество будущей ЭС.

Считается, что грамотно составленное ТЗ – это половина успеха в решении любой инженеринговой задачи (для ЭС, по мнению авторов, это треть успеха). На рис. 3.2 представлен распространенный в сети Internet шуточный пример процесса неправильной трактовки ТЗ на создание ИС и что в итоге из него получилось. Поэтому основная цель ТЗ – правильно перевести требования заказчика на язык предметной области, сформулировать поставленную задачу максимально полно и грамотно, обосновать необходимость и возможность ее практического научно-технологического решения.

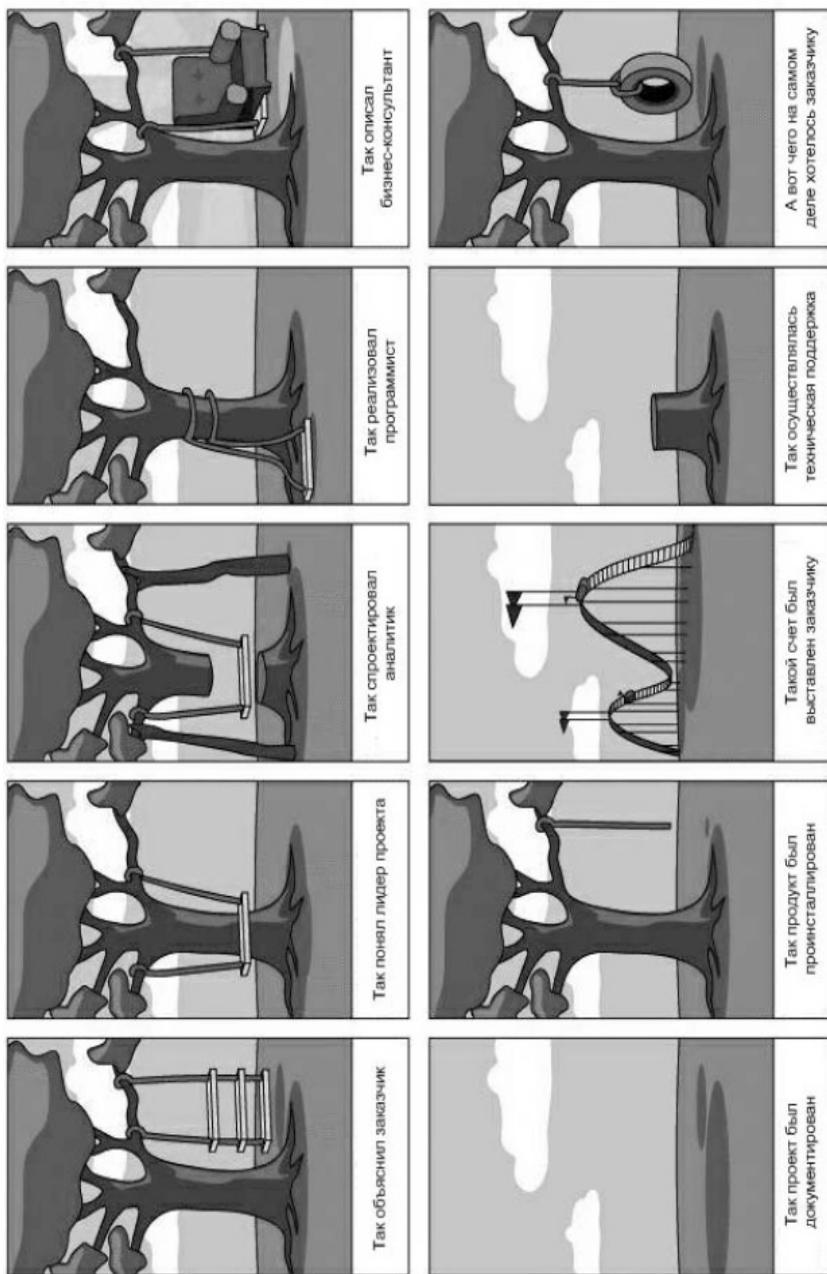


Рис. 3.2. Техническое задание на изготовление качелей

Как правило, ТЗ составляют на основе анализа результатов предварительных исследований и моделирования объекта, и чем полнее, обоснованнее будут сформулированы требования на начальном этапе, тем успешнее (быстрее и дешевле) обычно является последующий процесс создания системы.

Применительно к проектированию ЭС, подготовка ТЗ осуществляется исполнителем (инженером по знаниям) в тесном контакте с заказчиком, что фактически уже означает начало работы исполнителя над проектом. При разработке ТЗ необходимо решить следующие задачи:

- установить и документально зафиксировать цель создания ЭС, определить состав подсистем и функциональных задач;
- обосновать итоговые требования, предъявляемые к подсистемам и к ЭС в целом;
- обосновать уточненные требования, предъявляемые к информационной базе, математическому и программному обеспечению, а также комплексу технических средств;
- определить перечень задач создания ЭС и исполнителей;
- определить варианты концепции ЭС, удовлетворяющей требованиям пользователей;
- зафиксировать этапы создания ЭС и сроки их выполнения;
- провести предварительный расчет затрат на создание ЭС и определить уровень экономической (или социальной) эффективности ее внедрения.

Поскольку ТЗ является юридическим документом, который полностью определяет порядок и условия выполняемых работ, в том числе цель, задачи, принципы, ожидаемые результаты и сроки выполнения, все его положение (пункты и подпункты), а также все изменения, дополнения и уточнения формулировок обязательно согласуются с заказчиком и утверждаются им.

Это нужно потому, что, в случае обнаружения в процессе решения поставленной задачи какой-либо неточности или ошибочности исходных данных, возникает необходимость определения вины каждой из сторон, участвующих в разработке ЭС, а также распределения понесенных в связи с этим сторонами убытков.

Таблица 3.4. Типовые требования к составу и содержанию ТЗ согласно ГОСТ 34.602-89

№	Раздел	Содержание
1	2	3
1	Общие сведения	<ul style="list-style-type: none"> – полное наименование системы и ее условное обозначение; – шифр темы или шифр (номер) договора; – наименование предприятий разработчика и заказчика системы, их реквизиты; – перечень документов, на основании которых создается ИС; – плановые сроки начала и окончания работ; – сведения об источниках и порядке финансирования работ; – порядок оформления и предъявления заказчику результатов работ по созданию системы, ее частей и отдельных средств
2	Назначение и цели создания (развития) системы	<ul style="list-style-type: none"> – вид автоматизируемой деятельности; – перечень объектов, на которых предполагается использование системы; – наименование и требуемые значения технических, технологических, производственно-экономических и др. показателей объекта, которые должны быть достигнуты при внедрении ИС
3	Характеристика объектов автоматизации	<ul style="list-style-type: none"> – краткие сведения об объекте автоматизации; – сведения об условиях эксплуатации и характеристиках окружающей среды

Отметим также, что ТЗ может быть использовано и как эффективный инструмент коммуникации между заказчиком и проектной командой. В частности, с его помощью обе стороны имеют право и возможность:

- получить представление о будущей ЭС в готовом виде;
- при проведении приемо-сдаточных испытаний выполнить проверку готовности ЭС по всем пунктам ТЗ;
- уменьшить число ошибок, связанных с изменением требований в результате их неполноты или ошибочности.

Таблица 3.4 (продолжение)

1	2	3
4	Требования к системе	<p>Требования к системе в целом:</p> <ul style="list-style-type: none"> – требования к структуре и функционированию системы (перечень подсистем, уровни иерархии, степень централизации, способы информационного обмена, режимы функционирования, взаимодействие со смежными системами, перспективы развития системы; – требования к персоналу (численность пользователей, квалификация, режим работы, порядок подготовки); – показатели назначения (степень приспособляемости системы к изменениям процессов управления и значений параметров); – требования к надежности, безопасности, эргономике, транспортабельности, эксплуатации, техническому обслуживанию и ремонту, сохранности информации, защите от внешних воздействий, к патентной чистоте, по стандартизации и унификации). <p>Требования к функциям (по подсистемам):</p> <ul style="list-style-type: none"> – перечень подлежащих автоматизации задач; – временной регламент реализации каждой функции; – требования к качеству реализации каждой функции, к форме представления выходной информации, характеристики точности, достоверности выдачи результатов; – перечень и критерии отказов.

Заказчик в результате формирования ТЗ должен осознать, что именно ему требуется (получить не абстрактное представление, а описанное по пунктам предлагаемое решение). Исполнитель на основании ТЗ может спланировать выполнение проекта (определить затраты по трудоемкости, времени и цене), а также отказаться от выполнения работ, не указанных в задании. При разработке ТЗ, как правило, придерживаются стандарта ГОСТ 34.602-89. Согласно данному ГОСТ, разработанное ТЗ должно включать пункты, краткая характеристика которых приведена в таблице 3.4.

Таблица 3.4 (продолжение)

1	2	3
		<p>Требования к видам обеспечений:</p> <ul style="list-style-type: none"> – математическому (состав и область применения математических моделей и методов, типовых и разрабатываемых алгоритмов); – информационному (состав, структура и организация данных, обмен данными между компонентами системы, информационная совместимость со смежными системами, используемые классификаторы СУБД, контроль данных и ведение информационных массивов, процедуры придания юридической силы выходным данным); – лингвистическому (языки программирования, языки взаимодействия пользователей с системой, системы кодирования, языки ввода-вывода); программному (независимость программных средств от платформы, качество программных средств и способы их контроля, использование фондов алгоритмов и программ); – техническому; – метрологическому; – организационному (структура и функции эксплуатирующих подразделений, защита от ошибочных действий персонала); – методическому (состав нормативно-технической документации)
5	Состав и содержание работ по созданию системы	<ul style="list-style-type: none"> – перечень стадий и описание работ; – сроки исполнения – состав организаций-исполнителей работ; – вид и порядок экспертизы технической документации; – программа и обеспечение надежности; – программа метрологического обследования.
6	Порядок контроля и приемки системы	<ul style="list-style-type: none"> – виды, состав, объем и методы испытаний системы; – общие требования к приемке работ по стадиям; – статус приемной комиссии.

Таблица 3.4 (окончание)

1	2	3
6	Порядок контроля и приемки системы	– виды, состав, объем и методы испытаний системы; – общие требования к приемке работ по стадиям; – статус приемной комиссии.
7	Требования к составу и содержанию работ по подготовке системы в действие	– перечень подлежащих разработке документов; – перечень документов на машинных носителях
8	Источники разработки	Документы и информационные материалы, на основании которых разрабатывается ТЗ и система

3.4 Стадия проектирования ЭС

Разработка архитектуры ЭС. На предпроектной стадии во время анализа требований к ЭС основное внимание уделялось выяснению того, что должно быть сделано, вне зависимости от того, как это сделать. На этапе проектирования ЭС решается вопрос, как реализовать требования, сформулированные на этапе анализа.

Одним из важных шагов является разработка структуры (архитектуры) ЭС. Успешная разработка ЭС должна обеспечивать баланс компромиссов вследствие противоречащих требований заинтересованных участников; соответствовать принципам проектирования и рекомендованным методам, выработанным со временем; дополнять современное оборудование, сети и системы управления. Надежная архитектура разрабатываемого программного обеспечения требует значительного опыта в теоретических и практических вопросах, а также творческой составляющей, необходимой для преобразования бизнес-процессов и требований, которые могут казаться неясными, в надежные и практичные рабочие проекты.

Архитектура системы определяет ее разбиение на модули, задает контекст, в рамках которого принимаются проектные решения на следующих этапах разработки. Приняв решения о структуре системы в целом, разработчик системы производит ее разбиение на относительно независимые в реализации части (модули), разделяя

разработку между разработчиками выделенных модулей, что дает возможность расширить фронт работ, подключить к разработке системы новых исполнителей [24].

Архитектура программного обеспечения включает в себя определение структурированного решения, соответствующего всем техническим и рабочим требованиям, одновременно оптимизируя общие атрибуты качества, такие как производительность, безопасность и управляемость. Сюда входит серия решений, основанных на широком диапазоне факторов, и каждое из этих решений может значительно влиять на качество, производительность, программную поддержку и общий успех программного обеспечения.

Современное программное обеспечение редко бывает автономным. Как минимум, в большинстве случаев оно будет взаимодействовать с источником данных, например корпоративной ИС, предоставляющим информацию, с которой работают пользователи ЭС. Обычно современное программное обеспечение также должно взаимодействовать с другими службами и сетевыми функциями для выполнения проверки подлинности, получения и публикации информации и предоставления интегрированных сред работы пользователей. Без соответствующей архитектуры может быть довольно сложно, если вообще возможно, осуществить развертывание, эксплуатацию, обслуживание и успешную интеграцию с другими системами.

Со всеми этими факторы связаны некоторые компромиссы. Например, реализация наиболее безопасных механизмов с использованием сложного шифрования повлияет на производительность. Реализация множества параметров конфигурации и обновления может усложнить развертывание и администрирование. Кроме того, чем сложнее архитектура, тем дороже ее реализация.

Правильная архитектура должна обеспечивать баланс этих факторов с целью получения оптимального результата для определенного сценария. На самом высоком уровне проект архитектуры должен предоставлять структуру системы, но скрывать детали реализации; охватывать все случаи применения и сценарии; пытаться учитывать требования всех заинтересованных лиц; и удовлетворять настолько, насколько возможно, всем функциональные требования и требованиям к качеству.

На этапе конструирования системы ее разработчик должен принять следующие решения:

- определить разбиение системы на модули;
- определить состав вычислительного комплекса, на котором будет работать система;
- распределить компоненты системы по процессорам вычислительного комплекса и независимым задачам;
- организовать управление хранилищами данных и базой знаний;
- выбрать принципы реализации управления программным обеспечением;
- организовать управление пограничными ситуациями.

Выбор метода формализации знаний. При рассмотрении данной стадии создания ЭС более подробно остановимся на выборе метода формализации знаний, поскольку одним из важнейших групповых компонентов любой ИИС, во многом определяющим направления исследований в области ИИ, являются именно модели представления знаний. Данное направление, получившее, как уже было сказано, название «инженерия знаний», непосредственным образом связано и с теорией приобретения и формализации знаний, и с практикой разработки и проектирования БЗ в составе ЭС.

На сегодняшний день известно достаточно много моделей, принадлежащих теории знаний. Каждая из них обладает определенными достоинствами и недостатками, своими плюсами и минусами, поэтому для конкретной задачи можно и нужно выбирать свою модель – от которой будет зависеть не только эффективность выполнения поставленной задачи, но и возможность ее решения вообще. Модели представления знаний относятся к прагматическому направлению исследований в области ИИ, которое, что это тоже было отмечено, основано на предположении о том, что мыслительная деятельность ЛПР представляет собой «черный ящик». При таком подходе, напомним, не ставится вопрос об адекватности компьютерных моделей представления знаний тем моделям, которыми пользуется человек, а рассматриваются конечные результаты, полученные при решении конкретных задач.

При выборе модели представления знаний следует учитывать такие факторы, как однородность представления и простота понимания. **Однородность представления** приводит к упрощению механизма управления знаниями, **простота понимания** важна для

пользователей и экспертов, чьи знания закладываются в ИИС (если форма представления знаний будет трудна для понимания, это усложнит процессы приобретения и интерпретации знаний). Следует заметить, что одновременно выполнить оба указанных требования довольно трудно и проблематично.

Модели представления знаний обычно делят на логические (формальные) и эвристические (формализованные). В основе *логических моделей* лежит понятие формальной системы (теории знаний) – примерами здесь могут служить исчисление предикатов и любая конкретная система продукций. В логических моделях, как правило, используется исчисление предикатов первого порядка, дополненное рядом эвристических стратегий. При данном представлении знаний отношения, существующие между отдельными единицами знаний, выражаются только с помощью синтаксических правил используемой формальной системы.

Эвристические модели, в отличие от формальных, имеют существенно более разнообразный набор средств, передающих специфические особенности знаний в той или иной предметной области. Поэтому эвристические модели заметно превосходят логические как по возможности адекватно представлять проблемную среду, так и по эффективности используемых правил вывода. К эвристическим моделям можно отнести продукционные, сетевые и фреймовые модели. Следует отметить, что продукционные модели, используемые для представления знаний в ИИС, отличаются от формальных продукционных систем тем, что содержат эвристическую информацию о специфике проблемной среды, выражаемую часто в виде семантических структур [91]. Рассмотрим наиболее часто используемые модели представления знаний.

Логическая модель представления знаний предполагает унифицированное описание объектов и действий в виде предикатов первого порядка. В основе моделей такого типа лежит формальная система, задаваемая четверкой вида

$$M = \langle T, P, A, B \rangle, \quad (3.1)$$

где T – множество базовых элементов различной природы (алфавит), причем для T существует некоторый способ определения принадлежности или непринадлежности произвольного элемента к этому множеству; P – множество синтаксических правил; с их по-

мощью из элементов T образуют синтаксически правильные совокупности; A – аксиомы, то есть множество априорно истинных выражений; B – множество правил вывода, позволяющие расширять множество аксиом A за счет построения семантически правильных совокупностей, к которым снова можно применять правила из B .

Таким образом, формируется множество выводимых в данной формальной системе совокупностей. Если имеется процедура, с помощью которой можно определить для любой синтаксически правильной совокупности, является ли она выводимой, то соответствующая формальная система называется разрешимой. Это показывает, что именно правила вывода являются наиболее сложной составляющей формальной системы.

Для знаний, входящих в ЭС, можно считать, что множество A образуют все информационные единицы, которые введены в БЗ извне, а с помощью правил вывода из них выводятся новые производные знания. Другими словами, формальная система представляет собой генератор порождения новых знаний, образующих множество выводимых в данной системе знаний. Это свойство логических моделей делает их привлекательными для использования в БЗ, поскольку позволяет хранить в ЭС лишь те знания, которые образуют множество A , а все остальные знания получать из них по правилам вывода.

Положительными сторонами логических моделей являются единственность теоретического обоснования и возможность реализации системы формально точных определений и выводов [53]. Представление знаний в виде формул исчисления предикатов позволяет применить к ним формальные методы вывода. В частности, может быть использован метод резолюций, применяемый в системах тематического доказательства, обучения и автоматического синтеза программ.

К недостаткам этого способа представления знаний следует отнести:

- полный последовательный перебор всех возможных решений может приводить к комбинаторному взрыву, в результате чего поставленные задачи могут решаться недопустимо большое время;
- работа с неопределенностью знаний должна быть запрограммирована в виде самостоятельных метаправил, что на практике за-

трудняет разработку баз знаний с помощью логического формализма;

– сложность создания подсистемы объяснений – важной части интеллектуальной информационной системы.

Продукционная модель представления знаний. Используется для решения сложных задач, которые основаны на применении эвристических методов представления знаний. Эвристики позволяют настраивать механизм вывода на особенности проблемной области и учитывать неопределенность знаний. Термин продукция (от англ. production – «действие», «порождение») был введен в 1943 г. американским математиком Э. Постом. В общем виде под продукцией понимается выражение

$$(i); Q; P; A \Rightarrow B; N, \quad (3.2)$$

где i – **имя продукции**, с помощью которого данная продукция выделяется из всего множества продукций.

В качестве имени может выступать **лексема**, отражающая суть данной продукции, или порядковый номер продукции в их множестве, хранящемся в памяти системы. Элемент Q характеризует **сферу применения продукции**: разделение знаний на отдельные сферы позволяет экономить время при поиске нужных знаний.

Основным элементом продукции является ее **ядро**: $A \Rightarrow B$. Интерпретация ядра продукции может быть различной и зависит от того, что стоит слева и справа от знака секвенции \Rightarrow . Обычное прочтение ядра продукции выглядит так: ЕСЛИ A , ТО B , более сложные конструкции ядра допускают в правой части альтернативный выбор, например, ЕСЛИ A , ТО B_1 , ИНАЧЕ B_2 . Секвенция может истолковываться в обычном логическом смысле как знак логического следования B из истинного A (если A не является истинным выражением, то о B ничего сказать нельзя). Возможны и другие интерпретации ядра продукции, например A описывает некоторое условие, необходимое для того, чтобы можно было совершить действие B . Условие A и действие B , в свою очередь, могут рассматриваться как сложные высказывания, состоящие из ряда простых высказываний, соединенных логическими связками «И», «ИЛИ»: например, $A_1 \wedge A_2 \wedge \dots \wedge A_m \Rightarrow B_1 \vee B_2 \vee \dots \vee B_n$.

Элемент P есть *условие применимости ядра продукции*. Обычно P представляет собой логическое выражение (как правило, *предикат*). Когда P принимает значение «истина», ядро продукции активизируется. Элемент N описывает *постусловия продукции* – они актуализируются, если ядро продукции реализовалось. Постусловия продукции описывают действия и продукции, которые необходимо выполнить после реализации B – хотя выполнение N может происходить и не сразу же после реализации ядра продукции. Примером простейшего правила может быть: «ЕСЛИ оборудование не работает, ТО необходимо проверить наличие электропитания».

Набор продукции, который хранится в памяти системы, образует *систему продукции* – в этой системе должны быть заданы процедуры управления продукциями, с помощью которых происходит их актуализация и выбор для выполнения той или иной продукции из числа актуализированных.

На практике, чтобы описать достаточно сложные объекты (процессы, явления), всегда используют системы продукции, при работе с которыми неизбежно возникают вопросы полноты, непротиворечивости, расширения системы полученных знаний. Совокупность указанных продукции порождает структуру рассуждений типа «дерево», в котором начальная (корневая) вершина соответствует исходному высказыванию, а конечные (терминальные) вершины («листья») – результатам рассуждений, в каждом из которых должна быть истинной определенная группа высказываний, соответствующих промежуточным вершинам дерева, находящимся на пути от корневой к данной терминальной вершине.

Продукционная модель предполагает более гибкую организацию работы механизма вывода по сравнению с логической моделью. Отличительной особенностью продукционной модели является также способность осуществлять выбор правил из множества возможных на данный момент времени (из конфликтного набора) в зависимости от определенных критериев: например, по важности, трудоемкости, достоверности получаемого результата и других характеристик предметной области.

Популярность продукционных моделей определяется следующими факторами.

- подавляющая часть человеческих знаний может быть записана в виде продукций. Психологические исследования процессов принятия решений человеком показали, что, рассуждая, ЛПР использует правила, аналогичные продукциям, то есть вида «условие → действие» [91];
- системы продукций являются модульными: это элементарные единицы знаний, которые сравнительно легко могут быть добавлены в БЗ или исключены из нее. В каждой из продукций можно производить изменения, не затрагивая при этом содержания других продукций, также нет необходимости задавать в явной форме способ взаимодействия одного правила с другим;
- при необходимости системы продукций могут реализовывать любые алгоритмы и, следовательно, способны отражать любое процедурное знание, доступное ЭВМ;
- наличие в продукциях указателей на сферу применения продукций позволяет эффективно организовывать память ИИС, сокращая время поиска в ней необходимой информации. Классификация сфер может быть многоуровневой, что еще больше повышает эффективность поиска знаний, так как позволяет наследовать информацию в БЗ.

В то же время продукционные системы имеют, по крайней мере, два существенных недостатка:

- во-первых, при большом числе продукций становится сложной проверка непротиворечивости системы продукций – это заставляет при добавлении новых продукций тратить много времени на проверку непротиворечивости новой системы. Кроме того, затруднительна оценка целостного образа знаний, содержащихся в системе, а из-за присущей системе недетерминированности (неоднозначного выбора выполняемой продукции из фронта активизированных продукций) возникают трудности при проверке корректности работы системы. Считается, что если в ИИС число продукций достигает 10^3 , то мало шансов, что система продукций будет функционировать правильно – по этой причине число продукций в современных ИС, как правило, не превышает тысячи. Для компенсации указанных недостатков, наряду с продукциями, вводятся другие представления, вводящие структуру на множестве правил: это могут быть иерархия продукций, за-

дание списков свойств для параметров, фреймовые представления и т.п.;

- во-вторых, другая проблема практического применения продукционных моделей – переход от статических систем к динамическим, меняющим состав продукции в системе или перестраивающих алгоритм управления выбором продукции из фронта готовых продукции в зависимости от текущих ситуаций.

Сетевые модели. В основе моделей данного типа лежит конструкция, названная семантической сетью. Сетевые модели можно формально задать в виде

$$H = \langle I, C_1, C_2, \dots, C_n, G \rangle, \quad (3.3)$$

где I – множество информационных единиц; C_1, C_2, \dots, C_n – множество типов связей между информационными единицами; G задает связи между информационными единицами, входящими в I , которые входят в состав заданного набора типов связей.

Модель семантической сети – это модель, основой нотации которой является формализация знаний в виде ориентированного графа, отображающего смысл целостного образа. Вершинам графа соответствуют объекты или ситуации, дугам – отношения между ними. Моделируемые связи могут различаться между собой, описывая различные отношения: по наследованию; временные соотношения; пространственное расположение; причинно-следственные связи.

При построении семантической сети отсутствуют ограничения на число связей и на сложность сети. Для того, чтобы формализация оказалась возможной, семантическую сеть необходимо систематизировать: пример семантической сети представлен на рис. 3.3. Главным преимуществом данной модели является ее соответствие современным представлениям об организации долговременной памяти человека. По сравнению с другими методами представления знаний семантические сети более естественны, наглядны и понятны, они достаточно адекватно отражают взаимосвязь примитивных фактов. К достоинствам семантической сети также относят их наглядность (существует даже такое понятие как «семантическая визуализация» информации) и богатство средств отображения различных отношений между базовыми понятиями в рассматриваемой предметной области.



Рис. 3.3. Пример семантической сети

Недостаток модели – трудность поиска вывода на семантической сети, поскольку события, представленные в виде транзитивных формул, в виде сети представить еще сложнее. Кроме того, существует точка зрения, что семантические сети – это формализм, который существует для представления высказывания о внешнем мире, а не для представления проблемно-ориентированных моделей внешнего мира.

Фреймовые модели. Представления знаний являются развитием семантических сетей, где все атрибуты (поименованные отношения) объектов собираются в одну структуру данных, называемую фреймом. Данный тип моделей был предложен в 1974 г. профессором Массачусетского технологического института (США) М. Минским. Под фреймом (от англ. frame – «рама», «основа», «каркас») понимается некоторое минимальное описание объекта, то есть набор таких его признаков (атрибутов), без любого из которых описать этот объект с необходимой точностью невозможно.

В психологии и философии существует аналогичное понятие абстрактного образа: например, слово «комната» вызывает у слушателей образ типа: «жилое помещение с четырьмя стенами, полом, потолком, окнами и дверьми, площадью 6-20 м²».

Из этого описания в принципе ничего нельзя изъять: убрав, например, окна, мы получим уже чулан, а не комнату, но в нем имеются «дырки» или «слоты», – это незаполненные значения некоторых атрибутов – число дверей и окон, цвет стен, высота потолка, покрытие пола и др.

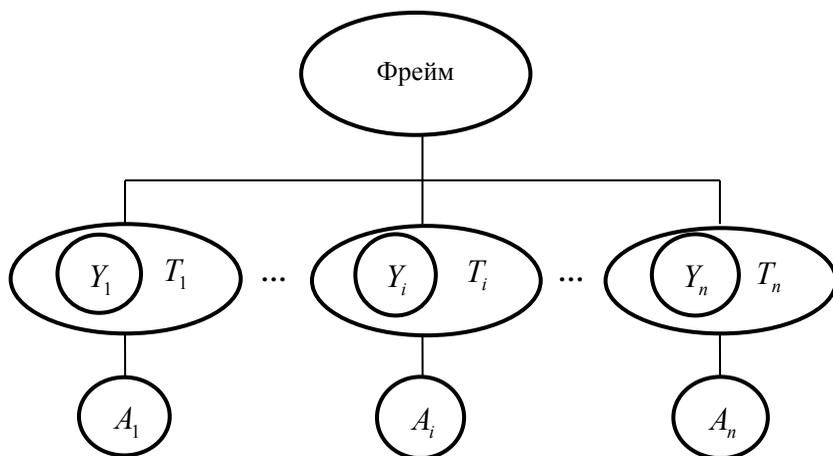


Рис. 3.4. Структура фрейма

Формально описание фрейма может быть представлено в виде:

$$\langle Z, (Y_1, T_1, A_1), \dots, (Y_n, T_n, A_n) \rangle, \quad (3.4)$$

где Z – имя фрейма; Y_i – имя слота; T_i – значение слота; A_i – имя присоединенной процедуры (необязательный параметр). В общем случае фрейм может быть представлен в виде структуры, показанной на рис. 3.4.

Важное свойство фреймов – их вложенность, то есть значение слота может, в свою очередь, представить собой набор слотов более низкого уровня и т.д. – по «принципу матрешки». Как и семантические сети, фреймовые модели представления знаний позволяют описать наиболее существенные отношения между атрибутами объекта, осуществлять быстрый вывод на основе принципа наследования. Сегодня фреймовые модели выступают как мощное и эффективное средство организации знаний с сохранением информации о структуре исследуемых объектов.

Над фреймами можно совершать некоторые теоретико-множественные операции – например, типа объединения и пересечения. При *объединении фреймов* в результирующем фрейме будут присутствовать все слоты, которые встречались в исходных фреймах. В слотах, не являющихся общими, будут сохранены исходные значения. Если в объединяемых фреймах были одноименные слоты, в результирующем фрейме останется один слот с таким же именем, значение которого определится в результате объединения значений одноименных слотов. При *пересечении фреймов* в результирующем фрейме будут присутствовать только те слоты, которые имелись во всех исходных фреймах. Вычислить результирующие значения можно двумя способами: первый способ состоит в том, что в результирующем фрейме присутствуют только те значения, которые совпали в исходных фреймах. При втором способе результирующие значения находят путем пересечения значений из исходных фреймов.

Важным свойством фреймов является заимствованное у семантических сетей наследование свойств: и во фреймах, и в семантических сетях наследование происходит по АКО-связям (A-Kind-Of = это). Слот АКО указывает на фрейм более высокого уровня иерархии, откуда неявно наследуются (переносятся) значения аналогичных слотов.

Основными *преимуществами фреймов* являются:

- способность отражать концептуальную основу организации памяти людей, а также ее гибкость и наглядность;
- модульность, поддержку возможности использования правил умолчания.

Основным недостатком фреймовых моделей является отсутствие механизмов управления выводом. Отчасти этот недостаток может быть устранен с помощью присоединенных процедур, используемых в качестве слотов-фреймов, которые реализуются пользователем системы. Однако при этом затрудняется управление завершенностью и постоянством целостного образа – по этой причине, в частности, существует большая опасность нарушения корректности присоединенных процедур.

Выбор формализма для представления знаний. При проектировании ЭС перед ее разработчиками всегда встает вопрос о выборе подходящего формализма для представления знаний. Решение

данного вопроса зачастую оказывает решающее влияние на итоговый результат – и хотя в настоящее время не существует однозначных правил по этому поводу, имеется ряд устоявшихся мнений и статистических рекомендаций [102]:

- ЕСЛИ понятия простые «И» отношения между ними выражены на языке исчисления предикатов «И» способ рассуждения дедуктивный, ТО целесообразно использовать *логические модели*;
- ЕСЛИ понятия являются в основном простыми «И» есть небольшое число отношений на понятиях «И» способ рассуждений индуктивный, ТО целесообразно использовать *индуктивные модели*;
- ЕСЛИ понятия устроены сложным способом «И» есть большое число отношений на понятиях «И» способ рассуждений – выдвижение гипотез, ТО целесообразно использовать *сетевые модели*;
- ЕСЛИ понятия устроены сложным образом «И» есть небольшое число отношений на понятиях «И» способ рассуждений дедуктивный, ТО целесообразно использовать *наследственно-конечные модели* (типа тезаурусов);
- ЕСЛИ понятия устроены сложным образом «И» есть большое число отношений на понятиях, ТО целесообразно использовать *фреймовые модели*;
- ЕСЛИ понятия устроены простым образом «И» есть большое число отношений на понятиях «И» способ рассуждений дедуктивный, ТО целесообразно использовать *продукционные модели*;
- ЕСЛИ понятия устроены сложным образом «И» структура многих понятий неясна «И» способ рассуждений – выдвижение гипотез, ТО целесообразно использовать *сетевые модели*;
- ЕСЛИ понятия устроены сложным образом «И» есть большое число отношений на понятиях «И» способ рассуждений индуктивный, ТО *подходящий формализм отсутствует*.

3.5. Разработка структуры БЗ и механизмов рассуждения

Структура БЗ зависит от выбранного метода формализации знаний. Так, структура БЗ продукционного либо логического типа может быть сформирована в виде объектов, атрибутов и условий.

При этом практические аспекты реализации БЗ зависят также и от языка разработки.

В качестве формальной основы проектируемых логико-семантических моделей БЗ ЭС используются графодинамические модели специального вида – семантические модели представления и обработки знаний, в основе которых лежат унифицированные семантические сети с базовой теоретико-множественной интерпретацией. Основным способом кодирования информации для таких сетей является SC-код (Semantic Code).

Для реализации фреймов в большей степени подходят объектно-ориентированные среды разработки, что также накладывает отпечаток на структуру БЗ. Так, фреймы расширяют возможности семантических сетей, позволяя организовать иерархию знаний. В сети каждое понятие представляется узлами и связями на одном и том же уровне детализации. Однако очень часто для одних целей объект необходимо рассматривать как единую сущность, а для других – учитывать детали его внутренней структуры.

Процедурные вложения являются важным свойством фреймов, поскольку они позволяют связать фрагменты программного кода с соответствующими сущностями фреймового представления. Например, в БЗ можно включить возможность генерировать графические образы. Для этого больше подходит графический язык, чем сетевой. Кроме того при проектировании БЗ необходимо решить вопрос об используемых механизмах рассуждения и аспектах управления выводом.

Управление выводом. Начнем с рассмотрения управления выводом, поскольку от выбранного метода поиска, то есть стратегии вывода, будет зависеть порядок применения и срабатывания правил. Процедура выбора сводится к определению направления поиска и способа его осуществления, ее суть – определение направления поиска знаний и способа его осуществления. Как правило, процедуры, реализующие поиск, «защиты» в механизм вывода. Поэтому в большинстве ЭС инженеры по знаниям не имеют к ним доступа. При разработке стратегии управления выводом, согласно [111], важны исходная точка в пространстве состояний, от выбора которой зависит метод осуществления поиска – прямой или обратный вывод; а также метод и стратегия перебора – в глубину, в ширину, по подзадачам и др.

При *обратном порядке* вывода сначала выдвигается некоторая гипотеза, а затем механизм вывода как бы возвращают назад – переходя к фактам и пытаясь найти среди них те, которые подтверждают гипотезу. Если гипотезу удалось подтвердить, она считается правильной и выбирается следующая гипотеза – детализирующая первую и являющаяся подцелью по отношению к ней. Далее отыскиваются факты, подтверждающие истинность подчиненной гипотезы и указанный процесс повторяется – его именуют выводом по управляемым целями, или выводом, *управляемым консеквентами*. Обратный поиск применяется в тех случаях, когда консеквенты (цели) известны или их сравнительно немного.

В системах с *прямым выводом* по известным фактам отыскивается заключение, которое из данных фактов следует – если такое заключение удалось найти, оно заносится в рабочую память. Прямой вывод часто называют выводом, управляемым данными, или выводом, *управляемым антецедентами*.

Рассмотрим пример [13] «ручного вывода» для фрагмента БЗ, описывающего ситуацию «Покупка легкового автомобиля» в виде шести правил, которые на естественном языке выглядят следующим образом.

П1: **ЕСЛИ** у человека много денег **«И»** ему удобно ездить на машине, **ТО** нужно ехать в автосалон.

П2: **ЕСЛИ** у человека мало денег **«И»** ему нужен автомобиль, **ТО** нужно ехать на рынок подержанных автомобилей.

П3: **ЕСЛИ** есть время **«И»** нужно ехать на авторынок, **ТО** необходимо определиться с маркой машины.

П4: **ЕСЛИ** есть время **«И»** нужно ехать на авторынок, **ТО** необходимо определиться с пробегом покупаемой машины.

П5: **ЕСЛИ** человек приехал в автосалон **«И»** определился с маркой, **ТО** нужно произвести покупку нового автомобиля.

П6: **ЕСЛИ** человек приехал на рынок **«И»** определился с пробегом, **ТО** нужно произвести покупку б/у автомобиля.

От пользователя в ЭС поступили следующие факты.

Начальное состояние РП:

«Много денег»

«Удобно ездить на машине»

«Есть время».

Рассмотрим основные шаги алгоритмов прямого и обратного вывода.

Прямой вывод. В этом случае необходимо просматривать все правила и выбирать те, у которых выполняются условия, тем самым формируя конфликтное множество. Затем, используя механизм разрешения конфликтов, выбирать подходящее правило, добавлять его заключение в рабочую память (за один проход можно добавлять только одно правило) и т.д.

Проход 1. Осуществляется просмотр правил, ищется то, у которого выполняются все условия. После того как такое правило найдено, факт из правой части добавляется в рабочую память, затем просматриваются все остальные правила, но со старой РП.

П1 – подходит (оба условия «Много денег» и «Удобно ездить на машине» выполнены, следовательно в рабочую память добавляется новый факт «Ехать в автосалон»).

П2 – не подходит.

П3 – не подходит.

П4 – не подходит.

П5 – не подходит.

П6 – не подходит.

Новое состояние РП:

«Много денег»

«Удобно ездить на машине»

«Есть время»

«Ехать в автосалон»

Проход 2. Выполняется аналогично первому, но с учетом того, что некоторые правила уже сработали, поэтому они пропускаются.

П1 – не подходит (сработало на первом проходе).

П2 – не подходит.

П3 – подходит (в РП добавляется новый факт «Определиться с маркой»).

П4 – не подходит.

П5 – не подходит.

П6 – не подходит.

Новое состояние РП:

«Много денег»

«Удобно ездить на машине»

«Есть время»

«Ехать в автосалон»
«Определиться с маркой»

Проход 3. Выполняется аналогично второму.

П1 – не подходит (сработало на первом проходе).

П2 – не подходит.

П3 – не подходит (сработало на втором подходе)

П4 – не подходит.

П5 – подходит (в рабочую память добавляется новый факт «Покупать новый автомобиль»).

П6 – не подходит.

Новое состояние РП:

«Много денег»

«Удобно ездить на машине»

«Есть время»

«Ехать в автосалон»

«Определиться с маркой»

«Покупать новый автомобиль»

Проход 4. Необходимо выполнить еще один проход, чтобы показать, что ни одно правило не срабатывает.

П1 – не подходит (сработало на первом проходе).

П2 – не подходит.

П3 – не подходит (сработало на втором подходе)

П4 – не подходит.

П5 – подходит (сработало на третьем подходе).

П6 – не подходит.

Таким образом, получена рекомендация ЭС – «Покупать новый автомобиль».

Обратный вывод. Здесь при просмотре правил необходимо искать то, у которого выдвинутая гипотеза (цель) находится в правой части. Затем, найдя такое правило, проверить наличие его условий из левой части в рабочей памяти. Если какого-либо условия там нет, то оно становится новой подцелью на текущей итерации (подцели выстраиваются в виде стека, т.е. при подтверждении последней из них происходит возврат к предыдущей).

Пусть выдвинута гипотеза (цель) «Покупка б/у автомобиля».

Начальное состояние РП:

«Мало денег»

«Нужен автомобиль»

«Есть время»

Проход 1. Просматриваются все правила, ищется цель «Покупка б/у автомобиля» в правой части.

П1 – не подходит.

П2 – не подходит.

П3 – не подходит

П4 – не подходит.

П5 – не подходит.

П6 – подходит (формулируется новая подцель «Ехать на авторынок»).

Проход 2. Ищется новая подцель «Ехать на авторынок» в правой части.

П1 – не подходит.

П2 – подходит (в рабочую память добавляется новый факт «Ехать на авторынок, все условия из левой части уже есть в РП, правило достоверно, пополняется РП, и так как подцель уже подтвердилась, последующие правила не проверяются).

Новое состояние РП:

«Мало денег»

«Нужен автомобиль»

«Есть время»

«Ехать на авторынок».

Проход 3. Выполняется для предыдущей подцели, которая в данном случае является главной – «Покупка б/у автомобиля»

П1 – не подходит.

П2 – не подходит.

П3 – не подходит

П4 – не подходит.

П5 – не подходит.

П6 – подходит (новая подцель «Определиться с пробегом»).

Проход 4. Ищется новая подцель «Определиться с пробегом» в правой части.

П1 – не подходит.

П2 – не подходит.

П3 – не подходит

П4 – подходит в рабочую память добавляется новый факт «Определиться с пробегом», все условия из левой части есть в РП, правило достоверно, пополняется РП).

Новое состояние РП:

«Мало денег»

«Нужен автомобиль»

«Есть время»

«Ехать на авторынок»

«Определиться с пробегом»

Проход 5. Выполняется для цели «Покупка б/у автомобиля»

П1 – не подходит.

П2 – не подходит.

П3 – не подходит

П4 – не подходит.

П5 – не подходит.

П6 – подходит (в рабочую память добавляется новый факт «Покупка б/у автомобиля», все условия есть в РП).

Таким образом, основная гипотеза подтвердилась: факты достоверны, цель достигнута.

3.6. Методы и стратегии перебора

В системах, БЗ которых насчитывает сотни правил, желательным является использование стратегии управления выводом, позволяющим минимизировать время поиска решения и тем самым повысить эффективность вывода. К числу таких стратегий относятся [61; 101; 104]:

- метод полного перебора в ширину;
- метод полного перебора в глубину;
- эвристические методы поиска решений;
- разбиение задач на подзадачи.

Прежде чем рассматривать данные стратегии, формализуем понятия «задача», «решение» и «алгоритм» [101]. При решении любой проблемы необходимо очертить *предметную область*, которую можно формализовать в виде модели предметной области

$$\langle E, P, R, Q \rangle, \quad (3.5)$$

где $E = (e_1, e_2, \dots, e_n)$ – множество имен объектов (предметов, сущностей и т.п. элементов предметной области), с которыми имеют

дело при решении задачи; $P = (p_1, p_2, \dots, p_m)$ – множество имен свойств (состояний) объектов, причем возможно, что $p_1 = (p_{11}, p_{12}, \dots, p_{1k})$, $p_2 = (p_{21}, p_{22}, \dots, p_{2s})$ и т.д. – отметим, что эти свойства могут меняться под действием некоторых операторов; $R = (r_1, r_2, \dots, r_l)$ – множество имен отношений, в которые могут вступать объекты моделируемой предметной области; $Q = (q_1, q_2, \dots, q_w)$ – множество имен операций (действий), которые допустимы с этими объектами через изменение их свойств и отношений между ними.

Выделенные множества E, P, R, Q позволяют представлять предметную область в виде некоторого языкового эквивалента, степень адекватности которого реальности зависит от множества факторов: сложности задачи, числа задач, сложности самой предметной области (число объектов, свойств, отношений, их семантика и т.п.). Говорят, что множества E, P, R, Q задают концептуальную модель предметной области и определяют ее статическую структуру. Для перехода к модели предметной области необходимо задать также пространство состояний.

Следует помнить, что состояние предметной области непостоянно, и может меняться с течением времени, поэтому нужно говорить о состоянии предметной области в данный момент времени:

$$S(t) : \langle E(t), P(t), R(t) \rangle. \quad (3.6)$$

В соответствии с (3.6), которое описывает ситуацию в предметной области и в зависимости от этой ситуации на момент $t \in (t_{\text{нач}}, t_{\text{окон}})$ человек выбирает из множества Q возможных операций именно те действия, которые необходимы для решения задачи. Если обозначить через F отображение $S(t)$ на множество Q , то можно записать:

$$F : (\langle E(t), P(t), R(t) \rangle \rightarrow Q), \quad (3.7)$$

Обозначим начальное состояние предметной области через $S_{\text{нач}}$ – тогда **решаемая задача** заключается в том, чтобы перевести предметную область из состояния $S_{\text{нач}}$ в некоторое заданное (целевое) состояние $S_{\text{цель}}$. Сделать это возможно, лишь применяя до-

пустимые в заданной предметной области действия из множества $Q = (q_1, q_2, \dots, q_w)$. Какие выбрать операции q_i и в какой последовательности их применять – неизвестно. Этот вопрос как раз и составляет **решение задачи**. Таким образом, схема решения задачи выражается формулой вида

$$Z = \left(S_{\text{нач}} \xrightarrow{Q} S_{\text{цель}} \right). \quad (3.8)$$

Допустим, что оператором $q_1 \in Q$, мы перевели состояние $S_{\text{нач}}$ в S_1 , но оно не совпало с состоянием $S_{\text{цель}}$, то есть $S_1 = q_1(S_{\text{нач}})$, $S_1 \neq S_{\text{цель}}$. Для перевода S_1 в состояние S_2 применим оператор $q_2 \in Q$: $S_2 = q_2(S_1)$ и, если S_2 снова не совпадет с $S_{\text{цель}}$, применим оператор q_3 , чтобы получить состояние S_3 и оценить его, сравнивая с $S_{\text{цель}}$ – будем продолжать данный процесс до тех пор, пока не найдется такой оператор q_i , что $S_j = q_j(S_{j-1})$ и $S_j = S_{\text{цель}}$. Описанный путь поиска решения отображается следующей цепочкой:

$$S_{\text{нач}} \xrightarrow{q_1} S_1 \xrightarrow{q_2} S_2 \xrightarrow{q_3} \dots \xrightarrow{q_j} S_j = S_{\text{цель}}. \quad (3.9)$$

То же самое можно представить и в другой форме: как $S_{\text{цель}} = q_i(q_{j-1}(\dots(q_3(q_2(q_1(S_{\text{нач}}))))))$, где последовательность $q_1, q_2, q_3, \dots, q_j$ будет представлять собой **алгоритм решения задачи**.

Если перевод $S_{\text{нач}} \rightarrow S_{\text{цель}}$ возможен не единственным способом, можно ставить задачу об оптимизации решения. Применение операторов к описанию начального состояния, а затем и ко всем промежуточным состояниям порождают множество возможных состояний предметной области, именуемое **пространством состояний**.

Процесс построения модели предметной области рассмотрим на примере известного решения задачи о бананах и обезьяне [11; 101]. Пусть в комнате, где находится обезьяна, имеются ящик и связка

бананов, причем бананы подвешены к потолку настолько высоко, что обезьяна не может дотянуться до них с пола. Какую последовательность действий должна совершить обезьяна, чтобы достать бананы? Можно предположить, что она должна догадаться подтащить ящик к бананам, взобраться на него и дотянуться до связки.

Построим модель данной предметной области.

E – имена предметов и объектов модели предметной области:

- *ОБЕЗЬЯНА* (далее *Об*),
- *ЯЩИК* (*Я*);
- *БАНАНЫ* (*Б*);

C – координаты *Об*, *Я*, *Б* в комнате – по горизонтали и вертикали. Значение координат – произвольное множество точек комнаты: a, b, c, \dots (допустим, что это множество конечно).

R – имена отношений между объектами: *РЯДОМ* (*Об*, *Я*), то есть обезьяна находится у ящика; *НА* (*Об*, *Я*) – обезьяна на ящике, наконец, *В* (*Об*, *Б*) – бананы в руках обезьяны. Очевидно, что возможны отрицания этих отношений: *НЕ-РЯДОМ* (*Об*, *Я*), *НЕ-В* (*Об*, *Я*), *НЕ-НА* (*Об*, *Я*) или иначе: $\overline{РЯДОМ}(Об, Я)$, $\overline{В}(Об, Я)$, $\overline{НА}(Об, Я)$.

Q – имена действий, которые может выполнить обезьяна:

- *ПОДОЙТИ*(*Об*, *Я*) = q_1 – обезьяне подойти к ящику;
- *ПЕРЕДВИНУТЬ*(*Об*, *Я*) = q_2 – обезьяне передвинуть ящик;
- *ВЗОБРАТЬСЯ*(*Об*, *Я*) = q_3 – обезьяне взобраться на ящик;
- *СХВАТИТЬ*(*Об*, *Я*) – обезьяне схватить бананы;
- *ПЕРЕЙТИ*(*Об*, d) = q_5 – обезьяне перейти в точку с двумерной координатой $d(e, f, m \dots)$.

Все действия имеют в качестве аргументов объекты модели предметной области, а субъект здесь единственный – обезьяна. Оператор q_5 отображает возможные перемещения обезьяны по комнате без ящика.

Теперь необходимо построить отображение $F : (S(t) \rightarrow Q)$, то есть множества состояний на множество действий. С этой целью

дадим описание начального и целевого состояний в модели предметной области. Допустим, что в $S_{нач}$ двумерная координата *Об* (обезьяны) точка *a*, координата *Я* – точка *b*, координата *Б* – точка *c*, расположенная непосредственно под бананами. Тогда начальное состояние будет удобно описать в виде

$$S_{нач} = (a, b, c, \overline{НА(Об, Я)}, \overline{В(Об, Я)}), \quad (3.10)$$

все объекты в начальных позициях, обезьяна *НЕ-НА ЯЩИКЕ*, бананы *НЕ-В* руках у обезьяны. Каждое из отношений *НА* и *В* может принимать лишь 2 значения *ДА* («1») и *НЕТ* («0»): или обезьяна *НА* ящике («1»), или *НЕТ* («0»); или *В* руках у нее бананы («1»), или *НЕТ* («0»).

Теперь можно упростить запись состояния (3.10):

$$S_{нач} = (a, b, c, 0, 0); \quad (3.11)$$

$$S_{цель} = (c, c, c, 1, 1). \quad (3.12)$$

Это означает, что все объекты в одной точке, обезьяна на ящике и схватила банан. Задача, таким образом, сводится к тому, как с помощью вышеперечисленных операторов q_1, q_2, q_3, q_4 из ситуации (3.11) перейти в ситуацию (3.12).

Применяем к $S_{нач}$ оператор q_1 – *ПОДОЙТИ*, имеем:

$$S_{нач} = (a, b, c, 0, 0) \xrightarrow{q_1} (b, b, c, 0, 0) = S_1.$$

(Обезьяна *ПОДОШЛА* к ящику, и их координаты совпали.)

Применяем q_2 – *ПЕРЕДВИНУТЬ*:

$$S_1 = (b, b, c, 0, 0) \xrightarrow{q_2} (c, c, c, 0, 0) = S_2.$$

Применяем q_3 – *ВЗОБРАТЬСЯ*:

$$S_2 = (c, c, c, 0, 0) \xrightarrow{q_3} (c, c, c, 1, 0) = S_3.$$

Применяем q_4 – *СХВАТИТЬ*:

$$S_3 = (c, c, c, 1, 0) \xrightarrow{q_4} (c, c, c, 1, 1) = S_{цель}.$$

Весь процесс можно отобразить формулой

$$S_{цель} = q_4(q_3(q_2(q_1(a, b, c, 0, 0))))). \quad (3.13)$$

Правила для описания ситуаций могут быть обобщены для случая, когда координаты нахождения обезьяны и ящика являются переменными. Обозначим их как x, y соответственно. Тогда каждое правило будет описывать класс состояний, к которым применим тот или иной оператор. Эти правила будут выглядеть следующим образом:

$$(x, y, c, 0, 0) \rightarrow q_1 \rightarrow (y, y, c, 0, 0); \quad (3.14)$$

$$(y, y, c, 0, 0) \rightarrow q_2 \rightarrow (c, c, c, 0, 0); \quad (3.15)$$

$$(c, c, c, 0, 0) \rightarrow q_3 \rightarrow (c, c, c, 1, 0); \quad (3.16)$$

$$(c, c, c, 1, 0) \rightarrow q_4 \rightarrow (c, c, c, 1, 1). \quad (3.17)$$

Подсчитаем число состояний (мощность N) пространства состояний, описываемого каждым правилом:

– для (3.14): $N_1 = 20 \times 20 \times 1 \times 1 \times 1 = 400$;

– для (3.15): $N_2 = 20 \times 1 \times 1 \times 1 \times 1 = 20$;

– для (3.16): $N_3 = 1 \times 1 \times 1 \times 1 \times 1 = 1$;

– для (3.17): $N_4 = 1 \times 1 \times 1 \times 1 \times 1 = 1$.

Здесь координаты x, y определены на множестве точек комнаты $n = 20$; « \times » – знак декартова произведения, поэтому пространство состояний имеет мощность, равную сумме мощностей подпространств, то есть 422 состояния.

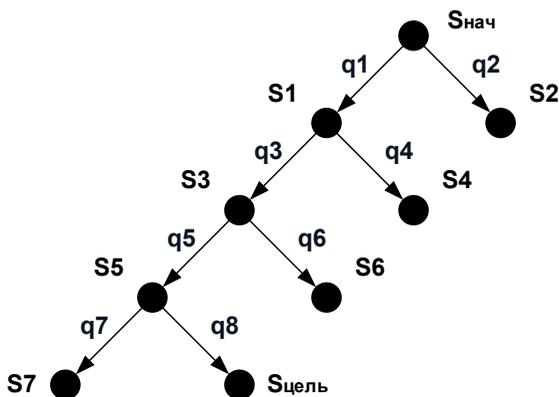
Отметим, что данную модель можно существенно усложнять и видоизменять: менять местоположение бананов и число ящиков, в комнате могут находиться другие предметы и т.п., но смысл преобразований при этом не изменится.

3.7. Методы поиска решений в пространстве состояний

Среди множества состояний данного пространства, если его модель построена корректно, рано или поздно появится одно состояние, которое будет соответствовать целевому, и тогда процесс поиска решения будет закончен. Однако процесс этот может быть длинным и сложным, поэтому возникает необходимость в едином способе представления множества состояний и поиска решений – таким способом стал метод, удобной графической моделью кото-

рого является граф. Если отождествить состояние $S_{нач}$ с корнем или начальной вершиной графа-дерева, то применяя к $S_{нач}$ какой-либо оператор $q_i \in Q$, мы порождаем новое состояние, образуя тем самым следующую вершину графа – см. рис. 3.5.

Эта новая вершина может быть промежуточной или целевой. Если вершина промежуточная, то процесс порождения новых вершин (с помощью операторов q_i) будет продолжен, пока не найдется целевая. Процесс применения оператора q_i к некоторой вершине называется *раскрытием вершины*. От каждой порожденной вершины к породившей ее расставляются указатели, которые позволяют найти путь назад, к начальной вершине, после того, как обнаружена целевая.



Путь решения задачи: $S_{цель} = G8(q6(q3(q1(S_{нач}))))$

Рис. 3.5. Граф решения задачи

Общая процедура построения дерева в пространстве состояний при этом выглядит следующим образом.

1. К корню дерева ($S_{нач}$) применяются операторы q_i из множества Q (их может быть несколько). Полученные при этом вершины образуют первый уровень новых вершин.

2. Каждая из вновь полученных вершин проверяется, не является ли она целевой. Если нет, то процесс продолжается по отноше-

нию к каждой из них: Образуется второй уровень вершин. Если к какой-либо вершине никакой оператор из Q неприменим, то эта вершина становится терминальной (конечной). Как видим, на каждом шаге проводятся две операции: порождение новой вершины и проверка, не является ли новая вершина целевой, т. е. совпадающей с целевым состоянием.

3. Когда целевая вершина найдена, в обратном направлении (от цели к началу) просматриваются указатели дуг и выделяется путь решения. Практически этот путь удобнее отображать посредством операторов, связанных с этими дугами (см. рис. 3.5).

В общем случае число вершин может быть большим и их последовательное раскрытие, анализ и пометка пути существенно осложняют задачу. Возникает проблема перебора вершин: в каком порядке они будут порождаться и анализироваться. Здесь возможны следующие варианты:

- если вершины раскрываются в том же порядке, в котором они порождаются, то такой перебор называется **полным перебором в ширину** (*breadth – first process*);
- на каждом шаге первой раскрывается вершина, которая была построена последней, такой процесс называется **полным перебором в глубину** (*depth – first process*). Отметим, что в данных процессах расположение целевой вершины не влияет на порядок раскрытия – поэтому их часто называют процессами **слепого перебора**;
- если имеется некоторая дополнительная (эвристическая) информация о предметной области, которая позволяет делать суждения о характере графа пространства состояний и расположения цели, то такой метод построения графа называется **эвристическим** – что в данном случае означает «служащий открытию». Эвристическая информация, опирающаяся, как правило, на предыдущий опыт, позволяет вести поиск в наиболее перспективных направлениях.

Метод полного перебора в ширину, основной алгоритм состоит в выполнении следующих действий.

1. Раскрывается начальная вершина $S_{нач}$. Она раскрывается до тех пор, пока ее можно раскрыть, применяя один и тот же оператор (или разные, смотря по условию). При этом образуются вершины

первого уровня: $S_1, S_2, S_3 \dots$. Они раскрываются в свою очередь, и образуются вершины второго уровня и т.д. (граф на рис. 3.6 может служить примером этого метода: здесь S_1 и S_2 – вершины первого уровня, S_3 и S_4 – вершины второго уровня, S_5 и S_6 – третьего и т.д.).

2. Расставляются указатели, ведущие от новых вершин к корню. Это могут быть условные имена, буквы, цифры, имена операторов и т. п. Но могут быть и реальные величины, например расстояния, стоимость, вес и т.д.

3. Проверяется, нет ли среди полученных вершин целевой. Если есть, то формируется решение на основе соответствующего оператора. Если целевых вершин нет, то рассматривается первая порожденная вершина и к ней применяется тот же алгоритм, после чего, переходят ко второй и т.д., пока среди получаемых вершин не окажется целевой.

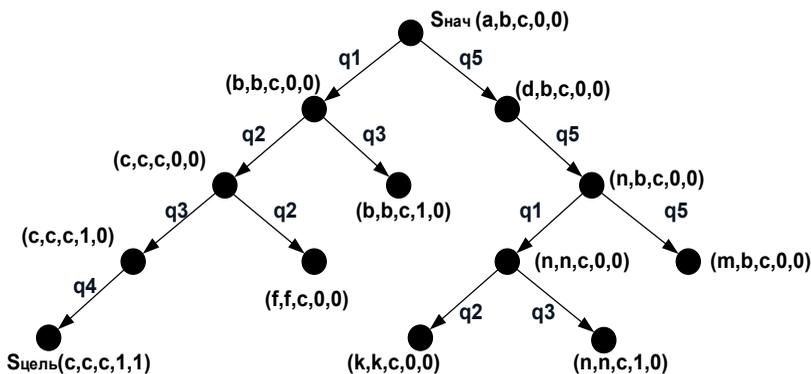


Рис. 3.6. Элемент дерева полного перебора в ширину для примера с бананами и обезьяной

Метод полного перебора в ширину гарантирует нахождение целевой вершины, так как перебор является полным. Путей достижения цели при этом, вообще говоря, может быть достаточно много – поэтому имеется возможность выбрать самый короткий из них (или самый дешевый, или самый легкий – в зависимости от выбранных критериев) путь. Однако граф поиска может оказаться бесконечным и тогда алгоритм не закончит свою работу никогда.

Таким образом, метод полного перебора гарантирует поиск оптимального решения, если дерево пространства состояний не бесконечно. Синонимами названия данного метода являются *метод грубой силы*, *метод проб и ошибок*. На рис. 3.6 показан элемент дерева полного перебора в ширину для примера с обезьяной и бананами (для полноты изложения здесь применяется дополнительный оператор q_5 , отображающий перемещение обезьяны по комнате из точки в точку).

Корень дерева совпадает с $S_{нач}(a,b,c,0,0)$. Точки d, f, k, n, t – координаты возможной миграции обезьяны по комнате. Таких точек, конечно, множество, но для примера выбрано лишь несколько точек. Они не приводят к решению, но теоретически вполне допустимы. Жирной линией показан путь до целевой вершины $S_{цель}(c,c,c,1,1)$, которому соответствует последовательность операторов q_4, q_3, q_2, q_1 . Алгоритм решения отобразится формулой: $S_{цель} = q_4(q_3(q_2(q_1(a,b,c,0,0))))$.

В отличие от метода перебора в ширину, *метод полного перебора в глубину* предлагает раскрывать прежде всего те вершины, которые были построены последними. Первой из них является корневая, процесс всегда будет идти по самой левой ветви вершин. Чтобы как-то ограничить перебор, вводится понятие глубины вершины в дереве перебора: считается, что глубина корня дерева равна нулю, а глубина любой последующей вершины равна единице плюс глубина вершины, непосредственно ей предшествующей. Отсюда следует, что наибольшую глубину всегда будет иметь та вершина, которая должна быть в этот момент раскрыта. Если образующийся путь оказывается бесполезным, то есть при заданной глубине раскрытия целевой вершины не получилось, необходимо вернуться в вершину, предшествующую раскрытой, и попытаться еще раз применить к ней операцию раскрытия – далее эта процедура многократно повторяется до тех пор, пока не будет получена целевая вершина.

Возврат осуществляется с помощью указателей: как только в процессе порождения вершин достигается заданная граничная величина, раскрывается вершина наибольшей глубины, не превыша-

ющая этой границы. Общая схема перебора в глубину показана на рис. 3.7.

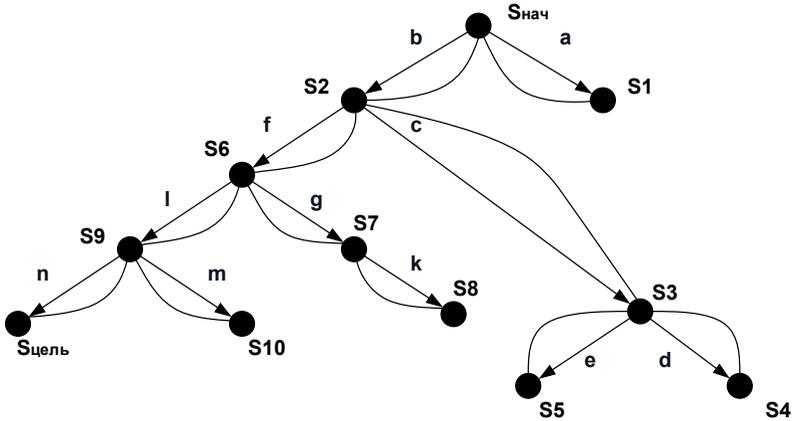


Рис. 3.7. Дерево полного перебора в глубину

Алгоритм перебора в глубину состоит в следующем.

1. Раскрывается начальная вершина, соответствующая начальному состоянию $S_{нач}$.

2. Раскрывается первая вершина, получаемая в результате раскрытия $S_{нач}$, ставится указатель.

3. Если она раскрывается, то следующей будет раскрываться вновь порожденная вершина. Если вершина не раскрывается, то процесс возвращается в предыдущую вершину.

4. По получении целевой вершины процесс раскрытия заканчивается и по указателям строится путь, ведущий к корню. Соответствующие дугам операторы образуют решение задачи.

5. Если для заданной глубины раскрытия целевая вершина не находится, то весь процесс повторяется снова, а в качестве новой вершины рассматривается самая левая из полученных на предыдущем этапе.

Так же, как и метод поиска «в ширину», этот метод относится к методам грубой силы. Он обеспечивает перебор всех состояний, если, конечно, прежде не встретит целевое.

Эвристические методы поиска в пространстве состояний.

Методы полного перебора гарантируют решение задачи, если оно

существует, а при наличии нескольких решений дают наилучшее из них – оптимальное в заданном смысле. Однако на практике все эти методы используются для решения только небольших по размерности графов состояний задач, тогда как для реальных случаев обычно используется дополнительная информация – основанная на предыдущем опыте или полученная на основе теоретических выводов. Такая информация по-прежнему (см. раздел 1.1) наносит название эвристической, а организованная в форме правил – именуется **эвристическими правилами** (эвристиками). Данные эвристики носят сугубо специальный характер и могут применяться лишь в рамках решения конкретной задачи, в лучшем случае – в рамках задач одного определенного класса.

Эвристическая информация превращает грубый перебор в упорядоченный – в подтверждение этого рассмотрим известное решение задачи о коммивояжере. Пусть некий коммивояжер должен построить свой маршрут так; чтобы побывать в каждом из n городов в точности по разу и возвратиться в исходный город. Желательно также, чтобы этот маршрут был минимальным по протяженности. Пусть в нашем случае $n = 5$; города обозначены как A, B, C, D, E ; а длина пути от города до города указана на рис. 3.8.

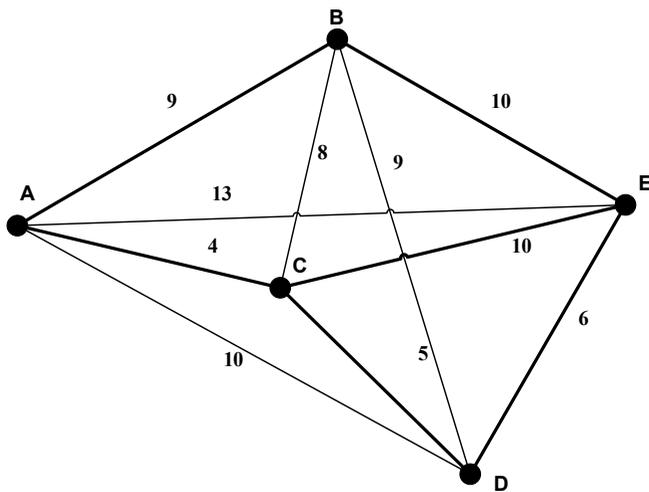


Рис. 3.8. Задача о коммивояжере

эвристических оценочных функций – которые в общем случае также не гарантируют нахождение оптимальных (наилучших или очень близких к ним) решений, но с их помощью достигается большая экономия в затратах на поиск решения, что является важным фактором.

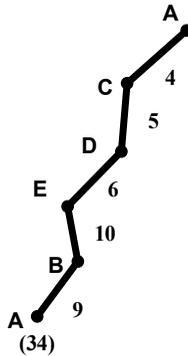


Рис. 3.10. Граф поиска решения задачи в глубину

Процедуру перебора на заданную глубину называют также **про-граммированием с обратным слежением** (Back-Track Programming). Отметим, что возможно использование комбинированных стратегий поиска: методами и в ширину, и в глубину; перебор этапами; ограничение числа дочерних вершин; двунаправленный перебор от $S_{нач}$ к $S_{цель}$ и обратно; методы ветвей и границ и динамическое программирование, применяемые в исследовании операций.

Поиск решения методом разбиения задач на подзадачи. Рассмотрим еще один распространенный способ решения методом разбиения задач на подзадачи. Идея данного метода состоит в следующем: если путь решения задачи неясен (или неизвестен), то нужно постараться найти какую-то исходную «точку опоры», начиная с которой (или от нее, или до нее) путь известен. Эта точка может быть в любом месте – в начале, в середине, в конце пути, но на практике она обязательно есть.

Очевидно, что данной точке будет соответствовать некоторое состояние S_i на пути от $S_{нач}$ до $S_{цель}$ – если точка S_i определена, то к ней уже можно применить какое-либо действие и этим изме-

нить состояние предметной области, то есть приблизиться к возможному решению. Для нового состояния также ищется точка опоры на оставшемся пути, к ней снова применяется какое-то (или то же самое) действие и т.д., пока не будет решена задача в целом.

Чтобы детально описать данный метод, необходимо вернуться к определению понятия задачи. Для этой цели воспользуемся понятием пространства состояний. В более общем виде некую задачу Z можно представить следующим образом:

$$Z = \langle S, Q, F \rangle, \quad (3.19)$$

где S – множество начальных состояний, Q – множество операторов, переводящих предметную область из одного состояния в другое, F – множество целевых состояний.

При таком обозначении промежуточные состояния S_i удобнее обозначать через f_i поскольку они теперь представляют собой как бы промежуточные цели. Конечной целью сведения задачи к подзадачам является получение таких *элементарных задач*, решения которых очевидны. Элементарными при этом считаются задачи, которые могут быть решены за один шаг, то есть за одно применение какого-либо оператора из множества Q .

Поиск решения методом разбиения задач на подзадачи включает два этапа:

- представление задачи в виде «И/ИЛИ» графа;
- сведение задачи к подзадачам.

Представление задачи в виде «И/ИЛИ» графа. Между полученными при разбиении подзадачами могут быть отношения согласованности (одновременности) их решения (отношение **И**) или отношение альтернативности (отношение **ИЛИ**). Рассмотрим это подробнее: допустим, что исходная задача Z может быть разбита на подзадачи A, B, C, D, E . Пусть при этом подзадачи A и B , а также C и D находятся в отношениях **И**, то есть должны решаться согласованно (одновременно), а множества $(A \text{ И } B)$, $(C \text{ И } D)$ и подзадача E находятся в отношении **ИЛИ**, то есть имеют альтернативный характер отношений. Все сказанное можно отобразить на графе как это показано на рис. 3.11.

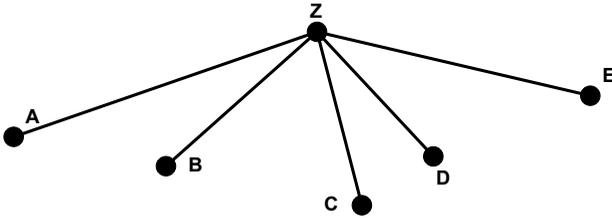


Рис. 3.11. Граф отношений для задачи Z

Отношения типа «И» будем отмечать двойной дугой, связывающей ребра графа. Для единообразия представления можно ввести дополнительные (фиктивные) вершины, которые группировали бы подзадачи типа «И» и ИЛИ под своей собственной родительской вершиной – см. рис. 3.12.

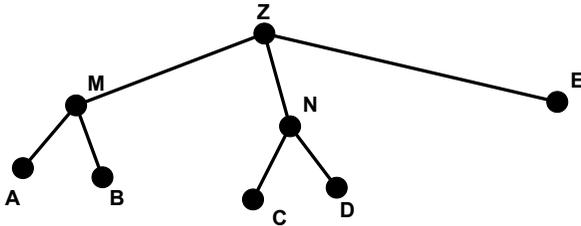


Рис. 3.12. И/ИЛИ граф задачи Z

Таким образом, исходная задача Z представляется подзадачами M, N и E, имеющими альтернативный ИЛИ-характер, а подзадачи M и N, в свою очередь, - подзадачами (A, B) и (C, D) с отношениями типа И. Альтернативные вершины M, N, E называются поэтому ИЛИ-вершинами, а вершины A, B, C, D – И-вершинами и помечаются двойной дугой. Структура типа изображенной на рис. 3.12 называется «И/ИЛИ» графом: можно видеть, что если вершина «И/ИЛИ» графа имеет непосредственно следующие за ней вершины, то все они являются либо ИЛИ - вершинами, либо «И» - вершинами. Если у вершины имеется только одна следующая за ней вершина, то ее можно считать как «И», так и «ИЛИ» - вершиной. Если вершин типа «И» вообще нет, то получается обычный граф, тот самый, который фигурировал в методе полного перебора при поиске в пространстве состояний.

Структура типа «И/ИЛИ» графа является удобной для представления дерева подзадач. При этом начальная вершина графа соответствует начальному состоянию $S_{\text{нач}}$ исходной задачи, а вершины, которые соответствуют описаниям элементарных подзадач, называются *заключительными*.

Основная цель поиска на «И/ИЛИ» графе – показать *разрешимость* вершины S_i . Вершина является разрешимой, если выполняется одно из следующих условий:

- вершина S_i является заключительной;
- следующие за S_i вершины являются вершинами типа «ИЛИ» и при этом хотя бы одна из них разрешима;
- следующие за S_i вершины являются вершинами типа «И» и при этом каждая из них разрешима.

Решающим графом называется подграф, состоящий из разрешимых вершин с корнем в начальной вершине. Если у вершины И/ИЛИ графа, не являющейся заключительной, нет следующих за ней вершин, такая вершина называется *неразрешимой*. Порождение новых вершин (данная операция называется *редукцией задачи*) выполняется путем применения обобщенного оператора Q (то есть каких-либо операторов из множества Q) сведения задачи к подзадачам. Применение оператора Q к описанию задачи порождает всю структуру И/ИЛИ графа (или иначе *графа редукции*).

Механизм сведения задачи к подзадачам. Напомним формулировку поставленной цели – построить алгоритм, который, будучи примененным чисто механически, обязательно даст результат, то есть требуемую структуру в виде И/ИЛИ графа с выделенным решающим подграфом. Такой алгоритм представляет собой один из механизмов планирования пути решения задачи.

Как следует из предыдущего раздела, исходная задача Z задается своим начальным описанием $Z = \langle S, Q, F \rangle$ – установлено, что свести ее к совокупности более простых задач в пространстве состояний можно, если нам удастся выделить основные промежуточные состояния $f_1, f_2, f_3, \dots, f_n$, каждому из этих состояний в соответствие можно поставить свое описание в виде троек

$$\langle S, Q, f_1 \rangle, \langle f_1, Q, f_2 \rangle, \langle f_2, Q, f_3 \rangle, \dots, \langle f_n, Q, F \rangle. \quad (3.20)$$

Решение совокупности подзадач (3.20) эквивалентно решению исходной задачи – на этом и построен рассматриваемый механизм, который включает следующие действия.

1. Выделяем по крайней мере один оператор $q_i \in Q$, который обязательно будет участвовать в решающей цепочке операторов – все операторы такого типа называются *ключевыми* (то есть это операторы, обязательно участвующие в решающей цепочке).

2. Для каждого из ключевых операторов (если их несколько) определяются промежуточные состояния, к которым они могут быть применимы в условиях задачи Z . Для оператора q_i это будет состояние f_i . – таких состояний может быть несколько, и тогда они образуют подмножество целевых состояний $F_{q_i} \in F$. После этого можно выделить подзадачу поиска пути от начала до состояния f_i (или до F_{q_i}) – другими словами, применение оператора q_i привело к первой подзадаче с описанием $\langle S, Q, f_i \rangle$ или $\langle S, Q, F_{q_i} \rangle$.

3. Как только такое описание найдено, может быть сформулирована вторая подзадача, соответствующая элементарной – поскольку состояние f_i соответствует оператору q_i , то его можно применить к f_i и получить, таким образом, новое состояние $q_i(f_i)$, приближающее нас к конечной цели – правда лишь на один шаг.

4. От вновь полученного состояния $q_i(f_i)$ до конечной цели F еще может быть еще очень неблизкий путь. Определить его – третья подзадача.

Таким образом, применение выбранного оператора q_i к задаче с описанием $Z = \langle S, Q, F \rangle$ позволяет выделить три подзадачи:

$$\langle S, Q, f_i \rangle, \langle f_i, q_i, q_i(f_i) \rangle, \langle q_i(f_i), Q, F \rangle, \quad (3.21)$$

одна из которых является элементарной. Сказанное иллюстрирует линейный график на рис. 3.13, где решение изображается отрезком,

который разбивается точкой f_i , соответствующей оператору q_i , на три указанные подзадачи.

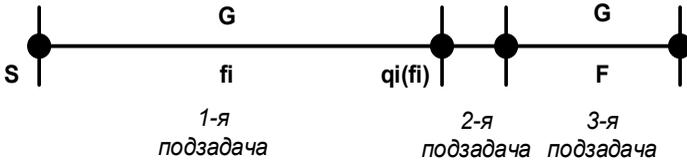


Рис. 3.13. Разбиение задачи на подзадачи

Данному разбиению соответствует И/ИЛИ граф, показанный на рис. 3.14. Элементарная подзадача второго типа для любой выбранной точки f пространства состояния решается всегда, поэтому ее можно не указывать.

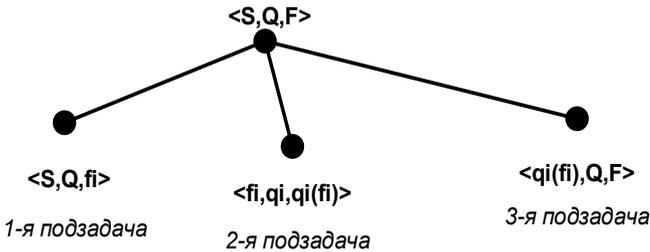


Рис. 3.14. И/ИЛИ граф разбиения на подзадачи для состояния f_i

Выбранная точка f_i является одной из возможных промежуточных целей $f_i \in F_{q_i}$ – к ней применяют оператор q_i и в итоге приходят к окончательному виду И/ИЛИ графа разбиения на подзадачи для одной точки, показанному на рис. 3.15.

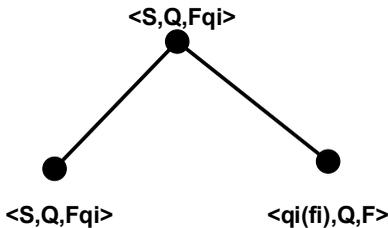


Рис. 3.15. Обобщенный И/ИЛИ граф для одной точки (f_i)

5. Каждая из полученных при разбиении подзадач, если она не является элементарной, может быть снова разбита на подзадачи аналогичным способом – возможно, при помощи другого оператора и т.д.

Таким образом, для разбиения задачи на подзадачи и построения соответствующего **И/ИЛИ** графа нужны ключевые операторы (обычно более одного). Один из способов нахождения операторов, могущих быть ключевыми» состоит в вычислении *различий* между состояниями по пути от $S_{\text{нач}} \in S$ к $S_{\text{цель}} \in F$. Каждому возможному различию ставится в соответствие оператор (или их множество), который это различие может устранить. Цепочка операторов, последовательно устраняющих различия между $S_{\text{нач}}$ и $S_{\text{цель}}$, называется решением задачи.

Проиллюстрируем метод разбиения задачи на подзадачи и поиска решений на **И/ИЛИ** графе на примере задачи с бананами и обезьяной. Исходное описание задачи представляет собой

$$S_{\text{нач}} = (a, b, c, 0, 0) \in S ;$$

$$S_{\text{цель}} = (c, c, c, 1, 1) \in F ;$$

$$Q = (q_1, q_2, q_3, q_4) .$$

Отметим, что оператор q_1 устраняет различие между состояниями $(a, b, c, 0, 0)$ и $(b, b, c, 0, 0)$, и аналогичным образом другие операторы:

$$q_2 \text{ – между } (b, b, c, 0, 0) \text{ и } (c, c, c, 0, 0),$$

$$q_3 \text{ – между } (c, c, c, 0, 0) \text{ и } (c, c, c, 1, 0),$$

$$q_4 \text{ – между } (c, c, c, 1, 0) \text{ и } (c, c, c, 1, 1),$$

Выделим подзадачи, применяя, например, оператор q_3 (выбор операторов может быть произвольным, поскольку речь идет о механической процедуре). Ему будут соответствовать подзадачи со следующими описаниями:

1. $(a, b, c, 0, 0)$, Q , $(c, c, c, 0, 0)$, так как оператору q_3 соответствует состояние $F_{q_3}(c, c, c, 0, 0)$.

2. $(c, c, c, 0, 0)$, q_3 , $(c, c, c, 1, 0)$. Здесь $(c, c, c, 1, 0) = q_3(c, c, c, 1, 0)$, то есть. результат применения ключевого оператора q_3 к состоянию $(c, c, c, 1, 0)$.
3. Подзадача $((c, c, c, 1, 0), Q, (c, c, c, 1, 1))$.

Мы видим, что первая подзадача не является заключительной (то есть допускает дальнейшее разбиение), тогда как третья подзадача является элементарной, то есть заключительной, если применить оператор q_4 . Вторая подзадача также является заключительной.

Разбиваем на вторичные подзадачи первую подзадачу. Одно из различий между начальным и целевым состояниями может быть устранено применением, например, оператора q_2 (возможно и q_1).

Получим новые подзадачи с описаниями:

- 1.1. $((a, b, c, 0, 0), Q, (b, b, c, 0, 0))$.
- 1.2. $((b, b, c, 0, 0), q_2, (c, c, c, 0, 0))$

Здесь мы видим, что подзадача 1.2 является элементарной, а различие в подзадаче 1.1 может быть устранено, если применить оператор $q_1 \in Q$, и тогда она тоже будет элементарной. Целиком граф «И/ИЛИ» приведен на рис. 3.16.

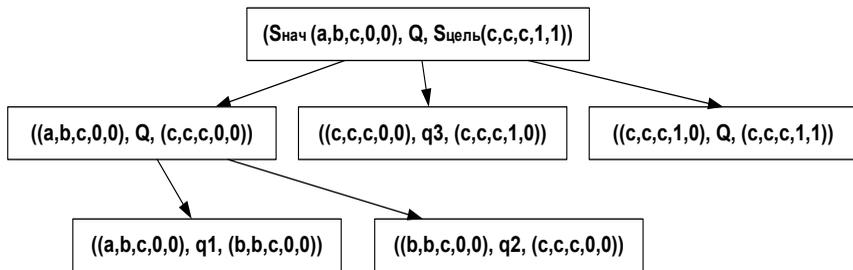


Рис. 3.16. И/ИЛИ граф для задачи с бананами и обезьяной
ключевой оператор – « q_3 »

Таким же образом можно было бы осуществить разбиение исходной задачи на подзадачи, применяя вначале любой другой оператор: q_4 , q_2 или q_1 . На рис. 3.17 показан граф для ключевого оператора q_4 .

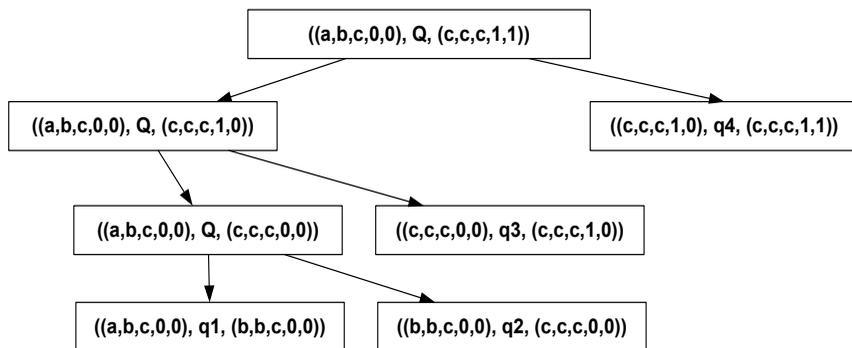


Рис. 3.17. И/ИЛИ граф для оператора q_4

3.8. Критерии оценки функционирования ЭС

Требования к современному программному обеспечению становятся все более сложными и изощренными, поскольку пользователи ожидают все больше и больше от предлагаемых приложений. Примером может служить концепция Internet-вещей, когда в мире высокоразвитой связи приложения должны взаимодействовать с другими приложениями и службами, а также работать в различных средах, например в облаке, и на портативных устройствах. Распространенные ранее монолитные архитектуры сменились сервис-ориентированным программным обеспечением, использующее платформы, операционные системы, хост-приложения и сети для реализации функций, о которых несколько лет назад ничего не было известно.

Как известно, совокупная стоимость программного обеспечения помимо затрат на его разработку, включает затраты, возникающие после развертывания и внедрения. Это влияет не только на архитектуру, но также и на процессы развертывания, обслуживания и администрирования программного обеспечения. Область инженерии знаний не является исключением, поэтому критерии оценки функционирования ЭС необходимо проработать и обсудить со всеми заинтересованными сторонами до начала активных действий по разработке ЭС, чтобы сократить затраты на разработку программного обеспечения.

Совокупная стоимость владения программным продуктом в основном состоит из затрат, возникающих после развертывания.

Приложение с хорошо продуманной архитектурой обеспечивает минимальную стоимость владения благодаря снижению затрат и времени, необходимых на развертывание приложения, обеспечение его работы, обновление для удовлетворения меняющимся требованиям и устранение проблем. Администрирование и поддержка пользователей также будут упрощены.

Анализ публикаций и практического состояния данного вопроса показывает, что на сегодня разработаны самые разнообразные подходы к оценке результатов создания и внедрения ИС. Каждому из этих подходов присущи как определенные преимущества, так и недостатки, собственные особенности применения, объем необходимых исходных данных, уровень обоснованности, глубины, достоверности предоставляемых пользователю результатов.

Ниже приведены наиболее часто используемые критерии (естественно, в каждом конкретном проекте по созданию ЭС данный перечень критериев может быть дополнен). Успешное программное обеспечение должно соответствовать нескольким важным критериям:

- пригодность ЭС для конечного пользователя, определяемая, как правило, удобством работы с ЭС и ее полезностью. Под полезностью ЭС обычно понимается ее способность в ходе диалога определять потребности пользователя, выявлять и устранять причины неудач в работе, а также удовлетворять указанные потребности пользователя, корректно решая поставленные задачи [91];
- удобство работы пользователя с ЭС, подразумевающее естественность взаимодействия пользователя с ЭС;
- гибкость ЭС – способность системы настраиваться на разные профили пользователей, а также учитывать изменения в квалификации одного и того же пользователя;
- должно обеспечивать безопасность, чтобы приложение и его данные были защищены от атак злоумышленников и случайных ошибок;
- должно быть устойчивым и надежным для минимизации сбоев и соответствующих затрат;
- должно работать с необходимыми параметрами в соответствии с требованиями пользователей, такими как максимальное время отклика или определенная рабочая нагрузка;

– должно быть простым в поддержке для снижения затрат на администрирование и поддержку и достаточно расширяемым для включения неизбежных изменений и обновлений, которые со временем потребуются.

Необходимо остановиться еще на одном универсальном критерии, который характерен не только для ИТ-проектов, но и для любых проектов. Это соблюдение рамок проекта (временных, бюджетных, содержания), получивший название «треугольник проекта»: ни один из углов треугольника не может быть изменен без оказания влияния на другие (например, чтобы уменьшить время, требуется увеличить стоимость и/или сократить содержание).

Оформление рабочего проекта. Все решения, принятые в ходе выполнения стадии проектирования, необходимо зафиксировать в едином документе, получившем название «Рабочий проект». Рабочий документ является документом, на основании которого будут проводиться все дальнейшие работы по созданию ЭС. Содержание рабочего документа, как правило, формируется согласно [95]. Пояснительная записка рабочего проекта содержит следующие разделы.

1. Общие положения

- 1.1. Наименование проектируемой автоматизируемой системы.
- 1.2. Документы, на основании которых ведется проектирование.
- 1.3. Организации, участвующие в разработке.
- 1.4. Стадии и сроки исполнения.
- 1.5. Цели, назначение и области использования.
- 1.6. Соответствие проектных решений нормам и правилам техники безопасности, пожарно- и взрывобезопасности.
- 1.7. Нормативно-технические документы.
- 1.8. НИР и изобретения, используемые при разработке системы.
- 1.9. Очередность создания системы.

2. Описание процесса деятельности

Отражает предметную область рабочего проекта.

3. Основные технические решения

- 3.1. Структура системы, перечень подсистем.
- 3.2. Способы и средства связи для информационного обмена между компонентами подсистем.

- 3.3. Взаимосвязь АС со смежными системами.
- 3.4. Режимы функционирования системы.
- 3.5. Численность, функции и квалификация персонала.
- 3.6. Обеспечение потребительских характеристик системы.
- 3.7. Функции, выполняемые системой.
- 3.8. Комплекс технических средств.
- 3.9. Информационное обеспечение системы.
- 3.10. Программное обеспечение системы.

4. Мероприятия по подготовке объекта автоматизации к вводу системы в действие

- 4.1. Приведение информации к виду, пригодному для обработки на ЭВМ.
- 4.2. Мероприятия по подготовке персонала.
- 4.3. Организация необходимых подразделений и рабочих мест.
- 4.4. Изменение объекта автоматизации.
- 4.5. Дополнительные мероприятия.

3.9. Стадия разработки ЭС

Целью данного этапа работ является создание одного или нескольких прототипов ЭС, решающих требуемые задачи. Содержание данной стадии напрямую зависит как от выбранного средства разработки и квалификации разработчиков, так и от особенностей предметной области и полученных от экспертов сведений.

Стадию разработки условно можно разделить на два процесса, осуществляющихся параллельно:

- работа инженера по знаниям с экспертов на предмет получения знаний для будущей работы ЭС;
- работа программистов по созданию работающего приложения.

При создании ЭС рекомендуют использовать технологию RAD (Rapid Application Development, быстрая разработка приложений) – это жизненный цикл процесса проектирования, созданный для достижения более высоких скорости разработки и качества ПО, чем это возможно при традиционном подходе к проектированию.

RAD предполагает, что разработка ПО осуществляется небольшой командой разработчиков за срок порядка трех-четырех месяцев путем использования инкрементного прототипирования с применением инструментальных средств визуального моделирования и разработки.

Применение технологии RAD целесообразно, когда [54]:

- требуется выполнение проекта в сжатые сроки. Быстрое выполнение проекта позволяет создать систему, отвечающую требованиям сегодняшнего дня. Если система проектируется долго, то весьма высока вероятность, что за это время существенно изменятся фундаментальные положения, регламентирующие деятельность организации, то есть, система морально устареет еще до завершения ее проектирования.
- нечетко определены требования к ПО. В большинстве случаев заказчик весьма приблизительно представляет себе работу будущего программного продукта и не может четко сформулировать все требования к ПО. Требования могут быть вообще не определены к началу проекта либо могут изменяться по ходу его выполнения.
- проект выполняется в условиях ограниченности бюджета. Разработка ведется небольшими RAD группами в короткие сроки, что обеспечивает минимум трудозатрат и позволяет вписаться в бюджетные ограничения.
- интерфейс пользователя (GUI) есть главный фактор. Нет смысла заставлять пользователя рисовать картинки. RAD технология дает возможность продемонстрировать интерфейс в прототипе, причем достаточно скоро после начала проекта.
- проект большой, но поддается разделению на более мелкие функциональные компоненты. Если предполагаемая система велика, необходимо, чтобы ее можно было разбить на мелкие части, каждая из которых обладает четкой функциональностью. Они могут выпускаться последовательно или параллельно (в последнем случае привлекается несколько RAD групп).

Основные принципы RAD можно сформулировать следующим образом:

- работа ведется группами;
- разработка базируется на моделях;
- использование итерационного прототипирования;
- RAD группа всегда работает только над одним прототипом;
- если проект сложный, то для него может быть выделено несколько RAD групп;
- обязательное использование инструментальных средств, автоматизирующих процесс разработки.

Как было указано выше, в рамках стадии разработки может быть создано несколько прототипов. Цель создания каждого прототипа: подтвердить, что прототип обеспечивает проверку адекватности идей, выбранных методов и способов представления знаний решаемым задачам. Создание первого прототипа должно подтвердить, что выбранные методы решений и способы представления пригодны для успешного решения, по крайней мере, ряда задач из актуальной предметной области, а также продемонстрировать тенденцию к получению высококачественных и эффективных решений для всех задач предметной области по мере увеличения объема знаний [91]. После разработки первого прототипа осуществляются тестирование работы ЭС. В результате анализ пожеланий и замечаний служит исходной точкой для создания второго прототипа или «расширения» первого прототипа. Данный процесс по созданию прототипов итеративный и может продолжаться довольно длительное время (зависит как от успешности проводимых работ, так и от имеющегося финансирования проекта).

В результате выполнения данного этапа должен быть получен конечный программный продукт, пригодный для работы.

Тестирование ЭС и подготовка документации. В ходе выполнения данного этапа осуществляют следующие шаги:

- провести оценку функционирования полученного прототипа;
- осуществить тестирование ЭС с помощью инженера по знаниям и эксперта;
- привлечь конечных пользователей к работе над прототипом ЭС;
- выявить и исправить найденные ошибки.

Как только создано детальное описание первого модуля системы к работе подключаются программисты и технические специалисты. Последние составляют всю необходимую документацию по работе с ЭС:

- руководство администратора, описывающего процедуру инсталляции ЭС;
- руководство инженера по знаниям, описывающего процедуру наполнения базы знаний;
- руководство пользователя, содержащее описание основных действий конечного пользователя при работе с ЭС.

В рамках тестирования эксперт и инженер по знаниям в интерактивном режиме с использованием диалоговых и объяснительных средств системы проверяют компетентность ЭС.

Специалисты выделяют три аспекта тестирования ЭС: тестирование исходных данных; логическое тестирование базы знаний; концептуальное тестирование прикладной системы.

- тестирование исходных данных: включает проверку фактографической информации, служащей основой для проведения экспертизы;
- логическое тестирование БЗ: заключается в обнаружении логических ошибок в системе продукций, не зависящих от предметной области;
- концептуальное тестирование: проводится для проверки общей структуры системы и учета в ней всех аспектов решаемой задачи.

При проведении тестирования могут быть использованы следующие виды тестирования:

- тестирование на основе концепции «черного ящика» – набор тестируемых ситуаций генерируется без учета используемых в системе методов решения задачи;
- тестирование на основе концепции «белого ящика» – тестируемые ситуации учитывают внутреннюю структуру системы в дополнение к входным и ожидаемым выходным наборам данных;
- тестирование полноты БЗ – тестирование правил на внутреннюю неполноту;
- тестирование целостности БЗ – тестирование правил на внутреннюю целостность

В ходе тестирования производится оценка выбранного представления знаний в ЭС в целом, для чего инженер по знаниям должен подобрать тестовые сценарии, которые должны обеспечить проверку функциональных возможностей разработанной ЭС.

Как правило, большая часть ошибок, выявляемых при тестировании ЭС, относится к одной из трех следующих категорий [60].

1. Неудачно подобранные тестовые сценарии. Так, например, тестовые примеры могут оказаться слишком однотипными и не охватывать всю предметную область либо могут вообще не относиться к предметной области.
2. Ошибки ввода-вывода, характеризуются данными, приобретенными в ходе диалога с экспертом, и заключениями, предъявлен-

ными ЭС в ходе объяснений. Методы приобретения данных могут не давать требуемых результатов, так как, например, задавались неправильные вопросы или собрана не вся необходимая информация. Кроме того, вопросы системы могут быть трудными для понимания, многозначными и не соответствующими знаниям пользователя. Наиболее распространенный источник ошибок в рассуждениях касается правил вывода. Важная причина здесь часто кроется в отсутствии учета взаимозависимости сформированных правил. Другая причина заключается в ошибочности, противоречивости и неполноте используемых правил.

3. Применяемые управляющие стратегии. Изменение стратегии бывает необходимо, например, если ЭС анализирует сущности в порядке, отличном от «естественного» для эксперта. Последовательность, в которой данные рассматриваются ЭС, не только влияет на эффективность работы системы, но и может приводить к изменению конечного результата. Кроме того, недостатки в управляющих стратегиях могут привести к чрезмерно сложным заключениям и объяснениям ЭС.

3.10. Стадии внедрения и опытной эксплуатации

Успех *внедрения проекта* напрямую зависит от работы команды внедрения. Рекомендуется назначать ответственных за внедрение лиц (или формировать так называемую команду внедрения). Команда внедрения должна включать трех групп специалистов. Первая – это представители команды ИТ-разработчиков; вторая – это функциональные специалисты (конечные пользователи системы), кто в дальнейшем будет непосредственно работать с разработанной ЭС; третья – это внешние консультанты (эксперты), не задействованные непосредственно при разработке ЭС, но которые могут дать объективную оценку о работе ЭС. Следует отметить, что поиск внешних экспертов может представлять трудность, поскольку исследуемая предметная область может быть слишком «узкоспециализированной».

Опытная эксплуатация включает следующие шаги (часть из них может быть опущена в зависимости от конкретного проекта по созданию ЭС):

- установка всех модулей системы у конечного пользователя;
- осуществление наладки и запуска ЭС у конечного пользователя;

- внесение изменений согласно предложениям, внесенным пользователями;
- проведение приемосдаточных испытаний и подписание необходимых отчетов и документации.

Поскольку при разработке ЭС необходимо соблюдать принцип модульности, то такая модульная структура позволит переносить ЭС частями, когда отдельные функциональные возможности будут доступны разным группам пользователей раньше, чем будет готова окончательная версия ЭС. Следовательно, конечные пользователи получают ЭС «как можно раньше» и смогут начать освоение (и параллельно, тестирование) системы постепенно. Подобный подход позволяет на ранних этапах, во-первых, выявить потенциальные ошибки, а, во-вторых, получить отзывы об удобстве работы системы и пожелания по доработке.

Перед непосредственной установкой на рабочие места пользователей система (и каждая ее часть) пройдет тестирование на соответствие требованиям и работоспособность. В случае положительного результата система будет введена в действие. При обнаружении ошибок (в зависимости от их серьезности) система либо будет отправлена на доработку, либо будет внедрена, а ошибки будут устранены с использованием, так называемых, обновлений. По окончании работ подписывается акт о результатах приемосдаточных работ. Перечень документов зависит от условий контракта.

Обучение пользователей является важнейшим этапом интеграции приложений и информационных систем. Основной целью обучения является формирование у пользователей необходимых знаний и навыков по работе с ЭС. Корректное обучение пользователей является одним из критериев успеха при внедрении и последующей эксплуатации ЭС.

Перед проведением непосредственного обучения необходимо провести **подготовительные мероприятия**:

- разработать программу обучения с указанием тем и краткого содержания по каждой теме;
- подготовить учебные примеры для демонстрации последовательности действий пользователя;
- сформировать группы обучаемых (при необходимости, возможно проведение индивидуального обучения);

- составить график обучения;
- выпустить приказ о проведении обучения, в котором явно указать место проведения; программу обучения; состав групп; график обучения;
- распространить среди обучающихся сотрудников «Руководство пользователя» и «Руководство инженера по знаниям по наполнения БЗ» (зависит от роли обучаемого пользователя).

Обучение пользователей может быть организовано с помощью следующих методов:

- отработка практических навыков в группах. Согласно графику обучения осуществляется знакомство каждой группы с функциональными возможностями работы системы путем демонстрации на экране компьютера и/или проектора учебных примеров. В ходе изложения материала обучаемые могут задавать все возникающие вопросы по теме. После чего каждый пользователь получает индивидуальное задание и пытается самостоятельно его пройти, а преподаватель оказывает помощь в случае затруднений.
- индивидуальная работа с ключевыми сотрудниками (как правило, инженером по знаниям, кто будет в дальнейшем заниматься пополнение БЗ). Этот метод обучения незаменим для специалистов, играющих ключевые роли в повседневном использовании ЭС в компании.
- вебинары помогают подготовить сотрудников территориальных подразделений. Этот метод обучения подходит для передачи теоретической базы. Так, при простом обучении в группе воспринимается и запоминается до 10% материала. При использовании наглядных пособий и звукового ряда восприятие увеличивается до 50-80% от общего объема материала. Поэтому использование различных презентаций, демо-роликов, инструкций намного эффективнее.

В целом каждый этап обучения рекомендуется проводить по следующему сценарию:

- теоретическая подготовка пользователей. Как правило, перед началом обучения обучаемым рекомендуется детально ознакомиться с руководствами пользователя, чтобы не тратить время на озвучивание основных положений работы ЭС.

- практическая подготовка пользователей. На данном этапе пользователям системы рекомендуется попробовать внести в систему тестовые данные в соответствии с шагами, описанными в обучающих презентациях и инструкциях.
- работа с реальными данными: пользователи вносят в систему реальные данные.

Внедрение любого инновационного продукта может сопровождаться противодействием со стороны конечных пользователей, которое может приводить к саботажу проекта. Подобное поведение вызвано следующими причинами:

- страх перед нововведением и консерватизм;
- опасение потерять рабочее место и/или свою незаменимость;
- боязнь возросшей ответственности за свои действия.

Для минимизации последствий подобного поведения руководитель компании (или подразделения), где планируется внедрение ЭС, должен вести разъяснительную работу с подчиненными кадрами, а также создать у сотрудников твердое ощущение неизбежности внедрения программного продукта.

Также на этапе внедрения может временно возрасти нагрузка на сотрудников компании, поскольку они, помимо выполнения обычных рабочих обязанностей, вынуждены осваивать новые технологии и новые знания. Кроме того, во время проведения опытной эксплуатации и при переходе к промышленной эксплуатации системы в течение некоторого времени приходится вести дела, как и в новой системе, так и продолжать ведение их традиционными способами. В связи с этим, отдельные этапы проекта внедрения системы могут затягиваться под предлогом того, что у сотрудников и так хватает срочной работы по прямому назначению, а освоение системы является второстепенным и отвлекающим занятием. В таких случаях руководителю предприятия, помимо ведения разъяснительной работы с уклоняющимися сотрудниками, необходимо:

- повысить уровень мотивации сотрудников к освоению системы в форме поощрений и благодарностей;
- принять организационные меры к сокращению срока параллельного ведения дел.

Результатом стадии обучения являются обученные сотрудники, то есть пользователи, которые работают с разработанной ЭС без постоянного наблюдения со стороны разработчиков. Вмешатель-

ство и помощь со стороны разработчиков происходит только по запросу пользователей.

3.11. Выводы

Процесс проектирования и разработки экспертных систем имеет ряд существенных особенностей, которые необходимо учитывать разработчикам:

- важнейшей задачей при этом является создание и наполнение базы знаний в составе ЭС;
- нужно учитывать эмпирический характер аксиологических знаний эксперта;
- процесс логического тестирования БЗ отличается сложностью и трудоемкостью и т.д.

Перечень проблем, с которыми приходится сталкиваться инженерам по знаниям и разработчикам, ведет к тому, что процесс проектирования ЭС зачастую ассоциируется у них исключительно с творчеством, поскольку извлечение и наполнение БЗ знаниями – это процесс действительно достаточно сложный и «ювелирный». Однако принадлежность ЭС к классу ИИС позволяет применить к процессу разработки ЭС базовые принципы и подходы проектирования «обычных» ИС. Поэтому при описании содержания этапов жизненного цикла ЭС в настоящем разделе были учтены основные положения методик по созданию ИС: ГОСТ 34.601-90, ISO/IEC 12207:1995, MSF, положения РМВОК и др. Рассмотрена проблема формирования проектной команды участников создания ЭС, определения их ролей и зон ответственности. Освещены вопросы об используемых механизмах рассуждения и аспектах управления выводом.

4. ИНТЕЛЛЕКТУАЛЬНАЯ ПОДДЕРЖКА УПРАВЛЕНИЯ БИЗНЕС-ПРОЦЕССАМИ

4.1. Квазиоптимальное управление иерархическими СС с позиций классической теории управления

Математические задачи, связанные с оптимизацией процессов управления, относятся к числу наиболее сложных экстремальных задач. В рамках классической теории вариационного исчисления их удалось применить к решению проблем оптимального управления благодаря принципу максимума Л.С. Понтрягина (1953 г.), суть которого состоит в следующем [81]. Если поведение объекта управления (ОУ) можно описать с помощью математической модели (ММ) в виде системы дифференциальных уравнений

$$\frac{dx_i}{dt} = f(x_1; x_2 \dots x_n; u_1; u_2 \dots u_n; t); \quad i [0; n], \quad (4.1)$$

где $x(x_1; x_2 \dots x_n; t)$ – вектор координат состояния ОУ в его фазовом пространстве; $u(u_1; u_2 \dots u_n; t)$ – вектор, отражающий координаты управляющего воздействия на ОУ, $f(x_1; x_2 \dots x_n; u_1; u_2 \dots u_n; t)$ – функционал качества работы ОУ; то **основная задача оптимального управления** формулируется так: среди всех допустимых управлений $u(t); t[t_1; t_2]$, способных перевести точку S , изображающую ОУ в его фазовом пространстве, из начального положения S_1 в конечное S_2 (предполагая, что данные управления существуют), необходимо найти такое управление, для которого функционал «пути» объекта

$$J(x; u) = \int_{t_1}^{t_2} f(x; u; t) dt, \quad (4.2)$$

соответствующий траектории перевода его качества из точки S_1 в точку S_2 , будет минимальным.

Теорема максимума Понтрягина гласит, что если такая оптимальная траектория перевода качества ОУ из точки S_1 в точку S_2 существует, то существует и необходимая для этого функция оптимального управления $u(t); t[t_1; t_2]$, которая должна удовлетворять условию максимума некоторой скалярной функции

$H(\psi; x; u; t)$, которая соответствует уравнению Гамильтона, именуется *гамильтонианом* системы (4.1) и имеет вид

$$H(\psi; x; u; t) = \sum_{i=1}^{n+1} \psi_i f_i, \quad (4.3)$$

где $\frac{d\psi_i}{dt} = -\sum_{j=1}^{n+1} \psi_j \frac{\partial f_j}{\partial x_i}$; $j \in [0; n]$; ψ_i – вспомогательная функция в составе гамильтониана, специальным образом вводимая при помощи системы «сопряженных» уравнений; аргументы функций ψ_i и f_i для краткости записи опущены.

Решение задачи оптимального управления предусматривает проведение следующих действий.

1. Запись системы уравнений, описывающих объект, включая уравнение для функционала качества (4.1).

2. Составление гамильтониана $H(\psi; x; u; t)$ и системы вспомогательных «сопряженных» уравнений для определения функций ψ_i .

3. Подстановка найденных значений ψ_i в гамильтониан и определение из условия его максимума $\partial H / \partial u_i = 0$ значений управляющего воздействия $u_i(t)$.

4. Подстановка найденных значений $u_i(t)$ в исходное уравнение (4.1) и определение из него значений функционала качества $f(x; u; t)$ и оптимальной траектории $x_1; x_2 \dots x_n$.

Таким образом, при использовании принципа максимума Понтрягина задача нахождения способа управления объектом, *экстремизирующего* (по максимуму или по минимуму) значение функционала $f(x; u; t)$, сводится к более простой задаче определения функции $u(t)$; $t \in [t_1; t_2]$, которая обеспечивает максимум гамильтониана $H(\psi; x; u; t)$. Достоинством данного метода являются его математическая строгость и абсолютный характер полученного результата – в том смысле, что найденный экстремум соответствует действительно оптимальному (по критерию «максимум-

максимумом» или «минимум-минимумом») значению многофакторного функционала качества (4.1).

Пример решения задачи оптимального управления. Задачу оптимального управления объектом рассмотрим на упрощенном примере динамического распределения из центра ресурсов $V_1(t) \geq 0$ и $V_2(t) \geq 0$, которые в данном случае выполняют роль управляющих воздействий, между двумя корпорациями, находящимися в фазовых состояниях $K_1(t)$ и $K_2(t)$, когда (4.1) можно записать как

$$\frac{dK_1}{dt} = -\mu_1 K_1(t) + V_1(t); \quad \frac{dK_2}{dt} = -\mu_2 K_2(t) + V_2(t); \quad t \in [0; T]; \quad (4.4)$$

$$V_1(t) + V_2(t) \leq V_m; \quad V_{1;2}(t) \in U,$$

а функционал пути

$$J(K; V) = \int_0^T [\alpha_1 V_1(t) + \alpha_2 V_2(t)] dt - \beta_1 K_1(t) - \beta_2 K_2(t) \rightarrow MIN, \quad (4.5)$$

где $K_1(0) = K_{10}$; $K_2(0) = K_{20}$; параметры $\mu_{1;2}$; $\alpha_{1;2}$ и $\beta_{1;2}$ соответствуют характеристикам корпораций в рассматриваемой предметной области. Составим гамильтониан в общем виде:

$$H(K; V; \psi) = \psi_1(-\mu_1 K_1 + V_1) + \psi_2(-\mu_2 K_2 + V_2) - \alpha_1 V_1 - \alpha_2 V_2. \quad (4.6)$$

Теперь необходимо найти функции управления $V_1(t)$ и $V_2(t)$, при которых имеет место максимум $H(K; V; \psi)$ по множеству U . Поэтому из (4.6) можно удалить слагаемые $\mu_1 K_1$ и $\mu_2 K_2$, которые не входят в данное множество:

$$H_V(K; V; \psi) = \psi_1 V_1 + \psi_2 V_2 - \alpha_1 V_1 - \alpha_2 V_2 \quad (4.7)$$

и представить (4.7) как

$$H_V(V; \psi; t) = \gamma_1(t) V_1(t) + \gamma_2(t) V_2(t), \quad = \quad (4.8)$$

где $\gamma_1 = \psi_1 - \alpha_1$; $\gamma_2 = \psi_2 - \alpha_2$; $t \in [0; T]$. Таким образом, от задачи максимизации гамильтониана $H(K; V; \psi)$ в (4.6) мы перешли к задаче линейного программирования: найти максимум целевой функции

$$H_V = \gamma_1 V_1 + \gamma_2 V_2, \quad (4.9)$$

при условиях $V_1(t) \geq 0$; $V_2(t) \geq 0$; $V_1(t) + V_2(t) \leq V_m$, которые на плоскости $V_{1,2}(t) \in U$ задают треугольник OAB , показанный на рис.

4.1. Решение этой задачи известно: функция H_V достигает максимума в одной из вершин треугольника OAB или на одной из его граней, а направление роста H_V определяется вектором \vec{u} с координатами γ_1 ; γ_2 . Причем решение не является единственным:

- при $\gamma_1 < 0$; $\gamma_2 < 0$ (когда вектор \vec{u} лежит в третьей четверти плоскости V_1OV_2 на рис. 4.1) точка S_2 максимума H_V совпадает с точкой O на рис. 1;
- при $\gamma_1 = 0$; $\gamma_2 < 0$ максимум H_V достигается на ребре OB ;
- при $\gamma_1 < 0$; $\gamma_2 = 0$ максимум H_V достигается на ребре OA ;
- при $\gamma_1 > \gamma_2$; $\gamma_1 > 0$, по аналогии, максимум H_V имеет место в точке $B(V_m; 0)$;
- при $\gamma_1 < \gamma_2$; $\gamma_2 > 0$ максимум H_V имеет место в точке $A(0; V_m)$;
- при $\gamma_1 = \gamma_2$, когда вектор \vec{u} ортогонален ребру AB , максимум H_V достигается в любой точке данного ребра.

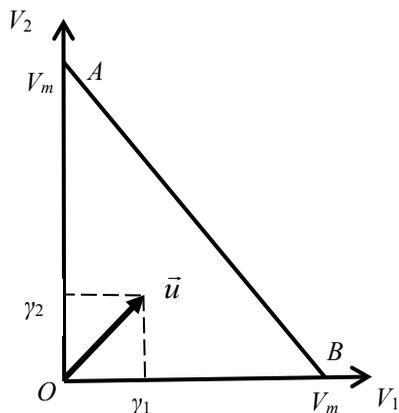


Рис. 4.1. К определению максимума целевой функции H_V

Частные выражения для оптимальных значений V_{10} и V_{20} имеют характер ограничений, аналогично K_{10} и K_{20} :

- $V_{10} = V_{20} = 0$ при $\gamma_1 < 0$; $\gamma_2 < 0$;
- $V_{10} = V_m$; $V_{20} = 0$ при $\gamma_1 - \gamma_2 > 0$; $\gamma_1 > 0$;
- $V_{10} = 0$; $V_{20} = V_m$ при $\gamma_1 - \gamma_2 \leq 0$; $\gamma_2 < 0$.

$$(4.10)$$

На этом поставленную задачу можно считать решенной, поскольку она сведена к известной математической задаче, где необходимо, согласно (4.9) с учетом (4.10), определить максимум H_V при γ_1 и γ_2 , соответствующих целевой функции центра управления, потенциалам корпораций и исходным условиям, в которых они находятся. Но трудности для ЛПР здесь фактически только начинаются, так как необходимо решить вопросы, связанные с определением начальных и промежуточных значений $\psi_{1,2}$; $\mu_{1,2}$; $\alpha_{1,2}$ и $\beta_{1,2}$, конкретизацией $K_1(t)$; $K_2(t)$ и т.д. Причем главным препятствием является невозможность формализовать поведение (динамику функционирования) ОУ, а также подсистем и элементов СУ, включая самих ЛПР – для которых нельзя разработать ММ в виде дифференциальных, разностных или интегральных уравнений, даже с применением современных компьютерных методов и средств вариационного исчисления, теории вероятности и математической статистики. Поэтому ценность (4.9)-(4.10) в практическом отношении невелика.

Изменения внешней среды и внутренних условий работы приводят к воздействию на ОУ и СУ, наряду с $u(t)$, различного рода непредсказуемых (стохастических, случайных по природе) возмущений. Эти возмущения (сигнальные, параметрические, структурные) представляют собой источник **неопределенности знаний** ЛПР о состоянии ОУ и СУ. Поэтому условия в задачах оптимального управления всегда являются в известной мере неопределенными (неточными, неадекватными, необъективными), а в ряде случаев – взаимно противоречивыми. В этой связи уместно вспомнить слова Н.Винера о том, что «приписывать таким неопределенным по самой своей сути величинам какую-то особую точность бесполезно... применение точных формул к этим слишком неточно опреде-

ляемым величинам есть не что иное, как обман и пустая трата времени».

Рефлекторные и нерелефлекторные СС. Принцип максимума Понтрягина сыграл выдающуюся роль в создании и развитии математической теории оптимального управления и оптимизационной идеологии, которые «глубоко проникли во все исследования конкретного характера и конструкторские разработки, и, можно сказать без преувеличения, что язык теории оптимальных процессов сделался общим языком современной теории управления» [81]. Вместе с тем Н.Н. Моисеев, с которым одному из авторов книги посчастливилось обсуждать эти вопросы, на основе представления (4.1) в виде

$$\frac{dx}{dt} = f(x; u; \xi) = Ax + u + \xi; \quad (4.11)$$

где ξ – случайное возмущение, также «управляющее» объектом, предложил разделить их на две группы: иерархические сложные системы (в наших обозначениях СС) технического типа, которые он назвал **рефлекторными**, и **нерелефлекторные** системы – организационно-технические, социально-экономические, эколого-эргономические, медицинские и т.п. [81]. Это разделение, как показал дальнейший опыт развития теории и практики создания СУ, также имеет глубокий смысл: поскольку реакция рефлекторных СС на возмущение ξ является однозначной и изучение процессов управления ими сводится к задачам оптимизации на основе принципа максимума Понтрягина без учета особенностей поведения звеньев (подсистем, элементов), входящих в состав СС первой группы.

Изучение нерелефлекторных СС второй группы, напротив, требует от ЛПР введения гипотез относительно поведения указанных подсистем и элементов: поскольку они способны самостоятельно максимизировать свои функционалы качества $f(x; u; t; \xi)$ при воздействии возмущений ξ (например, вследствие делегирования полномочий «сверху – вниз» при децентрализованном управлении СС), что ведет к конфликтным ситуациям и невозможности использовать принцип максимума Понтрягина при управлении ими. Иерархичность и нерелефлекторность возникают в СС, относящихся

ко второй группе, естественным путем – по мере роста, развития и усложнения любой простой системы.

Если управление звеньями в иерархической СС осуществляется по информации о состоянии этих звеньев, то есть $u_i(t) = u_i(x_i; t)$, можно записать управляющее воздействие на СС в матричном виде как

$$U(t) = B(t) u(t); \quad (4.12)$$

где $B(t)$ – блочно-диагональная матрица, подматрицы которой соответствуют подсистемам СС и подматрицам, входящим в состав $u(t)$. В рефлекторных СС матрица $B(t)$ распадается (декомпозируется) на подматрицы $b_i(t)$, выбор для которых оптимальных управлений $u_i(t)$ ведет к оптимизации процесса управления СС в целом. В неректорных СС матрица $B(t)$ не является блочно-диагональной, поэтому однозначно предсказать реакцию СС на ξ и аналогичным образом реализовать процесс управления не удастся.

Однако, рассматривая $f(x; u; t; \xi)$ как цель управления СС, о поиске минимального значения функционала пути к этой цели $J(x; u; t; \xi)$ можно говорить всегда. При этом, согласно [48], случайное возмущение ξ можно рассматривать как фактор неопределенности знаний ЛПР, которая осложняет его действия – ввиду отсутствия необходимой информации, присутствия помех, наличия нескольких целей и неопределенности намерений самого ЛПР, а также противодействия ему со стороны конкурента или злоумышленника, отсутствии согласованного взаимодействия с партнерами. Напомним, что знания ЛПР могут быть как *верифицированными*, так и *аксиологическими* – в последнем случае наибольший интерес при организации управления неректорными СС представляют *гипотезы поведения* звеньев, которые выдвигают ЛПР в процессе принятия решений.

В зависимости от характера данных гипотез формулируются соответствующие им математические задачи, важнейшими из которых являются:

- *игра с противоположными интересами* – антагонистическая игра Дж. фон Неймана [84] (например, в условиях конкуренции и рыночной борьбы), где выигрыш одного из игроков означает

проигрыш другого, обмен информацией между ними невозможен и т.д.;

- **игра с противоположными интересами** Ю.Б. Гермейера [25] – возникающая в условиях партнерства и сотрудничества ЛПР при достижении общих целей, когда обмен информацией не только возможен, но и целесообразен, действия ЛПР имеют согласованный системный характер и т.п.

Отметим, что в настоящее время нерелефторные СС второй группы (корпорации, компании, фирмы) часто относят к системам **холонического типа**, где структурными единицами являются **холоны** (например, предприятия, цеха, филиалы корпорации, то есть подсистемы и звенья СС), которые обладают свойством самоорганизации и наделены правом самостоятельно принимать управленческие решения, а элементы таких СС (в том числе ЛПР и другие сотрудники) именуют **акторами**. Холоны, относящиеся к разным уровням (стратам) иерархической СС, взаимодействуют с холонами верхнего и нижнего уровней, они способны принимать, обрабатывать и передавать дальше информационные и материальные потоки, потоки директивной (управленческой) информации, разного рода воздействия и т.д. Развитие методов и средств управления холоническими СС в настоящее время является актуальной научно-технологической задачей.

В этой связи Н.Н. Моисеевым применительно к нерелефторным СС были выдвинуты две идеи: первая идея о необходимости при разработке СУ для иерархических и многокритериальных СС каждый раз конкретизировать понятие оптимальности управления, чтобы «вложить разумный смысл в понятие оптимальности». Вторая идея сводится к предложению использовать для управления нерелефторными СС иерархические СУ.

Иерархические СУ для нерелефторных СС. Целью введения иерархии в СУ является распределение между ее звеньями функций обработки информации и принятие рациональных решений по выбору управлений $u_i(t)$. При этом, с одной стороны, существует угроза сужения множества допустимых управлений $u \in U$ ввиду дополнительных ограничений и усложнения технологии управления, с другой стороны – приходится считаться с тем, что «без разделения функций принятия решений система вообще не сможет

функционировать» [81]. Важным фактором при этом является связь между критерием качества работы СУ и требованиями к алгоритмам обработки информации: поскольку рост объемов информации ведет к необходимости «распараллелить» процессы ее сбора и обработки, что требует создания в составе СУ самостоятельных информационных подсистем, а вследствие этого – децентрализации всего процесса принятия управленческих решений.

Поскольку при децентрализованном принятии решений звеньям СУ необходимы существенно меньшие объемы информации, которые можно обработать быстрее и качественнее, в условиях меньшей неопределенности, это способствует повышению эффективности управления – но одновременно означает формирование в СУ иерархической структуры. Децентрализация управления становится источником новой неопределенности: связанной с тем, что у звеньев и подсистем СУ появляются собственные цели, потенциально нетождественные друг другу и интересам системы в целом, что характерно для нерелефторных СС. Поэтому возникает необходимость говорить и об оптимальной мере централизации и децентрализации при управлении СС, и об оптимальном и регулируемом распределении функций управления между звеньями и уровнями СУ.

Математическая задача в рассматриваемой ситуации соответствует игре Гермейера с противоположными интересами, где центр может управлять игроками как при помощи обязательных для исполнения команд, так и более «демократичными» способами: распределяя между ними ресурсы по своему усмотрению, используя систему штрафов и поощрений, ограничивая и расширяя права отдельных звеньев и т.д.

Примером *двухступенчатой иерархической СУ* является система «центр» – «производители» (далее без кавычек), где число последних равно N [36-38]. Роль центра может выполнять корпорация, роль отдельного производителя: n -ое предприятие, входящее в ее состав, где n [1; N]. Здесь налицо два игрока: центр и производитель, которые, во-первых, преследуют разные цели (хотя и неантагонистического характера), а во-вторых, неравноправны в правилах игры: поскольку центр имеет право первого хода. Его задача: распределять ресурсы, вводить штрафы и поощрения, ограничивать активность производителей путем введения квот и тарифов, регу-

лирования цен и т.д. таким образом, чтобы максимизировать общий доход

$$Q_N = Q[P_1; P_2; \dots; P_n \dots P_N; w_1(P_1); w_2(P_2); \dots; w_n(P_n) \dots w_N(P_N)], \quad (4.13)$$

где P_n и $w_n(P_n)$ – соответственно, объем реализованной продукции и поощрение (штраф) для n -го производителя.

В начальный момент игры центр сообщает производителям ее правила: плановые значения $P_n(0)$; условия получения $w_n(P_n)$ и т.д., после чего каждый производитель предпринимает усилия, чтобы максимизировать свой предполагаемый доход: $Q_n = Q(P_n) = MAX$, а центр регулирует их работу при помощи $w_n(P_n)$ таким образом, чтобы выполнилось условие (4.13). В итоге реальный доход n -го производителя равняется

$$Q_n = Q[P_n; w_n(P_n)], \quad (4.14)$$

и сложится ли по максимуму общий доход корпорации из доходов N предприятий, соответствующих (14), то есть выполнится ли условие

$$Q_N = \sum_{n=1}^N Q_n = MAX, \quad (4.15)$$

зависит как от активности производителей, так и от эффективности управления ими со стороны центра.

Поскольку в игре Гермейера конкуренция между производителями отсутствует, каждый из них заинтересован в том, чтобы центр успешно справился со своей задачей и условие (4.15) было достигнуто. Поэтому информация о темпах и результатах их текущей деятельности, уходящая «наверх», должна быть максимально полной и достоверной, а возможные штрафы не следует воспринимать как наказание. На всех этапах функционирования подсистем и элементов рассматриваемой СС: от составления и согласования планов каждого из N производителей с центром до совместных действий по сбыту продукции на рынке, их деятельность должна быть скоординирована и оптимизирована в конкретном «разумном смысле» – чтобы обеспечить выполнение (4.15). В теоретическом отношении при этом более уместно, на наш взгляд, говорить о **квази-оптимальном управлении** СС.

Идею квазиоптимального управления иллюстрирует рис. 4.2, где в графическом виде показана эффективность $\mathcal{E}(X)$ функционирования СС в зависимости от обобщенного аргумента X (функционала качества, многомерного показателя, многофакторного критерия и т.п.), определяющего эту эффективность. Теория оптимального управления ориентирует ЛПР на определение математическими методами максимально-возможного значения эффективности \mathcal{E}_{opt} , соответствующего на рис. 4.2 варианту « O » (далее без кавычек) и значению аргумента X_{opt} – если такой вариант существует в природе, что оговаривается специально. На отведенном ЛПР для принятия решения отрезке времени $t \in [0; T]$, таким образом, он должен не только определить значения \mathcal{E}_{opt} и X_{opt} , но и доказать, что найденный им вариант O действительно является оптимальным и лучшего варианта в природе быть не может.

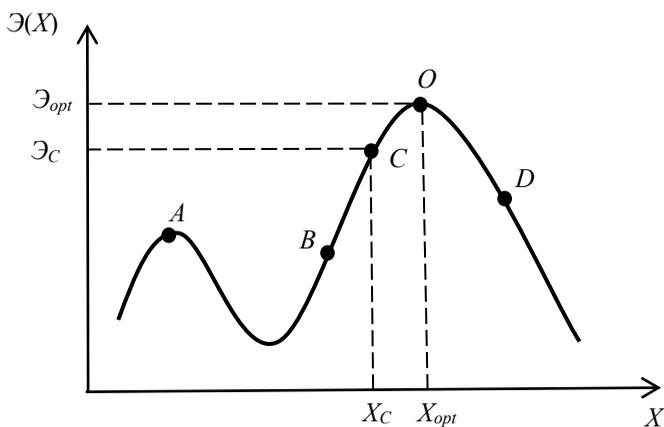


Рис. 4.2. К выбору квазиоптимального варианта принимаемого решения

С точки зрения практики это выходит за рамки здравого смысла, поскольку реальными и доступными для ЛПР являются, например, лишь варианты A ; B ; C и D , показанные на рис. 4.2, так что выбор наилучшего возможен только из их числа. Если $\mathcal{E}_C > \mathcal{E}_{A;B;D}$, то значения \mathcal{E}_C и X_C для него окажутся наилучшими и вариант управления C для него будет оптимальным – хотя субъективность данного выбора очевидна, поскольку объективно лучшим является вариант O .

Однако переход от оптимального управления СС второй группы к квазиоптимальному управлению требует пересмотра принципов разработки СУ и отказа не только от фундаментального принципа максимума Понтрягина, но и от идеологии моделирования СС на основе вариационного исчисления и теории дифференциальных уравнений в целом. В этой связи вновь обратимся к [81], где автор почти полвека тому назад рассматривал два перспективных пути проведения исследований:

- изучение *простейших модельных задач* с целью понимания (интуитивного, а затем все более осознанного) принципов, которые должны составить фундамент *информационной теории иерархических систем*, включающей широкий круг конкретных вопросов: от моделирования структуры штрафов и поощрений $w_n(P_n)$; $n [1; N]$; $N \gg 1$, до аппроксимации нерелекторных СС релекторными;
- использование *имитационных систем* (simulation systems), каждая из которых является наукоемким программным продуктом, который состоит из *системы моделей*, описывающих эволюцию изучаемого объекта; *системы процедур*, позволяющей объединить эти модели с ЛПР в диалоговом режиме; и специального *математического обеспечения* (в виде системы компьютерных языков и алгоритмов), необходимого для ее функционирования.

Представляют интерес три частных замечания [81]: первое касается «гипотез поведения» звеньев СУ, моделируемых с помощью «черного ящика», где функция выхода определяется не только входным воздействием, как это принято школой Форрестера-Медоуза, но и внутренним состоянием рассматриваемой СС (по нынешней терминологии это актуальная модель «серого ящика»).

Второе замечание относится к процедуре выработки компромиссных коллективных решений в игре N производителей: решение считается эффективным, если не существует никакого другого решения, которое было бы «лучшим» для любого из игроков, и устойчивым (равновесным), если отступление от него невыгодно, прежде всего, самому игроку-отступнику. В теории игр известна теорема о том, что в принципе «состояния равновесия существуют и они эффективны», однако реальность все же такова, что эффективные решения зачастую неустойчивы, а устойчивые решения неэффективны.

Третье замечание представляется также достаточно важным: особенности работы СС в условиях неопределенности таковы, что «в условиях неточного определения цели ... необходимо и оптимизацию производить с ошибкой». Другими словами, ЛПР не стоит напрасно терять ресурсы и время, максимизируя объем, достоверность и точность доступной ему информации, а также выбирая абсолютно лучший вариант предпринимаемых действий – если есть возможность вовремя ограничиться квазиоптимальным решением, которое впоследствии можно будет скорректировать и улучшить.

Применение метода СИМ по МДМ в иерархических СУ. Рассмотрим пример использования СИМ-моделей разного уровня в иерархической СУ дискретно-непрерывным технологическим процессом производства и обработки алюминиевых сплавов [36], актуальный применительно к бизнес-процессам в социально-экономических, организационно-технических и других современных СС. Надсистемой для **СУ на уровне предприятия** является корпоративная **отраслевая СУ**, в составе которой функционирует **моделирующий центр**, предназначенный для поддержки действий топ-менеджмента – ЛПР высшего уровня, возглавляющих процессы планирования бизнеса, стратегического и оперативного управления всеми N производителями в рамках игры с противоположными интересами Гермейера.

Будем исходить из того, что отраслевая СУ и моделирующий центр с его парком мощных ЭВМ и развернутым математическим и программным обеспечением все же не в состоянии самостоятельно обеспечить управление каждым из N игроков. Во многом это обусловлено тем, что производительность и экономические показатели каждого экземпляра рассматриваемого процесса на любом n [1; M] предприятии определяются целым рядом стохастических факторов, формирующих индивидуальные возмущения ξ_n :

- несвоевременное поступление сырья и ресурсов от поставщиков;
- случайные отклонения от технологии производства;
- случайные остановки и простои оборудования;
- случаи выпуска бракованной продукции;
- наличие внеочередных директивных заказов;
- ошибки в работе персонала и т.д.

Кроме того, уровень сложности моделируемой СС определяют следующие присущие ей свойства:

- высокие материалоемкость и стоимость предметов и продуктов труда;
- множество режимов работы при малом разнообразии оборудования;
- большое количество маршрутов металлопотока;
- наличие накопителей для полуфабрикатов;
- неравномерная загрузка оборудования;
- зависимость производства от внешней среды;
- большая номенклатура выпускаемых изделий;
- мелкосерийность выпускаемой продукции;
- необходимость корректировки планов запуска изделий в пределах установленной номенклатуры ввиду влияния возмущений и т.п.

Цель СУ на уровне предприятия, основным звеном которой является СИМ-модель рассматриваемого n -го экземпляра процесса – повышение эффективности планирования и регулирования производства. Назначение СИМ-модели – оказание информационной поддержки ЛПР в условиях неопределенности ситуации и случайных факторов, воздействующих на СС. Для обеспечения эффективности и адекватности действий ЛПР предусмотрено их взаимодействие с СИМ-моделью и другими элементами автоматизированной СУ в диалоговом режиме. Это позволяет ЛПР оперативно воздействовать на СС, используя, наряду с объективной информацией, полученной от СИМ-модели и других элементов СУ, собственный производственный опыт и субъективные аксиологические знания.

Рассмотрим в качестве примера функцию планирования производства [36]. Основой здесь является *директивный план*, сформированный в надсистеме СУ или в собственной подсистеме планирования (при делегировании ей соответствующих полномочий) при помощи экономико-математических методов, который представляет собой план выпуска изделий (портфель заказов) предприятия на конкретный плановый период (год, квартал, месяц).

Получив директивный план, СУ предприятия производит агрегированную оценку возможности его выполнения по всем многочисленным номенклатурным позициям. В случае отрицательных отклонений (при возможности невыполнения плана по отдельным

позициям – когда вероятность этого оказалась ниже установленно-го критерия), ЛПП обращается к СИМ-модели и с ее помощью ищет выход из положения: руководствуясь статистическими данными за предыдущее время и собственным опытом, производит перераспределение позиций директивного плана и вновь производит его оценку на СИМ-модели – до тех пор, пока не остановится на наиболее предпочтительном для себя варианте. О найденном квазиоптимальном решении он информирует центр, за которым в данной игре сохраняется право окончательного решения.

После согласования предложенного решения центром (при отсутствии центра необходима договоренность с другими $N - 1$ игроками) на предприятии разрабатывается план производства (также при помощи СИМ-модели) и выбирается наиболее предпочтительный его вариант, где устанавливаются последовательность запуска изделий с учетом размеров выпускаемых партий, приоритетов номенклатурных позиций, наличия директивных указаний, состояния оборудования, графика поступления ресурсов, а также фактических результатов работы за прошедший период. Для обеспечения указанных функций при помощи СИМ-модели ЛПП решает следующие задачи:

- генерация статистического плана выпуска изделий;
- оценка фактического состояния производства по выполнению текущего плана;
- формирование вариантов плана запуска на планируемый период;
- прогнозирование поведения объекта управления в плановом периоде;
- определение фактического выпуска изделий в плановом периоде;
- определение загрузки оборудования по отдельным позициям плана;
- определение вероятности выполнения плана по номенклатурным позициям для отдельных вариантов плана;
- определение ритмичности выпуска изделий в плановом периоде;
- выбор наиболее предпочтительного варианта плана запуска изделий с точки зрения выбранных критериев эффективности.

Решение комплекса данных задач в диалоговом режиме между ЛПП и ЭВМ обеспечивает достижение конечной цели – выбор наиболее предпочтительного (квазиоптимального) варианта запуска изделий в производство.

Указанный процесс планирования соответствует схеме на рис. 4.2 – где в общем случае также, во-первых, необходимо при помощи метода сценариев и метода функционально-стоимостного анализа разработать для рассматриваемой СС конкретные варианты *A*; *B*; *C* и *D*, о которых идет речь. Во-вторых, с применением СИМ-модели следует выбрать из них наилучший в данный момент времени вариант – с тем, чтобы СУ могла соответствующим образом скорректировать состояние ОУ. Сопровождающие квазиоптимальное управление вычислительные задачи (в отличие от задач оптимального управления) не отличаются повышенной сложностью, их решение возможно в автоматизированном режиме, с применением стандартных пакетов программ. «Центр тяжести» усилий ЛПР из области прикладной математики таким образом переносится в сферу профессиональных интересов и компетенций – что самым положительным образом сказывается на эффективности их труда.

4.2. Особенности управления знаниями

Форма представления знаний оказывает существенное влияние на характеристики ИИС, включая ЭС, где БЗ являются компьютерными моделями человеческих знаний. Однако, поскольку все знания, используемые в процессе решения сложных задач, смоделировать практически невозможно, в ИИС для поддержки управленческих действий ЛПР требуется разделить знания на те, которые предназначены для компьютерной обработки, и те, которые непосредственно используются ЛПР.

Вопросы приобретения и формализации знаний экспертов-профессионалов вызывают сегодня не только академический интерес, имеющий чисто познавательный гносеологический аспект. В рыночной экономике корпоративные знания являются одним из важнейших ресурсов компании – причем этот ресурс способен самым существенным образом повлиять на такие ее важнейшие качества, как конкурентоспособность и инвестиционная привлекательность. Целенаправленное использование и усовершенствование знаний высвобождает потенциалы экономии и роста, которые не могут быть реализованы с помощью традиционных концепций реорганизации и модернизации бизнеса. Никто из ЛПР сегодня с этим не спорит, и вопрос только в том, как лучше воспользоваться

корпоративными знаниями, так как большинство из них скрыто «в головах» у сотрудников компании.

Чтобы наглядно проиллюстрировать значимость задачи управления знаниями, напомним притчу, которая в разных вариантах часто встречается в зарубежной и отечественной литературе, посвященной проблемам *менеджмента знаниями* (knowledge management) [18]. В компании, производящей моющие средства, одна из автоматических линий неожиданно сначала стала функционировать неверно, а вскоре и вовсе остановилась. Тщательная диагностика системы и всех ее элементов ни к чему не привели – линия простаивает, фирма теряет миллионы долларов. В это время бывший ее служащий, а ныне пенсионер по фамилии Смит, приходит проведать старых друзей и узнает о возникшей проблеме. Он берет молоток, проходит вдоль линии и ударяет по одной из труб – после чего линия начинает функционировать в нормальном режиме. Смит не может логически объяснить правильность своих действий, но он говорит, что сорок лет назад был аналогичный случай и неполадку устранили именно таким способом. Менеджер спрашивает, сколько стоит его работа и получает ответ: «10 тысяч долларов, из расчета удар молотком – 1 доллар, знание, куда и с какой силой ударить – 9999 долларов».

Управление знаниями далеко не новая тема для стран с развитой рыночной экономикой: принципы управления знаниями были определены зарубежными специалистами к началу девяностых годов прошлого века. Передовые концепции менеджмента – «теория Z» У. Оучи [116] и модель «7-С» Т. Питерса и Р. Уотермена [89, 107] подчеркивали организационные отличия лучших компаний, отвергнувших механистические методы школы управления, авторитет которой к тому времени резко упал. Классиками данной теории можно назвать также японских исследователей I. Nonaka и H. Takeuchi .

Главной отличительной особенностью теории «Z» является обоснование коллективистских принципов *мотивации работников*, которая, согласно данной теории, должна исходить из ценностей «производственного клана», то есть предприятия как одной большой семьи, одного рода. Эти ценности требуют развития у всех сотрудников, в том числе ЛПР – с помощью соответствующей организации и стимулирования – отношений взаимного доверия,

солидарности, преданности коллективу и общим целям, удовлетворенности трудом и пребыванием на предприятии (в корпорации), что очевидным образом способствует росту производительности их совместного труда.

Название модели «7-С» (другой вариант: «счастливый атом») обусловлено тем, что ее авторы предложили рассматривать организацию как единство семи основных переменных: структуры (Structure), стратегии (Strategy), системы и процедур управления (Systems), совместных, то есть разделяемых всеми работниками, ценностных установок (Shared Values), совокупности приобретенных навыков, умения (skills), стиля управления (Style) и состава работников, то есть системы кадров (Staff).

В России данная тема сегодня активно обсуждается специалистами всех уровней управления, при этом целью управления знаниями считается необходимость объединить знания, накопленные корпорацией, со знаниями заказчика и использовать их для решения задач управления бизнесом.

В вопросах управления знаниями можно выделить две равнозначных сферы исследований:

- во-первых, это вопросы формирования архитектуры предприятия, ориентированной на знания. Здесь первоочередную роль играет процесс организации быстрого доступа к необходимым знаниям и повышение мотивации сотрудников к «опубликованию» собственных навыков и опыта, что позволяет значительно повысить качество ежедневных деловых процессов;
- во-вторых, это вопросы методологической организации выявленных знаний и закономерностей с целью их дальнейшей автоматизации: для того, чтобы «поместить» знания в ЭВМ, их нужно представить в виде определенных структур данных, соответствующих выбранной среде разработки ИИС (в настоящем разделе рассматриваются вопросы преимущественно из данного второго направления).

Культура любой корпорации, представляющей собой СС организационно-технического типа (предприятие, компания, фирма), концентрированно выражается в способе ее самоорганизации: чем совершеннее этот способ, тем больше шансов у такой СС на выживание. Одним из возможных взглядов на нее является представление в виде системы бизнес-процессов (БП) с последующим реше-

нием проблемы выживания СС путем развития корпоративных ИС. Класс ИС, ориентированных на управление БП, будем относить к АСУ БП (см. в разделе 2.2). Граница между ИС, ориентированными и не ориентированными на управление БП, определяется их функциональной возможностью поддерживать *реинжиниринг* БП и настраиваться на созданные проектировщиком организации новые модели БП. Отметим, что корпоративные ИС, обладающие закрытой архитектурой и ориентированные на решение функциональных, а не процессных задач, не могут быть отнесены к АСУ БП.

Кратко рассмотрим принципы управления СС организационно-технического (по терминологии [18] *холонического*) типа – социально-экономическими, экологическими, специального назначения и т.п., которые представляют для нас наибольший практический интерес. В соответствии с определением, СС холонического типа состоит из целостностей, которые включают в себя другие целостности и одновременно являются частями другого целого – на более высоком системном уровне. Это означает, что СС по принципу декомпозиции может быть разделена и детализирована вплоть до своих физически неделимых элементов – примерами которых являются отдельные ЛПР.

Однако для решения задач, связанных с работой *систем управления*, этого во многих случаях не нужно и масштаб детализации СС выбирается в соответствии с условием разумной достаточности в рамках достижения конкретных целей управления: подсистемами и элементами, или СС в целом. Например, если при исследовании и моделировании БП корпорации в целях управления ими необходима детализация до уровня отделов, а индивидуальные поведение ЛПР, составляющих эти отделы, учитывать не требуется – то эти отделы принимаются за элементы исследуемой холонической СС и осуществляется моделирование их поведения как некоторых неделимых целостностей.

Структура СС зависит от целей ее функционирования и может быть представлена тремя основными структурными «срезами» [19]:

- *функциональная структура*, которая представляет собой совокупность бизнес-процессов (производственно-технологических функций), обеспечивающих достижение целей СС;

- **организационная структура** – упорядоченное множество иерархических отношений между компонентами (подсистемами и элементами), необходимое для реализации функций СС;
- **техническая структура**, которую составляют множество технических элементов СС (компьютерные системы и сети, производственное оборудование, измерительные приборы, средства связи, здания, инженерные сооружения и т.п.).

В процессе своего существования структура СС претерпевает непрерывные изменения (как незначительные, так и в более редких случаях кардинальные), что обусловлено изменением целей и ведет к обновлению всех ее компонентов. Последнее связано с тем, что попытки реализовать новую функциональную структуру в рамках старой организации и (или) при устаревшем техническом обеспечении производства обречены на провал – поскольку приводят к неработоспособному или крайне неэффективному состоянию СС, в конце концов, к ее гибели (краху, банкротству).

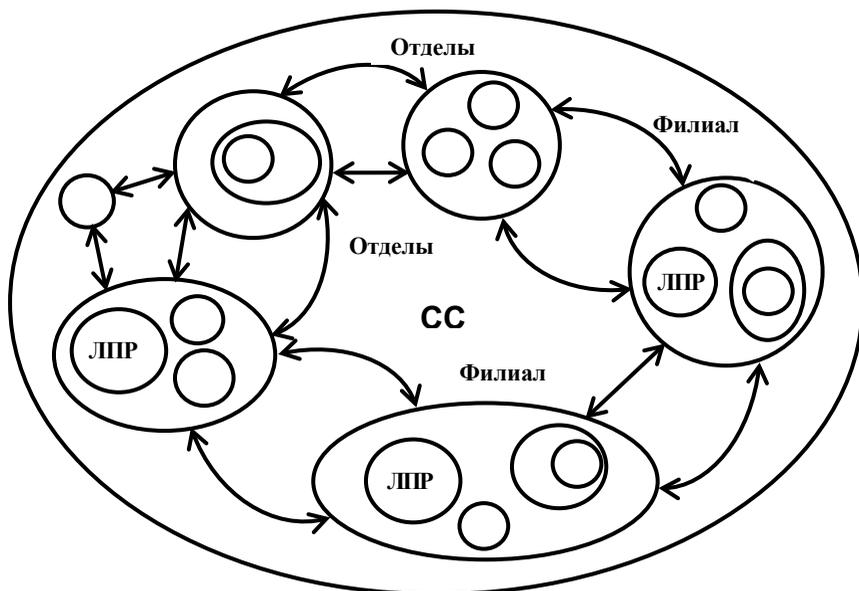


Рис. 4.3. Состав и структура холонической СС

Схему холонической структуры СС в символическом виде демонстрирует рис. 4.3, где овалами показаны **холоны**, соответству-

ющие структурные единицам СС (ЛПР и рядовые сотрудники, отделы, цехи, участки, филиалы рассматриваемой корпорации), с точки зрения системного анализа все они – системы, подсистемы и элементы СС различного назначения (технические, экономические, социальные, информационные и т.п.), относящиеся к разным уровням (*стратам*) ее иерархии. Каждый такой холон взаимодействует с соответствующими ему холонами верхнего и нижнего уровней, то есть может принимать, обрабатывать и передавать дальше информационные и материальные потоки, потоки директивной (управленческой) информации, разного рода воздействия и пр. Примеры взаимодействия между собой отдельных холонов показаны на рис. 4.3 стрелками.

Понятие *организации* применительно к данной СС также имеет свою специфику. Если в технических СС каузального типа свойства элементов и способ их соединения между собой однозначно определяют свойства СС в целом, то в холонических СС ситуация является существенно более сложной и организация СС, строго говоря, это характеристика, не тождественная ее структуре [21]. В то же время свойства СС зависят от ее организации, которая предполагает наличие целенаправленного поведения и т.д. – то есть общие закономерности поведения систем остаются теми же.

В этой связи ЛПР под корпорацией в ряде случаев имеют в виду *частично самоуправляемую* СС, которую характеризуют:

- способность СС выбирать направления своей деятельности, причем ответственность за этот выбор может быть распределена между ее элементами на основании выполняемых функций, местоположения и других признаков;
- коммуникация – в СС должны быть определены поведение и порядок взаимодействия между собой подсистем и элементов;
- выбор решений – участники этого важнейшего для самоуправляемой СС процесса должны распределить между собой соответствующие задачи и направления деятельности.

В теории управления выделяют два основных типа организационных СС: механистический (инжиниринговый) и органический [15]. *Механистический подход* сегодня распространен настолько широко, что в зарубежных вузах существует научная специализация: «*профессор инжиниринга*». Суть данного подхода заключается в том, что любая организация (социальная, экономическая,

экологическая, военная и т.п.), включая области медицины, искусства и религии, рассматривается как СС типа машины или механизма (примером здесь могут быть часы – как огромные башенные, так и в виде микрочипа), которая должна работать устойчиво и точно, по заранее установленному расписанию. Существуют и доступны для ЛПР способы текущего контроля состояния СС; оценки эффективности ее функционирования (через измеряемые показатели производительности, количества и качества продукции, услуг); а также «рычаги» воздействия на СС в желаемом направлении (с помощью положительных обратных связей – для поддержания позитивных тенденций развития, с помощью отрицательных обратных связей – для ослабления негативных явлений).

Если в работе СС наблюдаются сбои, с ней следует поступать, как поступают с часами: нужно остановить механизм, почистить шестеренки, подтянуть или заменить пружины, смазать специальным маслом, проверить по метроному и запустить вновь. Сигналом к началу решительных действий являются типовые *признаки-симптомы* – в причинах и природе которых опытный ЛПР досконально разбираться не будет. Даже если у него самого заболит голова, например, или желудок, он может принять соответствующее лекарство без консультации с врачом, поскольку неоднократно это делал и прогнозирует последствия своих действий: с вероятностью 0,90...0,95 боль должна пройти. Правда, с вероятностью 0,05...0,10 результаты такого самолечения могут потребовать вызова к нему неотложной медицинской помощи...

Специалисты в области социологии, медицины, экономики, экологии, эргономики, педагогики, психологии и т.д. порой пребывают в шоке от действий ЛПР, владеющих методологией инжиниринга – но практика показывает, что в типовых ситуациях (которым как раз и соответствуют вероятности 0,90...0,95) их простые, дешевые, субъективные, непроверенные теорией и т.п. предположения и предложения дают позитивный эффект. Хотя редкие сложные случаи, разумеется, требуют участия высококвалифицированных ЛПР с узкой профессиональной специализацией.

Технология выполнения работ в СС механистического типа в хорошем смысле рутинна: персонал выполняет повторяющиеся действия, доведенные до автоматизма операции, использует нормативные документы высокого уровня стандартизации и унифика-

ции – воздействие которых распространяется не только на выпускаемую продукцию, технологии, ресурсную базу и оборудование, но и на поведение людей – как ЛПР, так и рядовых исполнителей.

Управление организацией данного типа имеет в виду:

- четко определенные стандартные задачи;
- узкую специализацию выполняемых работ;
- централизованную структуру и строгую иерархию полномочий;
- преобладание в структуре СУ вертикальных связей;
- обезличенность человеческих отношений ввиду формализации рабочих правил и процедур;
- власть и влияние на ситуацию в СС в строгой зависимости от положения ЛПР в иерархии СУ;
- естественное сопротивление инновациям и изменениям в СС;
- строгую систему учета, контроля и проверки исполнения принятых решений.

В условиях рынка такие организации часто именуют бюрократическими со всеми вытекающими отсюда последствиями. Вместе с тем опыт показывает, что даже в странах со зрелой рыночной экономикой принципы механистического управления СС оправдывают себя – причем это касается не только кризисных периодов, но и рывков революционного развития при внедрении инноваций. Резервы СС механистического типа связаны с экономией рабочего времени и ресурсов; высокой производительностью и гарантиями качества выполняемых работ; выгодами от специализации труда; четким распределением функций и полномочий; рационализацией действий исполнителей на уровне операций – другими словами, с присущей им в целом высокой степенью организованности, надежности и порядка.

Вместе с тем такие СС неспособны постоянно приспосабливаться к изменяющимся условиям внешней среды; регулировать объем и качество своих предложений в соответствии с рыночным спросом; в кратчайшие сроки обновлять номенклатуру изделий; реагировать на действия конкурентов – поэтому их преимущества дают экономический эффект лишь в случае стабильного бизнес-окружения, при низком уровне неопределенности условий работы.

Органическая организация СС, напротив, специально предназначена для работы в условиях нестабильного бизнес-окружения, ей присущи эластичность, большая приспособляемость (адапта-

ция) к быстрым и частным внешним изменениям любого рода, а также следующие характерные признаки:

- универсальная гибкая структура, обеспечивающая динамику решаемых проблем;
- создание временных (в том числе виртуальных) трудовых коллективов (рабочих групп), нацеленных на решение проблем;
- децентрализация полномочий и ответственности;
- преобладание в структуре СУ горизонтальных связей;
- минимально-необходимое использование формализованных правил работы (инструкций) и процедур;
- власть и влияние на ситуацию в СС в зависимости от знаний и практического опыта, а не положения ЛПР в иерархии СУ;
- объективная готовность к инновационным изменениям;
- участие каждого сотрудника (элемента СС) в решении как групповых, так и общих задач организации;
- самоконтроль и контроль сотрудников со стороны коллег.

Органическое описание приводит к представлению СС в виде *живого организма* – к которому применимы принципы и законы сохранения и развития биологических СС, жизненный цикл которого включает все характерные стадии: от рождения до старения и обновления (или гибели). Чтобы приспособиться, выжить и устойчиво существовать в динамично изменяющейся внешней среде, СС «запускает» процессы саморегуляции, саморазвития, самоорганизации и самоуправления, позволяющие ей поддерживать внутреннее соответствие между подсистемами и элементами, обеспечивающие устойчивость выполнения всех основных функций, необходимых для решения поставленных задач и достижения намеченных целей. Важно, что органическая организация СС определяет наличие у нее холонической структуры и, соответственно, способная динамически изменяться холоническая структура не только не противоречит этому, но напротив – предполагает органический тип организации СС.

4.3. Принципы принятия решений в условиях неопределенности знаний

Рассмотрим более подробно особенности принятия решений в условиях неопределенности – задачи такого рода в теории управления именуются плохо поставленными (некорректными) и встре-

чаются на практике в самых разных предметных областях. Одним из существенных преимуществ ЭС над другими ИС как раз и является обладание методами, позволяющими «быть точными в отношении неточностей», подсказывать ЛПР выход из самых затруднительных ситуаций.

Вернемся к понятиям и определениям, введенным в разделе 2.7. Сегодня существуют два подхода к понятию «неопределенность». Основой первого из них можно считать знаменитое рассуждение Дж. Кейнса: «Под неопределенностью мы не имеем научного основания, мы просто не знаем. Однако потребность действовать и принимать решения заставляет нас, как практических людей, игнорировать этот неудобный факт и вести себя так, как если бы мы имели хороший способ вычисления предполагаемых вариантов (их преимуществ или недостатков) – каждый из которых должен быть умножен на соответствующие ожидаемые вероятности – которые только и ждут, чтобы мы их просуммировали». На данном фундаменте, как уже было сказано, в настоящее время успешно развиваются теории риска и ожидаемой полезности – объективной Дж. фон-Неймана – О. Моргенштерна [84] и субъективной Л. Сэвиджа [117].

Второй подход демонстрирует постоянное расширение и уточнение представлений ЛПР о спектре значений понятия «неопределенность», включающем целый ряд так называемых НЕ-факторов: недетерминированность, некаузальность (нежесткость), незнание, неизвестность, неполнота, неоднозначность, нечеткость, непредсказуемость и т.д. [12]. Источники неопределенности условно делят на четыре основные группы:

- факторы, относимые к объекту, который взаимодействует как с объектами внешней среды, так и с другими субъектами – обладающими собственными знаниями и знаниями в виде БЗ;
- факторы, относимые непосредственно к внешней среде – которыми обусловлена неопределенность ее воздействия на связанные с ней объекты;
- факторы, порождаемые неопределенностью, нечеткостью мышления и знаний ЛПР – это субъективная неопределенность, проявляющаяся при взаимодействии ЛПР с внешней средой;

– факторы, обусловленные неопределенностью, нечеткостью и противоречивостью знаний ЛПР, несовершенством и незавершенностью их действий.

Неопределенность среды характеризуют факторы, имеющие стохастическую и детерминистскую природу – среди последних выделяют факторы, определяющие природную, поведенческую и целевую неопределенность. В составе **субъективной неопределенности** (в зависимости от вида нечеткости мышления и знаний ЛПР) выделяют лингвистическую, ожидаемую, аксиологическую, многокритериальную и структурную неопределенность. **Логическая неопределенность** (непосредственно связанная с неопределенностью знаний и вывода в БЗ) проявляет себя в виде логической многозначности и неопределенности логического вывода, неполноте и противоречивости знаний в составе БЗ, а также в других НЕ-факторах, влияющих на действия ЛПР (также см. в разделе 2.7).

В практическом плане, таким образом, ясно, что неопределенность не означает отсутствие у ЛПР информации об СС и СУ – речь идет о неполноте, неточности, неадекватности и т.д. имеющихся сведений. Перед ЛПР фактически каждый раз возникает дилемма: действовать немедленно, но в условиях риска ввиду недостатка информации, или отложить решение до лучших времен, ожидая дополнительных данных. Однако в любом случае ЛПР надо без промедлений, активно и целеустремленно, исследовать ситуацию, добываясь уменьшения неопределенности всеми доступными способами – механизм этого уменьшения может быть линейным и нелинейным, аддитивным, синергетическим – чтобы правильно воспользоваться им, нужно знать специфику рассматриваемой СС, а также учитывать субъективный «человеческий фактор». Последнее объясняется тем, что роль личности при проведении такого рода исследований весьма велика: это касается и методики сбора и обработки первичных сведений, и выбора способов представления и критериев оценки конечных данных, и тем более – форм их последующего применения.

Отметим, что ЛПР здесь нужно иметь ввиду два обстоятельства: во-первых, «человеческий фактор» приводит к негативному результату, если пристрастия и предпочтения экспертов ведут к тому, что их групповое мнение (концептуальная платформа; по терминологии – **онтологическая модель ситуации**, далее ОМС) зависит

не от весомости аргументов, а от того, кто, что и как во время дискуссии говорил, а тем более подвел ее итоги. Кроме того, следует различать неопределенности, относящиеся к работе реальных СС, когда речь идет о действиях ЛПР в материальной среде, и неопределенности, сопровождающие разработку и проектирование новых СС в виртуальной среде моделирования.

Во-вторых, что самое главное, несмотря на то, что общий список НЕ-факторов достаточно велик, основным препятствием для ЛПР всегда являются либо недостаточное знание предметной области, либо недостаточная информация о текущем состоянии СС – то есть о сложившейся в данный момент ситуации. **Знание предметной области** может быть неясным или неполным, в эту область вообще могут входить объективно неполноценные знания: недостаточно четко сформулированные концепции, недостаточно изученные явления и т.д. Неопределенность знаний приводит к тому, что правила влияния даже в простых случаях не всегда дают корректные результаты. Располагая неполным знанием, ЛПР не может уверенно предсказать, какой эффект даст то или иное действие – но даже имея доступ к достаточно полной и надежной теории предметной области, он может решить, что ему лично лучше (проще, привычнее, эффективнее) использовать не точные, а эвристические методы.

Информация о ситуации может включать неточные или ненадежные данные о состоянии объекта по самым разным причинам: например, любой сенсор (датчик, источник первичной информации) имеет ограниченную разрешающую способность и отнюдь не стопроцентную надежность – поэтому при составлении отчетов могут быть допущены погрешности, в них могут попасть недостоверные сведения и т.п. Причиной ошибок может быть и недостаток времени – поскольку не всегда есть возможность быстро получить необходимые данные, хотя ситуация требует срочного решения.

Как уже было сказано в разделе 2.7, ЛПР (включая экспертов, участвующих в разработке ЭС) пользуются неточными методами по двум основным причинам: если точных методов не существует в природе или если они есть, но не могут быть применены на практике. Теоретическим фундаментом обработки неточной информации являются нечеткая логика Заде [100; 114] и теорема Байеса [83; 98], связанные с понятием вероятности – поэтому все извест-

ные в настоящее время подходы и методы так или иначе связаны с вероятностной концепцией.

Теорема Байеса во многих ЭС [83; 98] помогает как бы связать воедино информацию, поступающую из различных источников – на том основании, что для любого события (положения, утверждения) имеется, какой бы малая она ни была, априорная вероятность того, что данное событие истинно. В нашем случае байесовский подход предполагает начальное априорное задание предполагаемых гипотез (значений достигаемых СС целей), которые последовательно уточняются с учетом вероятностей свидетельств в пользу или против этих гипотез, в результате чего формируются апостериорные (условные) вероятности (2.1)-(2.3), отношение правдоподобия (2.8) и другие расчетные соотношения.

Добавим, что согласно [98] формулу Байеса целесообразно использовать в вероятностных терминах. Однако с вычислительной точки зрения более удобно работать с шансами (2.5)-(2.6). Поскольку шансы и вероятности связаны формулами

$O(H) = \frac{P(H)}{1 - P(H)}$; $P(H) = \frac{O(H)}{1 + O(H)}$, для гипотезы с вероятностью 0,5 шансы равны 1. А так как имеет место

$$O(H : E) = \frac{P(E : H)}{P(E : неH)} O(H), \text{ то} \quad (4.16)$$

$$\ln|O(H : E)| = \ln\left|\frac{P(E : H)}{P(E : неH)}\right| + \ln|O(H)|. \quad (4.17)$$

Другими словами, шансы являются простыми функциями $P(E : H)$ и $P(E : неH)$, и если перейти к логарифмам величин, а в БЗ хранить логарифм отношения $P(E : H) / P(E : неH)$, то все вычисления сведутся просто к суммированию.

Автор [83] возражает против использования шансов на том основании, что крайние значения для них равны $\pm \infty$, тогда как для вероятностей это, как известно, единица и ноль. Бесконечности находятся за пределами расчетных возможностей ЭВМ, поэтому вместе с крайними точками может теряться ценная информация – тогда как, например, учет событий с априорной вероятностью 0 и 1 может ускорить решение задачи.

Еще один важный вопрос связан с определением критерия, используемым системой при выработке решения. Обычно эксперт устанавливает два порога для вероятности гипотезы – верхний и нижний. Если вероятность превосходит верхний порог, то гипотеза принимается как основа для возможного заключения. Если гипотеза оказывается ниже нижнего порога, она отвергается как неправдоподобная. В качестве верхнего порога используют максимально достижимую для данной гипотезы апостериорную вероятность, которая соответствует случаю, когда каждый элемент данных, имеющийся в БЗ в данный момент, говорит в пользу гипотезы. В качестве нижнего порога вычисляют минимально возможную вероятность для такой гипотезы, предполагая, что все свидетельства, имеющиеся в БЗ, говорят против гипотезы. Естественно, что для разных гипотез такие значения будут разными.

Предположим, что для некоей гипотезы максимально возможная вероятность равна 0,5: если эксперт жестко установит верхний и нижний пороги равными 0,6 и 0,8, то эта гипотеза всегда будет отбрасываться как маловероятная. Для выхода из ситуации в [83] предлагается верхний и нижний пороги умножать на некоторые коэффициенты, в результате чего они будут специфическими для каждой конкретной гипотезы. Проблемой при этом, однако, становится определение данных весовых коэффициентов.

Применение *нечеткой логики* (fuzzy logic, называемой также логикой возможностей), как уже было также сказано в разделе 2.9, является более простым, но менее точным методом оценки достоверности знаний.

4.4. Снижение неопределенности выбора управленческих решений с помощью СИМ по МДМ

Развитие теории СУ иерархическими и многокритериальными СС холонического типа, неотъемлемыми компонентами которых являются ЛПР, до настоящего времени остается актуальной проблемой, которая имеет важное практическое значение [12; 28; 37; 44; 46]. Анализ и синтез таких СУ встречает ряд принципиальных трудностей: сложно, например, оптимизировать режим работы СС, компоненты которой (подсистемы и элементы) имеют возможность самостоятельно максимизировать свои функционалы. Поэтому, во-первых, в понятие оптимальности СУ, предназначенных для

управления СС, «необходимо вложить конкретный разумный смысл» [81]. Во-вторых, следует рассмотреть перспективы использования при решении задач управления ими новых информационных технологий – в частности, компьютерного метода СИМ [44]. В-третьих, чтобы избежать путаницы в терминах, целесообразно сразу сформулировать четыре следующих исходных тезиса.

1. Для достижения научно-исследовательских и коммерческих целей в настоящее время создаются «имитационные системы» [12; 37 и др.], каждая из которых представляет собой организационно-техническую (человеко-машинную, диалоговую) СС и состоит из моделей (математических или алгоритмических), которые описывают эволюцию рассматриваемого объекта или процесса; а также системы процедур, позволяющей объединить эти модели с ЛПР; и специального математического обеспечения, включающего языки общения с ЭВМ и систему алгоритмов для решения конкретных задач, в том числе оптимизационных.

2. Каждая такая имитационная (информационная) СС представляет собой индивидуальный наукоемкий проект, разработка и реализация которого требуют значительных усилий ЛПР разного профиля, базируются на результатах исследований фундаментального характера (в том числе в области искусственного интеллекта) [81].

3. Проблема обеспечения адекватности СИМ связана с учетом как детерминированных характеристик, так и достаточно большого числа (до 10^2 и более) стохастических факторов (параметров СИМ-модели), влияющих на рассматриваемый объект или процесс, тогда как проблема точности СИМ обусловлена тем, что полученные экспериментальные (численные) результаты носят случайный характер [37; 44].

4. Возможности метода СИМ позволяют строить, наравне со статическими, динамические модели объектов (СС; СУ), эффективность функционирования которых зависит от взаимодействия во времени указанного числа стохастических факторов (случайных параметров СИМ-модели по версии МДМ) [40-41].

Анализ предметной области СИМ показывает, что случайность параметров является характерной особенностью моделей большинства реальных СС. Несмотря на то, что эти модели имеют градацию влияющих факторов: от наиболее важных (безусловно включаемых в состав модели) к неопределенным и незначительным (ко-

торыми ЛПР стараются пренебречь), воспользоваться данной градацией не удастся, если это негативно сказывается на качестве модели – точности и адекватности полученных с ее помощью результатов. Наличие неопределенности параметров моделей СС и СУ, по нашему мнению, следует считать объективным – неустранимым и неотъемлемым фактором, существенно влияющим на качество решений по управлению СС.

Напомним, что СИМ-модели используются как для исследования и эффективного управления реальными (существующими и действующими) СС, так и при создании виртуальных СУ – предназначенных для управления новыми (проектируемыми, инновационными, разведанными) СС. От модели требуется способность воспроизводить или предсказывать все основные характеристики поведения СС: параметры режима работы, их изменения (флуктуации, колебания) и взаимосвязи во времени, тенденции к усилению или ослаблению внешних воздействий и т.п. При этом для СИМ-моделей точность прогноза выходных характеристик, и даже правильность воспроизведения последовательности специфических действий часто не являются решающим и единственно целесообразным стимулом для ее создания и успешного применения [12; 28; 40-41; 44; 46; 81].

Цель настоящего раздела – анализ перспектив и возможностей метода СИМ по уменьшению влияния неопределенности исходных данных и других параметров СС, СУ и их моделей (математических, компьютерных, имитационных) на качество управленческих решений. Напомним, что особенностями СИМ по МДМ являются системный; ориентация на управление процессами, происходящими в СС, а также использование знаний разного типа (верифицированных и аксиологических) при определении условий и исходных данных для проведения СИМ.

Схема применения СИМ-моделей в рамках МДМ для управления объектом показана на рис. 4.4. От традиционной СУ, включающей объект и субъект управления (в простейшем случае ЛПР), а также систему обмена информацией и систему управления состоянием объекта, она отличается наличием СИМ-модели объекта – которая по первому входу подключена к системе обмена информацией, по второму входу – к субъекту управления, а по выходу – к системе управления состоянием объекта.

Принцип работы СУ, показанной на рис. 4.4, подробно рассмотрен как в общем виде [44; 46; 77], так и применительно к условиям и целям функционирования конкретного объекта [88]. Поэтому ограничимся указанием на то, что данная схема содержит два контура управления: с помощью первого (внешнего) контура субъект может непосредственным образом (помимо СИМ-модели) осуществлять управление объектом, второй (внутренний) контур включает СИМ-модель и предназначен для выбора наилучших в данный момент времени управленческих решений – с учетом результатов, полученных при помощи МДМ.

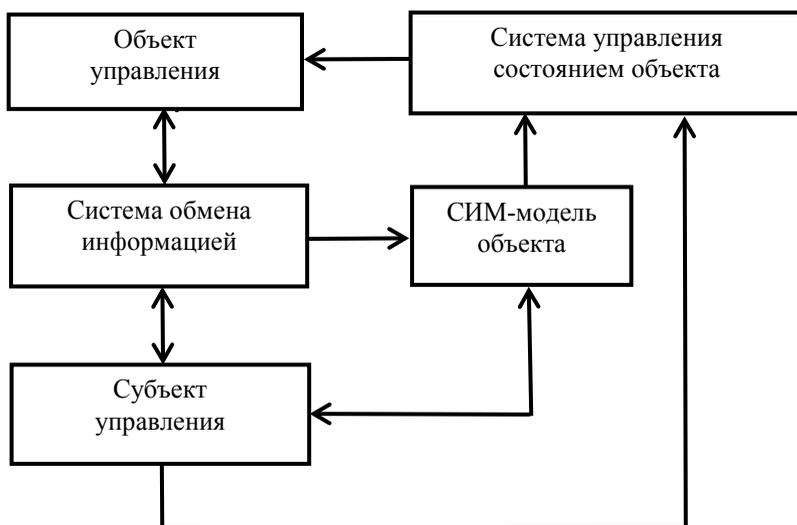


Рис. 4.4. Схема применения СИМ-модели для управления объектом

Анализ перспективности СИМ-модели предполагает, что мы располагаем информацией (ограниченной, частичной, неполной) о характеристиках частей (элементов и подсистем) СС, и эти известные данные, предполагаемые факты, а также другие известные закономерности, присущие предметной области, в целом достаточно правильно отражают характер поведения СС как объекта СИМ. В то же время наша способность интуитивно представлять взаимодействие частей СС менее надежна, чем знания о каждой из них – поэтому, построив статистическую имитационную модель и

наблюдая на ней взаимодействие разных факторов: как внешних, так и внутренних, мы можем промоделировать и исследовать с помощью МДМ и возможные варианты реализации СС, и режимы ее работы – с тем, чтобы выбирать наиболее подходящие из них в каждый заданный промежуток времени. Эффективная статистическая имитационная модель при этом должна выражать сущность СС и показывать, каким образом изменения режима функционирования или структуры приводят к изменению ее поведения. На модель также возлагается задача выявления внешних факторов и возмущений, к которым может быть чувствительна реальная СС.

Если СИМ-модель призвана предсказывать поведение СС, она обязана быть и достаточно точной (достоверной), и достаточно правильной (адекватной). Если же необходимо углубить знания об СС путем ее исследования, СИМ-модель может быть эффективной и в том случае, когда она правильно отражает лишь то, что мы посчитаем сущностью исследуемой СС. Возможности МДМ допускают использование в составе СИМ-модели до 10^3 переменных, а также всех ММ и других математических соотношений, которые представляются важными при вербальном (словесном) описании реальных СС.

Функция решения (математическая часть) СИМ-модели [40-41; 44; 77] может быть простым уравнением, но может представлять собой цепь соотношений, учитывающих целый ряд дополнительных условий и ограничений. Выбирая ограниченный и достаточно короткий интервал анализа, можно добиться того, что каждое отдельное конкретное решение практически не будет зависеть от других принимаемых в данный момент решений.

Данный принцип независимости решений служит основой для построения сложных СИМ-моделей и позволяет избежать продолжительных трудоемких вычислений. При решении уравнений не ставится задача повышения точности вычислений, если характер СИМ-модели делает ее нечувствительной (робастной) к принятым допущениям и ошибкам при округлении, сокращении и т.п. Ошибки и искажения в рамках МДМ можно вносить преднамеренно, чтобы проверить, насколько чувствительны к ним используемые вычислительные методы. Применение сложных методов обычно делает формулировку уравнений менее понятной, тогда как пре-

имуществами СИМ-модели, напротив, являются ее максимальная простота и наглядность [12; 28; 37; 40-41; 44; 77; 81].

Будущая СИМ-модель призвана функционировать в максимально широких границах изменения переменных и параметров: во-первых, потому что в будущем может понадобиться расширение пределов изменения условий работы реальной СС, во-вторых, поскольку нельзя предсказать заранее, какие значения примут по ходу СИМ различные переменные, в-третьих, ввиду необходимости выяснить, как будет работать и окажется ли полезной статистическая имитационная модель за пределами границ, которые имеются сегодня у рассматриваемой СС, – так как разработка новых СС обычно предполагает их действие вне рамок прежней практики [40-41; 44].

Возможны два способа повышения эффективности СИМ-моделей: *агрегирование* (объединение, группировка) однородных факторов, позволяющее упростить модель, избегая лишних деталей (хотя при этом есть опасность «потерять» ее существенные элементы), и *использование экзогенных переменных* (виртуальных проверочных сигналов), имитирующих помехи (случайные факторы), которые могут воздействовать как по внешним экзогенным входам СИМ-модели, так и являться эндогенными (внутренними) относительно ее функций решения. Испытанием практической значимости модели является проверка соответствия поведения СС ее поведению, предсказанному моделью. Суждение ЛПР о значимости СИМ-модели основывается на уверенности в том, что она отображает нужные особенности поведения СС, а также на соответствии общего поведения модели и реальной СС.

При проверке структуры и элементов СИМ-модели в отношении границ и взаимосвязи переменных (факторов, аргументов) модели, числовые значения их параметров берутся по статистическим оценкам результатов испытаний СС (подробнее см. ниже) – после того, как по методике МДМ установлены цели и задачи моделирования; определены границы СС; выбраны переменные и сформулированы гипотезы, определяющие их взаимодействие; приняты основанные на предварительных суждениях (соображениях) решения о параметрах, удовлетворяющих условиям статистических испытаний. Отметим, что если параметры и правила работы СС точно не определены и нет возможности их оценить, но они являются

управляемыми – с помощью имитационных моделей можно исследовать воображаемые ситуации и сценарии развития событий, несмотря на то, что они будут относиться к виртуальной, а не реальной среде, и речь будет идти о правдоподобии, а не о точности и адекватности (в строгом смысле слова) результатов СИМ по МДМ, предназначенных для управления СС.

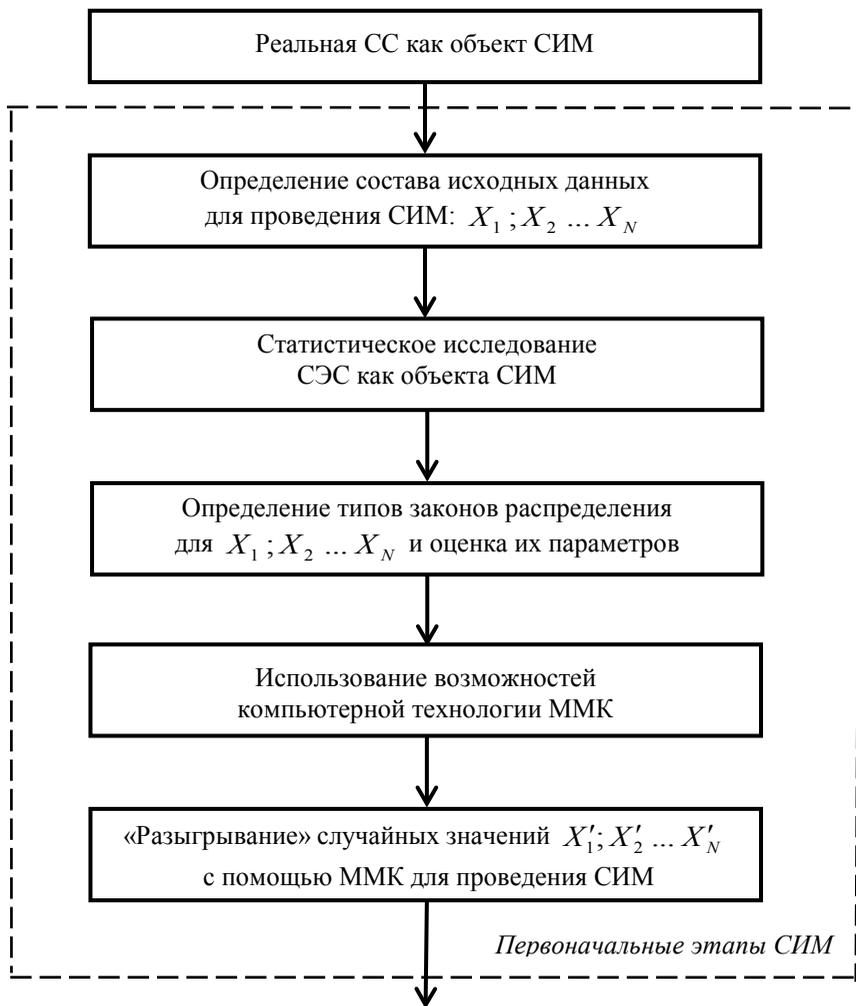


Рис. 4.5. Этапы, влияющие на точность и достоверность исходных данных для проведения СИМ по МДМ

Будем также иметь в виду, что в СС с обратной связью внешние и внутренние факторы способны как усиливать, так и уравновешивать друг друга: если отклонения, связанные с изменением одних факторов, компенсируются изменениями других факторов. В целом, чем более устойчивой и «жизнеспособной» является реальная СС, тем менее чувствительна она к указанным изменениям своих параметров – что и должна отражать ее СИМ-модель. Определить чувствительность СИМ-модели к вариациям параметров СС можно путем проведения соответствующих испытаний: зафиксировав параметр во времени и осуществляя его регулирование извне (при анализе чувствительности модели значения ее параметров должны увеличиваться не менее чем в два раза).

Методика СИМ по МДМ для СС открытого типа, подверженных воздействию (управляющему, исследовательскому, негативному и др.) со стороны внешней среды, достаточно подробно представлена в [48]. Там же отмечена важность первоначальных этапов СИМ (см. рис. 4.5), которые предусматривают определение состава исходных данных, влияющих на функционирование СС; проведение статистического исследования для сбора и обработки исходной информации о СС; идентификацию типов законов распределений исходных данных и расчет статистических оценок параметров этих законов; разработку математических моделей подсистем и элементов СЭС, и непосредственным образом влияют на адекватность и точность получаемых итоговых результатов.

Также были проанализированы особенности двух наиболее известных подходов к СИМ. Школа Форрестера-Медоуза полагает, что исходные данные можно задавать в первом приближении, а затем варьировать их, добиваясь необходимого исследовательского результата [109]. Большинство отечественных специалистов предполагает использование в интересах СИМ теории организации и планирования научного эксперимента, что позволяет оптимизировать объем и продолжительность расчетов – с учетом требуемой надежности их результатов, ориентированных на задачи управления СС [12; 28; 37; 40-41]. Достоверность всех фигурирующих при этом данных (исходных, промежуточных, итоговых) тесно связана с их кумулятивностью, под которой принято понимать свойство данных минимального объема адекватно отображать необходимые характеристики исследуемого объекта [77].

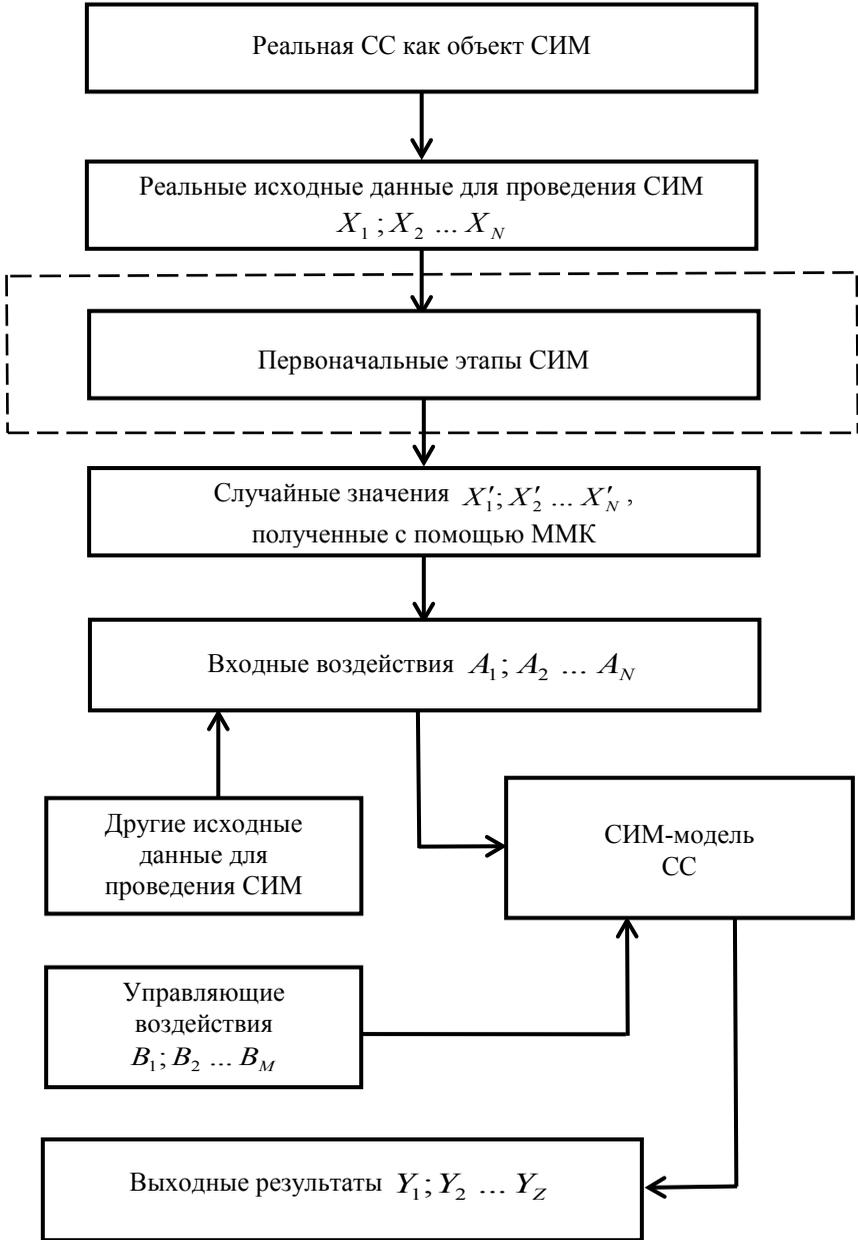


Рис. 4.6. Схема управления СИМ-моделью СС

Считается, что достичь кумулятивности можно с помощью упомянутого комплексного исследования объекта СИМ – то есть путем сбора и обработки необходимого объема информации на указанных первоначальных этапах.

При необходимости моделировать объекты с малой прецедентной базой (проектируемые, инновационные, разведанные) здесь возникают трудности, для преодоления которых МДМ использует возможности технологии ММК. Решение проблемы кумулятивности при этом сводится к рациональному содержательному описанию (вербальному моделированию) СС, в ходе которого важно найти правильное сочетание аналитических и численных, детерминированных и статистических методов, используемых при проведении СИМ.

Указанные особенности СИМ по МДМ иллюстрирует рис. 4.6, где штриховым контуром обозначены первоначальные этапы, определяющие точность и достоверность исходных данных для проведения СИМ. Поскольку реальная СС является основным источником информации о себе самой, определить состав исходных данных (соответствующие им характеристики и параметры СС обозначим как группу $X_1; X_2 \dots X_N$) можно в рамках проведения исследования СС, итогом которого являются статистические сведения, относящиеся к данной группе.

После анализа и обработки этих первичных данных (в виде гистограмм, графиков, численных оценок параметров и т.п.) определяются типовые законы распределения для $X_1; X_2 \dots X_N$ из числа известных и доступных для разработчиков СУ вероятностных функций (типовых законов), а также производится оценка их параметров – обычно отличающихся от полученных ранее первичных оценок. Затем с помощью компьютерной технологии метода Монте-Карло (ММК) «разыгрываются» массивы случайных числовых величин $X'_1; X'_2 \dots X'_N$, соответствующих им – которые, совместно с другими группами исходных данных, формируют состав воздействий $A_1; A_2 \dots A_N$ на входы СИМ-модели (см. рис. 4.6).

Помимо указанных этапов СИМ и входных воздействий $A_1; A_2 \dots A_N$ на схеме рис. 4.6 показаны воздействия $B_1; B_2 \dots B_M$, необходимые для управления СИМ-моделью, а

также выходные данные $Y_1; Y_2 \dots Y_Z$ – результаты СИМ. Отметим, что в адекватной СИМ-модели управляющие воздействия $B_1; B_2 \dots B_M$ должны в требуемой мере соответствовать факторам (методами и средствам), которые позволяют управлять рассматриваемой реальной СС. Мы видим (см. рис. 4.5), что неопределенности (неточности, неадекватности, нестационарности), присущие исходным данным $X_1; X_2 \dots X_N$, преобразуются в неопределенности входных воздействий $A_1; A_2 \dots A_N$ и результатов СИМ $Y_1; Y_2 \dots Y_Z$ – то есть непосредственным образом «участвуют» в формировании качества управленческих решений (см. рис. 4.6), потенциально ухудшая его. Уменьшение этих неопределенностей или снижение их влияния на результаты СИМ по МДМ представляют для разработчиков СУ не только теоретический, но и важный практический интерес.

Оптимизация управления СС при помощи СИМ-моделей.

Главная сложность оптимизации действий ЛППР при управлении СС обусловлена тем, что выступающие в качестве теоретических критериев оптимальности функционалы не поддаются вычислению не только аналитическими, но и численными методами [12; 28; 37; 81]. Использование СИМ-модели СС (см. рис. 4.4) способствует решению задачи оптимизации (упрощает поиск путей достижения цели, определение параметров и ограничений и т.п.), но здесь речь фактически идет о СИМ процесса квазиоптимизации работы СУ путем проведения статистического эксперимента на множестве реализаций модели СС с помощью целенаправленного изменения исходных данных и факторов внешней среды [42; 50]. Стратегия управления, как уже было сказано, состоит в том, чтобы поддерживать (усиливать, развивать) позитивные эффекты в СС с помощью имеющихся положительных обратных связей и останавливать (пресекать, ослаблять, ликвидировать) проявления негативных тенденций посредством отрицательных обратных связей – что даже с применением МДМ и ММК сделать не менее трудно, чем выполнить многокритериальную оптимизацию.

Рассмотрим варианты реализации наиболее простой дискретной схемы оперативного управления СС – при которой процесс управления сводится к выбору одного из M доступных состояний СС,

каждому из которых соответствует свой набор одних и тех же управляющих параметров.

1. Схема «слепого» порогового выбора имеет ввиду работу СС в текущем m -ом состоянии до тех пор, пока глобальный критерий управления W не станет меньше заданного порогового значения $W_{\text{п}}$: в случае $W < W_{\text{п}}$ автоматически срабатывает СУ и переводит СС в следующее $(m + 1)$ -ое состояние и т.д., пока в СС не будет иметь место $W > W_{\text{п}}$ – нетрудно видеть, что СУ в помощи СИМ-модели в данном случае не нуждается.

2. Схема модифицированного порогового выбора, отличающаяся тем, что при $W < W_{\text{п}}$ сработавшая СУ переводит СС в ближайшее состояние, соответствующее $W > W_{\text{п}}$ – для чего в текущем m -ом состоянии с помощью СИМ-модели проводится анализ эффективности работы СС в других ее состояниях.

3. Схема периодического выбора, при которой СУ на каждый последующий промежуток времени T переводит СС в состояние, признанное наилучшим по результатам работы с СИМ-моделью в течение предыдущего аналогичного периода времени.

4. Схема непрерывного (параллельного) выбора, где СУ с помощью СИМ-модели постоянно проводит анализ всех M доступных состояний СС и либо по условию $W < W_{\text{п}}$, либо через время T , либо по другому собственному решающему правилу переводит СС в состояние, признанное наилучшим по итогам работы с СИМ-моделью.

Приведенные схемы оперативного управления СС не гарантируют оптимизации управления в строгом смысле слова, поскольку не ориентированы на выполнение условия $W = W_{\text{МАКС}}$, где $W_{\text{МАКС}}$ – максимально-возможное значение критерия управления (глобальный максимум W). Однако они учитывают согласованные мнения экспертов и принимаются ЛПП к исполнению как наилучший вариант управленческих действий на ближайшее время. Их практическая эффективность определяются, с одной стороны, соответствием параметров СУ (выбранных значений $W_{\text{п}}$; T и др.) динамике режима работы СС; с другой стороны – обоснованным определением составляющих критерия управления W .

Для СС типа производственного предприятия задача оптимального управления может быть сформулирована следующим образом: «Необходимо определить управляющие воздействия на задан-

ном интервале времени T так, чтобы при соблюдении ограничений по производству и возможным вариантам управления СС величина критерия оптимальности W принимала оговоренное экстремальное (максимальное или минимальное) значение» [44]. Составляющими W при этом могут быть такие локальные критерии, как максимум вероятности выполнения плана; максимум загрузки оборудования; минимум суммарных затрат; минимум времени выполнения планового задания; минимум объема незавершенного производства и т.п.

4.5. Выбор и обоснование метода проектирования БЗ

База знаний, являющаяся ядром любой ЭС, представляет собой совокупность знаний о предметной области, записанных на машинный носитель в форме, понятной эксперту и пользователю (обычно на определенном языке, приближенном к естественному). Этап создания БЗ считается наиболее трудоемким, анализ [1; 11; 83; 86; 90; 91; 97; 106] показывает, что распространенными методами создания диагностических систем являются два метода: на базе РОЦ-технологии и основанный на построении дерева целей (дерева диагностики).

РОЦ-технология (от сокращения «Ресурс – Образование – Цель») [97] представляет собой совокупность процедур, выполнение которых позволит создать ЭС, позволяющие не только помочь в принятии решений, но и обеспечить воплощение этого решения в жизнь. В основе РОЦ-технологии лежат ключевые понятия обучения, ресурса и цели. Базой РОЦ-технологии является совокупность процедур, выполнение которых на систематической (полуформализованной) основе позволяет создавать ЭЭС, удовлетворяющие главному требованию: не только помочь ЛПР в принятии решений, но и обеспечить воплощение этих решений в жизнь. Ресурсно-целевой подход к построению ЭС не является новым в теоретическом отношении, однако эффективных, простых и легко внедряемых в практику методов разработки таких систем до сих пор чрезвычайно мало.

Процедуры РОЦ-технологии, которые ориентируются на представление целей ЛПР в виде дерева решений, либо в виде таблицы решений, можно разделить на три уровня, показанные на рис. 4.7. Верхний уровень, базирующийся на дереве целей, обеспечивает предварительное определение варианта принятия решения. Сред-

ний уровень ориентируется либо на дерево решений, либо на таблицу решений, либо на то и другое одновременно. Нижний уровень предназначен для обеспечения исполнения принятого решения путем его интеллектуальной информационной поддержки.

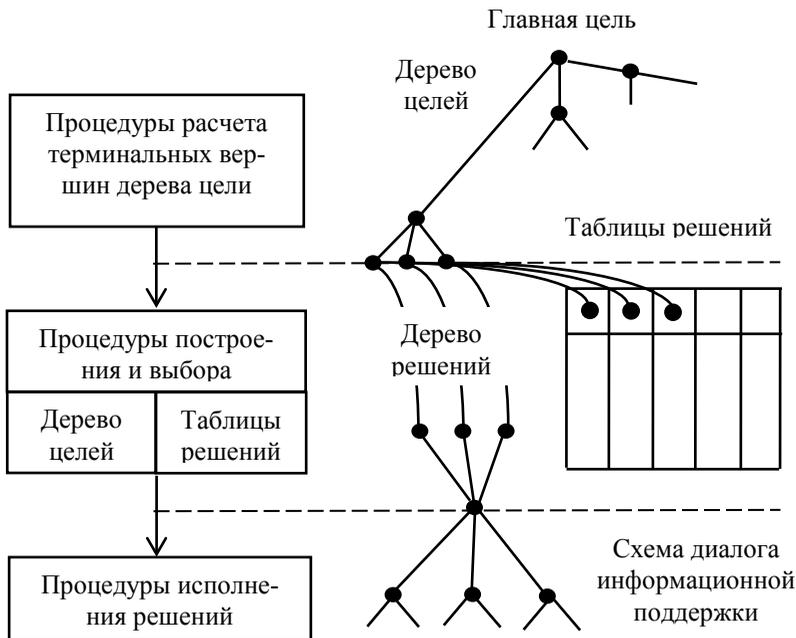


Рис. 4.7. Взаимосвязь уровней процедур РОЦ-технологии

РОЦ-технология ориентирована на понятие относительной важности целей. Числовая характеристика важности целей измеряется в шкале от нуля до единицы и называется коэффициентом относительной важности (КОВ). Значения этих коэффициентов дают возможность оценить, во сколько раз каждая цель превосходит другие по свойству важности, то есть данные КОВ – это относительные «веса» целей. Для измерения КОВ можно воспользоваться методами непосредственной оценки или последовательного парного сравнения. Выяснение отношения предпочтения по паре объектов производится достаточно точно. Пусть имеется n целей и выполнено их парное сравнение по важности. В результате получают матрицу, элементы которой имеют вид

$$Z_{ij} \begin{cases} 1, & \text{если } A_i \geq A_j; \\ 0, & \text{если } A_i < A_j. \end{cases} \quad (4.18)$$

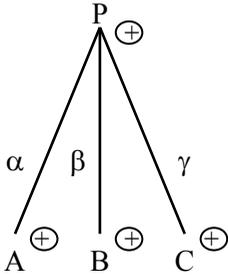
Элемент матрицы указывает с помощью единицы на цели, которая не менее важна по сравнению с другой. И, наоборот, если цель A_j строго предпочтительней цели A_i , то на месте элемента матрицы Z_{ij} указывается нуль. Если элементы Z_{ij} просуммировать по всем столбцам, то получим число целей, которые указывают на важность цели A_i . Таким образом, величина Z_{ij} указывает на число целей, по отношению к которым цель A_i более важна. Коэффициент относительной важности A_i равен:

$$KOB_I = \left(\sum_{j=1}^n Z_{ij} \right) / \left(\sum_{i=1}^n \sum_{j=1}^n Z_{ij} \right). \quad (4.19)$$

Информацию о направлении изменения целей и их относительной важности получают от экспертов согласно табличному описанию дерева целей. Измерение важности целей является прерогативой эксперта, так как, во-первых, оценка важности целей существенно влияет на конечные результаты, и, во-вторых, оцениваемые цели, как правило, связаны с факторами, не учитываемыми деревом целей (учесть которые либо затруднительно, либо невозможно). Довольно сложно учесть, например, законы развития общества, его задачи и цели, которые, несомненно, связаны с целями и задачами каждого конкретного предприятия. Однако всегда существуют важные и не очень важные цели, и эти понятия не абстрактны, а тесно связаны с интересами ЛПП предприятия.

Вторая часть данного этапа требует заполнения таблицы, где приводятся формулы расчета прироста или уменьшения для каждого графа «цель – показатель». Результатом выполнения данного этапа является набор формул для расчета приростов аргументов, размещаемых совместно с формулой расчета показателя – как это показано в таблице 4.1.

Таблица 4.1. Формулы для расчета прироста показателей без учета экспертных оценок

Номер	Граф цели	Формула для расчета
1		$P = A \cdot B \cdot C$ $X = \sqrt{\frac{P + \Delta P}{A \cdot B \cdot C}}$

Затем составляется матрица текущего состояния предприятия:

$$X = \left\{ \begin{array}{cccccc} x_{11} & x_{12} & \dots & x_{ij} & \dots \\ x_{21}^p - x_{21}^{p-1} & x_{22}^p - x_{22}^{p-1} & \dots & x_{2j}^p - x_{2j}^{p-1} & \dots \end{array} \right\}. \quad (4.20)$$

Здесь

$$x_{1j} = \begin{cases} 1, & \text{при } x_{2j}^p - x_{2j}^{p-1} > 0; \\ 0, & \text{при } x_{2j}^p - x_{2j}^{p-1} = 0; \\ -1, & \text{при } x_{2j}^p - x_{2j}^{p-1} < 0, \end{cases} \quad (4.21)$$

где x_{1j} – указатель состояния показателей (j -ый показатель за период p); $x_{2j}^p - x_{2j}^{p-1}$ – значения показателя j за текущий и предыдущий период. Общая оценка предприятия рассчитывается как

$$\tilde{X}^p = \sum_{j=1}^n x_{1j}.$$

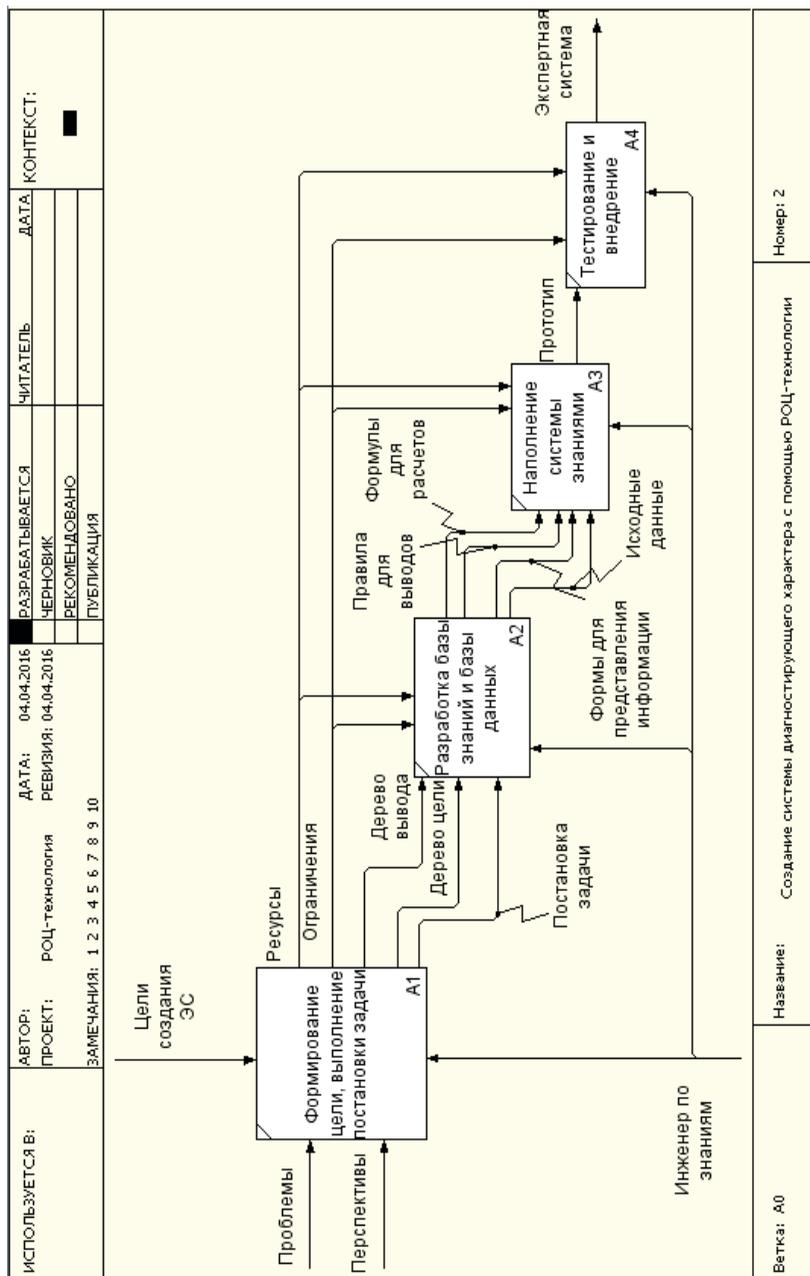
Диагноз формируется на основе таблицы диагноза (см. таблицу 4.2). Идентификация состояния предприятия с помощью таблицы 4.2 позволяет сгенерировать правила, согласно которым формируется диагноз.

Таблица 4.2. Таблица диагноза

Диагно- з	Период										
	$p-2$	$p-1$	p								
Неплатежеспособность DIA_1	$\tilde{X}^{p-2} < 0$	$\tilde{X}^{p-1} < 0$	$\tilde{X}^p \leq 0$								
				Реко- мен- дация R	Состояние						
					1	2	3	4	5	6	
				DIA_1	1	-	-	-	-	-	-
				DIA_2	1	-	-	+	-	-	-
				DIA_3	1	-	+	-	-	+	+
				DIA_4	2	+	-	-	-	-	-
				DIA_5	1	-	+	+	-	-	-
DIA_6	2	+	-	+	-	-	-				
DIA_7	1	-	+	+	-	-	-				
Состо- яние в норме	$\tilde{X}^{p-2} < 0$	$\tilde{X}^{p-1} < 0$	$\tilde{X}^p \geq 0$								
Тен- денции к улуч- шению	$\tilde{X}^{p-2} < 0$	$\tilde{X}^{p-1} \geq 0$	$\tilde{X}^p \geq 0$								
При- знаки ухуд- шения	$\tilde{X}^{p-2} \geq 0$	$\tilde{X}^{p-1} \geq 0$	$\tilde{X}^p < 0$								
Тен- денция к серь- езному ухуд- шению	$\tilde{X}^{p-2} \geq 0$	$\tilde{X}^{p-1} < 0$	$\tilde{X}^p < 0$								

Обобщенные этапы создания диагностирующей системы на основе РОЦ-технологии представлена на рис. 4.8. Рассмотрим каждый из комплексов более детально

Формулирование цели создания системы или гипотезы, которую следует доказать; постановка задачи. Постановка задачи должна содержать описание результирующей и входной информации, а также общее описание процедур (алгоритмы, формулы) преобразования входной информации в результирующую.



Ветка: A0 Название: Создание системы диагностического характера с помощью РОЦ-технологии Номер: 2

Рис. 4.8. Этапы создания диагностирующей системы на основе РОЦ-технологии.

Обычно выбор цели является следствием опроса руководителей и изучения показателей, представляющих собой информационную базу принятия решений. Цели формулируются исходя из анализа проблем, стоящих перед предприятием, перспектив развития и опыта, а затем представляются в графическом и табличном виде.

Таблица составляется с помощью опытного эксперта, так как она необходима для поиска путей выхода из возможных нежелательных состояний предприятия. Кроме того, таблица устанавливает связь между тактическими целями и экономическими показателями. Форма описания входной и результирующей информации является общепринятой, преобразование входа в выход описывается обобщенными процедурами, так как их конкретная форма отражается в базе знаний.

В результате выполнения первого этапа должны быть получены [86; 91]:

- дерево целей или дерево вывода;
- ограничения на используемые ресурсы;
- перечень исходной и выходной информации.

Разработка БЗ и структуры БД. База знаний содержит правила вывода или формулы расчета, а база данных - исходную информацию для решения задач. На данном этапе осуществляется выбор метода представления знаний. В рамках выбранного формализма осуществляется проектирование логической структуры БЗ. Процессу принятия решений присуща слабая формализуемость, поэтому этап формализации является достаточно творческим процессом. Используя текстовое описание предметной области, ключевые слова и связи между ними, полученные на предыдущих этапах, инженер по знаниям формулирует все возможные состояния объекта управления, а также производит группировку схожих ситуаций.

Наполнение системы значениями и данными. Данная процедура выполняется согласно инструкциям и макетам ввода знаний. Формулы для перечеа вводятся по специальным формам.

Тестирование и внедрение системы. Данный этап является заключительным и предполагает выполнение формальных процедур, ознакомиться с которыми можно в [8; 17].

В заключении, перечислим сильные и слабые стороны РОЦ-технологии. Основным достоинством данной методики создания

БЗ является возможность получение не только качественной, но и количественной оценки. Однако для формирования числовых оценок необходимо допустить ряд предположений, например, что связь между отдельными показателями выражается прямой. Если данные предположения неверны, то это приводит к формированию ложных рекомендаций.

К недостаткам следует отнести:

- дерево целей может быть сформулировано недостаточно четко, с учетом субъективных предпочтений ЛПР;
- как следствие, КОВ могут быть неудачно сформулированы;
- сложность корректировки и модификации правил БЗ;
- высокая, зачастую неоправданная трудоемкость;
- учет неопределенности осуществляется, только исходя из значений КОВ.

Метод на основе дерева диагностики. Экспертная система создается под конкретные цели и требования руководства предприятия. Центральным компонентом, определяющим уникальность каждой ЭС, является БЗ, так как специфика целей реализуется в структуре БЗ. На рис. 4.9 представлена общая последовательность выполнения этапов создания БЗ, каждый из которых делится на отдельные подэтапы.

Этап 1. Идентификации решаемых задач и текстового описание их предметной области.

Подэтап 1.1. Определение типа задачи, целей и подцелей ее решения. С точки зрения инженера по знаниям, важно правильно определить тип задачи. Специфика применения БЗ приводит к разделению задач на два типа. Первый тип характеризуется трудностями в формализации, неизвестностью последовательности шагов решения, отсутствием алгоритма, применением эвристических методов, большим объемом логических операций вывода, отсутствием или небольшим объемом операций вычислительного и информационно-поискового характера.

Второй тип отличается от первого наличием хорошо формализованных алгоритмов, использованием операций поискового и информационно-справочного характера, ориентацией на вычислительные процедуры, большим объемом обрабатываемых данных.

Первый тип задач наиболее подходит для применения БЗ, а второй – наименее. Поэтому соотнесение задачи к одному из типов полезно с точки зрения предварительного анализа эффективности использования тех или иных обеспечивающих средств.

Практика показывает, что довольно часто в той или иной степени задачи характеризуются свойствами обоих типов. К какому бы типу задача не относилась, необходимо выяснить цели и подцели ее решения. Информацией для этого служат проблемы, стоящие перед предприятием. Проблемы в свою очередь можно получить из отчетов о предпроектном обследовании предприятия. Тип задачи, как правило, зависит от целей ее решения.

Подэтап 1.2. Выявление состава и объемов, а также разработка формы представления результирующей информации.

Подэтап 1.3. Выявление состава и объемов входной информации.

Подэтап 1.4. Описание промежуточной (если таковая предусматривается) информации.

Выполнение подэтапов 1.2-1.4 предполагает выполнение процедур, подробное описание которых содержат [22; 23; 83; 86; 106]

Подэтап 1.5. Текстовое (словесное) описание предметной области. Используя одну или несколько стратегий получения знаний (приобретение, извлечение и формирование) [11; 86] необходимо получить вербальное описание предметной области задачи. Основой для описания служат цели, подцели, результирующая и исходная управленческая документация, а также опыт эксперта и специальная литература (методики, пособия, руководства, статьи, монографии, учебники).

Этап 2. Выделение и графическое представление понятий и их связей. Здесь производится первичная формализация знаний, полученных на первом этапе. Инженер по знаниям с помощью эксперта определяет ключевые понятия, отношения и характеристики, необходимые для описания процесса решения задачи. После выделения ключевых понятий необходимо определить связи между ними. Цели и связи между ними обычно представляют в виде дерева целей с указанием конъюнктивных и дизъюнктивных вершин.

Этап 3. Продукционное представление знаний о предметной области и решаемых задачах в терминах выбранной модели

представления знаний. На данном этапе осуществляется трансформация графического изображения БЗ в формальное представление. На основе дерева целей и сформированного словаря строятся правила БЗ.

Этап 4. Наполнение системы знаниями. Данный этап осуществляется с помощью инженера по знаниям и программиста. Результатом должен выступить прототип ЭС.

К достоинствам метода на основе дерева диагностики можно отнести:

- наглядность и простота в понимании;
- в процессе эксплуатации ЭС допускается настройка БЗ, программная реализация предусматривает интегрированное использование ЭС совместно с расчетными процедурами, использующими информацию непосредственно из БД;
- возможность учета неопределенности знаний.

Основным недостатком является сложность настройки БЗ, поскольку знания непосредственно включаются в программный код.

4.6. Информационные технологии организации удаленного доступа и выбор структуры распределенной БД

Построение распределенных корпоративных ИС не является тривиальной задачей и требует глубокого анализа и обобщенной проработки вопросов целесообразности, возможности создания системы, обладающей оптимальными параметрами технологического, информационного, финансового обеспечения и качественным уровнем функционирования [70; 74-75].

Важной функцией рассматриваемых систем является доставка запросов от удаленных пользователей к БД и ответов на запросы пользователям, пославшим эти запросы. Выбор способа взаимодействия между ними зависит от **конкретной задачи**, параметров предметной области (объем данных, частота обращений, требуемая скорость реакции на запрос и т.д.) и от возможности реализации вариантов технологии «клиент-сервер» на имеющейся технической и системной базе (наличие финансовых ресурсов, качество каналов связи, состав компьютерного парка и т.д.). Рассмотрим данную проблему на конкретном примере разработки распределенной БД **региональной организации по надзору за связью и информатизацией** и сети ее филиалов.

Средства создания систем, обеспечивающих взаимодействие компонентов БД в режиме реального времени, предпочтительны в следующих случаях:

- наличие устойчивой связи;
- необходимость высокой скорости реакции на запрос;
- высокая степень распределенности данных в узлах сети;
- наличие достаточных финансовых ресурсов.

Под информационной технологией организации *удаленного доступа* к БД будем понимать системно организованную совокупность средств передачи данных, информационных ресурсов, протоколов взаимодействия, аппаратно-программного и организационно-методического обеспечения, ориентированную на удовлетворение потребности пользователей в информационных ресурсах и услугах. Информационными ресурсами в нашем случае являются распределенные БД региональной организации по надзору за связью и информатизацией и ее филиалов, под основными информационными услугами понимаются:

- просмотр различными службами БД друг друга;
- фиксирование сведений в удаленных и централизованных БД;
- статистическая обработка данных распределенных БД региональных служб;
- удаленное обслуживание клиентов.

Все перечисленные услуги требуют подготовки запросов на обработку в стандартном формате; передачи запроса в центр обработки или на обрабатывающую компьютерную станцию; обработки запроса и прием ответа на станции, пославшей запрос.

Рассмотрим схему взаимодействия двух узлов компьютерной сети, не зависящую от конкретной структуры сети и маршрута передачи данных между узлами, показанную на рис. 4.10. В данном контексте под «узлом» понимается либо запрашивающая станция, либо станция, обрабатывающая запрос, либо почтовый сервер, если он присутствует в структуре сети.

В центральной части рис. 4.10 сплошным контуром обведены действия, необходимые в случае прерывания связи на время обработки запроса. Порядок приведенных на схеме действий соответствует их временной последовательности. Расположенные на одном уровне действия принимающего и передающего узла выполняются одновременно. В целях снижения уровня требований к

мощности ресурсов региональных компьютерных сетей обычно используются разные варианты технологии «клиент-сервер», позволяющие уменьшить общий трафик в сети за счет перераспределения функций обработки данных между ее узлами [9; 14; 16; 63].



Рис. 4.10. Схема взаимодействия удаленных узлов

В данном случае технология «клиент-сервер» – это модель взаимодействия элементов компьютерных сети, где эти элементы, как правило, не являются равноправными. Каждый из них имеет собственное назначение, играет свою роль – некоторые компьютеры «владеют» и распоряжаются информационно-вычислительными и сервисными ресурсами таких устройств, как процессоры, файловая система, почтовая служба, служба печати, БД. Другие компьютеры имеют возможность обращаться к этим службам, пользуясь услугами первых. Компьютер, управляющий тем или иным ресурсом, принято называть *сервером ресурса*, а компьютер, желающий им воспользоваться – *клиентом*. Конкретный сервер определяется видом ресурса, которым он владеет. Так, если ресурсом являются БД, то речь идет о сервере БД, назначение которого – обслуживать запросы клиентов, связанные с обработкой данных; если ресурс – это файловая система, то говорят о файловом сервере, или файло-сервере, и т.д.

В сети один и тот же компьютерный элемент может выполнять роль, как клиента, так и сервера. Например, в ИС, включающей персональные ЭВМ, большую ЭВМ и мини-компьютер под управлением UNIX, последний может выступать как в качестве сервера БД, обслуживая запросы от клиентов – персональных компьютеров, так и в качестве клиента, направляя запросы большой ЭВМ. Данный принцип распространяется на взаимодействие программ: если одна из них выполняет некоторые функции, предоставляя другим соответствующий набор услуг, то такая программа выступает в качестве сервера. Программы, которые пользуются этими услугами, являются клиентами. Так, например, ядро реляционной SQL-ориентированной СУБД часто называют сервером БД или SQL-сервером, а программу, обращающуюся к нему за услугами по обработке данных – SQL-клиентом.

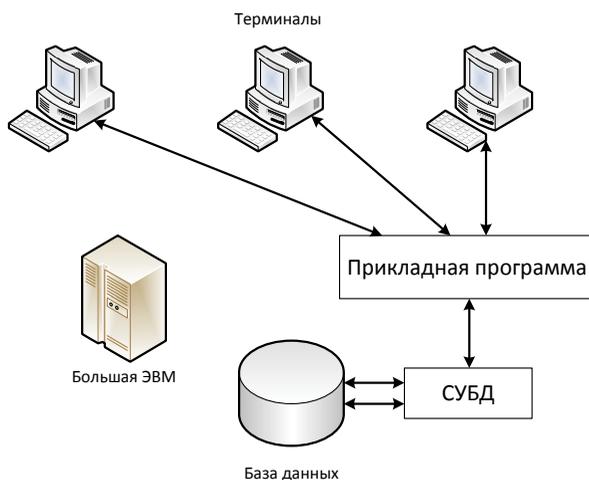


Рис. 4.11. Системы с централизованной архитектурой

Первоначально СУБД имели централизованную архитектуру, которая показана на рис. 4.11. В ней и СУБД, и прикладные программы, которые работали с БД, функционировали на центральном компьютере (большая ЭВМ или мини-компьютер), там же располагались БД. К центральному компьютеру были подключены терминалы, выступавшие в качестве рабочих мест пользователей. Все процессы, связанные с обработкой данных, а именно: поддержка

ввода, осуществляемого пользователем; формирование, оптимизация и выполнение запросов; обмен с устройствами внешней памяти и т.д. выполнялись на центральном компьютере, что предъявляло жесткие требования к его производительности.

Особенности СУБД первого поколения напрямую связаны с архитектурой систем больших ЭВМ и мини-компьютеров, они адекватно отражают все их преимущества и недостатки. Однако нас больше интересует современное состояние многопользовательских СУБД, для которых архитектура «клиент-сервер» стала фактическим стандартом и можно говорить о модели файлового сервера и модели сервера БД.

Технология «клиент-сервер», предполагает разделение функций приложения на три следующие группы.

1. Функции ввода-вывода и отображения данных.

2. Прикладные функции, характерные для данной предметной области.

3. Функции хранения и управления данными.

В соответствии с этим в любом приложении выделяются следующие логические компоненты:

- компонент представления для функций первой группы;
- прикладной компонент, реализующий функции второй группы;
- компонент доступа к информационным ресурсам, поддерживающий функции третьей группы.

Помимо этого, вводятся и уточняются соглашения о способах их взаимодействия (протокол взаимодействия). Различия в реализациях технологии «клиент-сервер» определяются, во-первых, тем, в какие виды программного обеспечения интегрированы каждый из этих компонентов. Во-вторых, какие механизмы программного обеспечения используются для реализации функций трех указанных групп. В-третьих, как логические компоненты распределяются между компьютерными элементами в сети. В-четвертых, какие механизмы используются для связи компонентов между собой.

Выделяют четыре подхода, реализованные в рассматриваемых моделях:

- модель файлового сервера File Server (FS);
- модель доступа к удаленным данным Remote Data Access (RDA);
- модель сервера БД DataBase Server (DBS);
- модель сервера приложений Application Server (AS).

Модель файлового сервера (FS-модель) является базовой для локальных сетей персональных ЭВМ. Первоначально она была популярна среди отечественных разработчиков, использовавших системы FoxPRO, Clipper, Clarion, Paradox. Суть данной модели проста: один из компьютеров в сети считается файловым сервером и предоставляет услуги по обработке файлов другим компьютерам. Файловый сервер работает под управлением сетевой операционной системы (например, Novell NetWare) и играет роль компонента доступа к информационным ресурсам (то есть к файлам). На других компьютерах в сети функционирует приложение, в кодах которого совмещены компонент представления и прикладной компонент (см. рис. 4.12). Протокол обмена представляет собой набор низкоуровневых вызовов, обеспечивающих приложению доступ к файловой системе на файл-сервере.

Запросы к файловой системе

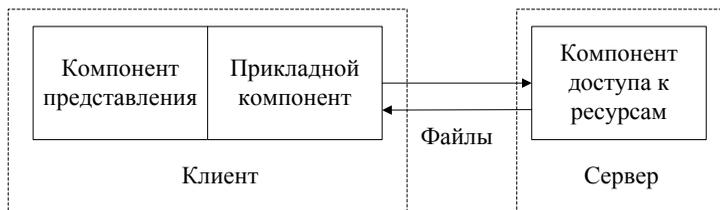


Рис. 4.12. Модель файлового сервера

Модель на рис. 4.12 послужила фундаментом для расширения возможностей персональных СУБД в направлении поддержки многопользовательского режима. В таких системах на нескольких ЭВМ выполняются как прикладная программа, так и копия СУБД, а БД содержатся в разделяемых файлах, которые находятся на файловом сервере.

Каждый раз, когда прикладная программа обращается к БД, СУБД направляет соответствующий запрос на файловый сервер: в запросе указаны файлы, где находятся нужные данные. В ответ на запрос файловый сервер направляет по сети требуемый блок данных, а СУБД, получив его, выполняет над ними действия, которые были декларированы в прикладной программе.

К технологическим недостаткам указанной модели относят высокий сетевой трафик (передача множества файлов, необходимых приложению), узкий спектр операций манипуляции с данными (здесь данные – это файлы), отсутствие адекватных средств безопасности доступа к данным (защита только на уровне файловой системы) и т.д. Перечисленные особенности не являются недостатками, но имеют место вследствие внутренне присущих FS-модели ограничений. Недоразумения возникают, например, когда FS-модель используют не по назначению: пытаются интерпретировать ее как модель сервера БД. Место FS-модели в иерархии «клиент-сервер» – это модель файлового сервера, поэтому попытки создания на ее основе крупных корпоративных ИС, которые предпринимались в прошлом и нередко предпринимаются в настоящее время, нельзя признать целесообразными.

Более технологичной является RDA-модель, которая существенно отличается от FS-модели характером компонента доступа к информационным ресурсам (как правило, это SQL-сервер). В RDA-модели коды компонента представления и прикладного компонента совмещены и выполняются на компьютере-клиенте, который поддерживает как функции ввода и отображения данных, так и чисто прикладные функции. Доступ к информационным ресурсам обеспечивается либо операторами специального языка (применительно к БД это язык SQL), либо вызовами функций из специальной библиотеки (если имеется соответствующий интерфейс прикладного программирования – API).

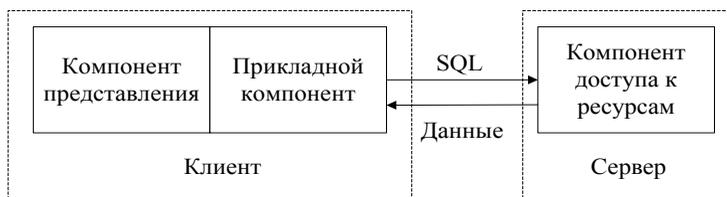


Рис. 4.13. Модель доступа RDA

Клиент направляет запросы к информационным ресурсам (например, к БД) по сети удаленному компьютеру, где функционирует ядро СУБД, которое обрабатывает запросы, выполняя предписанные в них действия, и возвращает клиенту результат, оформ-

ленный как блок данных (см. рис. 4.13). При этом инициатором манипуляций с данными выступают программы, исполняющиеся на компьютерах-клиентах, в то время как ядру СУБД отводится пассивная роль – обслуживание запросов и обработка данных. RDA-модель свободна от недостатков, присущих как системам с централизованной архитектурой, так и системам с файловым сервером: с одной стороны, перенос компонента представления и прикладного компонента на компьютеры-клиенты существенно разгружает сервер БД, сводя к минимуму общее число процессов операционной системы.

Сервер БД освобождается от несвойственных ему функций; процессор или процессоры сервера целиком загружаются операциями обработки данных, запросов и транзакций. Это становится возможным благодаря отказу от терминалов и оснащению рабочих мест компьютерами, которые обладают собственными локальными вычислительными ресурсами, полностью используемыми программами переднего плана. С другой стороны, резко уменьшается загрузка сети, так как по ней передаются от клиента к серверу не запросы на ввод-вывод как в системах с файловым сервером, а запросы на языке SQL, объем которых существенно меньше.

Основное достоинство RDA-модели – унификация интерфейса «клиент-сервер» в виде языка SQL. Взаимодействие прикладного компонента с ядром СУБД невозможно без стандартизованного средства общения – запросы, направляемые программой ядру, должны быть понятны обоим. Однако для этого их следует сформулировать на общем языке специального вида, а поскольку в СУБД уже существует язык SQL, целесообразно использовать его не только как средство доступа к данным, но и в качестве стандарта общения клиента и сервера.

Такое общение можно сравнить с беседой нескольких человек, когда один из них отвечает на вопросы других, а вопросы задаются одновременно и ответы на них даются так быстро, что время ожидания близко к нулю. Высокая скорость общения достигается при этом благодаря четкой формулировке вопроса, когда спрашивающему и отвечающему не нужны дополнительные консультации и собеседники обмениваются короткими однозначными фразами, содержание которых не нужно уточнять.

К сожалению, RDA-модель также не лишена недостатков: во-первых, взаимодействие клиента и сервера посредством SQL-запросов существенно загружает сеть. Во-вторых, удовлетворительное администрирование приложений в RDA-модели практически невозможно из-за совмещения в одной программе разных по своей природе функций (функции представления и прикладные). Поэтому, наряду с RDA-моделью, была предложена DBS-модель (см. рис. 4.14), которая сегодня реализована в реляционных СУБД Informix, Ingres, Sybase, Oracle и др. Ее основу составляет механизм хранимых процедур – средство программирования SQL-сервера. Процедуры хранятся в словаре БД, разделяются между несколькими клиентами и выполняются на том же компьютере, где функционирует SQL-сервер. Язык, на котором разрабатываются хранимые процедуры, представляет собой процедурное расширение языка запросов SQL и уникален для каждой конкретной СУБД.

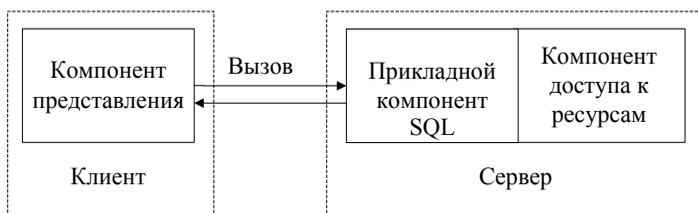


Рис. 4.14. Модель сервера базы данных DBS

В DBS-модели компонент представления выполняется на компьютере-клиенте, в то время как прикладной компонент оформлен как набор хранимых процедур и функционирует на компьютере-сервере БД. Там же выполняется компонент доступа к данным, то есть ядро СУБД. Достоинствами DBS-модели являются возможность централизованного администрирования прикладных функций; снижение трафика (вместо SQL-запросов по сети направляются вызовы хранимых процедур); возможность разделения процедуры между несколькими приложениями; экономия ресурсов компьютера за счет использования единожды созданного плана выполнения процедуры. К недостаткам можно отнести ограниченность средств, используемых для написания хранимых процедур – которые представляют собой расширения SQL, не выдерживающие

сравнения по изобразительным средствам и функциональным возможностям с языками третьего поколения, такими как С или Pascal. Сфера их применения ограничена СУБД, у большинства которых отсутствуют возможности отладки и тестирования разработанных хранимых процедур.

При наличии в сети нескольких серверов RDA- и DBS-модели не обеспечивают требуемой эффективности использования вычислительных ресурсов и в этом случае используют AS-модели. В AS-модели (см. рис. 4.15) процесс, выполняющийся на компьютереклиенте, «отвечает», как обычно, за интерфейс с пользователем (то есть осуществляет функции первой группы). Обращаясь за выполнением услуг к прикладному компоненту, этот процесс играет роль клиента приложения Application Client (AC). Прикладной компонент реализован как группа процессов, выполняющих прикладные функции, и называется сервером приложения Application Server (AS). Все операции над информационными ресурсами выполняются соответствующим компонентом, по отношению к которому AS играет роль клиента. Из прикладных компонентов доступны ресурсы различных типов – БД, очереди, почтовые службы и др.

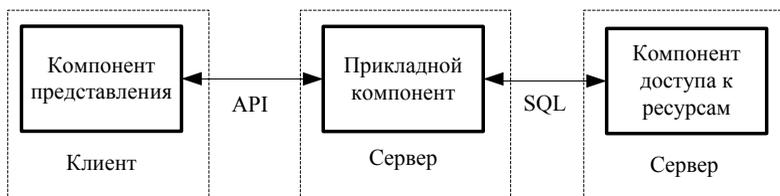


Рис. 4.15. Модель сервера приложений AS

RDA- и DBS-модели опираются на двухзвенную схему разделения функций. В RDA-модели прикладные функции приданы программе-клиенту, в DBS-модели ответственность за их выполнение несет ядро СУБД. В первом случае прикладной компонент сливается с компонентом представления, во втором случае – интегрируется в компонент доступа к информационным ресурсам.

В AS-модели реализована трехзвенная схема разделения функций, где прикладной компонент выделен как важнейший изолированный элемент приложения и для его определения используются универсальные механизмы многозадачной операционной системы,

а также стандартизованы интерфейсы с двумя другими компонентами. AS-модель не требует обеспечения миграции прикладных функций между серверами, что значительно облегчает администрирование системы в целом, однако для обеспечения достаточной скорости обработки данных сервер приложений и сервер БД должны находиться в одной локальной сети или должны быть соединены по выделенному каналу. Изложенное показывает, что в рассматриваемом случае для создания распределенной ИС лучше всего подходит AS-модель.

Трехзвенная модель сервера приложений позволяет сбалансировать нагрузку на разные узлы и сеть, а также способствует специализации инструментов для разработки приложений. Сбалансированная трехзвенная архитектура устраняет недостатки двухзвенной модели и обладает следующими достоинствами:

- четко разделены платформы и инструменты для реализации интерфейса и прикладной логики, что позволяет с наибольшей отдачей использовать их специалистам узкого профиля (каковыми и являются сотрудники предприятия – будущего владельца ИС);
- сервер приложений с помощью монитора транзакций обеспечивает интерфейс с клиентами и другими серверами, он может управлять транзакциями и гарантировать целостность распределенной БД путем двухфазной фиксации в неоднородной среде;
- изменения прикладной логики не затрагивают интерфейса, и наоборот.

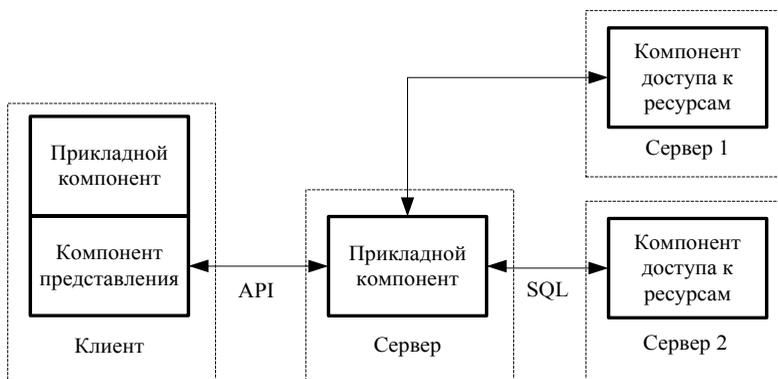


Рис. 4.16. Модификация модели сервера приложений

Для создания более гибкой и динамической системы, эту модель можно модернизировать по схеме, показанной на рис. 4.16. При этом поддержка целостности БД и простейшие прикладные функции поддерживаются хранимыми процедурами (DBS-модель), а более сложные функции реализуются непосредственно в прикладной программе, которая выполняется на компьютере-клиенте (RDA-модель). Предполагается, что для работы с удаленной БД компьютер-клиент вызывает существующую на сервере процедуру, либо поставляет ее в виде SQL-запроса или откомпилированного модуля, а также, что объем поставляемых процедур значительно меньше объема обрабатываемой БД.

При такой организации системы управление транзакциями осуществляют функции сервера приложений, скрывая используемый механизм от клиента. Обработку распределенного запроса (результат составляется из результатов обработки запроса на нескольких серверах) может выполнять клиент или сервер приложений в зависимости от архитектуры конкретной системы, однако второй вариант в большинстве случаев оказывается предпочтительней, вследствие снижения числа обращений клиента к серверу и (или) объема передаваемых между ними данных, а также уменьшения расходов на обработку транзакций за счет применения наиболее простых схем управления. Сервер приложений, в свою очередь, может быть клиентом другого сервера, в таких случаях говорят о многоуровневой системе, когда одни сервера обеспечивают собственно бизнес-приложения, другие – маршрутизацию в сети, третьи управление механизмом транзакций и т.д.

Компьютер-клиент и компьютер-сервер могут работать в условиях локальной вычислительной сети и быть абонентами глобальной сети, общаясь между собой по организуемому виртуальному каналу или используя для этого (при снижении требований к эффективности системы) электронную почту.

Из описания смешанной модели, обладающей свойствами моделей RDA, DBS и AS, видно, что существуют два типа запросов. Запросы первого типа требуют хранения процедур своей обработки на сервере. Запросы второго типа, представляющие собой процедуры обработки, формируются на узле-клиенте, отправляются на сервер и там выполняются. Первый тип запросов именуется *регламентными* или заранее определенными.

Второй – *нерегламентными*, формируемыми клиентом каждый раз, в зависимости от решаемой им задачи. В плане назначения эти типы запросов формально неразличимы, поскольку представляют собой допустимые процедуры обработки – они различаются только частотой выполнения и местом хранения.

Часть регламентных запросов может быть запрещена для исполнения конкретными клиентами – но запросы могут быть запрещены только все сразу, а прочие ограничения на доступ к данным реализуются средствами используемой СУБД или с помощью системы полномочий операционной системы сервера. Регламентные запросы (процедуры) создаются при проектировании системы и должны покрывать совокупность задач, определяемую инструктивными материалами заданной предметной области. При этом, естественно, должна быть предусмотрена возможность расширения состава запросов, в том числе за счет включения в общий список часто употребляемых нерегламентных запросов.

Поскольку в настоящее время общепризнанным считается факт необходимости использования в качестве языка запросов к БД SQL-языка, предполагается, что процедуры обработки регламентных и нерегламентных запросов реализуются на SQL-языке [31; 118] При этом достигаются:

- совместимость запросов с различными СУБД, располагающими SQL-средствам;
- высокая скорость выполнения и лаконичность запросов.

Обеспечение надежного функционирования распределенной ИС предприятия возможно лишь при использовании мощных реляционных СУ БД, таких как Oracle, Informix, SyBase, InterBase. Сравнительные характеристики реляционных СУБД представлены в таблице 4.3. Рассмотрим в качестве примера СУБД Oracle, которая поддерживает язык структурированных запросов SQL, может быть перенесена на различные платформы (операционные системы DOS, Unix, OS/2, Xenix, Macintosh), обладает СУ транзакциями, обеспечивающую целостность БД по чтению и записи, встроенное архивирование и восстановление. Кроме того, Oracle поддерживает архитектуру «клиент-сервер», обеспечивает обработку распределенных запросов, то есть позволяет единичному запросу обращаться к данным физически расположенным на разных компьютерах.

Таблица 4.3. Сравнительные характеристики наиболее распространенных реляционных СУБД.

Вид СУБД	Достоинства	Недостатки	Стоимость для одного пользователя
Informix	Надежность функционирования, динамическая масштабируемость ядра, параллелизм выполнения операций	Отсутствие SQL-версии для платформы DOSWindows.	1000-1500
Oracle	Реализация дополнительных типов данных, единство архитектуры от персонального сервера до сервера предприятия.	Сложность администрирования, большая ресурсоемкость, повышенная чувствительность к ошибкам	1200-2000
SyBaseSystem 11	Надежная система асинхронного тиражирования транзакций с использованием электронной почты, сфера применения от рабочих групп в среде DOS до географически распределенных корпораций и кластерных систем	Недостаточная поддержка симметричной асинхронной репликации	1000-1500

Серверы БД взаимодействуют между собой, вызывая у пользователей, удаленных в пространстве друг от друга, иллюзию работы с единой ИС. СУБД Oracle поддерживает асинхронное отображение данных посредством «снимков» БД, что позволяет хранить локальные копии используемой и удаленной информации с возможностью ее обновления через определенные интервалы времени. Такие моментальные снимки могут отображать отдельные таблицы или их подмножества, а также результаты сложных запросов к данным. Синхронное тиражирование данных позволяет удаленным друг от друга пользователям работать с единой актуальной инфор-

мацией. На основе Oracle успешно функционируют БД объемом в сотни ГБайт, объединяющие информационные потоки и позволяющие осуществлять доступ к любой информации – отчетам, документам, справочникам, планам, фотографиям, картам и схемам.

Следующий этап жизненного цикла ИС – разработка *клиентских приложений*, в результате чего создается готовый продукт, состоящий из ряда приложений, позволяющих пользователям вводить данные в таблицы, редактировать уже существующие данные, анализировать введенные данные, представлять их в более удобном для восприятия виде графиков, сводных таблиц или отчетов, а также сохранять их в виде твердых копий. Проектирования данных для реляционных СУБД является логическим процессом в том смысле, что подчиняется единой стандартной методологии. Это дает низкую степень зависимости последовательности выполняемых действий как от того, какое именно средство проектирования применяется, так и от того, применяется ли оно вообще. Поэтому все применяемые средства проектирования данных в большей или меньшей степени сходны по своему интерфейсу – который, по сути, отражает процесс вычерчивания моделей данных на бумаге.

Напротив, процесс создания клиентских приложений, работающих с БД, в виде подобной универсальной последовательности действий описать достаточно сложно – поскольку логика каждого конкретного приложения зависит от логики моделируемого бизнес-процесса. Средства разработки приложений как категория программных продуктов существуют гораздо дольше, чем средства проектирования данных, они более разнообразны: от компилятора, запускаемого из командной строки, до инструментов, где приложение собирается «мышью» из готовых компонентов, а код генерируется автоматически.

При таком многообразии средств разработки предпочтение следует отдавать средствам разработки с развитыми визуальными инструментами, позволяющими буквально «рисовать» интерфейс, частично стирая различия между работой программиста и пользователя, удешевляя конечный продукт за счет привлечения к проектированию интерфейса разработчиков, обладающих не самой высшей квалификацией. Именно данная категория средств разработки чаще всего применяется при создании клиентских приложений – среди которых к наиболее популярным продуктам относятся

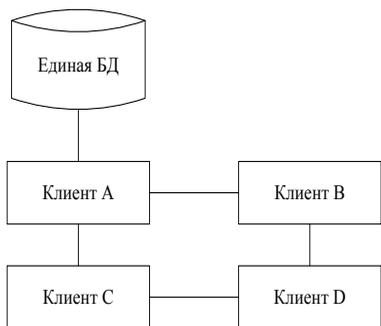
Microsoft Visual Basic, Borland Delphi, Sybase PowerBuilder и Borland C++Builder. Они похожи вплоть до устанавливаемого «по умолчанию» расположения окон на экране: как правило, среда разработки такого продукта содержит:

- «заготовку» проектируемой формы (аналог окна);
- отдельную панель с пиктограммами элементов пользовательского интерфейса и других используемых в приложении объектов, которые можно выбрать и помещать на форму;
- окно, где отображаются и редактируются свойства одного из выбранных на форме элементов (при необходимости и список событий, на которые реагирует данный элемент);
- окно редактора кода, где можно вводить фрагменты кода, связанные с обработкой тех или иных событий, а также код, реализующий логику работы данного приложения и т.д.)отметим, что, как правило, современные средства разработки данного класса позволяют создавать простейшие приложения для ввода и редактирования данных практически без написания кода).

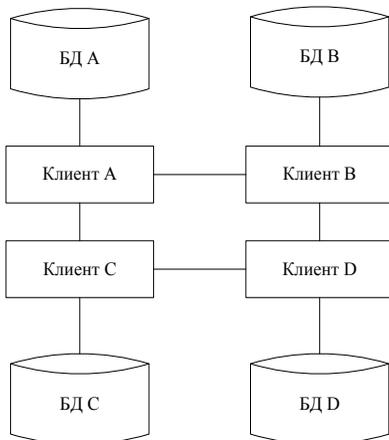
Одно из популярных средств визуальной разработки приложений, использующих БД – продукт Borland Delphi, который является логическим развитием Borland Pascal и, соответственно, использует Object Pascal в качестве языка для написания кода. Доступ к данным в Delphi осуществляется с помощью универсальных механизмов, призванных облегчить создание приложений, и на рынке компонентов Delphi имеется широкий выбор компонентов доступа, реализованных разными производителями.

Выбор оптимальной структуры распределенной БД. При разработке ИС [70; 74- 75] выяснилось, что в рассматриваемой региональной организации действуют свыше 30 отдельных, разрозненных и разнородных БД, предназначенных для решения отдельных мелких задач управления. Поэтому возникла необходимость разработки единой БД для ИС управления по надзору за связью и информатизацией. При этом возникла необходимость рассмотреть вопрос об организации обработки данных, что зависит от способа их распределения. Существуют централизованный, децентрализованный и смешанный способы определения данных [6; 10; 82]

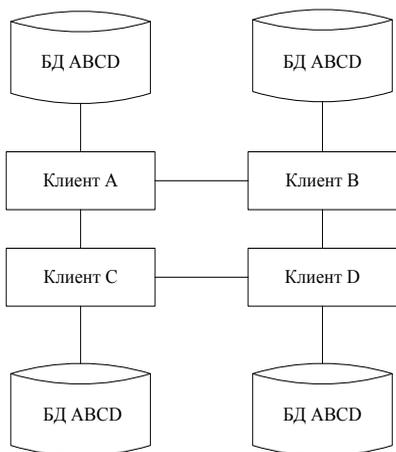
Централизованную организацию данных, наиболее простую для реализации и эксплуатации, иллюстрирует рис. 4.17а.



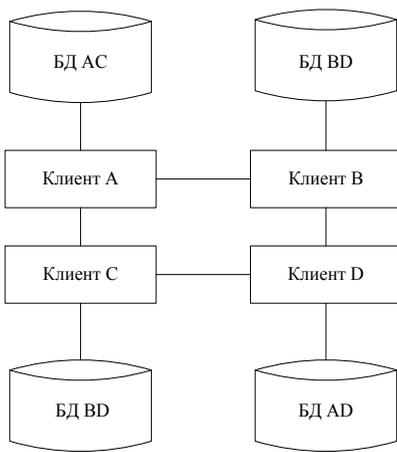
а)



б)



в)



г)

Рис. 4.17. Способы обработки данных в распределенных БД:
 а) централизованная организация данных;
 б) децентрализованная организация способом распределения;
 в) децентрализованная организация способом дублирования;
 г) смешанная организация данных

Согласно рис. 4.17а, на выбранном сервере находится единственная копия БД и все операции с ней обеспечиваются этим сервером – доступ к данным осуществляется посредством удаленного запроса или удаленной транзакции. Недостатками являются ограничение размера БД ввиду ограниченного объема внешней памяти, а также невозможность параллельной обработки – так как все запросы направляются к единственному серверу с соответствующими затратами на стоимость связи и временную задержку. База может быть общедоступной для удаленных пользователей при появлении ошибок связи, но полностью выходит из строя при отказе центрального сервера.

Децентрализованная организация данных предполагает разбиение информационной базы на несколько физически распределенных: при этом клиент пользуется своей БД, которая может быть либо частью общей базы (см. рис. 4.17б), либо копией информационной базы в целом (см. рис. 4.17в), что приводит к возможности ее дублированию для каждого клиента.

При распределении данных путем разделения БД она размещается на нескольких серверах, причем существование копий отдельных частей БД также допустимо.

Достоинства данного метода:

- большинство запросов удовлетворяются локальными базами, что сокращает время ответа;
- увеличиваются доступность данных и надежность их хранения;
- стоимость запросов на выборку и обновление снижается по сравнению с централизованным определением;
- система остается частично работоспособной при выходе из строя отдельного сервера и т.п.

Недостатки связаны с тем, что часть удаленных запросов или транзакций может потребовать доступ ко всем серверам, а это увеличивает время ожидания и цену обслуживания; кроме того, необходимо иметь сведения о размещении данных в различных БД. Однако доступность и надежность БД в схеме, показанной на рис. 17б, по сравнению со схемой на рис. 17а существенно возрастают, поэтому разделенные БД наилучшим образом соответствуют случаю совместного использования локальных и глобальных сетей ЭВМ. При дублировании БД (см. рис. 17в) в каждом сервере компьютерной сети размещается полная БД, что обеспечивает

наибольшую надежность хранения данных. Главное достоинство такого способа: все запросы выполняются локально, что обеспечивает быстрый доступ к данным. Недостатки: повышенные требования к объему внешней памяти; усложнение корректировки информационного содержания баз, так как необходима синхронизация процессов их работы для согласования копий. Данный способ используется, когда фактор надежности является критическим, БД небольшая по объему и интенсивность ее обновления невелика.

Возможна также смешанная организация хранения данных, которая объединяет разбиение и дублирование БД (см. рис. 4.17*з*), и наследует как преимущества, так и недостатки обоих способов. Необходимость хранить информацию о том, где находятся данные в сети, требует компромисса между объемом памяти под базу в целом и под базу в каждом сервере, чтобы обеспечить надежность и эффективность ее работы; но зато легко реализуется параллельная обработка (обслуживание любого распределенного запроса или транзакции). Несмотря на гибкость смешанного способа организации данных, остаются в силе проблемы взаимозависимости факторов, влияющих на производительность системы, а также требования к надежности и объему памяти БД.



Рис. 4.18. Схема организации БД ИС регионального управления по надзору за связью и информатизацией и его филиалов

Считается, что смешанный способ практически оправдан лишь при наличии сетевой СУБД, поэтому для ИС рассматриваемой организации был выбран вариант децентрализованной организации хранения данных способом дублирования (см. схему на рис. 4.18). Объективная необходимость такой формы организации данных обусловлена требованиями, предъявляемыми конечными пользователями:

- централизованное управление рассредоточенными информационными ресурсами;
- повышение эффективности управления базами и банками данных при максимальном уменьшении времени ожидания доступа к информации;
- поддержка целостности, непротиворечивости и защиты данных;
- обеспечение приемлемого уровня в соотношении «цена-производительность-надежность».

Функцию дублирования данных – под которой понимается асинхронный перенос изменений объектов исходной БД (Source Database) в БД, принадлежащие различным узлам распределенной ИС – выполняет специальный модуль СУБД: сервер тиражирования данных, называемый репликатором (Replicator).

Его задачей является обеспечение идентичности данных в принимающих БД (Target Database) данным в исходной БД. Сигналом для запуска репликатора служит срабатывание правила, перехватывающего подобные изменения тиражируемого объекта БД. Возможно и программное управление репликатором посредством сигнализаторов о событиях в распределенной БД.

В качестве базиса для дублирования выступает транзакция к БД. В то же время возможен перенос изменений группами транзакций, периодически или в нужный момент времени, что позволяет исследовать текущее состояние принимающей БД. Подробности процесса тиражирования данных полностью скрыты от прикладной программы – ее функционирование никак не зависит от работы репликатора, который целиком находится в ведении администратора БД. Следовательно, для переноса программы в распределенную среду с тиражируемыми данными не требуется ее модификации.

4.7. Принципы построения расчетно-диагностических ЭС

Как уже было отмечено, советующая ИС создается под конкретные цели и требования руководства корпорации. Центральным компонентом, определяющим уникальность каждой такой ИС, является БЗ, структура которой реализует специфику целей управления. На рис. 4.19 представлена последовательность выполнения этапов создания БЗ, каждый из которых делится на подэтапы.

Этап 1. Идентификация решаемых задач и текстовое описание их предметной области.

Определение типа задачи, целей и подцелей ее решения. Важным с точки зрения инженера по знаниям является определение типа каждой решаемой задачи. Спецификой применения БЗ обусловлено деление задач на два типа. **Первый тип** характеризуется трудностями в формализации, неизвестностью последовательности шагов решения, отсутствием алгоритма, применением эвристических методов, большим объемом логических операций вывода, отсутствием или небольшим объемом операций вычислительного и информационно-поискового характера. **Второй тип** отличается от первого наличием хорошо формализованных алгоритмов, использованием операций поискового и информационно-справочного характера, ориентацией на вычислительные процедуры, большим объемом обрабатываемых данных.

Соотнесение задачи к одному из указанных типов полезно с точки зрения предварительного анализа эффективности использования конкретных средств, обеспечивающих ЭС. Однако практика показывает, что реальные задачи часто характеризуются признаками, присущими обоим типам одновременно – например, в экономической ЭС диагностирующего характера при расчете величин финансовых показателей преобладают признаки задач второго типа, а при формировании рекомендаций – признаки задач первого типа. Однако вне зависимости от типу задачи, необходимо прежде всего выяснить цели и подцели ее решения – информацией для этого служат проблемы, стоящие перед корпорацией. Эти проблемы можно выявить из отчетов о предпроектном обследовании рассматриваемого объекта: предприятия или организации. Результаты обследования оформляются в виде специальной проектной таблицы, которая сдержит перечень проблем и соответствующие им цели и подцели решения каждой задачи.



Рис. 4.19. Технология проектирования БЗ

Выполнение стандартных процедур по созданию любой ИС (см. в разделе 3) предполагает проведение работ по следующим подэтапам:

- выявление состава и объемов, а также разработка формы представления результирующей информации;
- выявление состава и объемов входной информации;
- описание промежуточной (если таковая предусматривается) информации.

Текстовое (словесное, вербальное) описание предметной области. Используя одну или несколько стратегий получения знаний (приобретение, извлечение и формирование) [22-23; 86; 97] необходимо получить вербальное описание предметной области решаемой задачи. Основой для данного описания служат цели и подцели, исходная и результирующая управленческая документация, включая непосредственный опыт ЛПР, выступающего в роли эксперта, а также специальная литература (методики, руководства, рекомендации, статьи, монографии, учебники и учебные пособия).

Этап 2. Выделение и графическое представление понятий и их связей. На данном этапе происходит первичная формализация знаний, полученных на предыдущем этапе. Здесь предусматривается выполнение следующих подэтапов.

Выделение из предложений текста описания понятий, отражающих объекты, процессы и состояния. На данном подэтапе инженер по знаниям с помощью эксперта определяет ключевые понятия (слова), отношения и характеристики, необходимые для описания процесса решения задачи.

Определение связей между понятиями. После выделения ключевых понятий необходимо определить связи между ними, что необходимо для последующего формирования дерева целей, а также лучшего понимания существующей взаимосвязи между ключевыми понятиями.

Графическое представление понятий и связей. Цели и связи между ними обычно представляют в виде дерева целей с указанием конъюнктивных и дизъюнктивных вершин.

Этап 3. Представление знаний о предметной области и решаемых задачах в терминах выбранной модели представления знаний. Данный этап, на котором осуществляется трансформация графиче-

ского изображения БЗ в формальное представление, включает следующие подэтапы.

Представление символьно-логического пути вывода для каждого запроса по принципу «от обратного». На основе дерева целей строится цепочка правил логического вывода заключения.

Представление на полужормальном языке правил вывода. Используя цепочку правил логического вывода заключения, составляется таблица правил (продукций) с указанием коэффициентов уверенности.

4.8. Выводы

Принцип максимума Понтрягина, который сыграл выдающуюся роль в создании и развитии теории оптимального управления, а также самой оптимизационной идеологии, неприменим к нерелефторным СС иерархического типа (корпорации, компании, фирмы и другие элементы рыночной среды). Это требует не только пересмотра принципов разработки соответствующих СУ, но и отказа от указанной идеологии моделирования СС на основе вариационного исчисления и теории дифференциальных уравнений. На их место приходят методы и средства управления СС с применением новых информационных технологий – в том числе метода СИМ по версии МДМ, специально предназначенный, в том числе, для квазиоптимального управления нерелефторными СС, которые подвержены внутренним и внешним возмущениям случайного (стохастического) характера.

Случайные возмущения можно рассматривать как фактор неопределенности знаний ЛПР (верифицированных и аксиологических), осложняющий его действия – ввиду объективных (недостаток информации, присутствие помех, наличия нескольких целей и т.п.), и субъективных (неопределенность намерений самого ЛПР, противодействия со стороны конкурента или злоумышленника, отсутствии взаимодействия с партнерами) обстоятельств. Наибольший интерес при управлении нерелефторными СС представляют аксиологические знания в виде гипотез поведения их звеньев, которые выдвигают ЛПР в условиях антагонистической (с противоположными интересами) игры фон Неймана и игры с противоположными интересами Гермейера. Введение гипотез поведения звеньев позволяет в ряде случаев изучать нерелефторные СС и

управлять ими по аналогии с рефлекторными, вкладывая специально оговоренный смысл в понятие оптимальности.

Для реализации квазиоптимального управления нерелекторными СУ необходимы иерархические СУ, определение целевых функций для звеньев (подсистем и элементов) которых представляет собой актуальную комплексную проблему – мощным средством информационно-технологической поддержки процесса решения которой является метод СИМ по версии МДМ.

В традиционной системе обработки данных системы понятий, относящиеся к предметной области и к формальной ММ, положенной в основу ИС, как правило, не совпадают. Это различие является основной причиной затруднений, возникающих при взаимодействии ЛПР с компьютером в процессе решения задачи. В этой связи разработчики ИИС и ЭС, как важной составляющей актуальных ИТ-решений, ставят своей целью изменение традиционных подходов к системе взаимодействия ЛПР и ЭВМ, обеспечение удобства и комфорта пользователя, повышения эффективности взаимодействия подсистем и элементов СУ. Особенно эффективны ИИС применительно к слабоструктурированным задачам, где отсутствует строгая формализация и применяются эвристические процедуры, позволяющие в большинстве случаев получить удовлетворительное с точки зрения практики решение. Диапазон применения ИИС необычайно широк: от управления технологическими процессами в до оценки последствий игры на фондовой бирже.

По мере совершенствования принципов логического и правдоподобного вывода, применяемых в ИИС за счет использования нечеткой, модальной, временной логики и байесовских сетей вывода, ИС начинают проникать в высокоинтеллектуальные области, связанные с разработкой стратегических решений по совершенствованию деятельности СУ организационно-технического типа (предприятий, корпораций, фирм и т.д.). Этому способствуют как современные алгоритмы анализа и синтеза предложений естественного языка, облегчающие общение пользователя с ИС, так и средства ИТ, необходимые для организации удаленного доступа к распределенным БД. Наличие в составе ИИС объектно-ориентированных БД позволяет обеспечить хранение и актуализацию как фактов, так и знаний, необходимых для эффективного управления объектами различного назначения.

5. ПОСТАНОВКА И ПРИМЕРЫ РЕШЕНИЯ ЗАДАЧ РЕАЛИЗАЦИИ ЭКСПЕРТНЫХ И СОВЕТУЮЩИХ СИСТЕМ

5.1. Предметная область ЭС диагностики неисправностей оборудования ИКК

Диагностика является одной из традиционных областей применения ЭС. Под диагностикой (от греч. *diagnostikos* – способный распознавать) в общем случае принято понимать процесс распознавания и оценки свойств, особенностей и состояний СС как объекта исследования, который заключается в целенаправленном изучении, истолковании полученных результатов и их обобщении в виде заключения (диагноза).

Актуальной тенденцией при организации такого рода работ является создание и внедрение ЭС диагностического типа, что ведет к экономии фонда заработной платы (дешевле приобрести эффективную ЭС, чем содержать многочисленный высококвалифицированный персонал – экспертов и ЛПР); повысить эффективность проводимой диагностики (отсутствует влияние «человеческого фактора»); организовать СУ знаниями, накапливая в едином хранилище информацию об инцидентах и способах их устранения и т.д. В то же время полная замена персонала ЭС невозможна по следующим причинам:

- любая ЭС отражает конкретный (субъективный) взгляд на процедуру диагностики СС, соответствующий мнению одного или нескольких экспертов;
- мощность БЗ, даже в идеальном случае, равна, а на самом деле всегда меньше суммы знаний, которыми обладают эксперты.

Однако ЭС диагностического типа вполне способна корректно обрабатывать и диагностировать наиболее характерные, часто возникающие проблемы в той или иной предметной области. Также следует отметить распространенное терминологическое расхождение: наряду с понятием «экспертная система» в литературе часто встречаются такие синонимичные словосочетания, как «советующие системы», «системы, основанные на знаниях», «системы, основанные на прецедентах» и др. Данная ситуация вызвана как обилием классификационных признаков ИИС (подробнее см. в первом разделе), так и неточностью перевода зарубежных публикаций.

Наиболее востребованными направлениями использования ЭС в области диагностики является медицина и поиск неисправностей оборудования – хотя возможности применения ЭС диагностического типа ими сегодня далеко не исчерпываются. В данной главе представлены типовые примеры реализации ЭС диагностического характера, ориентированных на применение в деятельности ИКК. По каждому из примеров приводится описание этапов проектирования и разработки ЭС, которые были рассмотрены в третьем разделе настоящей монографии.

Инфокоммуникационная отрасль, с точки зрения роли влияния на нее ИТ, представляет значительный интерес, поскольку успешность развития ИКК существенно зависит как от общего мирового прогресса в данной области, так и от специфики их применения в конкретных ИКК – что также стимулирует развитие ИТ [38; 85]. Решение задачи эффективного биллинга, например, поначалу сводилось к накоплению данных о разговорах клиентов и привело к использованию автоматизированных средств обработки информации, а затем – к необходимости анализа и получения скрытых знаний из данных, что стимулировало развитие таких направлений, как OLAP и BigData. Таким образом, совершенствование ИТ-сферы – это одно из ключевых направлений развития ИКК, которое не только обеспечивает деятельность офисных ИТ-приложений, но и влияет на результаты бизнеса в целом, что требует внедрения отраслевых ИТ-продуктов, которые в ряде случаев являются уникальными и единичными для российского рынка.

Для решения проблемы управления ИТ-ресурсами и повышения эффективности их использования предназначена библиотека ITIL (Information Technology Infrastructure Library), которая содержит все лучшие на сегодняшний день способы организации работ в подразделениях и компаниях, занятых предоставлением услуг в области ИТ. Сегодня ITIL фактически является стандартом в области организации и управления ИТ, причем использованный в данной библиотеке процессный подход полностью соответствует стандартам серии ISO 9000 (ГОСТ Р ИСО 9000). Методология ITIL применяется для оптимизации процессов обслуживания ИС в интересах управления СС.

Основные принципы библиотеки ITIL – сервисная модель работы, измерения и контроль качества сервисов, а также построение

отношений между бизнесом и ИТ на базе договоров об уровне сервисов (Service Level Agreement – далее SLA). Для реализации этих принципов ИТIL рекомендует использование специализированных процессов управления ИТ, которые нацелены не просто на обеспечение бесперебойной работы всех компонентов ИТ-инфраструктуры, но в гораздо большей степени ориентированы на выполнение требований пользователей и заказчика. В соответствии с содержанием ИТIL, выделяют десять базовых процессов, указанных в таблице 5.1.

В рамках «Управления инцидентами» нормальный процесс обслуживания оборудования ИКК определяется как обслуживание, соответствующее по эффективности целям соглашения об уровне услуг (см. SLA), соглашения об уровне работоспособности (Operational Level Agreement – OLA) и договора поддержки (Underpinning Contract – UC).

Таблица 5.1. Процессы ИТIL

Процесс	Цель
Управление инцидентами (incident management)	Устранение инцидентов в предельно сжатые сроки. Инцидент – это любое событие (сбой, запрос на консультацию и т. п.), которое может привести к снижению качества предоставления услуги. Для успешного управления инцидентами необходимо создание диспетчерской службы (service desk), которая должна являться единой точкой контакта с пользователями и координировать вопросы устранения инцидентов
Управление проблемами (problem management)	Сокращение количества инцидентов за счет выявления и устранения причин их возникновения
Управление конфигурациями (configuration management)	Создание и поддержание в актуальном состоянии логической модели инфраструктуры компании
Управление изменениями (change management)	Проведение изменений с наименьшим риском возникновения инцидентов, вызванных изменениями

Таблица 5.1. Процессы ИТІЛ (окончание)

Процесс	Цель
Управление релизами (release management)	Обеспечение работоспособности производственной среды при внедрении изменений
Управление уровнем сервиса (service level management)	Определение состава и уровня сервиса на основании требований потребителей и поставщиков ИТ-сервисов, контроль за достижением установленного уровня сервиса
Управление финансами (financial management for IT services)	Обеспечение стабильного финансирования всех прочих процессов
Управление мощностью (capacity management)	Исключение инцидентов, возникающих по причине недостаточной мощности ИТ-инфраструктуры компании, причем без неоправданных затрат на приобретение излишних, неиспользуемых мощностей
Управление непрерывностью (IT service continuity management)	Обеспечение гарантированного восстановления уровня ИТ-сервисов до величины, необходимой для продолжения бизнес-операций, за определенный промежуток времени в случае чрезвычайной ситуации (пожара, отключения электроэнергии и пр.)
Управление доступностью (availability management)	Обеспечение согласованного уровня доступности сервиса, а также оценка текущей доступности сервиса и планирование действий, направленных на его дальнейшее улучшение

Отчет и регистрация инцидентов производятся персоналом технической поддержки, который может направить уведомление в службу поддержки пользователей в случае обнаружения той или иной неисправности. Отметим, что не все события регистрируются в качестве инцидентов – поскольку многие классы событий не связаны с нарушениями, но являются индикаторами нормального функционирования оборудования ИКК или просто предоставляют информацию о них. Процесс «Управление инцидентами» включает регистрацию, исследование, диагностику и устранение инцидентов.

5.2. Предпроектная стадия создания ЭС диагностики неисправностей оборудования ИКК

Предпроектная стадия разработки ЭС начинается с формирования схемы бизнес-процесса (БП) по устранению инцидента, приведенной на рис. 5.1. Информация о наступлении инцидента может поступать в ЭС одним из двух способов: как входящая заявка от клиента через систему Help-desk, либо непосредственно от системы мониторинга сети. Заявка регистрируется дежурным диспетчером, который на каждый инцидент открывает так называемый Trouble-ticket – где фиксирует имеющуюся информацию об инциденте, на основании которой осуществляется классификация его типа – сбой или обращение, что определяет выбор способа обработки заявки.

Классифицировать инцидент дежурный может как самостоятельно, основываясь на собственном опыте, так и с помощью ЭС диагностики неисправностей. Обязанности по устранению сбоя возложены на дежурных инженеров, которые в интерактивном режиме локализуют зону аварии, определяют тип аварии и получают информацию о схожих инцидентах, имевших место быть в СС, а также рекомендациях по устранению сбоя. Если для устранения аварии необходим выезд к клиенту или на узел связи, ЭС выдает рекомендации по устранению аварии непосредственно у клиента или на данном узле. Если инцидент не имеет аналогов и представляет собой достаточно сложный случай, к вопросу устранения сбоя подключаются эксперты – сотрудники технических отделов ИКК.

После устранения инцидента информация о нем анализируется инженером по знаниям – который по итога анализа вносит изменения и добавления в БЗ, добавляя или корректируя наборы соответствующих правил. Кроме того, ЭС можно «дооснастить» элементами CRM-System [39; 43-44], например, если клиенту выгоднее общаться с одним и тем же оператором в целях экономии времени, чтобы не объяснять каждую свою типовую проблему по многу раз; с другой стороны, оператор службы технической поддержки не в состоянии удержать в памяти всю информацию о пользователях и проблемах.

При поступлении заявки от клиента целесообразно перенаправить звонок на «индивидуального» оператора».

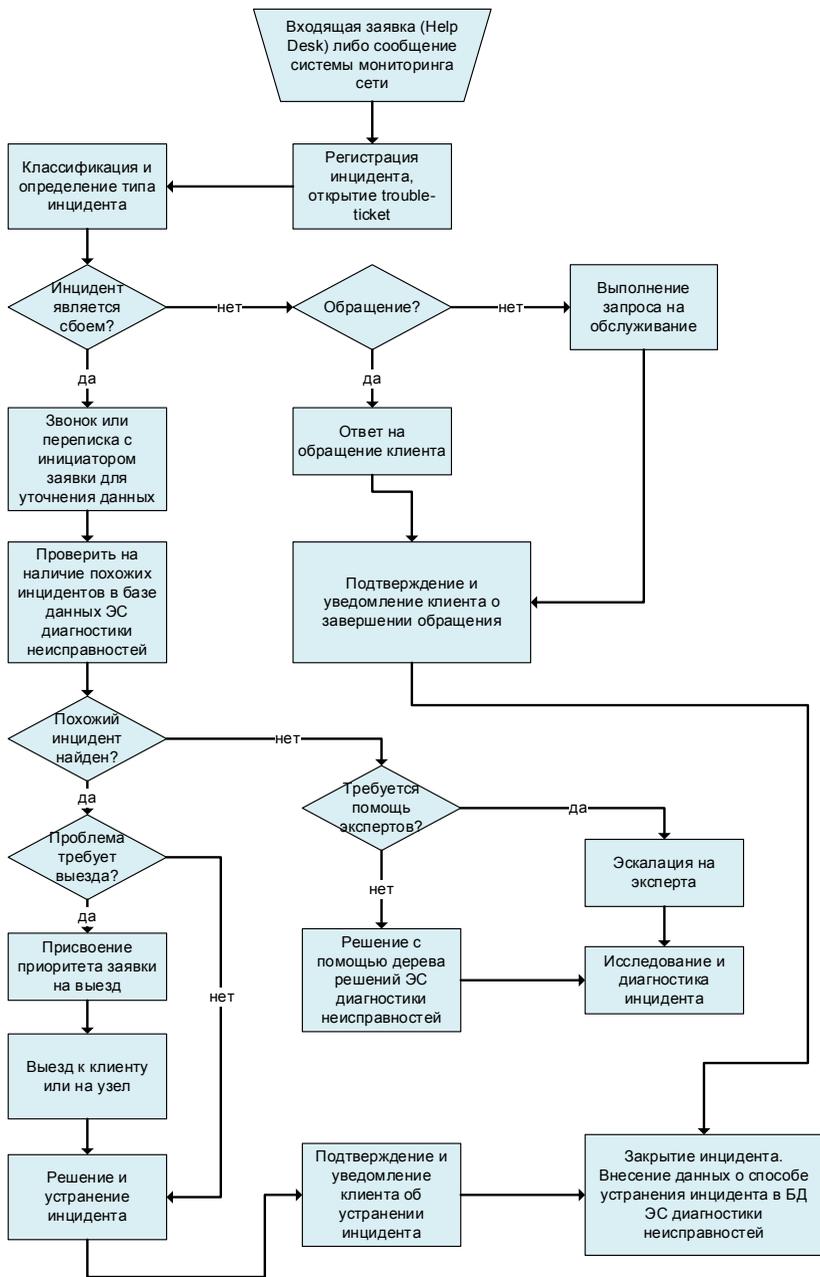


Рис. 5.1. Схема регламента работы с инцидентом

Таким образом, в данном случае ЭС предлагается использовать как центральное звено СУ, диагностирующее сбои и повреждения оборудования ИКК, а также предлагающее рекомендации по их устранению. Поскольку диагностику инцидентов сможет осуществлять наименее квалифицированный персонал, это позволит сэкономить на фонде заработной платы, так как ИКК будет дешевле содержать «смешанный» штат центра технической поддержки.

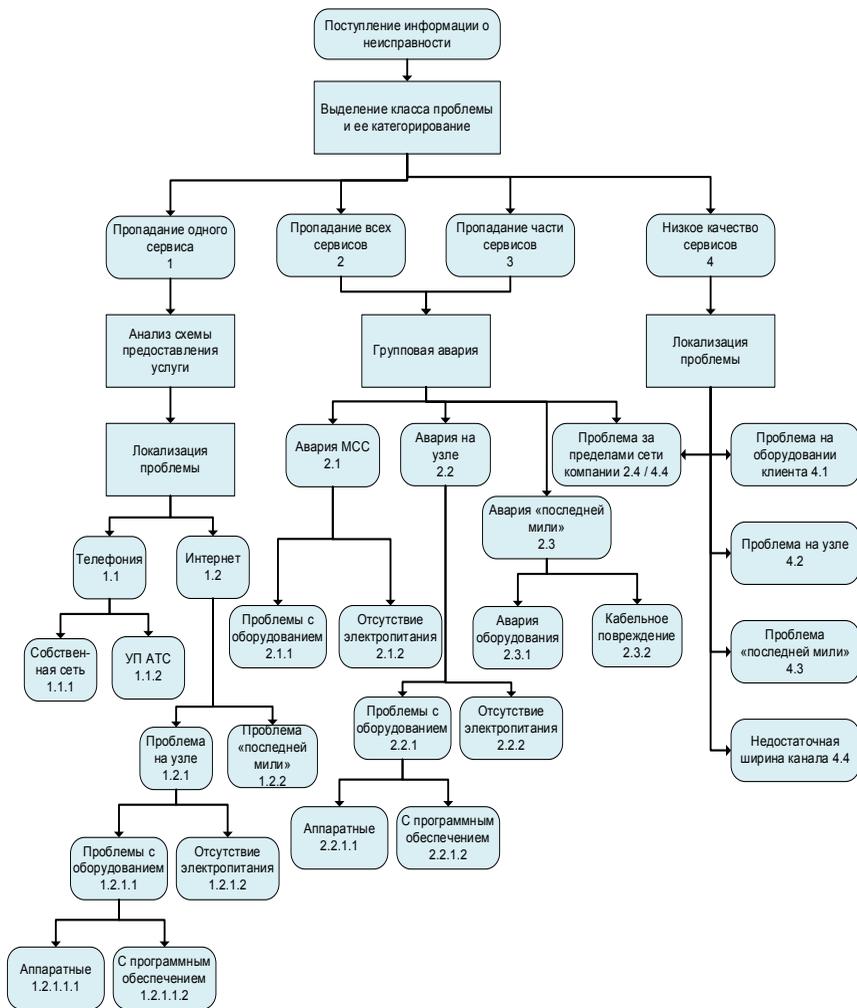


Рис. 5.2. Классификация неисправностей в сети ИКК

5.3. Стадия проектирования ЭС диагностики неисправностей оборудования ИКК

Стадия проектирования ЭС начинается с разработки протокола действий эксперта. В рамках стадии проектирования интерес представляет процесс формирования БЗ. На основании заключений экспертов телекоммуникационной компании, а также с учетом общепринятой практики диагностики и устранения отказов была сформирована общая схема диагностики отказов (см. рис. 5.2).

В дальнейшем при разработке протокола действий эксперта необходимо разработать деревья диагностики неисправностей.

Все возникающие проблемы были разделены на три класса:

- пропадание одного сервиса (см. рис. 5.3);
- пропадание всех сервисов (см. рис. 5.4);
- низкое качество сервисов (см. рис. 5.5).

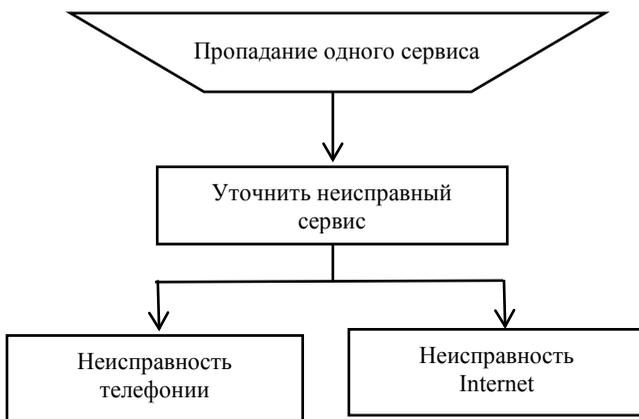


Рис. 5.3. Пропадание одного сервиса

Такая проблема, как пропадание одного сервиса, подразумевает пропадание у клиента либо телефонии, либо Internet. Пропадание всех сервисов подразумевает пропадание у клиента одновременно и телефонии и Internet. Рассмотрим действия эксперта при возникновении проблем каждого класса.

Неисправность телефонии. В данном случае первый шаг эксперта – уточнить узел подключения клиента. Если клиент подключен через аналоговую АТС (АТС старого типа), необходимо сооб-

щить клиенту о невозможности использования телефонного аппарата с тональным набором. Если клиент использует современную АТС, необходимо уточнить наличие офисной АТС.

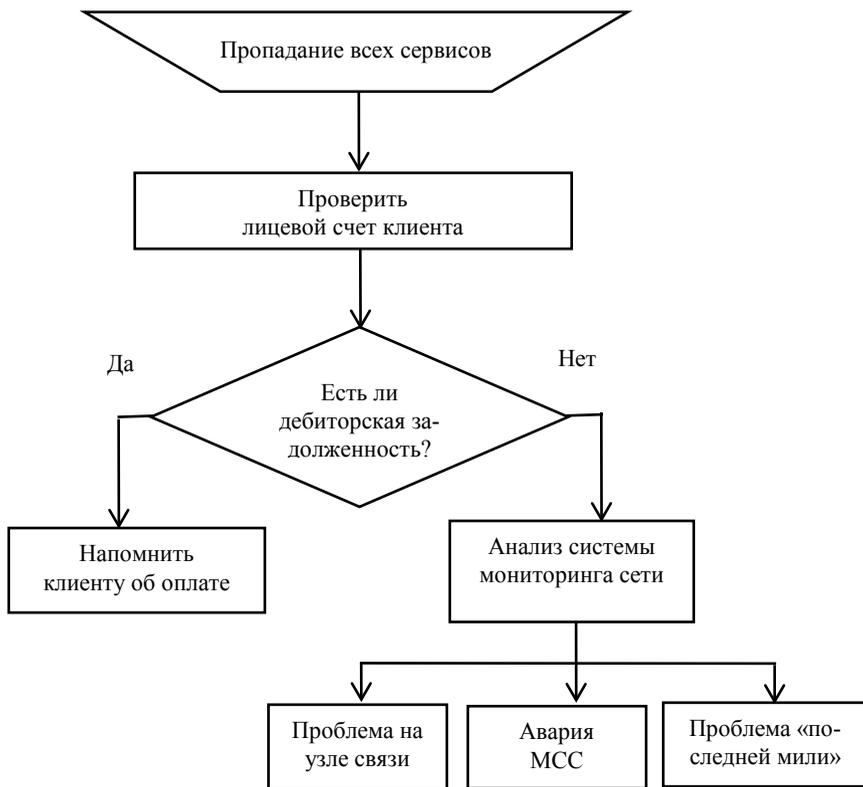


Рис. 5.4. Пропадание всех сервисов

При наличии у клиента офисной АТС необходимо узнать (у него самого или по имеющимся БД), под чьим управлением она находится. Если станция находится не под управлением ИКК, в техническую службу которой обратился клиент, то эксперт рекомендует клиенту решить проблему самостоятельно или обратиться в организацию, обслуживающую данную офисную АТС. Если станция находится под управлением компании сотовой связи (КСС), необходимо попросить клиента проверить, включено ли

владельцем сети оборудование. В случае положительного ответа производится уточнение проблемы.

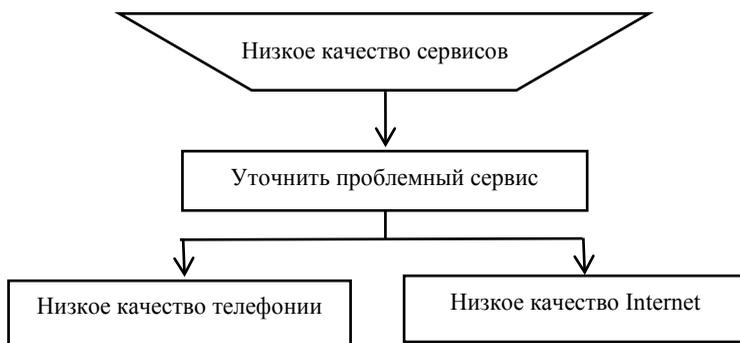


Рис. 5.5. Низкое качество сервисов

Проблемы с телефонией делятся на три группы:

- нет выхода на междугородную (или международную) сеть (см. рис. 5.7);
- нет выхода на городскую линию (см рис. 5.8);
- одновременно нет выхода на междугородную (международную) сеть и на городскую линию (см. рис. 5.9).

Анализ данных проблем начинается с уточнения у клиента наличия зуммера ответа станции. Если у клиента зуммер ответа станции есть, необходимо проверить лицевой счет клиента. При наличии дебиторской задолженности, эксперт напоминает клиенту об оплате услуг. В случае, когда клиент жалуется на отсутствие выхода на междугородную (международную) сеть, эксперт проверяет, подключен ли клиент к данной услуге. Если клиент подключен, необходимо проверить правильность набора клиентом кодов выхода в междугородную (международную) сеть и правильность набора конкретных кодов городов или стран.

Правильность набора номера на зону необходимо проверять, используя соответствующий список кодов. Если у клиента нет выхода на городскую линию, необходимо проверить соседние городские номера. В случае если соседние номера работают, то проблема на стороне вызываемого абонента.

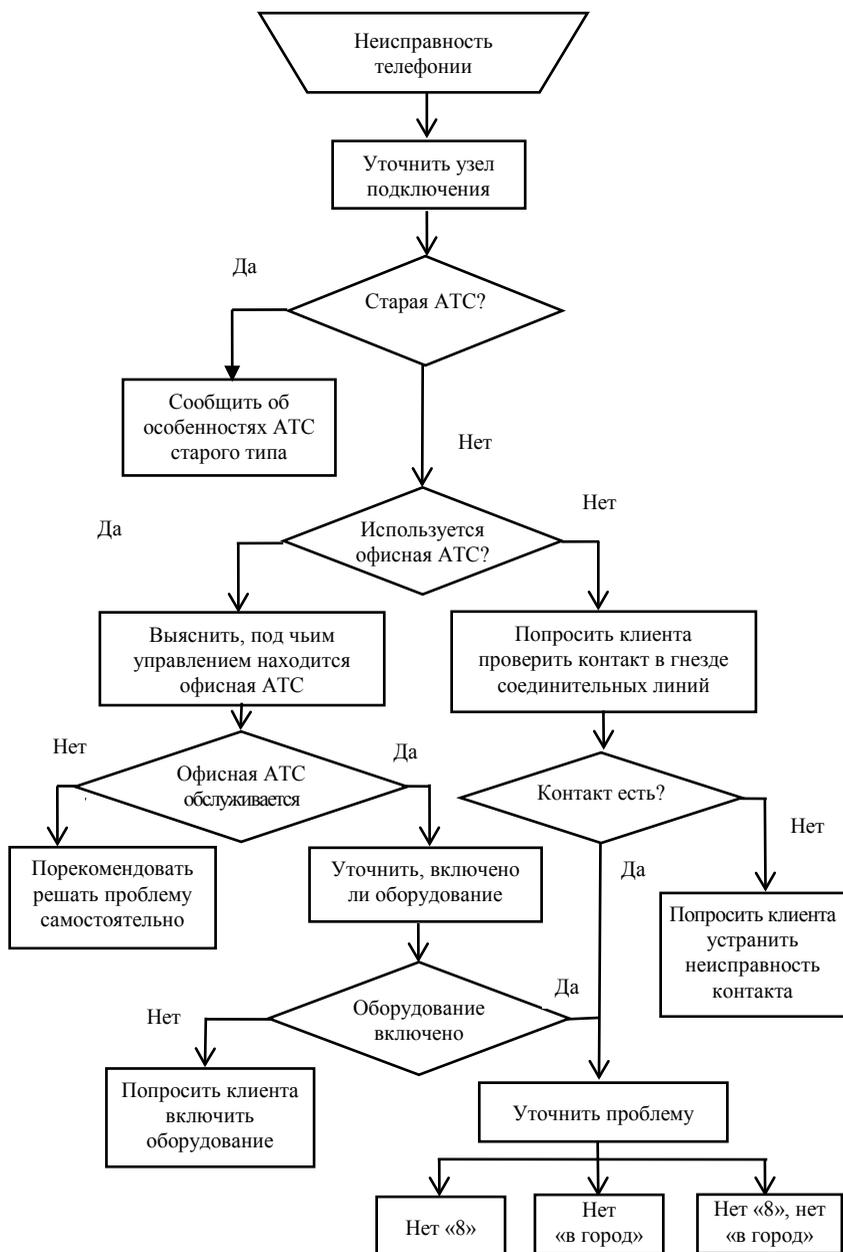


Рис. 5.6. Неисправность телефонии

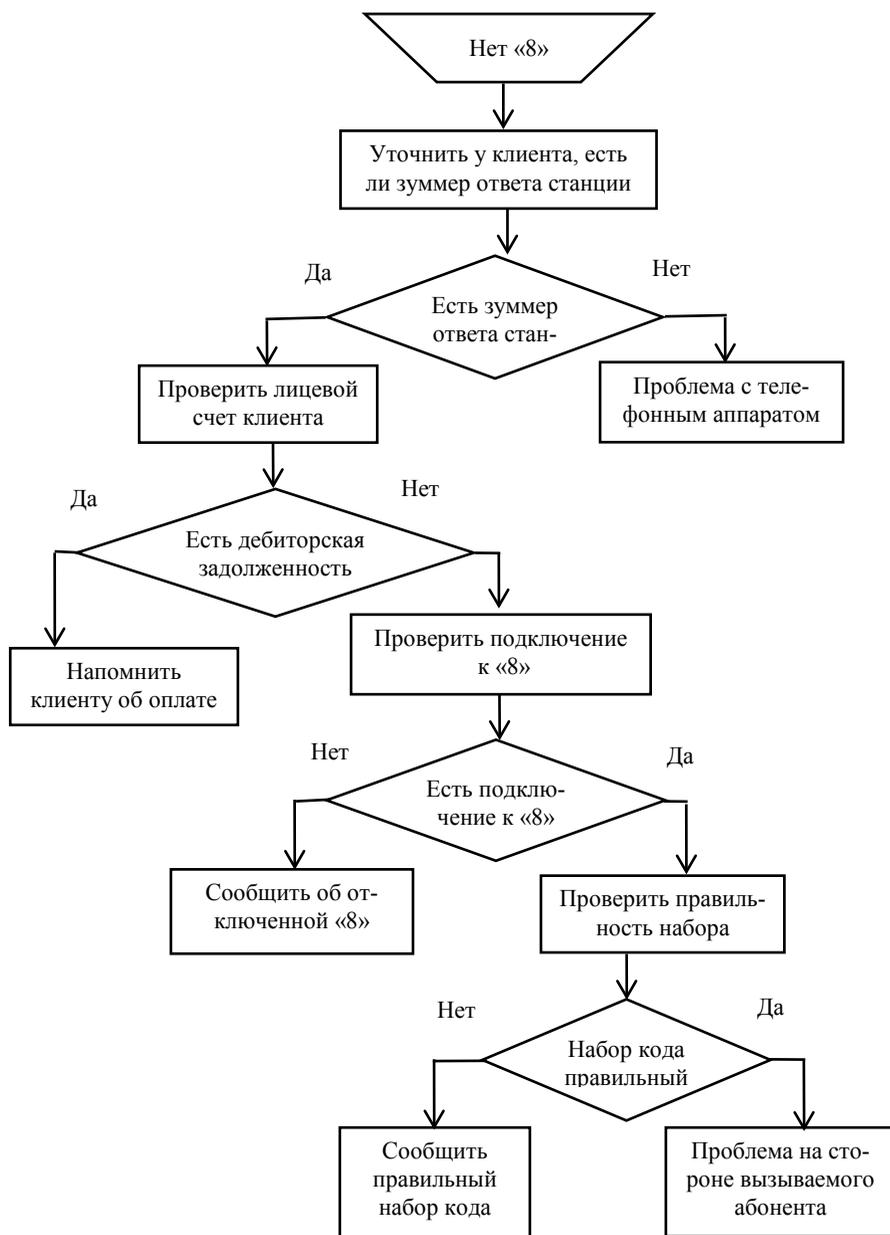


Рис. 5.7. Неисправность выхода на междугородную (международную) сеть

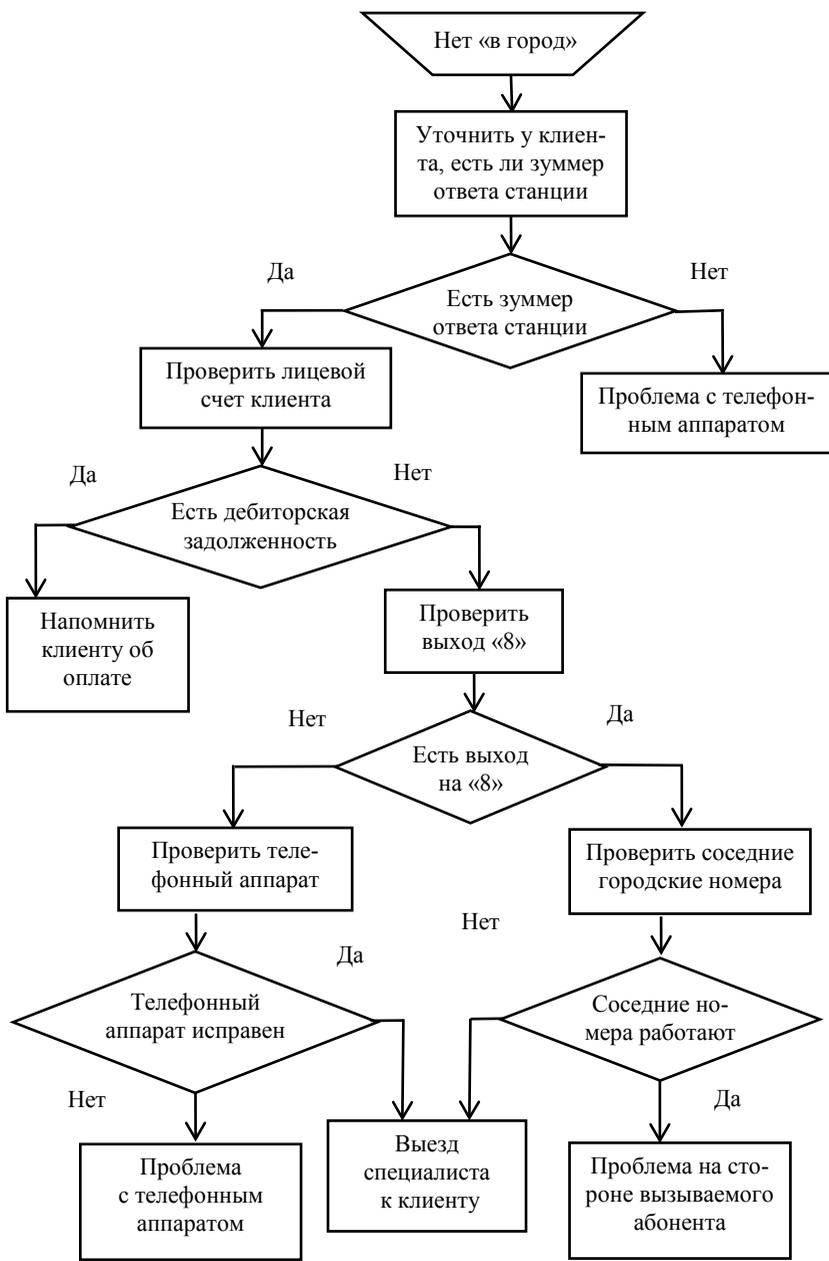


Рис. 5.8. Неисправность выхода на городские номера

В случае, когда у клиента одновременно нет выхода на между-городную (международную) сеть и на городскую линию и нет дебиторской задолженности, эксперт просит клиента проверить телефонный аппарат.

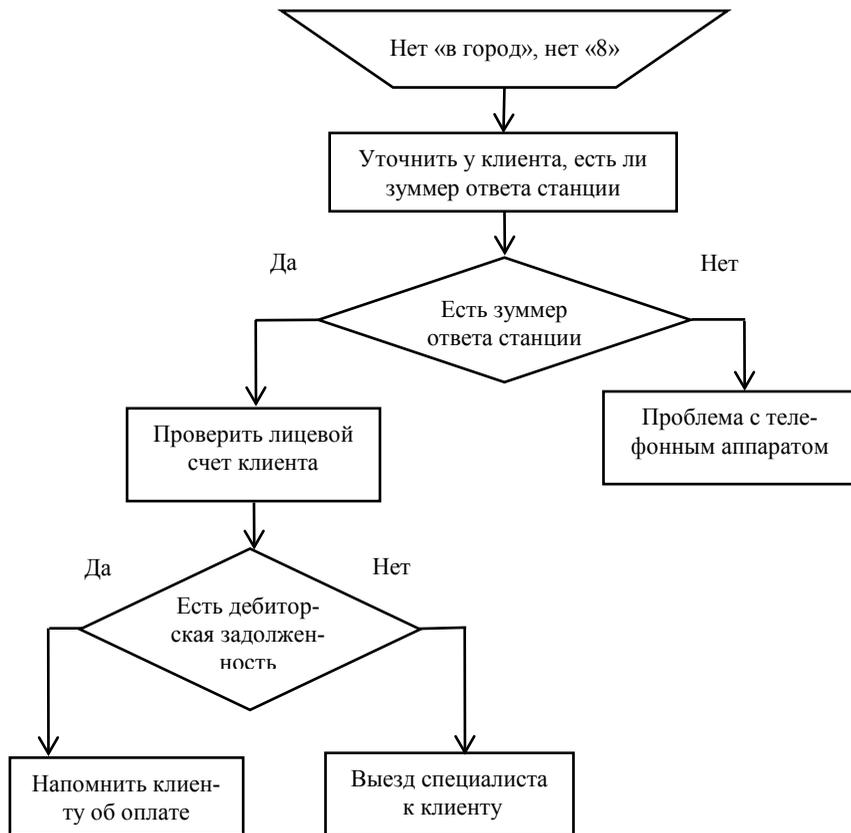


Рис. 5.9. Неисправность выхода на городские и междугородние (международные) номера

Пропадание сети Internet. Для анализа проблемы необходимо определить IP адрес серийного интерфейса на роутере (оконечное оборудование) клиента и сделать на него трейсруут. В случае если трейсруут не доходит до необходимого адреса, эксперт должен проанализировать систему мониторинга сети. Если система не выявила наличие проблемы, требуется выезд специалиста к клиенту.

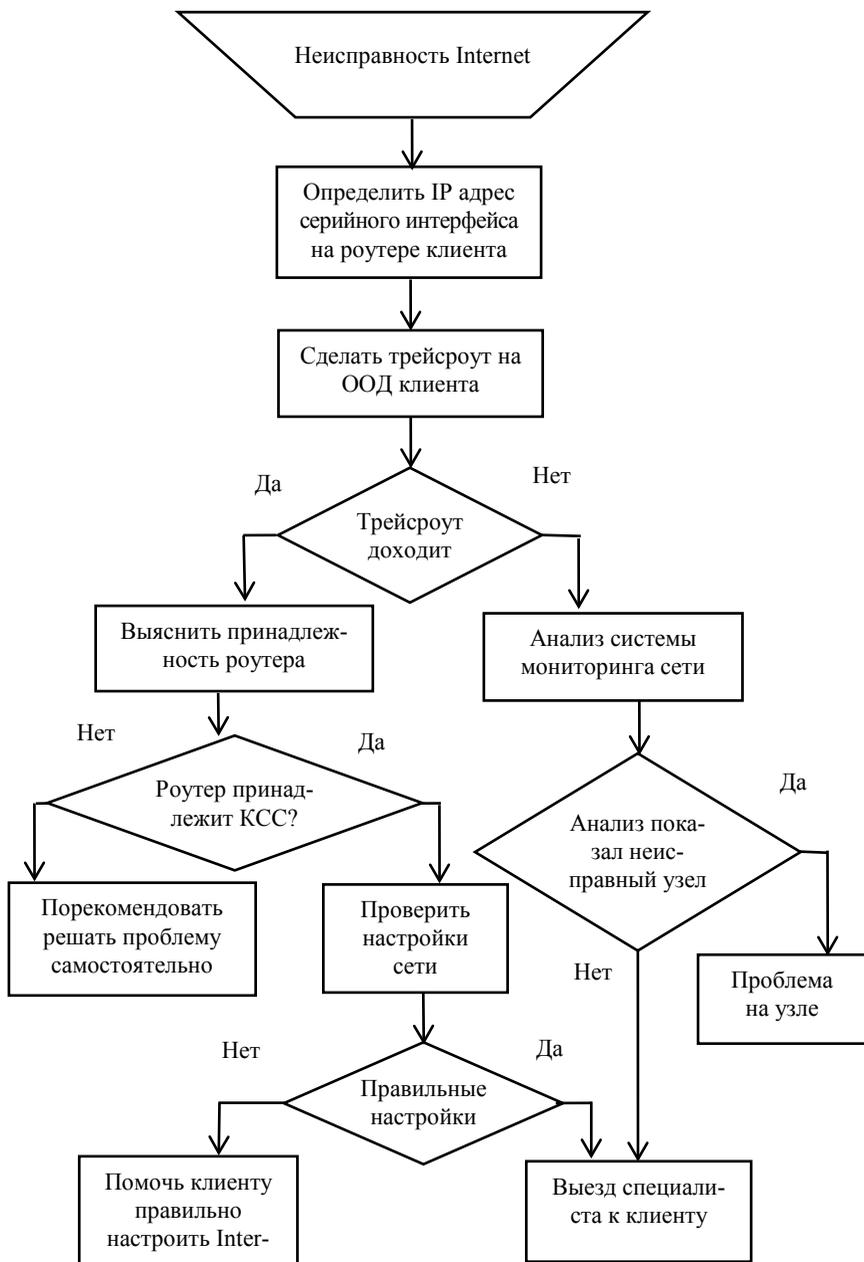


Рис. 5.10. Пропадание Internet

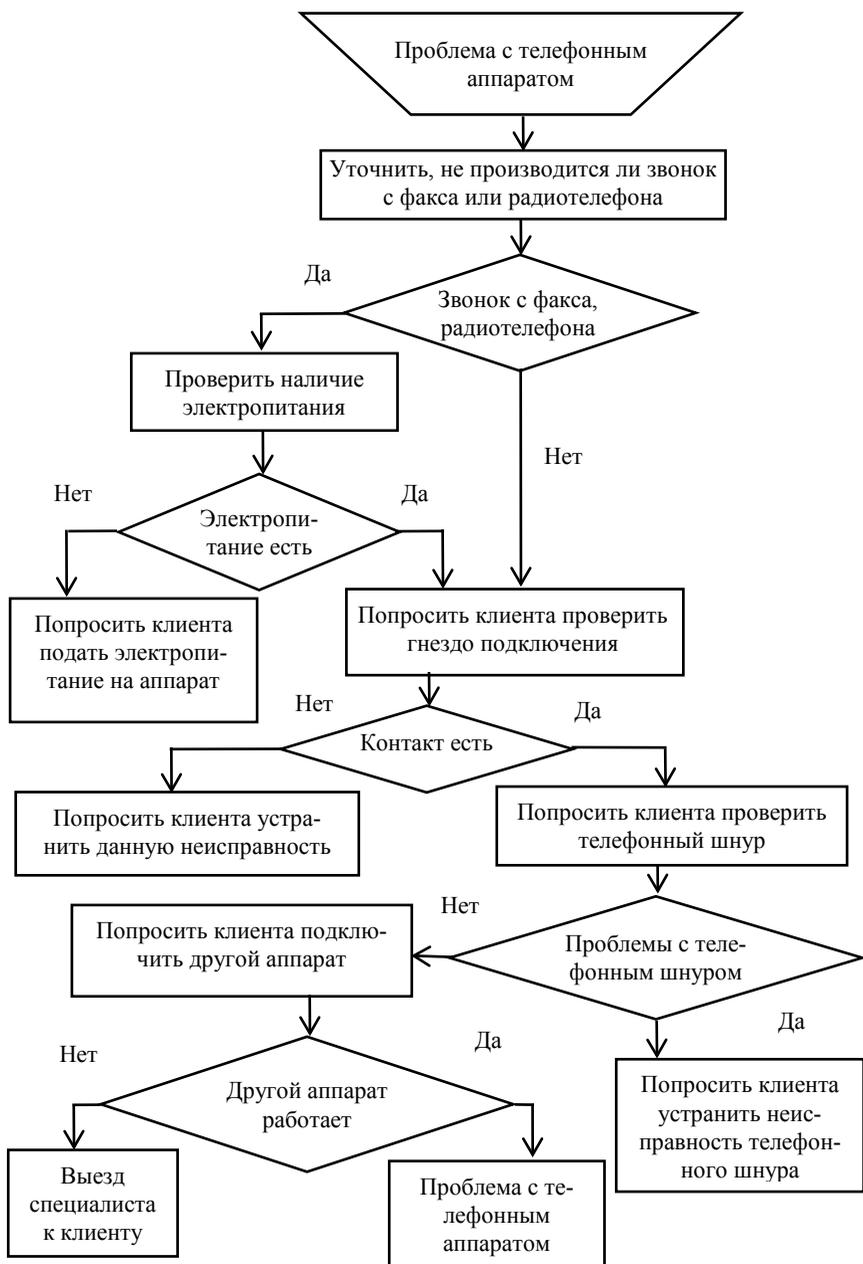


Рис. 5.11. Проблема с телефонным аппаратом

Проблема с телефонным аппаратом. Если у клиента обнаружена проблема с телефонным аппаратом, необходимо уточнить вид телефонного аппарата, с которого производится попытка дозвона. Если звонок производится с факсимильного аппарата или радиотелефона, необходимо попросить клиента проверить подачу электропитания на данный телефонный аппарат. Далее эксперт просит клиента проверить гнездо подключения и телефонный шнур. Если у клиента имеется другой аппарат, эксперт просит клиента подключить к линии другой аппарат. В случае, когда данные действия не привели к решению проблемы, необходим выезд специалиста к клиенту.

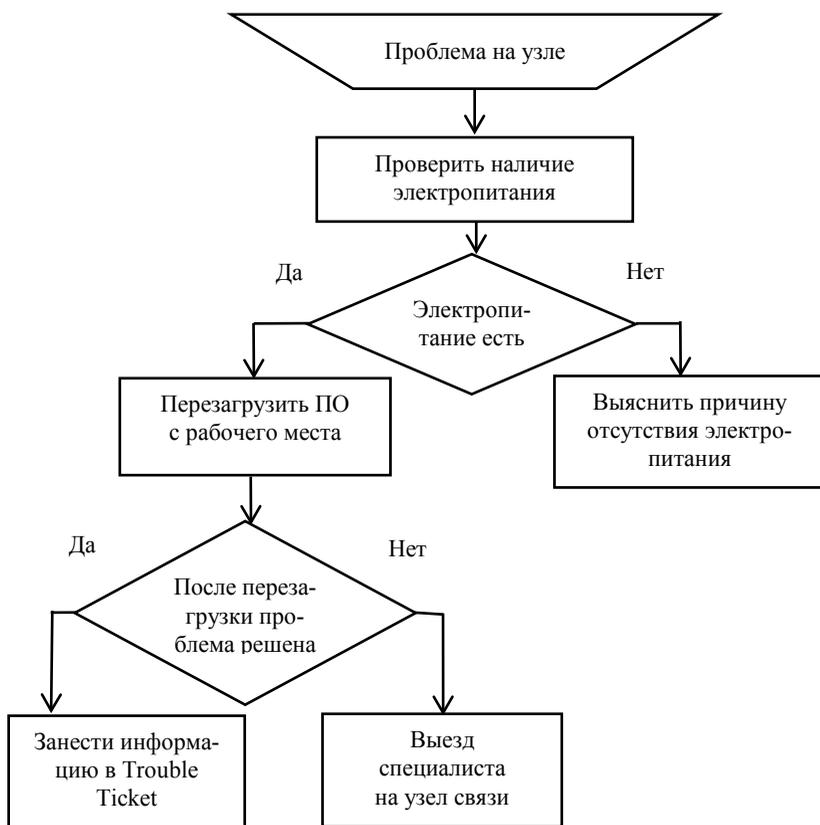


Рис. 5.12. Проблема на узле связи

Проблема на узле связи. В том случае, когда система мониторинга сети показала, что проблема возникла по причине неисправности на узле (см. рис. 5.12), эксперту необходимо в первую очередь проверить наличие там электропитания. Затем эксперт перезагружает узловое программное обеспечение с рабочего места. Если после перезагрузки проблема не решена, специалист выезжает на место возникновения неисправности.

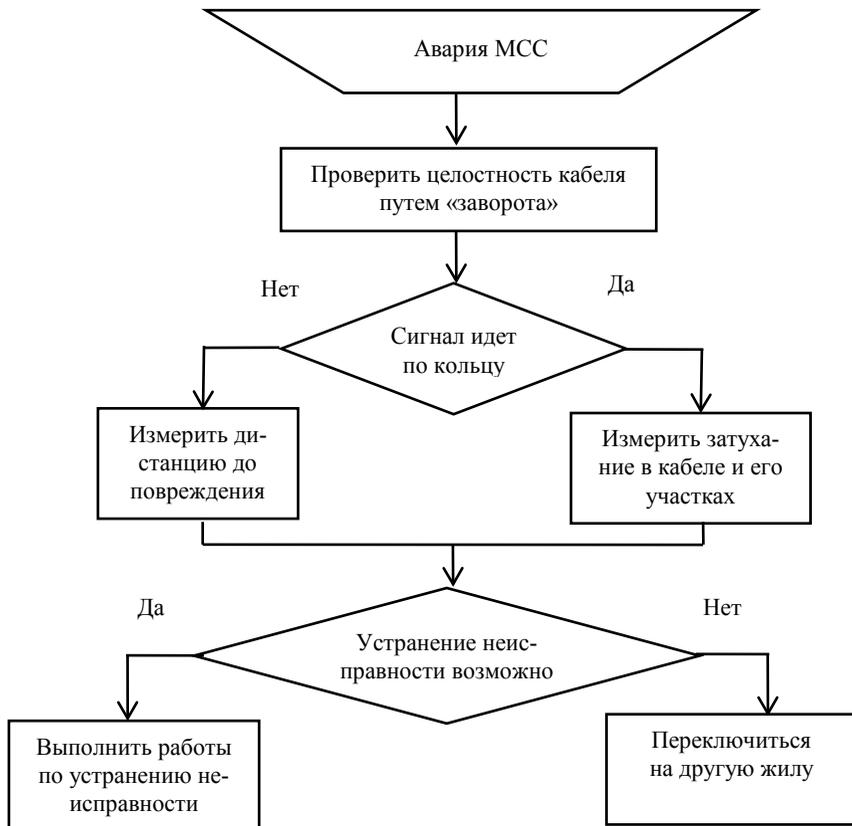


Рис. 5.13. Авария межстанционных связей

Авария межстанционных связей – эти аварии видны на системе мониторинга связей. Обычно при аварии на потоке проверяют целостность кабеля (или его жил) путем так называемого «заворота». То есть на дальнем конце замыкают «прием» на «передачу».

Таким образом, если сигнал идет по кольцу, то есть возвращается – проблема не в кабеле. Если сигнал проходит, но с большими потерями – измеряют затухание в кабеле и его участках. Если ничего не помогает – переключаются на другую жилу. Если повреждение в кабеле – измеряют дистанцию до повреждения.

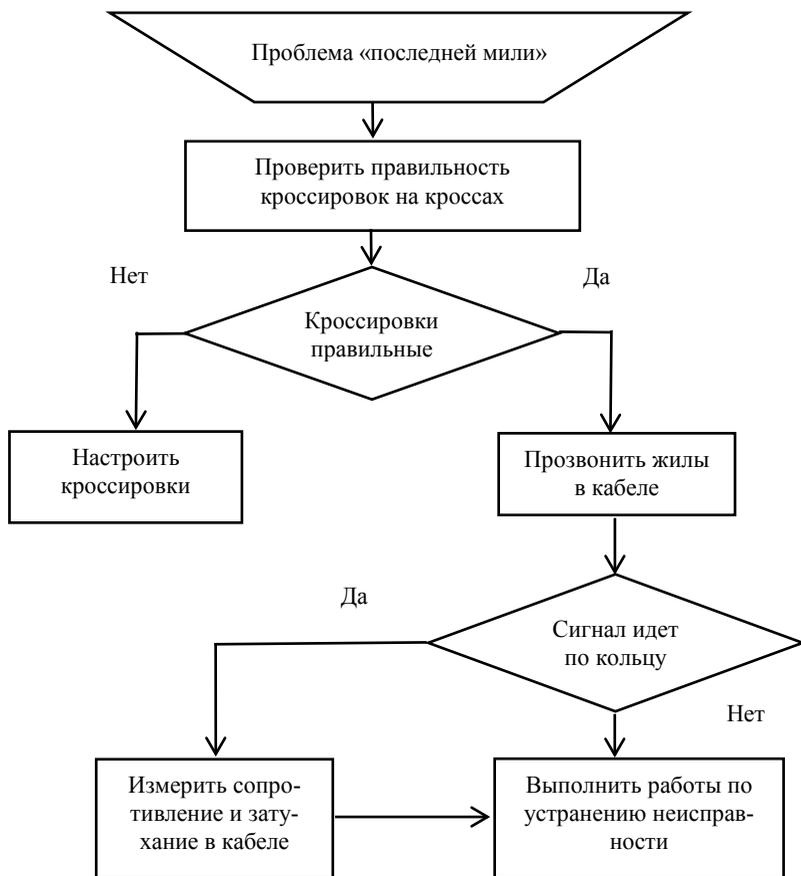


Рис. 5.14. Проблема «последней мили»

Проблема «последней мили». Авария на «последней миле» – это обычно повреждение медного кабеля и (или) неправильная его кроссировка на кроссах. Эксперты обычно проверяют правильность кроссировок, и если там все нормально – «прозванивают»

жины в кабеле, измеряют сопротивление и затухание (см. рис. 5.14), после чего устраняют проблему.

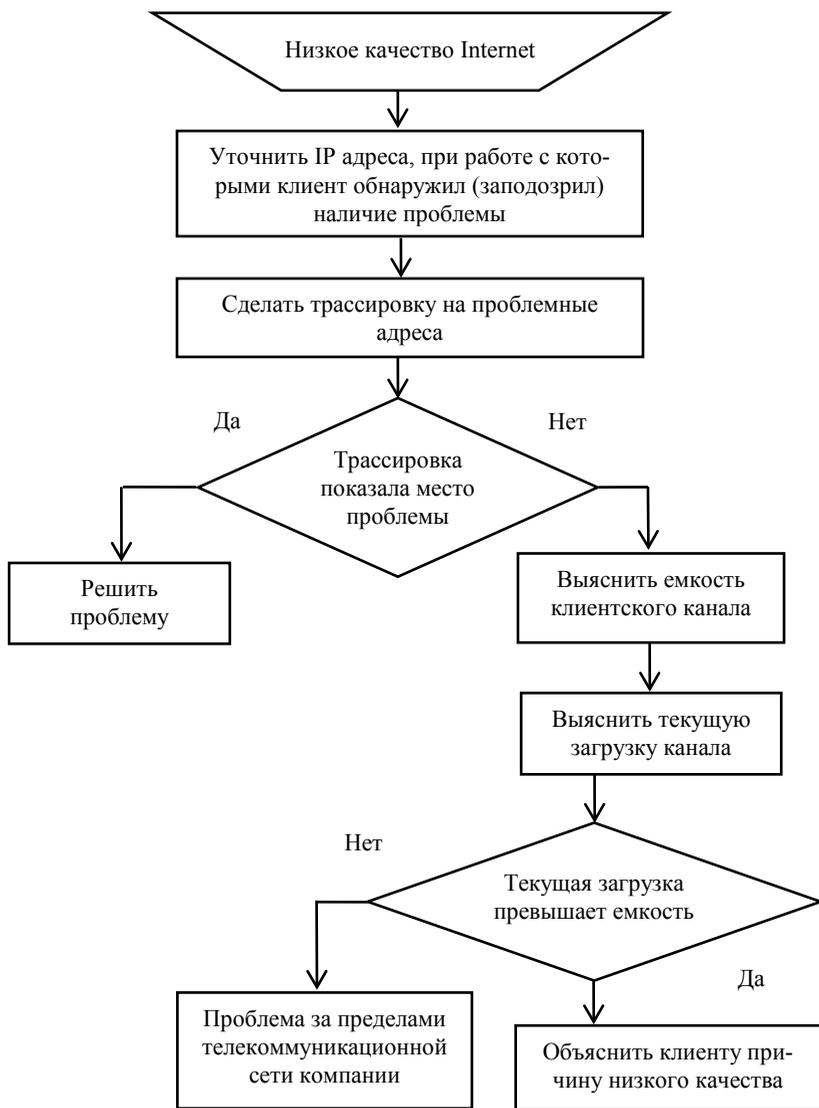


Рис. 5.15. Низкое качество Internet

Низкое качество Internet – при возникновении данной неисправности необходимо уточнить IP адреса, при работе с которыми клиент обнаружил или заподозрил наличие проблемы, после чего сделать трассировку и пинг на данные адреса. Далее необходимо выяснить емкость клиентского IP канала и загрузку канала в текущий момент времени.

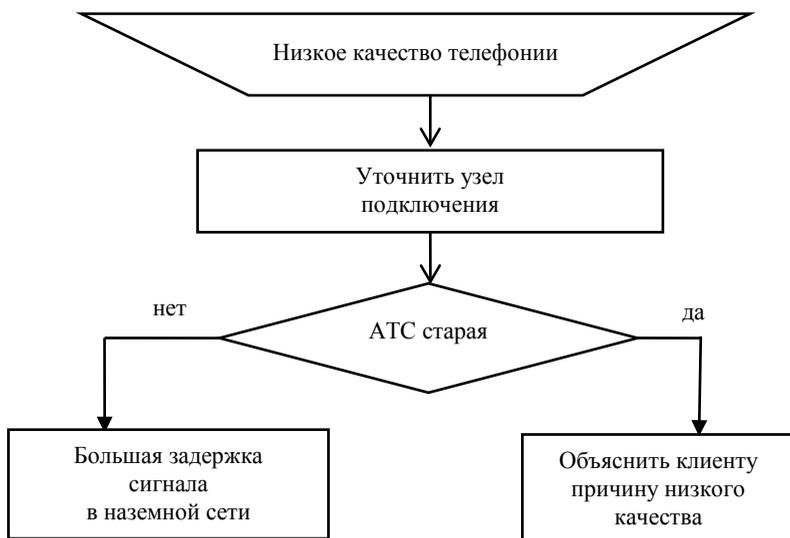


Рис. 5.16. Низкое качество телефонии

Низкое качество телефонии. Одна из основных проблем здесь – возникновение «эхо» при голосовой связи, которая относится к разряду чрезвычайно трудноустраняемых. Эффект «эхо» имеется всегда, вне зависимости от типа используемых соединительных линий и телефонов.

Рассмотрим **особенности продукционного представления знаний** в разрабатываемой ЭС. Основными задачами в процессе формализации являются проблемы структуризации исходной задачи и знаний в выбранном (разработанном) формализме. Структуризация знаний достигается на этапе составления **словаря системы**, который имеет как бы «двустороннюю» сущность и назначение: с одной стороны, это набор слов, фраз, кодов, наименований, ис-

пользуемых экспертом для обозначения условий, заключений и гипотез.

С другой стороны, это тот набор символов и обозначений, которым владеет система и благодаря которому пользователь понимает результаты ее работы.

Таблица 5.2 Словарь условий

Номер	Обозначение ключевого слова	Характеристика
1	2	3
У1	Старый тип АТС	Вероятнее всего помехи вызваны объективными факторами, то есть техническая составляющая
У2	Дисковый телефон	Вероятнее всего помехи вызваны объективными факторами (несоответствие используемого аппарата и узла подключения)
У3	офисная АТС	Вероятнее всего помехи вызваны неисправностью офисной АТС
У4	Офисная АТС обслуживается КСС?	Проблема имеет место, если УАТС находится под управлением КСС
У5	Оборудование включено	Проблема может быть вызвана отключенным оборудованием
У6	Трейсруот доходит	Проблема может быть связана с плохим качеством линии соединения
У7	Роутер принадлежит КСС	Вероятнее всего помехи вызваны неисправностью роутера
У8	Правильные настройки	Проблема может быть связана с неправильными настройками сети
У9	Анализ показал неисправный узел	Проблема заключается в неисправности на узле
У10	Есть зуммер ответа станции	Проблема может быть связана с наличием неисправного телефонного аппарата
У11	Есть дебиторская задолженность	Проблема связана с наличием дебиторской задолженности клиента
У12	Есть подключение к «8»	Проблема может быть связана с тем, что клиент не имеет выход на междугородную (международную) сеть

Таблица 5.2 Словарь условий (продолжение)

1	2	3
У13	Набор кода правильный	Проблема может быть связана с неправильным набором кодов
У14	Есть выход на «8»	Проблема связана с отсутствием у клиента выхода на междугородную/международную сеть
У15	Телефонный аппарат исправен	Неисправность может заключаться в неисправности телефонного аппарата
У16	Соседние номера работают	Вероятнее всего проблема на стороне вызываемого абонента
У17	Звонок с факса, радиотелефона	Вероятнее всего неисправность связана с использованием факсимильного аппарата или радиотелефона
У18	Контакт есть	Проблема может быть связана с отсутствием контакта
У19	Проблемы с телефонным шнуром	Проблема может быть связана с плохим качеством телефонного шнура
У20	Другой аппарат работает	Вероятнее всего проблема с телефонным аппаратом
У21	Электропитание есть	Неисправность может быть вызвана отсутствием электропитания
У22	После перезагрузки проблема решена	Проблема вызвана отказом или ошибкой программного обеспечения
У23	Сигнал идет по кольцу	Причина неисправности может заключаться в неисправности кабеля
У24	Устранение неисправности возможно	Неисправность может быть устранена линейно-кабельным отделом
У25	Кроссировки на кроссах правильные	Вероятнее всего неисправность вызвана неправильными кроссировками на кроссах
У26	Трассировка показала место проблемы	Причина неисправности может быть выявлена после трассировки
У27	Текущая загрузка превышает емкость	Низкое качество может быть связано с тем, что текущая загрузка клиентского канала превышает его номинальную емкость
У28	АТС старая	Вероятнее всего помехи вызваны объективными факторами (старая аналоговая АТС)

Составление словаря – важная работа, так как четко сформулированные условия и ответы значительно повышают эффективность эксплуатации ЭС. В нашем случае словарь системы включает в себя две части:

- словарь условий (см. таблицу 5.2);
- словарь заключений (см. таблицу 5.3).

Таблица 5.3 Словарь заключений

Номер	Обозначение ключевого слова	Характеристика
1	2	3
Z1	Идентифицировать абонента	Спрашивается название организации клиента, номер контракта, контактные данные клиента (имя, телефон)
Z2	Выделить класс проблемы	На начальном этапе выделяется класс проблемы (пропадание одного сервиса, пропадание всех сервисов, низкое качество сервисов)
Z3	Уточнить неисправный сервис	Выясняется неисправный сервис (пропадание телефонии, пропадание Internet)
Z4	Уточнить узел подключения	Вероятно, что неисправность вызвана особенностями узла подключения
Z5	Уточнить, какой телефонный аппарат используется	Проблема может возникать из-за несоответствия работы телефонного аппарата и узла подключения
Z6	Сообщить об особенностях АТС старого типа	Вероятнее всего помехи вызваны объективными факторами, то есть наличием технической составляющей
Z7	Выяснить, нет ли офисной АТС	Помехи могут быть связаны с наличием офисной АТС
Z8	Выяснить под чьим управлением находится офисная АТС	Проблема имеет место, если офисная АТС находится под управлением данной КСС
Z9	Порекомендовать решать проблему самостоятельно	Проблема находится в зоне ответственности клиента
Z10	Уточнить, включено ли оборудование	Проблема может быть связана с выключенным оборудованием

Таблица 5.3 Словарь заключений (продолжение)

1	2	3
Z11	Попросить клиента включить оборудование	Необходимо попросить клиента проверить оборудование
Z12	Уточнить проблему	Происходит уточнение проблемы (нет «8», нет «в город», нет «8» и нет «в город»)
Z13	Попросить клиента проверить контакт в гнезде соединительных линий	Клиенту необходимо визуально проверить контакт в гнезде соединительных линий
Z14	Попросить клиента устранить неисправность контакта	Клиенту необходимо наладить контакт в гнезде соединительных линий
Z15	Определить IP адрес серийного интерфейса на роутере клиента	Для анализа такой неисправности, как отсутствие доступа в Internet, необходимо определить IP адрес серийного интерфейса на роутере клиента
Z16	Сделать трейсруут на ООД клиента	Первичный результат при анализе такой неисправности, как отсутствие доступа в Internet, можно получить, сделав трейсруут на ООД клиента
Z17	Выяснить принадлежность роутера	Если роутер находится под управлением клиента, необходимо попросить тех. специалиста клиента, ответственного за эксплуатацию роутера, решить проблему самостоятельно
Z18	Проверить настройки сети	Проблема может заключаться в неправильных настройках сети
Z19	Помочь клиенту правильно настроить Internet	Проблема может быть решена после правильной настройки
Z20	Анализ системы мониторинга сети	Локализовать аварию (проблему) можно посредством анализа системы мониторинга сети
Z21	Выезд специалиста	Проблема не может быть решена на рабочем месте, необходимо присутствие специалиста на месте возникновения неисправности

Таблица 5.3 Словарь заключений (продолжение)

1	2	3
Z22	Уточнить у клиента, есть ли зуммер ответа станции	Отсутствие зуммера ответа станции может означать неисправность телефонного аппарата
Z23	Проверить лицевой счет клиента	Необходимо использование базы данных расчетного отдела
Z24	Напомнить клиенту об оплате	Проблема с лицевым счетом клиента
Z25	Проверить подключение к «8»	Необходимо использование БД абонентского отдела
Z26	Сообщить об отключенной «8»	Проблема связана с отсутствием подключения клиента к «8»
Z27	Проверить правильность набора кода	Необходимо использование БД кодов
Z28	Сообщить правильный набор кода	Необходимо сообщить правильную последовательность набора кодов
Z29	Проблема на стороне вызываемого абонента	Проблема находится в зоне ответственности клиента
Z30	Проверить выход «8»	Проблема может быть связана с отсутствием выхода «8»
Z31	Проверить телефонный аппарат	Проблема может быть связана с неисправностью телефонного аппарата
Z32	Проверить соседние городские номера	Действие поможет выявить неисправные номера
Z33	Уточнить, не производится ли звонок с факса или радиотелефона	Звонок с факсимильного аппарата или радиотелефона обладает некоторыми особенностями
Z34	Проверить наличие электропитания	Проблема может быть связана с отсутствием электропитания
Z35	Попросить клиента подать питание на аппарат	Попросить клиента подать электропитание
Z36	Попросить клиента проверить телефонный шнур	Попросить клиента потеревить телефонный шнур

Таблица 5.3 Словарь заключений (продолжение)

1	2	3
Z37	Попросить клиента подключить другой аппарат	Если у клиента имеется другой аппарат, необходимо попросить его подключить к линии другой аппарат
Z38	Попросить клиента устранить неисправность телефонного шнура	Клиент может либо поменять телефонный шнур, либо устранить недостатки имеющегося
Z39	Выяснить причину отсутствия электропитания	Вероятнее всего, отсутствие электропитания не зависит КСС
Z40	Перезагрузить ПО с рабочего места	Проблема может быть решена без выезда на место возникновения неисправности
Z41	Проверить целостность кабеля путем «заворота»	Данное действие выполняет линейно-кабельная группа на месте повреждения
Z42	Измерить дистанцию до повреждения	При наличии повреждения в кабеле, необходимо найти конкретное место повреждения (выполняет линейно-кабельная группа)
Z43	Измерить затухание в кабеле и его участках	Выполняет линейно-кабельная группа
Z44	Выполнить работы по устранению неисправности	Выполняет линейно-кабельная группа
Z45	Переключиться на другую жилу	В случае если устранение неисправности невозможно, необходимо переключиться на другую жилу
Z46	Проверить правильность кроссировок на кроссах	Вероятнее всего неисправность вызвана неправильными кроссировками на кроссах
Z47	Настроить кроссировки	Для устранения неисправности необходимо правильно настроить кроссировки
Z48	Прозвонить жилы в кабеле	При наличии повреждения кабеля, необходимо найти конкретное место повреждения (выполняет линейно-кабельная группа)

Таблица 5.3 Словарь заключений (окончание)

1	2	3
Z49	Уточнить проблемный сервис	Уточняется, низкое качество телефонии или Internet
Z50	Уточнить IP адреса, при работе с которыми клиент обнаружил (заподозрил) наличие проблемы	Необходимо для дальнейшего анализа проблемы
Z51	Сделать трассировку на проблемные адреса	Сначала необходимо провести анализ данных IP адресов
Z52	Выяснить емкость клиентского канала	Необходимо для дальнейшего анализа проблемы
Z53	Выяснить текущую загрузку канала	Необходимо для сравнения емкости и текущей загрузки
Z54	Проблема за пределами КС компании	Проблема находится в зоне ответственности клиента
Z55	Уточнить «проблемный» номер	Действие поможет выявить неисправный номер
Z56	Большая задержка сигнала в наземной сети	Данный аспект является причиной возникновения «эхо» (относится к разряду трудноустраняемых)
Z57	Занести информацию в ГТ	После устранения неисправности необходимо занести информацию в БД

5.4. Описание БЗ на выбранном языке представления знаний

Описание БЗ разрабатываемой ЭС на выбранном языке представления знаний иллюстрирует таблица 5.4. В нашем случае используется продукционный метод представления знаний, где основной единицей знаний служит правило в следующем виде:

«Если <посылка>, то <заключение>».

Правила в рассматриваемой БЗ имеют одну из двух форм:

- 1) ЕСЛИ (У) = ЛОЖЬ, ТО (Z);
ЕСЛИ (У) = ИСТИНА, ТО (Z).
- 2) ЕСЛИ (У) = ЛОЖЬ, ТО (Z1) И (Z2);
ЕСЛИ (У) = ИСТИНА, ТО (Z1) И (Z2).

Таблица 5.4. Описание БЗ на выбранном языке представления знаний

Номер	ЕСЛИ < >	Значение	ТО < >
1	2	3	4
1	Корпоративный клиент	ЛОЖЬ	Сообщить звонящему, что его звонок будет переадресован в службу поддержки физических лиц
2	Корпоративный клиент	ИСТИНА	Идентифицировать абонента: узнать название организации, номер контракта, контактные данные (имя, телефон)
3	Названные данные совпадают с данными абонентской базы	ЛОЖЬ	Идентифицировать абонента: узнать название организации, номер контракта, контактные данные клиента (имя, телефон)
4	Названные данные совпадают с данными абонентской базы	ИСТИНА	Выделить класс проблемы
5	Пропадание одного сервиса	ИСТИНА	Уточнить неисправный сервис
6	Неисправность телефонии	ИСТИНА	Уточнить узел подключения
7	Узел АТС старого типа	ЛОЖЬ	Выяснить, нет ли офисной АТС
8	Узел АТС старого типа	ИСТИНА	Сообщить об особенностях АТС старого типа
9	Офисная АТС	ИСТИНА	Выяснить под чьим управлением находится офисная АТС
10	Офисная АТС	ЛОЖЬ	Попросить клиента проверить контакт в гнезде соединительных линий
11	Офисная АТС обслуживается КСС	ЛОЖЬ	Порекомендовать решать проблему самостоятельно
12	Офисная АТС обслуживается КСС	ИСТИНА	Уточнить, включено ли оборудование

Таблица 5.4. Описание БЗ на выбранном языке представления знаний (продолжение)

1	2	3	4
13	Оборудование включено	ИСТИНА	Уточнить проблему
14	Оборудование включено	ЛОЖЬ	Попросить клиента включить оборудование
15	Контакт есть	ИСТИНА	Уточнить проблему
16	Контакт есть	ЛОЖЬ	Попросить клиента устранить неисправность контакта
17	Неисправность Internet	ИСТИНА	Определить IP адрес серийного интерфейса на роутере клиента И сделать трейсрут на ООД клиента
18	Трейсрут доходит	ИСТИНА	Выяснить принадлежность роутера
19	Трейсрут доходит	ЛОЖЬ	Анализ системы мониторинга сети
20	Роутер принадлежит КСС	ИСТИНА	Проверить настройки сети
21	Роутер принадлежит КСС	ЛОЖЬ	Порекомендовать решать проблему самостоятельно
22	Правильные настройки	ЛОЖЬ	Помочь клиенту правильно настроить Internet
23	Правильные настройки	ИСТИНА	Выезд специалиста
24	Анализ показал неисправный узел	ЛОЖЬ	Выезд специалиста
25	Анализ показал неисправный узел	ИСТИНА	Проблема на узле
26	Нет «8»	ИСТИНА	Уточнить у клиента, есть ли зуммер ответа станции
27	Есть зуммер ответа станции	ИСТИНА	Проверить лицевой счет клиента
28	Есть зуммер ответа станции	ЛОЖЬ	Проблема с телефонным аппаратом
29	Есть дебиторская задолженность	ИСТИНА	Напомнить клиенту об оплате

Таблица 5.4. Описание БЗ на выбранном языке представления знаний (продолжение)

1	2	3	4
30	Есть дебиторская задолженность	ЛОЖЬ	Проверить подключение к «8»
31	Есть подключение к «8»	ИСТИНА	Проверить правильность набора
32	Есть подключение к «8»	ЛОЖЬ	Сообщить об отключенной «8»
33	Набор кода правильный	ИСТИНА	Проблема на стороне вызываемого абонента
34	Набор кода правильный	ЛОЖЬ	Сообщить правильный набор кода
35	Нет «в город»	ИСТИНА	Уточнить у клиента, есть ли зуммер ответа станции
36	Есть зуммер ответа станции	ИСТИНА	Проверить лицевой счет клиента (пр. №23)
37	Есть зуммер ответа станции	ЛОЖЬ	Проблема с телефонным аппаратом (пр.№24)
38	Есть дебиторская задолженность	ИСТИНА	Напомнить клиенту об оплате (пр. №25)
39	Есть дебиторская задолженность	ЛОЖЬ	Проверить выход «8»
40	Есть выход на «8»	ИСТИНА	Проверить соседние городские номера
41	Есть выход на «8»	ЛОЖЬ	Проверить телефонный аппарат
42	Телефонный аппарат исправен	ИСТИНА	Выезд специалиста
43	Телефонный аппарат исправен	ЛОЖЬ	Проблема с телефонным аппаратом
44	Соседние номера работают	ИСТИНА	Проблема на стороне вызываемого абонента
45	Соседние номера работают	ЛОЖЬ	Выезд специалиста
46	Нет «в город» , нет «8»	ИСТИНА	Уточнить у клиента, есть ли зуммер ответа станции
47	Есть зуммер ответа станции	ИСТИНА	Проверить лицевой счет клиента (пр.№27)

Таблица 5.4. Описание БЗ на выбранном языке представления знаний (продолжение)

1	2	3	4
48	Есть зуммер ответа станции	ЛОЖЬ	Проблема с телефонным аппаратом (пр.№28)
49	Есть дебиторская задолженность	ИСТИНА	Напомнить клиенту об оплате (пр.№29)
50	Есть дебиторская задолженность	ЛОЖЬ	Выезд специалиста
51	Проблема с телефонным аппаратом	ИСТИНА	Уточнить, не производится ли звонок с факса или радиотелефона
52	Звонок с факса, радиотелефона	ИСТИНА	Проверить наличие питания
53	Звонок с факса, радиотелефона	ЛОЖЬ	Проверить гнездо подключения
54	Электропитание есть	ИСТИНА	Проверить гнездо подключения
55	Электропитание есть	ЛОЖЬ	Попросить клиента подать питание на аппарат
56	Контакт есть	ИСТИНА	Попросить клиента проверить кабель
57	Контакт есть	ЛОЖЬ	Попросить клиента устранить неисправность контакта
58	Проблемы с телефонным шнуром	ИСТИНА	Попросить клиента устранить неисправность телефонного шнура
59	Проблемы с телефонным шнуром	ЛОЖЬ	Попросить клиента подключить другой аппарат
60	Другой аппарат работает	ИСТИНА	Проблема решена
61	Другой аппарат работает	ЛОЖЬ	Выезд специалиста
62	Проблема на узле	ИСТИНА	Проверить наличие электропитания
63	Электропитание есть	ИСТИНА	Перезагрузить ПО с рабочего места
64	Электропитание есть	ЛОЖЬ	Выяснить причину отсутствия

Таблица 5.4. Описание БЗ на выбранном языке представления знаний (продолжение)

1	2	3	4
65	После перезагрузки проблема решена	ИСТИНА	Занести информацию в Trouble Ticket
66	После перезагрузки проблема решена	ЛОЖЬ	Выезд специалиста на узел
67	Пропадание всех сервисов	ИСТИНА	Проверить лицевой счет клиента
68	Есть дебиторская задолженность	ИСТИНА	Напомнить клиенту об оплате (пр. №26)
69	Есть дебиторская задолженность	ЛОЖЬ	Анализ системы мониторинга сети
70	Авария МСС	ИСТИНА	Проверить целостность кабеля путем «заворота»
71	Сигнал идет по кольцу	ИСТИНА	Измерить затухание в кабеле и его участках
72	Сигнал идет по кольцу	ЛОЖЬ	Измерить дистанцию до повреждения
73	Устранение неисправности возможно	ИСТИНА	Выполнить работы по устранению неисправности
74	Устранение неисправности возможно	ЛОЖЬ	Переключиться на другую жилу
75	Проблема «последней мили»	ИСТИНА	Проверить правильность кроссировок на кроссах
76	Кроссировки на кроссах правильные	ИСТИНА	Прозвонить жилы в кабеле
77	Кроссировки на кроссах правильные	ЛОЖЬ	Настроить кроссировки
78	Сигнал идет по кольцу	ИСТИНА	Измерить сопротивление и затухание в кабеле и выполнить работы по устранению неисправности
79	Сигнал идет по кольцу	ЛОЖЬ	Выполнить работы по устранению неисправности

Таблица 5.4. Описание БЗ на выбранном языке представления знаний (окончание)

1	2	3	4
80	Низкое качество сервисов	ИСТИНА	Уточнить проблемный сервис
81	Низкое качество Internet	ИСТИНА	Уточнить IP адреса, при работе с которыми клиент обнаружил / заподозрил наличие проблемы и сделать трассировку на проблемные адреса
82	Текущая загрузка превышает емкость	ЛОЖЬ	Проблема за пределами сети компании
83	Низкое качество телефонии	ИСТИНА	Уточнить узел подключения
84	АТС старая	ИСТИНА	Объяснить клиенту причину низкого качества
85	АТС старая	ЛОЖЬ	Большая задержка сигнала в наземной сети

В процессе работы программы, введенные правила связываются в сеть вывода:

П20: ЕСЛИ (У8) = ИСТИНА, ТО (Z17);

П21: ЕСЛИ (У8) = ЛОЖЬ, ТО (Z20);

П22: ЕСЛИ (У9) = ИСТИНА, ТО (Z18);

П23: ЕСЛИ (У9) = ЛОЖЬ, ТО (Z9) и т.д.,

где П_n – номер правила (см. таблицу 5.3); У_n – номер условия (см. таблицу 5.1); Z_n – номер заключения (см. таблицу 5.2).

Пример сети вывода иллюстрирует рис. 5.17.

5.5. Стадия разработки ЭС диагностики неисправностей оборудования ИКК

На стадии разработки ЭС была разработана диаграмма классов, (см. фрагмент на рис. 5.18), которая демонстрирует основные классы системы, их атрибуты, методы и взаимосвязи между ними.

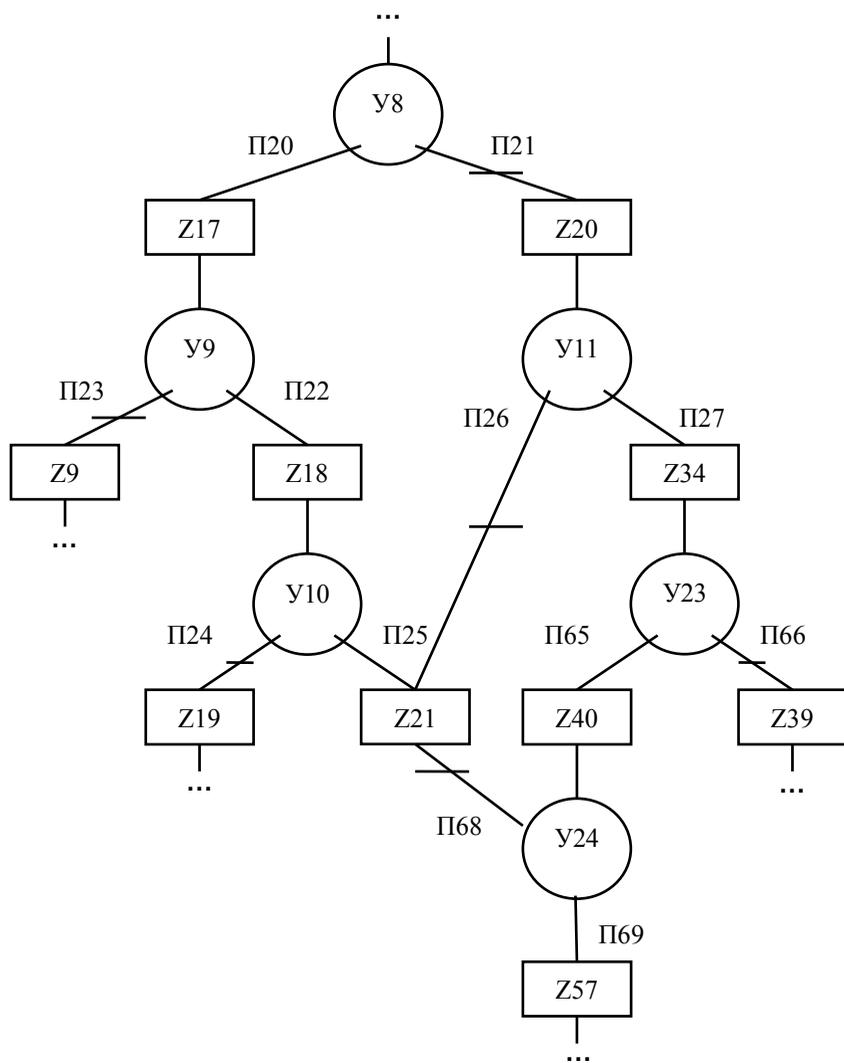


Рис. 5.17. Пример реализации сети вывода

Основные экранные формы прототипа ЭС приведены на рис. 5.19-5.21. При поступлении заявки от клиента, оператор, после его идентификации, принимает решение об открытии инцидента (см. рис. 5.19). В обязательном порядке должны быть указаны тип инцидента, описание, приоритет и задействованные участники.

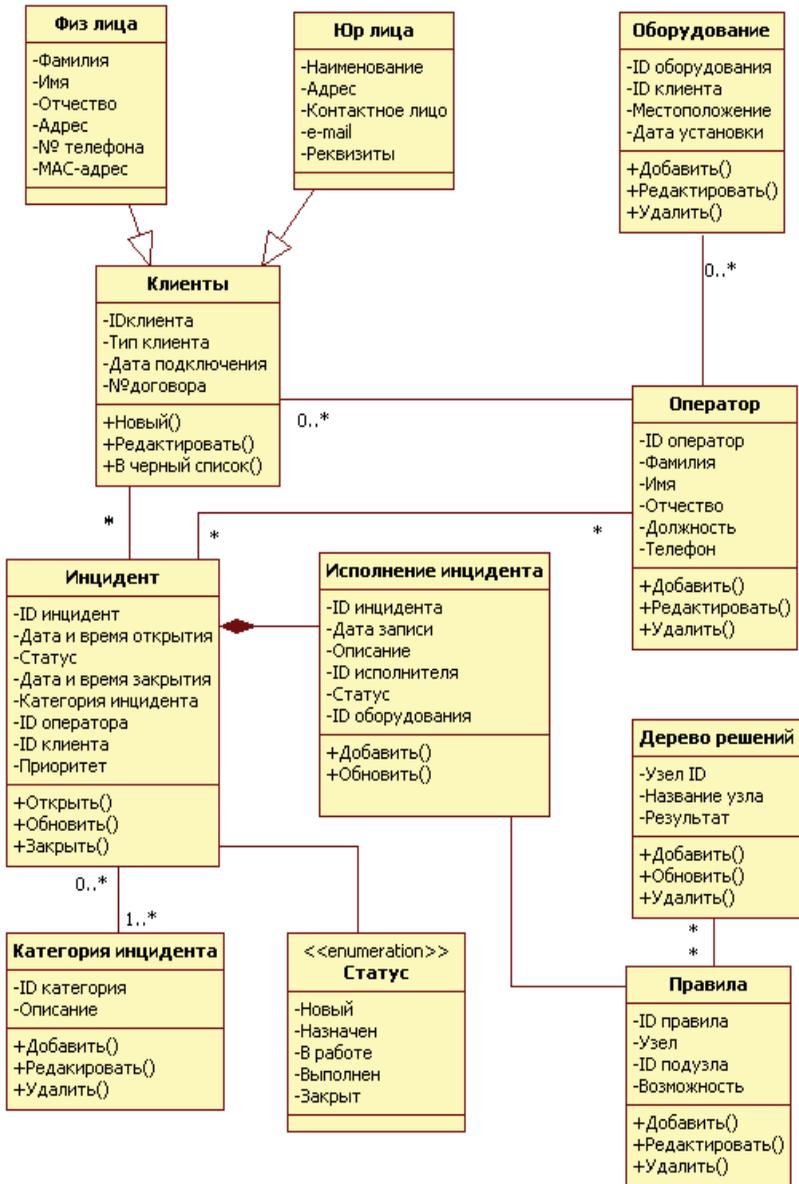


Рис. 5.18. Диаграмма классов проблемной области

Указание предварительного типа инцидента позволяет задействовать дерево диагностики, «отвечающее» за определение конкретной неисправности (см. рис. 5.20). Оператор, осуществляет опрос клиента по ветвям дерева диагностики, отмечая с помощью чек-боксов ответы. В зависимости от каждого последующего ответа осуществляется локализация проблемы.

Инциденты

Инцидент ID 618 Дата открытия 16.10.2014 15:02

Клиент ООО "Росвет" Дата закрытия

Тип инцидента Неисправность Интернет

Приоритет средний

Участники

- служба поддержки
- технический отдел
- отдел по работе с клиентами

Описание отсутствует соединение с Интернет

Добавить инцидент Редактировать Сохранить

Рис. 5.19. Регистрация инцидента

Дерево диагностики неисправностей

Тип инцидента авария МСС

- Проверить целостность кабеля путем "заворота"
- Сигнал идет по кольцу
- Сигнал по кольцу не проходит
- Измерить затухание в кабеле и его участках
- Измерить дистанцию до повреждения

Назад Определить Далее

Рис. 5.20. Определение неисправности оборудования ИКК

При необходимости оператор может осуществить поиск по БЗ, введя ключевые слова запроса клиента и получив ранее встречавшиеся похожие инциденты и способы их устранения (см. рис. 5.21).

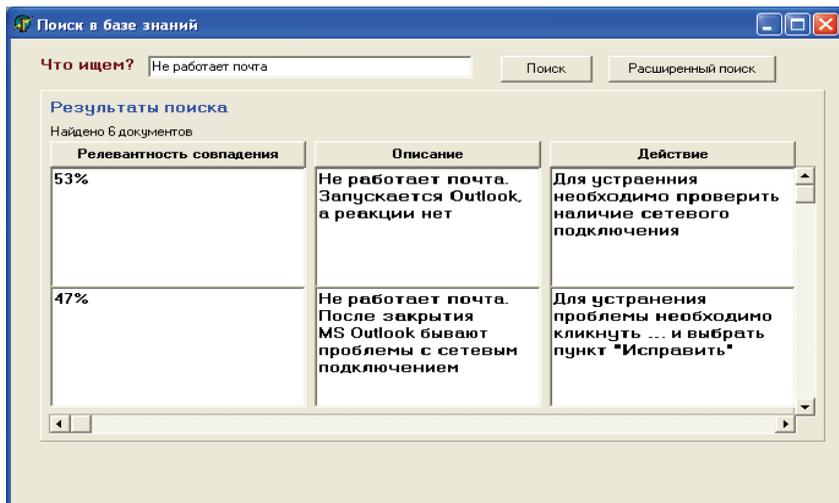


Рис. 5.21. Поиск в БЗ

5.6. Особенности предметной области экономической ЭС расчетно-диагностического типа

Предприятие как социально-экономическая СС не только производит определенную продукцию или услуги и распоряжается необходимыми средствами производства, но и присваивает часть создаваемого продукта с целью удовлетворения коллективных и личных интересов работников компании. Поэтому каждая организация заинтересована в том, чтобы производить высококачественную продукцию, пользующуюся спросом, которую можно продавать по ценам, отражающим общественно необходимые затраты на ее изготовление. Другими словами, предприятие (корпорация, компания, фирма) как экономически самостоятельная СС призвано обеспечивать высокую эффективность своего бизнеса – поскольку лишь в этом случае оно может быть конкурентоспособным на внутреннем и внешнем рынках товаропроизводителей.

От того, насколько эффективно и целесообразно осуществляется процесс управления с точки зрения обеспечения устойчивости финансового состояния и прибыльности предприятия, зависит не только финансовое благополучие данного хозяйствующего объекта, но также стабильность и эффективность функционирования региона, отрасли, экономики страны в целом.

Одним из главных инструментов обеспечения научной обоснованности принимаемых управленческих решений является финансово-экономический анализ (ФЭА). Финансовое состояние является критерием деловой активности и надежности предприятия, определяющим его конкурентоспособность и потенциал в реализации экономических интересов всех участников хозяйственной деятельности. Важной задачей финансового учета является формирование общей картины положения организации, определение конечных результатов ее хозяйственной деятельности для составления финансовой отчетности организации. Сопоставление с другими предприятиями аналогичного профиля дает важную информацию о деятельности конкурентов, их финансовой политике и финансовом состоянии, ситуации на рынке в целом. Знание финансовых слабостей других предприятий позволяет получить конкурентные преимущества, а анализ их сильных сторон способствует поиску рецептов укрепления собственного положения (более подробно о задачах и целях ФЭА см. в разделе 5.8).

Экономические ЭС данного типа относятся к классу наиболее сложных и полезных для практики ИИС, поскольку способны оперативно ставить диагноз и давать рецепты от «болезни» предприятия. Однако ввиду своей сложности данные ЭС не получили пока широкого практического распространения. Переходя к созданию ЭС данного типа следует отметить, что многие предприятия сегодня используют собственные методы оценки и анализа экономической ситуации, имеют свои устоявшиеся мнения и соображения относительно поиска способов ее улучшения – поэтому разработать типовую универсальную ЭС достаточно трудно.

Для большинства ЛПР наиболее важными исходными данными и требованиями при проектировании ЭС расчетно-диагностического типа являются:

– главная и подчиненная цели, которые ЛПР преследуют в процессе выполнения своих функций;

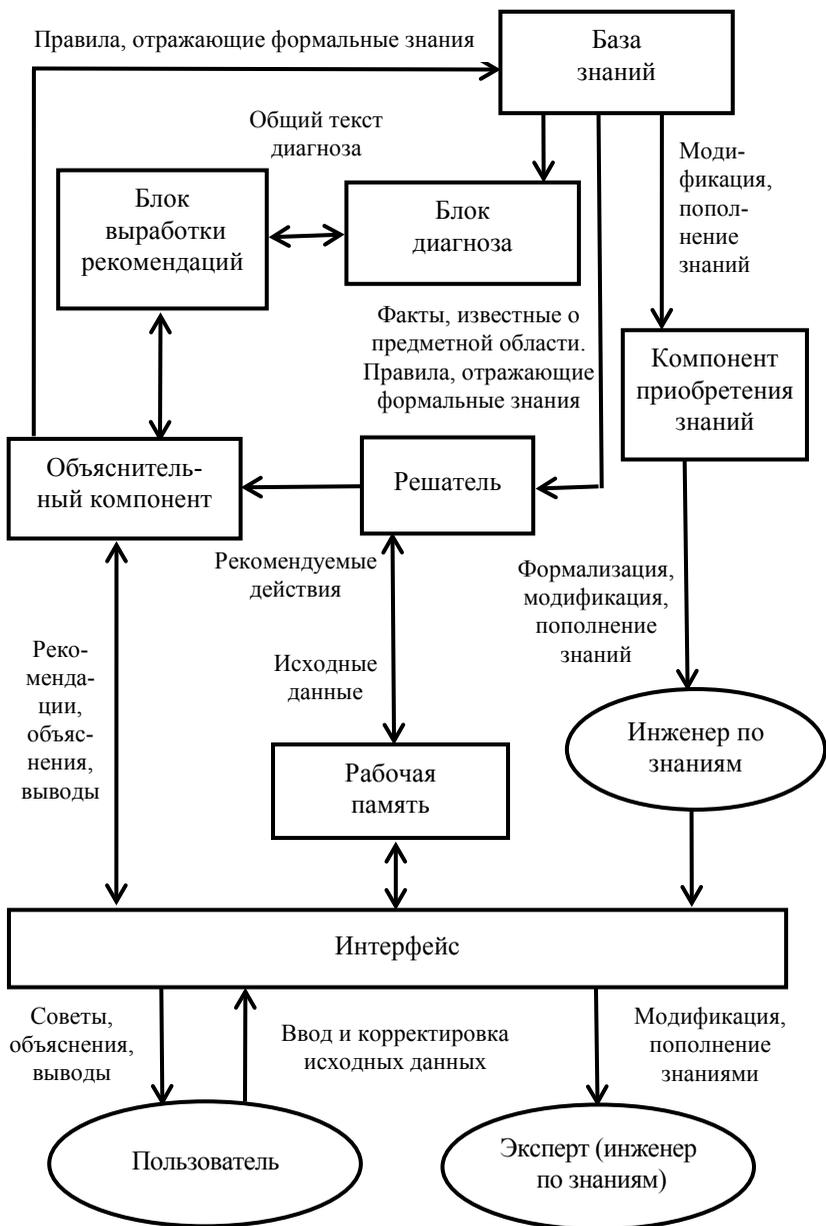


Рис. 5.22. Схема взаимосвязей основных блоков ЭС расчетно-диагностического типа

- оценочные показатели, выбираемые ЛПР для выяснения истинного состояния предприятия;
- принятые правила анализа финансовых ситуаций, а также оценка доверия к ним;
- процедуры формирования путей выхода из создавшихся ситуаций и оценка доверия к полученным вариантам принимаемых решений;
- методы оценки качества системы учета;
- объективность оценки своих действий за отчетный период.

Перечисленные особенности накладывают свой отпечаток на архитектуру ЭС расчетно-диагностического типа. Эти изменения касаются прежде всего компонента БЗ схемы на рис. 5.22, в составе которой целесообразно выделить три блока:

- собственно БЗ (блок правил);
- блок диагноза – который отвечает за процедуры манипулирования БЗ и рабочей памятью, выполняемые с помощью индикаторных таблиц, таблиц формирования диагноза, таблиц формирования текста рекомендаций и т.п.;
- блок разработки рекомендаций, который функционирует на основании информации, получаемой от блока диагноза, его назначение – выдача качественных характеристик состояния предприятия и направлении выхода из создавшейся ситуации.

5.7. Технологические принципы создания экономической ЭС расчетно-диагностического типа

Как правило, ЭС используются для информационной поддержки действий ЛПР, которые характеризуются следующими свойствами.

- подавляющее большинство управленческих решений имеют многошаговый характер и служат скорее соблюдению корпоративных интересов, чем достижению конкретных целей;
- управленческие процессы протекают в условиях несогласованности целей и интересов участников принятия решений. Отсюда немалая часть решений направлена на урегулирование и разрешение корпоративных (внутрифирменных) конфликтов;
- процессы, протекающие вне компании (во внешней среде СС), характеризуются случайными событиями, что создает трудности их учета;

– экспертные знания профессионалов носят субъективный характер, что приводит к противоречиям при их интеграции в БЗ.

Второе и третье из указанных свойств вынуждают вводить дополнительные средства, с помощью которых можно свести к минимуму трудности, мешающие ЛПР принимать правильные решения. Форма представления и способ обработки знаний зависят от целей и задач ЛПР – ЭС расчетно-диагностического типа, например, характеризуются четко поставленными целями, вполне определенными алгоритмами их достижения, значительным объемом расчетных процедур.

Соответственно БЗ в них отражается в форме дерева целей, синтезированных с графом показателей. Системы, ориентированные на приближенные рассуждения, напротив, характеризуются «размытостью» целей, именуемых гипотезами, нечеткими алгоритмами («мягкими» вычислениями) при относительно небольшом объеме исходных данных. В таких системах БЗ содержит деревья вывода (деревья «И-ИЛИ») и базу фактов, представляемых обычно в форме семантической сети.

Акценты в рассматриваемых системах различны: ЭС первого вида требуют повышенного внимания к процедурам расчета и БД (процедуры ввода и корректировки данных); ЭС второго вида – к БЗ (семантическим сетям и правилам вывода). Общие этапы и процедуры, присущие всем советующим СУ, положены в основу технологий их проектирования: при их разработке можно использовать также технологии, специально ориентированные на ЭС экономического профиля.

Отличительной чертой данных технологий является наличие средств поддержки исполнения принимаемых решений – в случае затруднений ЛПР, располагающие необходимыми сведениями о целях и ресурсах своих предполагаемых действий, могут обратиться за помощью к ЭС. Наличие функции поддержки исполнения решений дает возможность разработчикам ЭС по-новому взглянуть на проблему их создания и вообще на технологию принятия управленческих решений – описание РОЦ-технологии см. в разделе 4.5 [86; 97]. Возможная технология проектирования ЭС расчетно-диагностического типа, включающая четыре основных этапа, представлена на рис. 5.23.

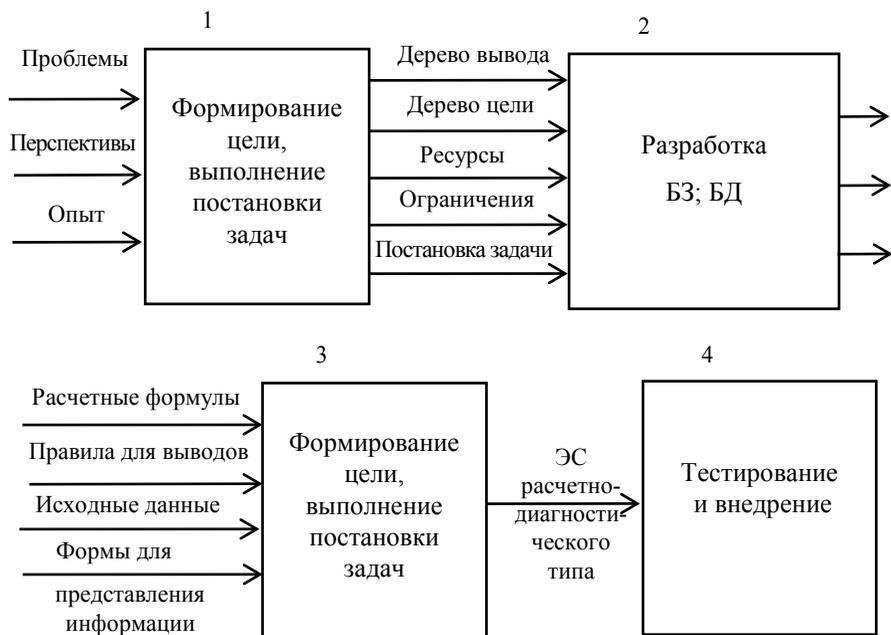


Рис. 5.23. РОЦ-технология создания ЭС расчетно-диагностического типа

Указанные этапы выполняются в соответствии с общими правилами создания ИС и ИИС, но с учетом ряда специфических особенностей – которые иллюстрирует рис. 5.24, где отражены параллели между рассматриваемой РОЦ-технологией и традиционной технологией создания ЭС.

В соответствии с рис. 5.23, на схеме рис. 5.24 использованы следующие числовые обозначения: 1 – дерево вывода; 2 – дерево целей; 3 – ресурсы; 4 – постановка задачи; 5 – ограничения; 6 – расчетные формулы; 7 – правила для выводов; 8 – исходные данные; 9 – формы для представления информации; 10 – прототип соответствующей СУ.

Блоки (подсистемы и элементы) схемы на рис. 5.24 рассмотрим более подробно. Формулировка цели создания ЭС или гипотезы, которую следует доказать (блок 1) подразумевает два этапа.

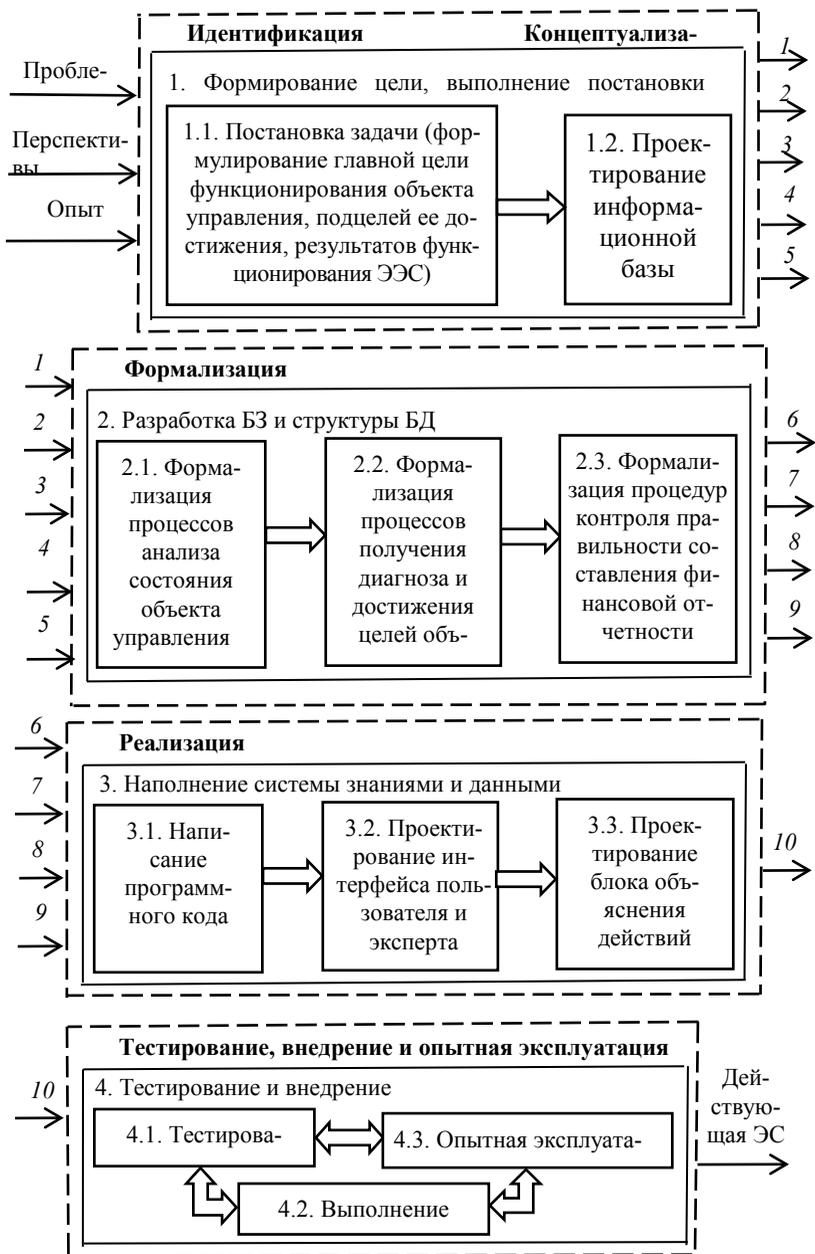


Рис. 5.24. Взаимосвязь РОЦ-технологии с этапами создания ЭС

Постановка задачи (блок 1.1) – на данном этапе выявляется основная цель функционирования объекта управления и выполняется ее декомпозиция на составляющие. Обычно выбор цели является следствием запроса руководителя и изучения показателей, представляющих собой информационную базу принятия решений. Цели формулируются, исходя из анализа проблем, стоящих перед корпорацией (компанией, предприятием, фирмой), перспектив развития и опыта ЛПР, а затем представляются в графическом и табличном виде. Таблица (см. как пример таблицу 5.4), которая составляется при помощи квалифицированного и опытного эксперта, необходима для поиска путей выхода из возможных нежелательных состояний бизнеса предприятия, кроме того, она устанавливает связь между тактическими целями и экономическими показателями.

Таблица 5.4. Табличное описание дерева целей

Номер уровня (подуровня)	Код цели (подцели)	Цель (подцель)	Номер уровня (подцели)	Код цели (подцели)	Цель (подцель)
1	А	Обеспечить высокий уровень А	2	В	Обеспечить эффективное использование В
			2	С	Обеспечить высокий уровень С
2	В	Обеспечить эффективное использование В	3	Д	Уменьшить коэффициент Д
			3	Е	Стабилизировать значение Е
			3	F	Увеличить коэффициент F
3	С	Обеспечить высокий уровень С	3	G	Повысить коэффициент G
			3	Н	Снизить значение Н

Пример графа целей показан на рис. 5.26. Согласно [64; 98], постановка задачи должна содержать описание входной, промежуточной и выходной (результатирующей) информации, а также общее описание процедур (алгоритмы, формулы) преобразования входной информации в результирующую. Состав и содержание результирующей информации вытекают из целей, для достижения которых создается ЭС. Форма описания входной и выходной информации должна соответствовать принятым требованиям [95].

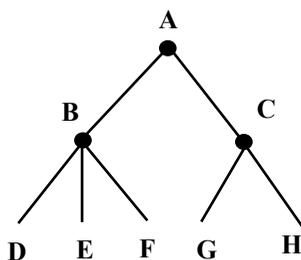


Рис. 5.25. Представление дерева целей

Проектирование информационной базы (см. блок 1.2 на рис. 5.24) – здесь изучаются первичные документы, составляющие основу данной базы, при этом состав входной и результирующей информации зависит от цели создания ЭЭС расчетно-диагностического характера и определяется по итогам тестирования финансовой документации и диагностирования бизнеса предприятия. Для этого используются бухгалтерская и финансовая документация, а также таблица возможных тактических целей.

Далее описываются результаты функционирования системы (таблицы, диаграммы, ведомости) и составляется список входных документов, используемых для ввода оперативной информации. Указываются периодичность и сроки ввода оперативных данных, а также методы их контроля. В результате выполнения первого этапа должны быть получены [86; 97]:

- дерево целей или дерево вывода;
- ограничения на используемые ресурсы;
- перечень исходной и выходной информации.

Разработка БЗ и структуры БД (блок 2) включает три этапа.

Формализация процессов анализа состояния объекта управления задачи (блок 2.1) – на данном этапе производится выбор метода представления знаний и в рамках выбранного формализма осуществляется проектирование логической структуры БЗ. процессу принятия решений присуща слабая формализуемость, поэтому этап формализации является креативным (в известной мере творческим) процессом. Используя текстовое описание предметной области, ключевые слова и связи между ними, полученные на предыдущих этапах, инженер по знаниям формулирует все возможные состояния объекта управления, а также производит группировку схожих ситуаций.

Формализация процессов получения диагноза и путей достижения целей объекта управления (блок 2.2) может производиться на основе стратегий локального и комплексного диагностирования. В первом случае предполагается локализация, а затем ликвидация причин и следствий прежде всего тех факторов, которые привели к ухудшению экономического положения корпорации. Все внимание сосредоточено на их, а в наиболее простых случаях пользователю выдаются заранее подготовленные типовые рецепты-рекомендации (указания) с расчетными данными, характеризующими конкретные процессы или объекты.

Достоинство данной стратегии в том, что, поскольку общее число значимых показателей и связей существенно сокращается и акцент делается на самых «узких» местах, лаконичность изложения сути вопроса о затруднительном положении корпорации и путях выхода из него становится важнейшим критерием выбора стратегии реализации данного этапа проектирования ЭС.

Поиск факторов, вызвавших ухудшение состояния предприятия, осуществляется путем сравнения плановых и фактических значений экономических показателей, после чего на основе факторного анализа выявляются те показатели, негативное влияние которых на результирующие данные может оказаться настолько значительным, что это приведет к недопустимому отклонению от нормативных величин. Если обнаружены факторы, отклонения уже превысили норму, информация о них служит для ЭС сигналом для выработки в оперативном порядке соответствующих рекомендаций (советов, указаний) ЛПР. Таким образом, выявленные локальные причины ухудшения показателей эффективности бизнеса являются

основой для установления локального диагноза состояния предприятия с последующей выдачей рекомендаций по устранению этих причин.

Во втором случае имеют место комплексный логический анализ динамики экономических показателей, выявление их отклонений от нормы и формирование диагноза с выдачей рецепта, в котором излагается перечень мероприятий, необходимых для выхода из создавшейся ситуации. Комплексность данного анализа заключается в выявлении таких состояний бизнеса, которые при автономном рассмотрении отдельных показателей не идентифицируются и нежелательными не считаются, поскольку значения последних порознь удовлетворяют допустимым нормам. В основе комплексного анализа лежат правила «ЕСЛИ_ТО», позволяющие установить диагноз системы. Например, запись вида

$$\begin{aligned} \text{DIA}_5: & A(UM \leq 5\%), B(UB > 10\%), C(UM); \\ & A(UM \leq 5\%), B(UB > 20\%), C = 20\%; \\ & A(UM), B(UM), C(UB > 5\%), \end{aligned} \quad (5.1)$$

представляет правило для доказательства диагноза за номером 5, обозначенное как DIA_5, которое следует читать таким образом:

ЕСЛИ показатель А уменьшился на 5%, И показатель В увеличился на 10%, ИЛИ показатель С уменьшился, ИЛИ показатель А уменьшился на 5%, И показатель В увеличился на 20%, ИЛИ показатель С равен 20%, ИЛИ показатель А уменьшился, ИЛИ показатель В уменьшился, ИЛИ показатель С увеличился на 5%, ТО имеет место ДИАГНОЗ DIA_5

Правило DIA_5 получают из таблиц, составленных на основе консультаций с экспертами.

Формально число диагнозов равно числу сочетаний всех возможных ситуаций на объекте управления, однако часть этих ситуаций группируется и формирует классы, указывающие на наличие сходных ситуаций. Поскольку меры доверия к связям между ситуациями указаны, то диагноз носит нечеткий характер – но, указав или рассчитав надежность каждого входящего в него условия, можно установить общую надежность диагноза.

Пусть, например, надежность каждого из условий правила

$$\begin{aligned} \text{DIA}_7: & A(UM > 2\%), B(UM < 15\%), 0 < C < 100; \\ & A(UM > 5\%), 10 < B < 20; C(UB > 10\%), \end{aligned} \quad (5.2)$$

равна, соответственно, 0,64 и 0,8.

Тогда общая мера доверия к такому диагнозу, который состоит из указанных двух условий, будет равна

$$\begin{aligned} \text{МД}[\text{DIA}_7] &= \text{МД}[\text{ПР1}] + \text{МД}[\text{ПР2}] (1 - \text{МД}[\text{ПР1}]) = \\ &= 0,64 + 0,8 (1 - 0,64) = 0,93. \end{aligned}$$

Установив, таким образом, диагноз, далее необходимо перейти к расчету и оформлению рекомендаций, в которых должны быть предусмотрены конкретные мероприятия. Оформление рекомендаций в соответствии с выбранной стратегией анализа экономического положения предприятия может ориентироваться на следующие варианты обслуживания конечного пользователя.

1. Предварительно заполненные «бланки», выдаваемые без изменений по мере подтверждения соответствующего диагноза: этот вариант возможен при идентификации относительно несложных диагнозов, где рекомендации являются типовыми (общеизвестными) – данный вариант не всегда эффективен, так как ЛПР желательно знать не только о том, что нужно делать в принципе – но и о конкретных деталях, количественных параметрах рекомендуемых ЭС управленческих действий.

2. Составление рекомендаций в зависимости от результатов проведенного анализа и выполненных расчетов – данный вариант является более гибким, поскольку позволяет составлять рекомендации в динамическом режиме: используя правила «ЕСЛИ_ТО» и оперативно учитывая дополнительные качественные и количественные изменения, происходящие на предприятии.

Однозначно установить диагноз без дополнительных процедур и вопросов к ЛПР удастся лишь в простейших случаях – обычно же расчетно-диагностическая ЭС формирует несколько гипотез и подготавливает ряд вопросов для их уточнения. Гипотезы группируются в соответствии с полученными от ЛПР ответами и предварительно выработанными ЭС критериями – каждая из гипотез проверяется и, в зависимости от результатов проверки, принимается или отвергается. При этом сами гипотезы формируются на основании иерархически связанных между собой потенциальных диагнозов.

Диагнозы можно группировать и связывать по степени общности: по отношению «род – вид». Например, на первом уровне диагнозами могут быть «низкая платежеспособность», «низкая креди-

тоспособность», «низкая прибыльность», «высокая стабильность» и т.д. В свою очередь первый диагноз: «низкая платежеспособность» может быть связан с более детальными прогнозами: такими как низкие показатели ликвидности, высокие показатели активности капитала и т.п. Четкая классификация возможных «болевых точек» бизнеса позволяет ЭС увереннее диагностировать экономическое состояние предприятия и способствует более эффективной поддержке действий ЛПП.

Формализация процедур контроля правильности составления финансовой отчетности (блок 2.3) – на данном этапе осуществляется контроль достоверности бухгалтерской и финансовой отчетности, который может быть выполнен на трех уровнях:

- сопоставление значений показателей из различной отчетной документации, что необходимо, поскольку при проектировании ЭС могут использоваться документы двух подуровней, в которых показатели могут частично дублироваться или противоречить друг другу: документация, которая предоставляется по требованию контролирующих организаций, и документация, отражающая рабочие операции производственно-хозяйственной деятельности предприятия. Исходной информацией в обоих случаях служат таблицы, где указаны координаты сравниваемых показателей (номера форм, строк, граф) и правила проверки, а также документ с текущей финансово-бухгалтерской отчетностью;
- перерасчет наиболее важных показателей (налоги, доходы, прибыль, затраты) и сравнение их с отчетными данными – этот этап предполагает разработку графа показателей, отражающего основные характеристики предприятия, с помощью которого производится повторный расчет проверяемых показателей и сравнение его результатов с данными отчетных документов;
- логический анализ правильности составления отчетной документации, который заключается в проверке правильности списания затрат, начисления заработной платы, определения себестоимости продукции и т.п. Вычисления здесь минимальны по объему (почти отсутствуют), так как основное внимание уделяется выявлению действий администрации, не согласующихся с законодательством.

Наполнение системы значениями и данными (блок 3) состоит из трех следующих этапов.

Написание программного кода (блок 3.1) предполагает выполнение операций и процедур, зависящих от выбранной среды разработки. Каких-либо специфических особенностей в рассматриваемом случае они не имеют.

Проектирование интерфейса пользователя и эксперта (блок 3.2) – в основу создания данного интерфейса должны быть положены международные стандарты, в соответствии с которыми подавляющее большинство ЭС использует диалоговые системы [91]. Диалог между ЛПР и ЭВМ можно определить как последовательный обмен информацией между программой и пользователем, осуществляемый с помощью интерактивного терминала по определенным правилам. В настоящее время существуют три формы такого диалога .

1. Диалог «Вопрос – ответ», который допускает вывод только одного вопроса, на который следует дать четкий ответ. В зависимости от содержания полученного ответа, ЭС решает, каким должен быть следующий вопрос – чаще всего эти вопросы формируются из одного слова с ограниченной грамматикой и подсказкой формата ответа (типа «Номер накладной XXX?») Данный диалог удобен в том случае, когда диапазон входных сообщений велик, а последующий вопрос зависит от ответа на предыдущий вопрос.

2. Диалог типа «Меню» – суть которого заключается в том, что пользователю на разных этапах диалога предлагаются возможные варианты ответа для его продолжения, такой диалог целесообразен, если пользователю необходимо видеть все возможные варианты ответов или если диапазон возможных ответов невелик.

3. Диалог «Экранная форма» – который предполагает получение от пользователя сразу всей информации для выбора режима решения задачи. Этот тип диалога используется в том случае, когда последующий ответ не зависит от предыдущего вопроса, он эффективен, если известна последовательность независимых выходных сообщений.

4. Диалог типа «Команда» – применяемый при манипулировании функциональными клавишами, когда пользователь, нажав нужную клавишу, запускает соответствующую процедуру. Диалог данного типа целесообразен, если число входных сообщений для

ЛПР достаточно мало и число ответов для идентификации задачи ограничено.

На практике в рамках решения одной экономической задачи разные типы диалога, как правило, сочетаются. Проектирование диалогового режима включает разработку его динамики и экранных форм, отражающих состояние диалога. Динамика диалога может отражаться с помощью направленного графа, вершины которого соответствуют определенным состояниям диалога (этапам), а дуги указывают условия перехода от одной вершины к другой, или с помощью диаграммы структуры специализированного программного приложения типа SSD (System Structure Diagram) [64].

Проектирование блока объяснений действия ЭС (блок 3.3) является одной из наиболее сложных и трудоемких задач при создании советующей СУ. Поскольку для пользователя важно знать, почему и как ЭС пришла к тому или иному результату расчета или выводу, ее проектировщик должен разработать средства отслеживания промежуточных дедуктивных выводов и сохранять их до конца сеанса работы с ЭС. Если у пользователя возникнут сомнения, например, относительно того, нужно ли увеличивать стоимость активной части основных фондов, чтобы повысить рентабельность производства – можно будет обратиться к соответствующему правилу в таблице, где детально описывается сложившаяся экономическая ситуация, чего, как правило, вполне достаточно, чтобы убедить ЛПР в правильности предложенного ему ЭС решения. Осторожности при этом требует неопределенность информации, выражаемой мерами доверия – здесь часто требуются дополнительные усилия и затраты на сохранение промежуточных вычислительных мер и коэффициентов доверия, поэтому нечеткая логика в пояснениях используется сравнительно редко.

Тестирование и внедрение ЭС (блок 4) состоит из трех этапов, которые часто выполняются параллельно друг другу.

Тестирование (блок 4.1) – данный этап фактически включается в каждую стадию прототипирования ЭС, хотя формально тестирование относится к заключительной стадии процесса ее разработки. Операционное прототипирование, которое характеризуется возможностью изменения целей проектирования и, соответственно, уточнения спецификаций ЭЭС, предъявляет особые требования к

доказательству корректности (верификации – verification) и соответствия разрабатываемой ЭС предъявляемым требованиям (концептуальное тестирование – validation) – обе эти задачи должны выполняться параллельно, то есть одновременно с процессом разработки советующей СУ. По аналогии с технологией тестирования традиционных программных средств процесс верификации (логического тестирования ЭС) можно интерпретировать как альфа-тестирование ИС, а концептуальное тестирование – как бета-тестирование, хотя тестирование ЭС все же существенно отличается от тестирования традиционных ИС. В то время как достаточно строгие предварительные спецификации традиционной ИС позволяют программисту осуществлять эти работы (особенно верификацию программ) самостоятельно, для тестирования ЭС необходимо привлекать экспертов в предметной области.

Согласно [17] можно выделить три аспекта тестирования ЭС:

- тестирование исходных данных – проверка фактографической информации, которая служит основой для проведения экспертизы в заданной предметной области;
- логическое тестирование – обнаружение логических ошибок, не зависящих от предметной области: таких как избыточность и цикличность;
- концептуальное тестирование – проверка общей структуры ЭС и учета в ней всех аспектов решаемой задачи, на этом этапе проведение тестирования невозможно без привлечения конечных пользователей.

Выполнение (блок 4.2) – цель данного этапа состоит в создании одного или нескольких прототипов ЭС расчетно-диагностического типа, решающей поставленные задачи в заданной предметной области. С учетом результатов тестирования и опытной эксплуатации на данном этапе создается ЭС как конечный программный продукт, пригодный для промышленного использования. Разработка прототипа ЭС состоит в программировании его компонентов (или выборе их из числа уже имеющихся у ЛППР инструментальных средств) и наполнении БЗ.

Опытная эксплуатация (блок 4.3) – на данном заключительном этапе проверяется пригодность системы для конечного пользователя: здесь ЭС занимается решением конкретных (по возможности самых разных по тематике и уровню сложности) задач при

взаимодействии с разными пользователями – эту работу целесообразно организовать не на стенде разработчика ЭС, а на рабочем месте ее пользователя.

5.8. Использование функций полезности при управлении выводом в ЭС расчетно-диагностического типа

Алгоритм работы интерпретатора в обобщенном виде может быть описан следующим образом: по имеющимся в БЗ данным ЭС вычисляются цены свидетельств, затем определяется свидетельство с наибольшей ценой и, с использованием соответствующего вопроса из БЗ, выявляются (узнаются) у пользователей соответствующие этому свидетельству данные; затем проводится анализ накопленных данных на предмет возможности формулировки окончательного заключения (принятия решения), или, если решение не принимается, то пересчитываются цены свидетельств, после чего процесс повторяется и т.д.

Задача диагностирования ЭС формально может быть представлена парой $\{X; f^x\}$, где X – множество альтернатив (типов состояний), содержащихся в БЗ, а $f^x = (f_1^x, \dots, f_m^x)$ – m -мерный вектор свидетельств (симптомов), характеризующих конкретную альтернативу $x \in X$: например, f_1^x – коэффициент ликвидности ниже нормативного значения и т.д. Нечеткие правила продукционного типа могут быть записаны в форме «если $Z^j(f^{xj}) \geq C_x$, то x'' », где $f^{xj} = (f_{i1}^{xj}, \dots, f_{it_j}^{xj})$ – «подвектор» (далее без кавычек) вектора f^x , $t_i \leq m_x$; $Z^j(f^{xj})$ – функция принадлежности, характеризующая степень истинности альтернативы $x \in X$ при реализующихся симптомах f^{xj} . При этом функции принадлежности Z^j определяются в общем случае на основе исчисления нечеткой логики с использованием значений функции принадлежности $Z_i(f_i^x)$, на области симптомов f_i^x , C_x – порог истинности альтернативы $x \in X$.

В нашем случае нечеткие подмножества описываются лингвистическими переменными, определяющими симптомы или собственно диагнозы $x \in X$. В первом случае областями определения

функций принадлежности $Z_i(f_i^x)$, $i = 1; \dots, m_x$, являются множества d_i^x возможных значений свидетельств f_i^x , во втором случае областями определения функций принадлежности $Z^j(f^{xj})$ являются множества $D_{xj} = d_{i_1}^x \times \dots \times d_{i_{t_j}}^x$ – декартовы произведения областей возможных значений соответствующих свидетельств $f_{il}^{xj} \in f^{xj}$, $l = 1; \dots, t_j$, так, например, нечеткое подмножество, описываемое понятием (лингвистической переменной) «низкий коэффициент ликвидности» может определяться на множестве значений $f_i^x \in [0; 0,6]$. В качестве лингвистических переменных, относительно которых определяются функции принадлежности $Z^j(f^{xj})$, выступают характеристики состояния.

Рассмотрим возможность принятия решения на основе **полезностной интерпретации** функций принадлежности. Любая скалярная функция $\mu(y)$ одной или нескольких переменных определяет на области значений своего аргумента $y \in Y$ отношение предпочтения в соответствии с соотношением $\mu(y') \geq \mu(y'') \rightarrow y' \succ y''$. Данное утверждение справедливо и для функций принадлежности $Z_i(f_i^x)$, $i = 1; \dots, m_x$; $Z^j(f^j)$, $j \leq m_x$, используемых в нечетких правилах продукционного типа в задаче $\{X; f^x\}$.

Уточним сущность отношений предпочтения, описываемых функциями принадлежности $Z_i(f_i^x)$ и $Z^j(f^{xj})$, $i = 1; \dots, m_x$; $j \leq m_x$; $\forall x \in X$, для этого, наряду с задачей $\{X; f^x\}$, рассмотрим совокупность задач $\{F^{xj}, S, R\}$, где $\dim S = 1$; F^{xj} – некоторое (любое) подмножество возможных реализаций подвектора свидетельств f^{xj} , относящихся к альтернативе $x \in X$; R – принцип оптимальности, отвечающий целевой установке выбора такой реализации $f_k^{xj} \in F^{xi}$ подвектора f^{xj} свидетельств, которая наилучшим образом выражает состояние проблемы $x \in X$.

Под свойствами – критериями реализаций $f^{xj} \in F^{xi}$ при этом понимаются соответствующие лингвистические переменные, а под

их оценками – значения свидетельств $f_i^x \in f^{x_i}$, то есть тождественные отображения $f_i^x \rightarrow f^x$. Тогда, согласно смысловой сущности функций принадлежности при фиксированном $x \in X$, принципу оптимальности R в задаче $\{F^{x_j}, S, R\}$ соответствует отношение предпочтения, моделируемое функцией принадлежности $Z^j(f^{x_j})$. Следовательно, рассматриваемые отношения предпочтения изоморфны отношениям \succ , то есть системе предпочтений эксперта-специалиста в задачах $\{F^{x_j}, S, R\}$, $\forall x \in X$.

В свою очередь, из того, что предпочтения ЛППР в задачах $\{F^{x_j}, S, R\}$ описываются функциями полезности $v_i(f_i^x)$; $\varphi_i(f^{x_j})$ и $\varphi(f^x)$ при $\dim f^{x_j} = m_x$, следует изоморфность (в смысле описания системы предпочтений ЛППР и измерения степени принадлежности значений свидетельств соответствующим нечетким подмножествам) указанным функциям полезности и функций принадлежности $Z_i(f_i^x)$; $Z_i(f^{x_j})$ и $Z(f^x)$. Последнее определяет полезностную интерпретацию функций принадлежности в задаче $\{X; f^x\}$. Рассмотрим одну из версий (возможностей реализации) управления выводом в ЭС, которые обеспечиваются полезностной интерпретацией функций принадлежности в нечетких правилах продукционного типа.

1. Положить $I_x = \{1, \dots, m_x\}$ для всех $x \in X$.

2. Для каждой альтернативы $x \in X$ найти оценку $\varphi_x(f^x)$ путем извлечения из БЗ наличных данных о значениях свидетельств f_i^x , $i = 1; \dots, m_x$. Отсутствие данных по некоторым (или по всем) свидетельствам приводит к значениям $v_i(f_i^x) = Z_i(f_i^x) = 0$ соответствующих базовых функций полезности.

3. Проверить, существует ли альтернатива $x^* \in X$, для которой $\varphi_{x^*}(f^{x^*}) \geq \tilde{C}_{x^*}$ – если такая альтернатива имеется, необходимо выдать ее пользователю как решение задачи $\{X; f^x\}$ и завершить вывод.

4. Убрать из X все альтернативы, для которых $I_x = 0$, так как это условие означает, что обо всех свидетельствах, касающихся

альтернативы x , пользователь уже опрошен. Если $X = 0$, то перейти к шагу 9.

5. Найти $x' = \arg \min_{x \in X} (\tilde{C}_x - \varphi_x(f^x))$.

6. Найти $i_0 = \arg \min_{i \in I_{x'}} \left\{ \left| \frac{\partial \varphi_{x'}(f^{x'})}{\partial f^{x'}} \right| \right\}$.

7. Используя вопрос из БЗ, расспросить пользователя по поводу свидетельства $f_{i_0}^{x'}$, которое в общем случае может характеризоваться симптомом не только явления x' , но и других состояний проблемы $x \neq x'$, $x \in X$.

8. Для всех $x \in X$, в состав характеристик которых входит свидетельство $f_{i_0}^{x'}$, убрать из множества I_x номер, соответствующий свидетельству $f_{i_0}^{x'}$, перейти к шагу 2.

9. Выдать пользователю сообщение о том, что диагностируемый случай не относится ни к одному из состояний проблемы $x \in X$.

5.9. Реализация ЭС для управления региональной КСС

Этап идентификации. Условием эффективного управления и залогом успешного развития современной корпорации является научно обоснованный, оперативный и достоверный анализ ее финансово-экономического состояния – на котором должны базироваться действия ЛПР. В связи с этим существенно возрастают приоритетность и роль ФЭА (см. в разделе 5.5) – как в отношении методов и средств комплексного изучения состояния бизнеса рассматриваемого предприятия и факторов его формирования, так и с точки зрения оценки финансовых рисков и прогнозирования уровня доходности капитала.

Рассмотрим в качестве конкретного примера крупного оператора сотовой связи регионального уровня – условимся далее для краткости именовать эту СС организационно-технического типа КСС. Финансовое состояние КСС оценивает система показателей, отражающих наличие, размещение и использование его финансовых ресурсов (денежных средств), оно характеризуется обеспеченностью ресурсами, необходимыми для нормального функционирования; целесообразным их размещением и эффективным использо-

ванием; денежными взаимоотношениями с другими юридическими и физическими лицами; платежеспособностью и финансовой устойчивостью предприятия. Финансовое состояние отражает, по сути дела, конечные результаты деятельности КСС – которые интересуют не только работников данного предприятия, но и его партнеров по экономической деятельности, государственные, финансовые и налоговые органы. С точки зрения теории СС, ФЭА можно разделить на две части: внутренний и внешний анализ, которые существенно различаются между собой по целям и содержанию.

Внутрифирменный ФЭА – это процесс исследования механизма формирования, размещения и использования капитала с целью поиска резервов укрепления финансового состояния, повышения доходности и наращивания собственного капитала предприятия как субъекта хозяйственной деятельности. Внешний ФЭА представляет собой процесс исследования финансового состояния субъекта хозяйствования с целью прогнозирования степени риска инвестирования капитала и уровня его доходности.

Согласно [86; 106], схема ФЭА включает следующие основные блоки (подсистемы ЭС).

1. Оценка имущественного положения и структуры капитала:

- анализ источников формирования капитала;
- анализ размещения капитала.

2. Оценка эффективности и интенсивности использования капитала:

- анализ рентабельности (доходности) капитала;
- анализ оборачиваемости капитала.

3. Оценка финансового состояния предприятия:

- оценка динамики состава и структуры активов, их состояния и движения;
- оценка динамики состава и структуры собственного и заемного капитала, их состояния и движения;
- анализ величин и динамики абсолютных и относительных показателей финансовой устойчивости;
- анализ платежеспособности и ликвидности активов баланса.

4. Оценка кредитоспособности и риска банкротства.

Из приведенной структуры ФЭА следуют цели, а также задачи ЭС расчетно-диагностического типа, предназначенной для проведения комплексной оценки финансового состояния предприятия.

Можно выделить две группы задач, которые призвана решать данная ЭС:

- анализ финансовой устойчивости предприятия;
- анализ ликвидности предприятия.

Наряду с этим, ЭС решает целый ряд других задач:

- ввод и проверка правильности составления бухгалтерской отчетности;
- расчет аналитических коэффициентов, необходимых для дальнейшего проведения анализа;
- выработка рекомендаций, направленных на более эффективное использование финансовых ресурсов и укрепление финансового состояния предприятия.

Как уже было сказано, конечной целью ФЭА является получение минимально необходимого числа ключевых (*кумулятивных* – наименьших по объему и наиболее информативных) параметров, дающих объективную и точную картину финансового состояния предприятия, его прибылей и убытков, изменений в структуре пассивов и активов, расчетов с дебиторами и кредиторами. По итогам ФЭА для ЛПП представляют интерес как текущее состояние предприятия, так и его «проекция» на ближайшую или более отдаленную перспективу – то есть ожидаемые параметры финансового состояния КСС.

Описание (вербальная модель) предметной области для ЭС расчетно-диагностического типа, цель которой состоит в обеспечении заданного уровня финансовой стабильности, может быть следующим: финансовое состояние предприятия может быть отличным, хорошим, удовлетворительным или неудовлетворительным. Для получения итоговой оценки финансового состояния требуется проанализировать показатели финансовой устойчивости и показатели ликвидности и платежеспособности.

Оценка ликвидности (платежеспособности) производится как по обобщенным показателям в результате их проверки на соответствие нормативным данным, так и на основе соотношения статей актива и пассива баланса (ликвидности баланса). Оценка финансовой устойчивости формируется из оценок трехкомпонентного показателя типа финансовой ситуации, определяющего покрытие основных и оборотных средств собственными и заемными оборотными финансовыми источниками, а также оценок коэффициентов

устойчивости по сравнению с их нормативными значениями. При нахождении значений сформулированных подцелей и оценке финансового состояния в целом следует принимать во внимание невозможность точно установить, насколько удовлетворительными (или неудовлетворительными) являются те или иные конкретные значения показателей. Кроме того, нормативные значения финансовых показателей, на основании которых строятся конечные выводы, могут зависеть от других (предварительных, исходных, базовых) показателей. Поэтому, несмотря на то, что все возможные исходы решения проблемы с помощью правильно выбранной системы показателей могут быть описаны всегда, на деле они могут быть описаны лишь с некоторой степенью уверенности (неопределенности).

Рассмотрим особенности реализации ЭС для управления КСС. Оценка финансового состояния компании может быть дана в виде одного из следующих вариантов:

- отличное состояние (наилучший, редко встречающийся результат – это абсолютно устойчивое финансовое состояние, при котором все показатели удовлетворяют нормативным ограничениям, ухудшения финансового состояния по сравнению с предыдущими периодами деятельности не отмечено);
- хорошее состояние (нормальное финансовое состояние – чаще встречающийся на практике результат, при котором большинство показателей удовлетворяют нормативным значениям и платежеспособность предприятия гарантирована);
- удовлетворительное состояние (часто встречающийся результат, когда большинство показателей удовлетворяют нормативным значениям или имеют место незначительные отклонения, причем может быть отмечена тенденция к ухудшению финансового состояния, но финансовая устойчивость удовлетворительная);
- неудовлетворительное состояние в двух его вариантах:
 - а) в допустимых пределах (большинство показателей не удовлетворяет нормативным значениям, однако не замечено тенденции к общему ухудшению финансового состояния и налицо лишь незначительные ухудшения по отдельным показателям, причем возможно постепенное улучшение положения);
 - б) кризисное состояние (предприятие на грани банкротства: баланс абсолютно неликвиден, финансовые показатели не

удовлетворяют нормативным значениям, необходимы немедленные меры по санации предприятия).

При проведении ФЭА необходимо учитывать ряд важных факторов, оказывающих влияние на состояние предприятия:

- уровень и состав (по срокам погашения) обязательств;
- уровень и состав (в разрезе ликвидности) собственных средств;
- тенденция развития предприятия (изменения в структуре отчетности по сравнению с прошлыми периодами деятельности);
- особенности вида деятельности (отрасли) предприятия;
- общеэкономические факторы (например, уровень инфляции).

Для получения итоговой оценки состояния предприятия необходимо выделить промежуточные этапы (подцели) проводимого анализа. Указанными подцелями ФЭА в данном случае являются:

- оценка ликвидности – анализ платежеспособности КСС с точки зрения краткосрочной перспективы, оценка средств предприятия, достаточных для уплаты долгов по всем краткосрочным обязательствам и одновременного бесперебойного осуществления процесса производства и реализации продукции;
- оценка финансовой устойчивости – анализ на более долгосрочный период, оценка финансовой обеспеченности процесса деятельности на перспективу, с учетом степени зависимости КСС от внешних кредиторов, инвесторов и партнеров. Финансовую устойчивость предприятия определяет соотношение основных и оборотных средств (запасов и затрат), а также объемов собственных и заемных источников их формирования.

Основным источником информации при проведении ФЭА служит бухгалтерский баланс по форме № 1, определяющий состояние дел предприятия на момент его составления (образец данного баланса в укрупненном виде демонстрирует таблица 5.5).

Таблица 5.5. Укрупненный бухгалтерский баланс КСС

Наименование	Код строки	Обозначение
1	2	3
АКТИВ		
I. Внеоборотные активы (110+120+140+150)	190	ВА
Нематериальные активы	110	<i>НМА</i>

Таблица 5.5. Укрупненный бухгалтерский баланс КСС (продолжение)

1	2	3
Основные средства	120	<i>ОС</i>
Долгосрочные финансовые вложения	140	<i>ФВ_{долг}</i>
Прочие внеоборотные активы	150	<i>ВА_{пр}</i>
II. Оборотные активы (210+230+240+250+260+270)	290	<i>ОА</i>
Запасы	210	<i>З</i>
Дебиторская задолженность	230+240	<i>ДЗ</i>
Долгосрочная	230	<i>ДЗ_{долг}</i>
Краткосрочная, в том числе	240	<i>ДЗ_{крат}</i>
задолженность участников по взносам в уставный капитал	244	<i>ДЗ_{уст. кап.}</i>
Краткосрочные финансовые вложения, в том числе	250	<i>ФВ_{крат}</i>
Собственные акции, выкупленные у акционеров	252	<i>ФВ_{акции}</i>
Денежные средства	260	<i>ДС</i>
Прочие оборотные активы	270	<i>ОА_{пр}</i>
БАЛАНС	190+290	<i>БН</i>
ПАССИВ		
III. Собственный капитал	490+460	<i>СК</i>
Капитал и резервы	490	<i>К</i>
Нераспределенная прибыль	460	<i>НП_р</i>
IV. Заемный капитал	590+690	<i>ЗК</i>
Долгосрочные обязательства	590	<i>ДО</i>
Краткосрочные обязательства (610+620+630+640+650+660)	690	<i>КО</i>
Кредиты и займы	610	<i>К_{редит}</i>

Таблица 5.5. Укрупненный бухгалтерский баланс КСС (окончание)

1	2	3
Кредиторская задолженность	620	<i>КЗ</i>
Задолженность участникам по выплате доходов	630	<i>З_{дох}</i>
Доходы будущих периодов	640	<i>ДБП</i>
Резервы предстоящих расходов	650	<i>РПР</i>
Прочие краткосрочные обязательства	660	<i>КО_{пр}</i>
БАЛАНС	460+460+590+690	<i>БН</i>

Этап концептуализации. На данном этапе происходит выделение ключевых понятий и определение их отношений. Ключевыми словами являются:

- финансовое состояние предприятия;
- финансовая устойчивость;
- ликвидность и платежеспособность;
- коэффициенты устойчивости;
- коэффициент автономии;
- коэффициент соотношения собственных и заемных средств;
- коэффициент маневренности;
- коэффициент обеспеченности запасов;
- коэффициент абсолютной ликвидности;
- коэффициент ликвидности;
- коэффициент покрытия;
- излишек (недостаток) собственного оборотного капитала;
- излишек (недостаток) собственного оборотного и долгосрочного заемного капитала;
- излишек (недостаток) общей величины всех основных источников для формирования запасов;
- абсолютная устойчивость финансового состояния;
- нормальная устойчивость финансового состояния;
- неустойчивое финансовое состояние;
- кризисное финансовое состояние;
- наиболее ликвидные активы;
- быстрореализуемые активы;

- медленно реализуемые активы;
- труднореализуемые активы;
- наиболее срочные обязательства;
- краткосрочные пассивы;
- долгосрочные пассивы;
- постоянные пассивы.

Напомним, что любая ЭС создается с определенной целью, достижение которой зависит от достижения подчиненных целей. Эти цели формулируются, исходя из анализа проблем, стоящих перед предприятием, перспектив развития и опыта производственного коллектива. Зависимость главной цели от подцелей удобно представлять с помощью дерева целей (см. рис. 5.26) или в табличном виде – см. таблицу 5.6.

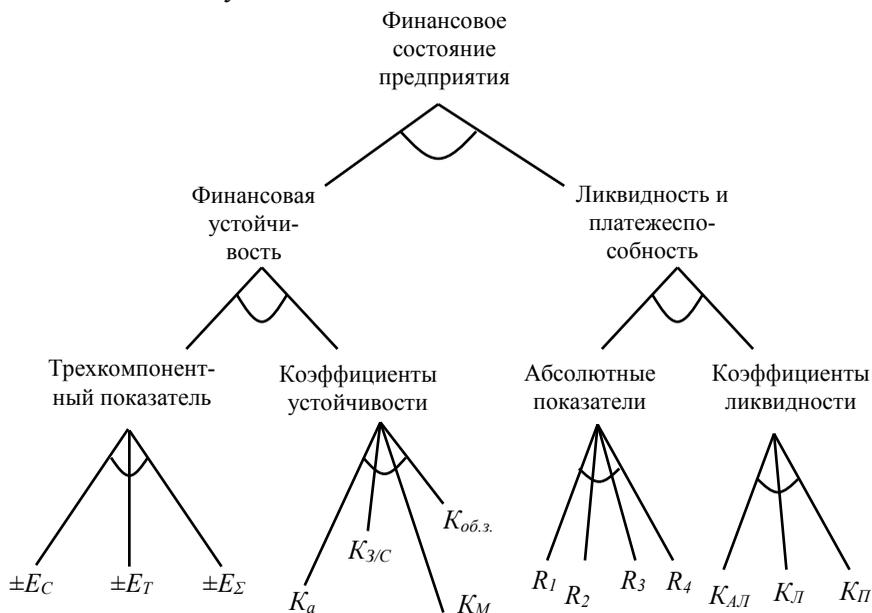


Рис. 5.26. Дерево целей оценки финансового состояния КСС

Финансовая устойчивость базируется на оптимальном соотношении между отдельными видами активов организации (оборотными или внеоборотными активами с учетом их внутренней структуры) и источниками их финансирования (собственными или привлеченными средствами) [1; 106].

Таблица 5.6. Представление дерева принятия решений для оценки финансового состояния КСС

Номер уровня	Код цели (подцели)	Наименование цели (подцели)	Номер уровня	Код цели (подцели)	Наименование цели (подцели)
1	2	3	4	5	6
1	Финансовое состояние предприятия	Обеспечить устойчивое финансовое состояние предприятия	2	Финансовая устойчивость	Обеспечить высокий уровень финансовой устойчивости
			2	Ликвидность и платежеспособность	Обеспечить высокую ликвидность и платежеспособность предприятия
2	Финансовая устойчивость	Обеспечить высокий уровень финансовой устойчивости	3	Трехкомпонентный показатель	Обеспечить значимые значения трехмерного показателя равно $S=(E_C >0 \text{ и } E_T >0 \text{ и } E_\Sigma >0)$
			3	Коэффициенты устойчивости	Обеспечить коэффициенты устойчивости, соответствующие нормативному значению
3	Трехкомпонентный показатель	Обеспечить значение трехмерного показателя равно $S = (E_C >0 \text{ и } E_T >0 \text{ и } E_\Sigma >0)$	4	$\pm E_C$	Обеспечить излишек собственного оборотного капитала
			4	$\pm E_T$	Обеспечить излишек собственного оборотного и долгосрочного заемного капитала
			4	$\pm E_\Sigma$	Обеспечить излишек величины основных источников для формирования запасов

Таблица 5.6. Представление дерева принятия решений для оценки финансового состояния КСС (продолжение)

1	2	3	4	5	6
3	Коэффициенты устойчивости	Обеспечить уровень коэффициентов устойчивости, соответствующий нормативному значению	4	K_a	Обеспечить увеличение коэффициента автономии
			4	$K_{з/с}$	Обеспечить стабилизацию коэффициента соотношения собственного и заемного капитала
			4	K_M	Обеспечить уменьшение коэффициента маневренности
			4	$K_{об.з.}$	Обеспечить увеличение показателя обеспеченности запасов
2	Ликвидность	Обеспечить высокую ликвидность предприятия	3	Абсолютные показатели ликвидности	Обеспечить соответствие между активами и платежными обязательствами предприятия по срокам
			3	Коэффициенты ликвидности	Обеспечить уровень коэффициентов ликвидности, соответствующий их нормативному значению
3	Абсолютные показатели ликвидности	Обеспечить соответствие между активами и платежными обязательствами предприятия по срокам	4	R_1	Обеспечить превышение величины наиболее ликвидных активов над величиной наиболее срочных обязательств
			4	R_2	Обеспечить превышение величины быстро реализуемых активов над величиной краткосрочных пассивов

Таблица 5.6. Представление дерева принятия решений для оценки финансового состояния КСС (окончание)

1	2	3	4	5	6
			4	R_3	Обеспечить превышение величины медленно реализуемых активов над величиной долгосрочных пассивов
			4	R_4	Обеспечить уменьшение величины трудно реализуемых активов
4	Коэффициенты ликвидности	Обеспечить уровень коэффициентов ликвидности, соответствующий их нормативному значению	4	$K_{АЛ}$	Обеспечить увеличение коэффициента абсолютной ликвидности
			4	$K_{Д}$	Обеспечить увеличение коэффициента быстрой ликвидности
			4	$K_{Д}$	Обеспечить увеличение коэффициента покрытия

Финансовая устойчивость КСС характеризуется системой абсолютных и относительных показателей, приведенных в таблицах 5.7-5.9. Наиболее общим показателем данной устойчивости является соответствие (либо несоответствие – излишек или недостаток) величины источников средств для формирования запасов.

Определение показателей обеспеченности запасов источниками их формирования позволяет классифицировать финансовые ситуации по степени устойчивости. В экономической литературе рекомендуется использовать трехмерный (трехкомпонентный) показатель S вида

$$S = \{S_1(\pm E_C), S_2(\pm E_T), S_3(\pm E_{\Sigma})\}. \quad (5.11)$$

Пользуясь показателем (5.11), выделяют четыре типа финансовых ситуаций, показанные в таблице 5.8.

Таблица 5.7. Абсолютные показатели финансовой устойчивости

Наименование показателя	Формула расчета	Что означает
Собственные оборотные средства	$E_C = CK - BA$	Отражают степень охвата разных видов источников
Собственные оборотные и долгосрочные заемные средства	$E_T = E_C + ДО$	
Все основные источники средств для формирования запасов	$E_{\Sigma} = E_T + КО$	
Излишек (недостаток) собственных оборотных средств	$\pm E_C = E_C - З$	Характеризуют обеспеченность запасов источниками их формирования
Излишек (недостаток) собственных оборотных и долгосрочных заемных средств	$\pm E_T = E_T - З$	
Излишек (недостаток) общей величины основных источников средств для формирования запасов	$\pm E_{\Sigma} = E_{\Sigma} - З$	

Таблица 5.8. Типы финансовых ситуаций

Трехмерный показатель	Финансовая ситуация
$E_C > 0$ и $E_T > 0$ и $E_{\Sigma} > 0$	Абсолютно устойчивое финансовое положение: запасы минимальны
$E_C < 0$ и $E_T > 0$ и $E_{\Sigma} > 0$	Нормальное финансовое состояние: нормальные величины запасов
$E_C < 0$ и $E_T < 0$ и $E_{\Sigma} > 0$	Неустойчивое состояние финансовой положение: избыток запасов. Такое состояние можно восстановить путем привлечения долгосрочных и среднесрочных кредитов и заемных средств или обоснованным снижением уровня запасов
$E_C < 0$ и $E_T < 0$ и $E_{\Sigma} < 0$	Кризисное финансовое состояние: чрезмерная величина неподвижных и малоподвижных запасов. В этом случае денежные средства, краткосрочны ценные бумаги и дебиторская задолженность не покрываю даже кредиторской задолженности и просроченных ссуд

При нахождении значений сформулированных подцелей и оценке финансового состояния в целом следует принимать во внимание, что невозможно точно установить, насколько удовлетворительными (неудовлетворительными) являются те или иные значения показателей. Кроме того, сами нормативные значения некоторых финансовых показателей, на основании которых стоят выводы, зависят от множества обстоятельств: структуры баланса, особенностей деятельности предприятия, экономической ситуации в стране.

Поэтому, несмотря на то, что все возможные исходы решения проблемы могут быть описаны, они оцениваются с некоторой степенью уверенности (достоверности). При этом задача оценки финансового состояния предприятия приобретает нечеткий характер. Основное влияние на финансовую устойчивость при этом оказывает оценка трехмерного показателя примерно в соотношении 2:1.

Таблица 5.9. Относительные показатели финансовой устойчивости

Наименование показателя	Формула расчета	Нормативное значение	Уверенность	Что означает
1	2	3	4	5
Коэффициент автономии	$K_a = SK/БН$	$\leq 0,2$	«Неудовлетворительно» с уверенностью 100%	Рост коэффициента свидетельствует об увеличении финансовой независимости предприятия, снижения риска финансовых затруднений в будущие периоды, повышает гарантии погашением предприятием своих обязательств. Характеризует независимость от заемных средств. Показывает долю собственных средств в общей сумме всех средств предприятия
		$> 0,2;$ $\leq 0,5$	«Удовлетворительно» с уверенностью 75%	
		$> 0,5$	«Удовлетворительно» с уверенностью 100%	

Таблица 5.9. Относительные показатели финансовой устойчивости (окончание)

1	2	3	4	5
Коэффициент соотношения заемных и собственных средств	$K_{зс} = (ДО + КО)/СК$	$< 0,7$	«Удовлетворительно» с уверенностью 100%	Рост коэффициента отражает превышение величины заемных средств над собственными источниками их покрытия. Показывает, сколько заемных средств привлекло предприятие на 1 руб. вложенных в активы собственных средств
		$\geq 0,7$	«Неудовлетворительно» с уверенностью 100%	
Коэффициент маневренности	$K_M = (СК - ВА)/СК = E_d/СК$	$= 0,5$	«Удовлетворительно» с уверенностью 100%	Рост коэффициента говорит о снижении финансовой устойчивости, показывает способность поддерживать уровень оборотного капитала и пополнять его за счет собственных источников
		$\diamond > 0,5$	«Неудовлетворительно» с уверенностью 100%	
Коэффициент обеспеченности запасов и затрат собственными источниками финансирования	$K_{об.з} = (СК - ВА)/З = E_d/З$	$\leq 0,6$	«Неудовлетворительно» с уверенностью 100%	Является показателем устойчивости финансового состояния
		$> 0,6;$ $\leq 0,8$	«Удовлетворительно» с уверенностью 100%	
		$> 0,8$	«Удовлетворительно» с уверенностью 80%	

Под **ликвидностью организации** понимается ее способность покрывать свои обязательства активами, срок превращения которых в денежную форму соответствует сроку погашения обязательств. Ликвидность означает безусловную платежеспособность организации [1; 106]. **Платежеспособность организации** – внешний признак финансовой устойчивости и обусловлена степенью обеспеченности оборотных активов долгосрочными источниками.

Она определяется возможностью организации наличными денежными ресурсами своевременно погашать свои платежные обязательства (вовремя удовлетворять платежные требования поставщиков техники и материалов в соответствии с хозяйственными договорами, возвращать кредиты, производить оплату труда персонала, вносить платежи в бюджет). Поэтому анализ платежеспособности необходим не только для самой организации, но и для внешних партнеров и потенциальных инвесторов.

Таблица 5.10. Абсолютные показатели ликвидности

Наименование показателя	Как рассчитывается
Наиболее ликвидные активы	$A_1 = ДС + ФВ_{крат} - ФВ_{акции}$
Быстро реализуемые активы	$A_2 = ДЗ + ОА_{пр} - ДЗ_{уст. кап.}$
Медленно реализуемые активы	$A_3 = З + ДЗ_{уст. кап.} + ФВ_{долг}$
Трудно реализуемые активы	$A_4 = ВА - ФВ_{долг}$
Наиболее срочные обязательства	$П_1 = КЗ + З_{дох} + КО_{пр}$
Краткосрочные пассивы	$П_2 = К_{редит}$
Долгосрочные пассивы	$П_3 = ДО$
Постоянные пассивы	$П_4 = СК$

Понятия платежеспособности и ликвидности по смыслу близки, но второе считается более емким: поскольку ликвидность означает не только текущее состояние расчетов, но и характеризует соответствующие перспективы. Для оценки изменения степени платежеспособности и ликвидности нужно сравнивать показатели балансового отчета предприятия по различным группам активов и обязательств. На основе этого сравнения определяют абсолютные и относительные аналитические показатели (см. таблицы 5.9 и 5.10).

В общей оценке платежеспособности наибольшее значение придается оценке баланса по сравнению с оценкой коэффициентов ликвидности: например, факторы уверенности назначаются в соотношении 2:1. Оценку баланса ликвидности можно представить в виде матрицы – в качестве примера см. таблицу 5.11.

Таблица 5.11. Матрица для оценки баланса ликвидности

$R_1 = A_1 - П_1$	$R_2 = A_2 - П_2$	$R_3 = A_3 - П_3$	Оценка	Уверенность
> 0	> 0	> 0	«Удовлетворительно»	1
> 0	> 0	< 0	«Удовлетворительно»	0,8
> 0	< 0	> 0	«Удовлетворительно»	0,75
< 0	> 0	> 0	«Удовлетворительно»	0,7
< 0	< 0	< 0	«Неудовлетворительно»	1
< 0	< 0	> 0	«Неудовлетворительно»	0,75
< 0	> 0	< 0	«Неудовлетворительно»	0,7
> 0	< 0	< 0	«Неудовлетворительно»	0,6

В таблице 5.11. необходимым условием абсолютной ликвидности является выполнение первых трех неравенств, где R_1 – платежный излишек (недостаток) наиболее ликвидных активов; R_2 – платежный излишек (недостаток) быстро реализуемых активов; R_3 – платежный излишек (недостаток) медленно реализуемых активов. Четвертое неравенство носит балансирующий характер: его выполнение свидетельствует о наличии у предприятия собственных оборотных средств.

Таблица 5.12. Относительные показатели финансовой ликвидности

Наименование показателя	Формула расчета	Нормативное значение	Уверенность	Что означает
1	2	3	4	5
Коэффициент абсолютной ликвидности	$K_{ал} = (ДС + ФВ_{крат}) / КО$	< 0,2	«Неудовлетворительно» с уверенностью 50%	Отражает возможности предприятия погасить в ближайшее время какую то часть краткосрочной задолженности
		> 0,2; ≤ 0,7	«Удовлетворительно» с уверенностью 30%	
		> 0,7; ≤ 1,0	«Удовлетворительно» с уверенностью 50%	
		> 1,0	«Удовлетворительно» с уверенностью 30%	

Таблица 5.12. Относительные показатели финансовой ликвидности (окончание)

1	2	3	4	5
Коэффициент ликвидности	$K_{л} = \frac{ДС + ФВ_{крат} + ДЗ + ОА_{пр}}{КО}$	< 0,6	«Неудовлетворительно» с уверенностью 60%	Отражает прогнозируемые платежные возможности предприятия при условии своевременного проведения расчетов с дебиторами.
		> 0,6; ≤ 0,8	«Удовлетворительно» с уверенностью 40%	
		> 0,8; ≤ 1,0	«Удовлетворительно» с уверенностью 60%	
		> 1,0	«Удовлетворительно» с уверенностью 40%	
Коэффициент покрытия	$K_{л} = ОА/КО$	≤ 1,0	«Неудовлетворительно» с уверенностью 70%	Показывает платежные возможности предприятия, оцениваемые при условии не только своевременных расчетов с дебиторами и благоприятной реализации готовой продукции, но и продажи в случае необходимости прочих элементов материальных оборотных средств
		> 1,0; ≤ 2,0	«Удовлетворительно» с уверенностью 50%	
		> 2,0; ≤ 3,0	«Удовлетворительно» с уверенностью 70%	
		> 3,0	«Удовлетворительно» с уверенностью 50%	

Этап формализации. Используя схему, показанную на рис. 5.26 и таблицы 5.6-5.12, можно сформировать цепочку правил (фрагмент см. на рис. 5.27), на основании которой формулируются правила БЗ. Оценки показателей платежеспособности и финансовой устойчивости корректируются в зависимости от оценки тенденции развития предприятия (динамики соответствующих показателей). При этом производится сравнение показателей отчетного периода со средней величиной этих показателей за предшествующий период деятельности предприятия с учетом инфляционных процессов. В случае улучшения значений показателей коэффициент уверенности удовлетворительной оценки увеличивается, в случае ухудшения коэффициент уверенности, соответственно, уменьшается.

Проверка ограничений на значения отдельных показателей и их последующая оценка задаются в виде правил БЗ по следующей форме:

ЕСЛИ: <посылка> ТО: <заключение> Уверенность <значение>

Например:

Если $R_1 > 0$ И $R_2 > 0$ И $R_3 > 0$

ТО Ликвидность баланса = «Удовлетворительна» Уверенность 100

В случае независимого воздействия на оценку некоторой целевой переменной нескольких показателей (соответственно нескольких правил, оценивающих эту переменную) коэффициент уверенности (*cf*) формируется на основании формул (2.15)-(2.17).

Например, коэффициент уверенности оценки ликвидности на основе значений финансовых коэффициентов ликвидности рассчитывается следующим образом:

$K_{АЛ} = 1,1 \rightarrow$ Ликвидность коэффициентов = «Удовл» Уверенность 0,3

$K_{Л} = 1,51 \rightarrow$ Ликвидность коэффициентов = «Удовл» Уверенность 0,4

$K_{П} = 3,1 \rightarrow$ Ликвидность коэффициентов = «Удовл» Уверенность 0,5

$sf = 0,3 + 0,4 + 0,5 - 0,3 \cdot 0,4 - 0,4 \cdot 0,5 - 0,3 \cdot 0,5 + 0,3 \cdot 0,4 \cdot 0,5 = 0,79$.

5.10. Принципы эксплуатации и пример работы ЭС

расчетно-диагностического типа

В итоговом виде схема функционирования экономической ЭС расчетно-диагностического типа представлена на рис. 5.28. Работа системы начинается с ввода исходных данных. В демонстрационном прототипе все данные вводятся вручную пользователем, но предполагается, что баланс КСС может также заноситься в БД ЭС автоматически из других программ, используемых на предприятии, или из единой БД предприятия. Вся информация вводится один раз, и в дальнейшем, при необходимости повторного ФЭА, пользователь может использовать ее повторно. После того как все необходимые данные занесены в БД, управление передается модулю расчетов. По завершении расчетов управление передается модулю формирования диагноза, который идентифицирует состояние предприятия с одним из правил на основании доказательства (распознавания) соответствующих условий. Диагноз формируется на нескольких уровнях. На самом нижнем уровне рассчитанные показатели сравниваются с их нормативными значениями, а также определяется тенденция изменения этих показателей. Результаты этого сравнения сопоставляются с условиями правил из БЗ – таким образом и определяется финансовое состояние предприятия.

Рекомендации, как и диагноз, строятся по иерархическому принципу. Сначала формируются общие рекомендации на уровне ликвидности и финансовой устойчивости. Они вырабатываются на основе сформированного диагноза и правил из БЗ: условия правил сопоставляются с полученными оценками ликвидности и финансовой устойчивости. По итогам данного сопоставления формируются общие цели дальнейшей деятельности КСС по улучшению ее финансового состояния. На втором уровне (на уровне показателей) формируются конкретные рекомендации по достижению поставленных целей – они вырабатываются на основе анализа рассчитанных показателей ликвидности и финансовой устойчивости и правил из БЗ.

Кратко остановимся на принципах действия *блока объяснений*. В ЭС, основанных на правилах, объяснения получают обычно пу-

тем повторного прослеживания тех шагов рассуждений, которые привели к тому или иному выводу.

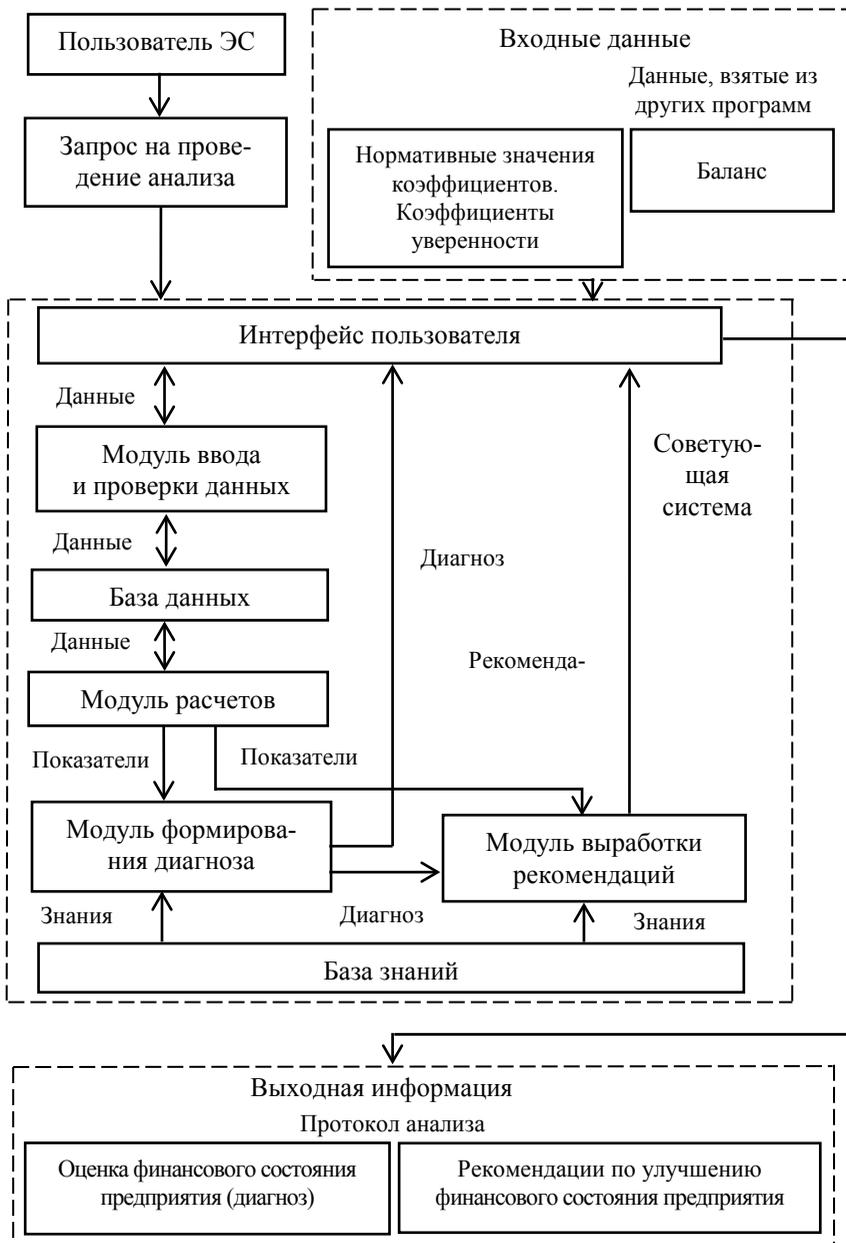


Рис. 5.28. Схема функционирования ЭС расчетно-диагностического типа

Однако такой способ объяснения, на наш взгляд, не является в достаточной мере убедительным, поскольку правила нередко дают непонятное (ненаглядное) объяснение в силу специфичности языка инструментального средства. Для формирования блока объяснений более приемлемым представляется графический метод объяснения, который, во-первых, позволяет пользователю выбрать необходимый уровень детализации показателей, во-вторых, демонстрирует в графическом виде и в соответствии с выбранным уровнем, какие правила задействовала БЗ. При реализации данного способа выбор уровня детализации осуществляется путем выбора соответствующего показателя на дереве целей (см. схему на рис. 5.26). На основании выбранного показателя формируется цепочка правил, применимая к рассматриваемой ситуации.

В процессе эксплуатации ЭС допускается настройка БЗ на особенности предприятия путем интерактивного введения нормативных ограничений и оценивающих коэффициентов уверенности. Программная реализация предусматривает интегрированное использование ЭС совместно с расчетными процедурами, использующими необходимую информацию непосредственно из БЗ.

Актив баланса		Пассив баланса	
	На начало периода	На конец периода	
I. Внеоборотные активы			
Нематериальные активы	76000	78000	
Основные средства	45000	46000	
Долгосрочные финансовые вложения	1500	1600	
Прочие внеоборотные активы	4500	5000	
Итого внеоборотных средств	127000	130600	
II. Оборотные активы			
Товарно-материальные запасы	4500	2700	
- аппараты	4300	2500	
- аксессуары	200	200	
Дебиторская задолженность	3500	3500	
Долгосрочная	2500	2500	
Краткосрочная, в т.ч.	1000	1000	
задолженность участников по взносам в уставный капитал	0	0	
Краткосрочные финансовые вложения, в т.ч.	1500	1600	
собственные акции, выкупленные у акционеров	0	0	
Денежные средства и их эквиваленты	6500	6600	
Прочие оборотные активы	5000	5000	
Итого оборотных средств	21000	19400	
Итого активы баланса	148000	150000	

Рис. 5.29. Окно для просмотра баланса предприятия

Пример работы системы. При загрузке системы на экране появляется окно с ниспадающим меню. Работа ЭС начинается с ввода исходных данных. Для ввода данных пользователь должен выбрать соответствующий пункт меню, при этом загрузится окно для просмотра, ввода и корректировки информации. Окно для просмотра баланса предприятия представлено на рис. 5.29.

Показатели платежеспособности

Показатели платежеспособности | Диаграмма - Абсолютные показатели | Диаграмма - Относительные показатели

Абсолютные показатели платежеспособности и ликвидности

Абсолютные показатели платежеспособности и ликвидности	На начало пер.	На конец пер.
Наиболее ликвидные активы, А1	8000	8200
Быстро реализуемые активы, А2	8500	8500
Медленно реализуемые активы, А3	6000	4300
Трудно реализуемые активы, А4	129500	129000
Наиболее срочные обязательства, П1	7500	7600
Краткосрочные пассивы, П2	1500	1900
Долгосрочные пассивы, П3	6000	6000

Относительные показатели платежеспособности и ликвидности

Относительные показатели платежеспособности и ликвидности	На начало пер.	На конец пер.
Коэффициент абсолютной ликвидности	0,444444444444	7,22105263157
Коэффициент ликвидности	0,918666666666	0,87894736842
Коэффициент покрытия	1,166666666666	1,02105263157

Рассчитать | Рекомендации | Закрыть

Рис. 5.30. Коэффициенты платежеспособности и ликвидности

Показатели финансовой устойчивости

Коэффициенты | Диаграмма - Абсолютные показатели | Диаграмма - Относительные показатели

Абсолютные показатели финансовой устойчивости

Абсолютные показатели финансовой устойчивости	На начало пер.	На конец пер.
Излишек/Недостаток собственных оборотных средств	-6500	-7300
Излишек/Недостаток собственных оборотных и долгосрочных заемных средств	-1500	-2300
Излишек/Недостаток общей величины основных источников средств для формирования запасов	18500	16700

Относительные показатели финансовой устойчивости

Относительные показатели финансовой устойчивости	На начало пер.	На конец пер.
Коэффициент автономии	0,84459459459	0,84
Коэффициент соотношения заемных и собственных средств	0,184	0,1904761904
Коэффициент маневренности	-0,016	-0,0365079368
Коэффициент обеспеченности запасов и затрат собственными источниками финансирования	-0,444444444444	-1,7037037037
Коэффициент обеспеченности оборотных активов собственными оборотными средствами		

Рассчитать | Рекомендации | Закрыть

Рис. 5.31. Коэффициенты финансовой устойчивости

После того как все необходимые данные занесены в БД, управление в ЭС передается модулю расчетов. Модуль расчетов осуществляет расчет показателей, на основе которых производится анализ финансового состояния КСС.

Обозначение коэффициентов	
Показатели финансовой устойчивости	
Излишек/Недостаток собственных оборотных средств	Характеризует обеспеченность запасов источниками их формирования
Излишек/Недостаток собственных оборотных и долгосрочных заемных средств	Характеризует обеспеченность запасов источниками их формирования
Излишек/Недостаток общей величины основных источников средств для формирования запасов	Характеризует обеспеченность запасов источниками их формирования
Коэффициент автономии	Характеризует независимость от заемки средств. Показывает долю собственных средств в общей сумме всех средств предприятия
Коэффициент соотношения заемных и собственных средств	Показывает, сколько заемных средств привлекло предприятие на 1 руб. вложенных в активы собственных средств
Коэффициент маневренности	Показывает способность предприятия поддерживать уровень собственного оборотного капитала и пополнять оборотные средства за счет собственных источников
Коэффициент обеспеченности запасов и затрат собственными источниками финансирования	Является показателем устойчивости финансового состояния
Показатели ликвидности и платежеспособности	
Коэффициент абсолютной ликвидности	Отражает платежные возможности предприятия, т.е. какую часть краткосрочной задолженности предприятие может погасить в ближайшее время
Коэффициент ликвидности	Отражает прогнозируемые платежные возможности предприятия при условии своевременного проведения расчетов с дебиторами
Коэффициент покрытия	Показывает платежные возможности предприятия, оцениваемые при условии не только своевременных расчетов с дебиторами и благоприятной реализации готовой продукции, но и продажи в случае необходимости прочих элементов материальных оборотных средств. характеризует запас прочности, возникающий

Рис. 5.33. Показатели финансовой устойчивости, ликвидности и платежеспособности

База знаний				
Платежеспособность		Финансовая устойчивость		Общие правила
Значение показателя	Краткий диагноз	Расширенный диагноз		Коэффициент уверенности
Если R1>0 и R2>0 и R3>0 ТО	удовлетворительно	Предприятие может рассчитаться по всем своим	+	1
Если R1>0 и R2>0 и R3<0 ТО	удовлетворительно	Предприятие не может рассчитаться лишь по	-	0,8
Если R1>0 и R2<0 и R3>0 ТО	удовлетворительно	Предприятие не может рассчитаться лишь по	-	0,75
Если R1<0 и R2>0 и R3>0 ТО	удовлетворительно	Предприятие может рассчитаться по всем,	+	0,7
Если R1<0 и R2<0 и R3<0 ТО	неудовлетворительно	Предприятие не в состоянии рассчитаться по	-	1
Если R1<0 и R2<0 и R3>0 ТО	неудовлетворительно	Предприятие в состоянии рассчитаться лишь по	-	0,75
Если R1<0 и R2>0 и R3<0 ТО	неудовлетворительно	Предприятие в состоянии рассчитаться лишь по	-	0,7
Если R1>0 и R2<0 и R3<0 ТО	неудовлетворительно	Предприятие в состоянии рассчитаться лишь по	-	0,6
Если Kал < 0,2	неудовлетворительно	У предприятия недостаточно средств, чтобы	-	0,5
Если Kал > 0,2 < 0,7	удовлетворительно	У предприятия достаточно средств, чтобы	+	0,3
Если Kал > 0,7 < 1,0	удовлетворительно	У предприятия достаточно средств, чтобы	+	0,5
Если Kал > 1,0	удовлетворительно	У предприятия достаточно средств, чтобы	+	0,3
Если Kл < 0,6	неудовлетворительно	Предприятие может рассчитаться по своим	-	0,6
Если Kл > 0,6 < 0,8	удовлетворительно	Прогнозируемые платежные возможности	+	0,4
Если Kл > 0,8 < 1,0	удовлетворительно	Кредитоспособность предприятия, при условии	+	0,6
Если Kл > 1,0	удовлетворительно	У предприятия достаточно средств, чтобы	+	0,4
Если Kл < 1,0	неудовлетворительно	У предприятия недостаточно средств, чтобы	-	0,7
Если Kл > 1,0 < 2,0	удовлетворительно	У предприятия достаточно средств, чтобы	+	0,5
Если Kл > 2,0 < 3,0	удовлетворительно	У предприятия достаточно средств, чтобы	+	0,7
Если Kл > 3,0	удовлетворительно	У предприятия достаточно средств, чтобы	+	0,5

Рис. 5.34. Значения показателей финансовой устойчивости
и платежеспособности

Эта информация является промежуточной и не хранится в БД, но пользователь при необходимости может просмотреть ее, выбрав соответствующий пункт меню ЭС (см. рис. 5.31-5.32). Способ организации БЗ иллюстрируют рис. 5.33-5.34.

5.11. Выводы

Построение ЭС диагностического характера представляет собой достаточно сложный процесс, требующий больших затрат компьютерных и интеллектуальных ресурсов корпорации. Практическая ценность от внедрения в практику управления ЭС состоит в существенном повышении производительности труда ЛПР, задействованных в бизнес-процессах компании, что достигается за счет следующих факторов.

Во-первых, улучшается качество решений, принимаемых менеджментом корпорации, поскольку возможность использования рассуждений на основе сценария «ЧТО – ЕСЛИ» позволяет проводить диагностику той или иной конкретной проблемы, охватывая наиболее часто встречавшиеся в прошлом (то есть наиболее вероятные в настоящее время) причины ее возникновения. Поскольку БЗ ЭС содержит независимо приобретенные аксиологические (опытные, субъективные) знания разных экспертов, анализ проблемы может быть проведен одновременно как бы с нескольких точек зрения – что приближает указанные знания к объективным верифицированным, и, в конечном итоге, позволяет получить не просто подходящее, а наилучшее в заданном смысле (квазиоптимальное на данный момент) решение.

Во-вторых, в ходе работы ЭС происходит расширенное использование – своего рода тиражирование знаний высококвалифицированных экспертов. При этом достигнутая компетентность ЛПР не утрачивается, она может документироваться в системе, передаваться, воспроизводиться и наращиваться. Применение ЭС позволяет специалистам с относительно низкой квалификации и малым производственным опытом получать качественные результаты, а компании в целом – сокращать затраты путем экономии на фонде заработной платы. Также ЭС может быть использована как обучающий тренажер, когда начинающий пользователь с ее помощью осваивает азы диагностики разного рода конкретных проблем на

алгоритмах, которые реализованы в ЭС, сверяя свои заключения с заключениями системы, и достаточно быстро достигая уровня, заложенного в ЭС ее разработчиками.

В-третьих, имеет место максимальная автоматизация рутинной работы ЛПР, что позволяет минимизировать влияние человеческого фактора (поскольку отсутствуют эмоциональные и другие факторы, определяющие «ненадежность» организма человека), улучшить оперативность и качество принимаемых решений. Освобождение персонала корпорации от монотонной и утомительной работы способствует творческой заинтересованности работников в результатах их труда.

В-четвертых, если ЭС является открытой системой, то ЛПР с достаточным уровнем квалификации и соответствующим правом доступа может самостоятельно вносить накапливаемую им аналитическую информацию в БЗ, что в идеале позволяет пользователям ЭС постоянно повышать эффективность их «совместной» работы.

В-пятых, появляется возможность совершенствовать структуру потоков информации и системы документооборота в корпорации за счет создания единого хранилища (репозитория) данных на базе ЭС и т.д.

Хотя на сегодняшний день существует достаточно много ЭС диагностического типа в самых разных предметных областях, их расширенное применение в интересах отечественного бизнеса затруднено по причинам, которые, по мнению авторов, сами заслуживают отдельного критического анализа при помощи ЭС. Тем не менее, они безусловно уверены в том, что скоро положение изменится в существенно лучшую сторону и число таких ЭС будет неуклонно возрастать.

ЗАКЛЮЧЕНИЕ

Настоящая книга завершает трехтомную монографию «Новые информационные технологии: подготовка кадров и обучение персонала», первая часть которой носит название «Реинжиниринг и управление бизнес-процессами в инфокоммуникациях» [46]; вторая часть – «Имитационное моделирование и управление бизнес-процессами в инфокоммуникациях» [44]; третья часть – «Интеллектуальные информационные системы и управление бизнес-процессами в инфокоммуникациях». В разное время в состав авторского коллектива входили Димов Э.М., Диязитдинова А.Р., Маслов О.Н., Новаковский В.Ф., Пчеляков С.Н. и Скворцов А.Б., которые на паритетных началах представляют как Высшую школу в лице кафедры Прикладной информатики (до 2017 г. – кафедра Экономических информационных систем) Поволжского государственного университета телекоммуникаций и информатики (г. Самара), так и инфокоммуникационный бизнес России в лице трех успешных выпускников данной кафедры.

С одной стороны, это отражает стремление авторов продвинуть свой труд как можно дальше в сферу реального бизнеса, что невозможно сделать без активной поддержки его заинтересованных представителей. С другой стороны, главной целью при этом (к сожалению или к счастью, сказать трудно) остается применение знаний в учебном процессе технических университетов – поскольку российский бизнес даже в такой продвинутой сфере, как инфокоммуникации, к расширенному применению новых ИТ пока не готов.

«Мы потеряли мудрость ради знаний. Мы растеряли знания в потоке информации» – эти знаменитые слова Т. Элиота, как никогда актуальны в начале нынешнего XXI века, экономику которого именуют сегодня инновационной, информационной, сетевой, цифровой, но, в конечном счете, всегда – экономикой знаний. Призывы к практической востребованности выпускников ведут вузы к принуждению передачи им постоянно обновляемой информации, которой якобы обладают представители современного бизнеса – особенно из числа окончивших зарубежные колледжи или прошедших стажировку в столичных тренинг-центрах. Однако формализация и структурирование учебного процесса, оперативная корректировка тестов и контрольных заданий, внедрение дистанционного образования и т.п. при этом лишь удаляют будущих ЛПП от подлинных

знаний: как верифицированных, так и аксиологических, не говоря уже о недоступном понятии «мудрость» – под которой современному менеджеру следует понимать творческую способность (включая внешнюю экономическую возможность и собственную внутреннюю готовность) самостоятельно и эффективно решать инновационные задачи, достигать креативные цели.

Неуважение общества и власти к людям умственного труда проявляется сегодня в том, что за решение проблем образования и науки берутся специалисты, бесконечно далекие от них, в лучшем случае подглядевшие и подслушавшие что-то за рубежом. Когда столичный менеджер влиятельной ИКК трактует ключевой для него термин «компетенция» как способность работника справляться с порученным делом, он переводит его с американского языка на русский, не подозревая, что латинское *competentia* **означает широкий круг вопросов**, в котором ЛПР обладает необходимыми теоретическими знаниями и практическим опытом. Для приобретения которых явно недостаточно дистанционного образования, тестирования и умения пользоваться услугами Internet.

По мнению авторов монографии, спасительным средством от всей этой напасти является развитие гуманистических традиций отечественного народного образования, которое в условиях хронического пренебрежения со стороны псевдоинтеллектуалов и отсутствия необходимых средств сначала всегда выдерживало, а затем преодолевало усилия либералов, направленных на его уничтожение под видом реформ. Поэтому сверхзадачей авторов в каждой части трехтомного издания была демонстрация широких возможностей, которые дает применение новых ИТ, в том числе реализованных в виде ИС и ИИС, для **комплексной поддержки действий ЛПР** – включая технологии реинжиниринга БП, имитационного моделирования процессов функционирования организационно-технических СС, разработку АСУ, СППР, CRM-Systems – направленных на повышение эффективности производственной и социально-экономической деятельности современных корпораций, включая отечественные ИКК. Стоит подчеркнуть еще раз: цель в **поддержке рациональных и творческих усилий** ЛПР, этого «человеческого фактора» в составе современных СС организационно-технического типа, а не в подмене его методами и средствами теории ММ или за счет разработок в области ИИ.

Оперативность и востребованность информации о технологических особенностях конкретных ИТ при этом как бы отходит на второй план – тем более, что за темпами ее обновления, которые опережают даже процесс издания литературы, призванной ее зафиксировать, гнаться бессмысленно. Тем более, что **принципы создания, использования и анализа эффективности** даже самых новых ИТ относительно стабильны в течение продолжительных периодов времени – поэтому знания в данной предметной области, сначала аксиологические, а затем и верифицированные, для рыночной экономики имеют непреходящую ценность. Именно на платформе этих знаний наиболее склонные к дальнейшему творчеству ЛПР призваны развивать свою креативную «мудрость» по Т. Элиоту. Именно эти знания должны сохранять и транслировать последующим поколениям ЛПР формируемые сегодня БЗ в составе ИИС и ЭС, о которых идет речь в настоящей книге.

Последние двадцать лет экономика России твердо стоит на месте. За аналогичный период после второй мировой войны наша страна не только восстановила свой потенциал, но и сделала колоссальный рывок вперед и вверх: вышла в космос, нарастила оборонную мощь, освоила целинные земли и т.д. Нынешние достижения настолько малы, что даже самые идейные демократы и либералы, похоже, засомневались: все ли у нас в порядке, поскольку списать на коммунистическое прошлое новые и старые проблемы в глазах общественности больше не удается.

На Петербургском экономическом форуме 2016 г. во время пленарной сессии «Технологии – пропуск в завтра. Изменись или умри», которую патронировал Сбербанк России, глава этой финансовой организации Г. Греф продолжил тему реформирования науки и образования, обратившись к аудитории с наивным вопросом: «Сможет ли Россия конкурировать?» Продолжил, потому что на предыдущем Гайдаровском форуме уже говорил о том, что вся модель образования, от детских садов до вузов, должна быть радикально изменена, нынешнее онлайн-образование – это «лузер», а старая советская система совершенно неприемлема, образование должно быть связано с наукой и практикой, экспорт мозгов – это самая большая потеря для России и т.п. Повторять свои резкие высказывания перед международной аудиторией банкир не решился,

а как бы попросил совета у авторитетных экспертов, один из которых – профессор Массачусетского технологического института Л. Грэхем – дал ему ответ, интересный не только для экономистов, но и для университетского сообщества в целом.

Профессор Грэхем обозначил разницу между новым научным продуктом (открытием, изобретением) и инновацией – успешным коммерческим применением этого продукта в условиях рынка. По его мнению, если опустить реверансы в сторону изумительных русских изобретателей и первооткрывателей, беда в том, что России не удастся выстроить общество, которому эти научные продукты будут жизненно нужны – создать социально-экономическую среду, которая способствует развитию и коммерциализации инноваций. В этой связи название его книги о российской науке: «Одинокие идеи», красноречиво само говорит за себя. Если затем опустить упреки в том, что у нас запрещают демонстрации, подавляют оппонентов и поддерживают авторитарные режимы, конечный вывод эксперта: «Вам нужно молоко без коровы» не понять и не принять трудно. Хотя его самарский вариант был таков: «Ведь прежде чем корову доить, ее досыта кормят – а у нас кормильцев нет, зато дояров пруд пруди» [76] – и не знает ли об этом кто-нибудь из руководящих коллег, а тем более членов сената, парламента или кабинета министров-капиталистов? Конечно, знает, не может не знать, но что толку...

Проекты типа Сколково и «Жигулевской долины», Олимпиады в Сочи и чемпионата мира по футболу, клонируют иностранные аналоги, но не продвигают развитие России в практически заметных масштабах. Поскольку для этого инновации должны быть повседневным массовым делом, а сказочные условия Сколково – на рабочих местах тех самых изумительных изобретателей и первооткрывателей, о которых говорил Л. Грэхем. Для кластерной организации наукоемкого бизнеса нужны три составляющие: вузовская наука (которая еще сохранилась с советских лет), внедренческие фирмы (о которых больше говорится, чем делается), и, что самое главное – крупные корпорации, рыночные игроки, способные принять на себя риски и другие неизбежные при неудаче издержки. Одному из авторов довелось узнать об этом еще десять лет назад, когда делегацию Торгово-промышленной палаты Самарской обла-

сти в Париже коллеги знакомили с принципами организации и результатами выполнения первых кластерных проектов во Франции.

В роли *гаранта инноваций* у нас выступает государство, то есть общество в целом – которое «нанотехнологи» и «фундаментальные исследователи» то и дело подставляют и обируют. Впрочем, проблемы возникают во всех компонентах кластерной триады: российским ученым тупо не хватает денег (объем только частных пожертвований университетам США в 2015 г. составил \$ 40,3 млрд.), кроме того, ввиду юридической беззащитности на них часто списывают издержки и злоупотребления более предприимчивых ЛПР. Внедренческие фирмы не имеют опыта, их представители некомпетентны, склонны к аферам, коррумпированы и криминализированы. По данным Internet, проект «Роснано» под названием «Унирем» специализировался «на выпуске инновационного асфальтового покрытия с добавлением частиц автомобильных покрышек» и в 2010 г. получил на эти цели 301,5 млн. рублей. Через три года созданная в рамках проекта компания «Новый каучук» была ликвидирована, а ее менеджеров осудили по статье 159 УК РФ «Мошенничество».

Причина всего этого в том, что представители власти, как правило, малосведущи в предметной области, поэтому поручают партнерам самостоятельно формулировать научно-технологические задачи и планировать ход их выполнения – хотя затем пытаются спрашивать за невыполнение графиков, а также подозревают во всем и вся – причем то с пристрастием требуют отчета за каждый потраченный рубль, то «прощают» серьезные финансовые грехи. Отметим также, что у нас в области поиски прототипа для технопарка привели губернских ЛПР в аграрную Португалию, после чего создание «Жигулевской долины» тоже пошло по самому скромному и наименее эффективному пути.

Несмотря на отмеченные издержки, жизнь не стоит на месте и вопрос о том, какими будут направления дальнейшего развития и применения новых ИТ в России, на наш взгляд, имеет сегодня не только академический, но и немалый практический интерес. Можно предположить, что одним из таких направлений будет *кибернетический подход* к управлению СС на основе конвергенции теории СУ и теории знаний. Согласно определению, кибернетика

(от греч. kibernetike – искусство управления) представляет собой науку о закономерностях процессов управления и обмена информацией в СС различного назначения. Соответственно, кибернетический подход к исследованию СС состоит в том, что всякое целенаправленное поведение ЛПР рассматривается как управление ими.

Область применения данного подхода постоянно расширяется, и кибернетика сегодня подошла к порогу, за которым не только высококвалифицированные ЛПР, но и, например, рядовые пользователи Internet в формируемой ими виртуальной среде ведут себя по-хозяйски. Попытки истолковать и смоделировать эти реалии XXI века приводят к необходимости формировать ключевые научные теории в двух вариантах: *объективном* (по принципу «как это должно быть») и *субъективном* («как это обычно бывает») у людей. Причем круг задач, решаемых методами и средствами именно *субъективных теорий*: вероятностей, риска [76], ожидаемой полезности [113] постоянно растет, выходя за границы разработок в сфере ИИ. Важное значение в этой связи имеет разработка компьютерных критериев для оценки ожидаемой полезности – в частности, функционала ожидаемой полезности [78-79], естественным образом сочетающего достоинства объективного и субъективного подходов к оценке эффективности СС.

С точки зрения теории СС, наблюдаемая тенденция закономерна, особенно если речь идет о *нерефлекторных системах* [72; 81], неотъемлемой частью которых является упомянутый «человеческий фактор» – в виде самих ЛПР, а также их партнеров, клиентов, абонентов, конкурентов, злоумышленников и т.п. В монографии показано, во-первых, что изучение нерефлекторных СС (социально-экономических, экологических, военных, специального назначения [45; 50; 76]), требует от ЛПР *введения гипотез* (версий, предположений) относительно поведения указанных подсистем и элементов при наличии воздействий и возмущений – поскольку эти СС способны максимизировать свои функционалы качества самостоятельно, что ведет к невозможности использовать принцип максимума Понтрягина и конфликтным ситуациям.

Во-вторых, что случайные воздействия и возмущения на объекты управления можно рассматривать как фактор *неопределенности знаний* ЛПР о свойствах, состоянии и поведении СС [47-48],

которые могут быть как объективными верифицированными, так и субъективными аксиологическими – причем последние представляют наибольший интерес для организации управления нерелекторными СС, поскольку ЛПР при выдвижении гипотез о поведении звеньев часто руководствуются именно ими.

В-третьих, что в зависимости от характера гипотез, важнейшими из соответствующих им задач теории СУ *являются игры*: антагонистическая игра Дж. фон Неймана [84] и игра с противоположными интересами Ю. Гермейера [25] – детальное прогнозирование хода которых компьютерными средствами СИМ по МДМ, а также методами сценариев и функционально-стоимостного анализа, приобретает первостепенное значение.

В-четвертых, что сочетание объективных и субъективных теорий приводит к целесообразности *квазиоптимального управления* СС – что на практике означает нежелание ЛПР тратить ресурсы и время, максимизируя объем, достоверность и точность получаемой информации, выбирая абсолютно лучший вариант предпринимаемых действий, если есть возможность вовремя ограничиться квазиоптимальным (лучшим из числа доступных в данный момент) решением, которое впоследствии можно будет скорректировать и улучшить. Разработка методов и средств компьютерной поддержки квазиоптимальных решений (как статических, так и динамических) имеет в настоящее время важное практическое значение [3; 47-49].

В-пятых, что традиционной формой представления иерархической СС организационно-технического типа является *многоуровневая пирамида*, где верхний (топовый) уровень соответствует элите рассматриваемой СС (властной, финансовой, интеллектуальной), а нижний уровень – ее элементарным функционерам (избирателям, потребителям, клиентам, студентам). В этой связи добавим, что *социально-экономический кризис* наносит удар одновременно по всем уровням (слоям, эшелонам) этой пирамиды, и если элита стремится переложить максимум проблем на промежуточный уровень (менеджеров среднего звена, малый бизнес, преподавателей и сотрудников вузов), дискредитируя и сокращая их количество, то образуется негативный разрыв между верхним и нижним уровнями (средний уровень «исчезает») и корпорация становится нежизнеспособной.

Это объясняется тем, что для функционеров нижнего уровня доступ к топ-уровню невозможен, а малочисленная элита высшего уровня неспособна управлять ими без посторонней помощи – в результате складывается революционная ситуация (верхи не могут руководить по старому, низы не могут жить по старому), то есть возникает проблема киевского «майдана». Разрешить эту проблему можно, изменяя уровни ЛПР и приводя в соответствие множества руководителей и руководимых ими элементов путем дробления пирамиды на подсистемы исходной СС (менеджеров верхнего уровня заменяют более многочисленные и менее требовательные менеджеры промежуточного уровня). Архитекторы данного процесса уповают на то, что со временем указанные подсистемы образуют новые СС, руководить которыми будет проще и экономически выгоднее. Согласование управленческих действий они надеются обеспечить путем замены иерархических СС «плоскими» гетерархическими структурами – по типу Евросоюза взамен «Европы наций» и раздела крупных стран по аналогии с распадом Югославии. Однако естественное стремление сохранить иерархичность СС ее собственниками и ЛПР неизбежно ведет к конфликтам между ними и владельцами СС, образующих новые надсистемы гетерархического типа, а также к обратным структурным преобразованиям СС.

За время работы над трехтомником в ПГУТИ сформировались две активно действующие научные школы: «Имитационное моделирование и управление сложными процессами в организационно-технических и социально-экономических системах» (основатель и бессменный руководитель Димов Э.М.) и «Моделирование стохастических процессов взаимодействия источников электромагнитных полей с окружающей средой» (Маслов О.Н.), в рамках которых получили развитие идеи выдающихся отечественных специалистов в области СИМ и теории управления Н.Н. Моисеева, Н.П. Бусленко, Д.И. Голенко и др. Одним из важнейших достижений в данной области представляется доказательство практической эффективности применения СИМ в интересах квазиоптимального интеллектуального управления нерелекторными СС, целесообразность которого обсуждалась одним из авторов монографии с

Н.Н. Моисеевым еще в 1979 году, а также разработка технологии СИМ по версии МДМ, специально предназначенной для этой цели.

Следует отметить, что новые научные результаты, так или иначе связанные с СИМ по МДМ, сегодня получили признание в виде двух докторских и почти уже двадцати кандидатских диссертаций, подготовленных и успешно защищенных представителями указанных научных школ в университетах Москвы, Уфы, Самары и Курска. Регулярными стали публикации преподавателей, научных сотрудников и аспирантов кафедры «Прикладная информатика» ПГУТИ в ведущих отечественных журналах по широкому спектру актуальных приложений ИТ: от разработки иерархических СС управления ИКК и нефтедобывающими компаниями на основе СИМ по МДМ, до интеллектуальной поддержки решений по обеспечению информационной, экологической и экономической безопасности корпораций с применением критериев риска и ожидаемой полезности.

...Изложенное показывает, что проблема применения новых ИТ во всех сферах жизни информационного общества XXI века, где присутствуют СС организационно-технического типа, касается целого ряда актуальных научных проблем, которые заслуживают подробного анализа и повседневного изучения. Авторы надеются, что материалы монографии помогут специалистам в продолжении исследований по данной тематике.

ЛИТЕРАТУРА

1. Абдикеев Н.М. Проектирование интеллектуальных систем в экономике. – М.: «Экзамен», 2004. – 528 с.
2. Андрейчиков А.В., Андрейчикова О.Н. Интеллектуальные информационные системы. – М.: «Финансы и статистика», 2004. – 424 с.
3. Ануфриев Д.П., Димов Э.М., Маслов О.Н., Халимов Р.Р. Сравнительная эффективность методов и средств информационной поддержки управленческих решений // Инфокоммуникационные технологии. Т.12, №1, 2014. – С. 54-67.
4. Архипенков С.Я. Лекции по управлению программными проектами. – М.: Самиздат, 2009. – 128 с.
5. Атцик А., Гольдштейн А., Сизихин К. Путь NGOSS: Дао Телекома // URL: <http://niits.ru/public/2009/2009-010.swf>. – Загл. с экрана.
6. Банковские информационные системы. Под ред. Дика В.В. – М.: «Маркет ДС», 2010. – 816 с.
7. Бирюков А.Н. Процессы управления информационными технологиями: Конспект лекций. – М.: ИНУИТ // URL: <http://www.intuit.ru/studies/courses>. (д.о. 16.10.2016)
8. Благодатских В.А., Енгибарян М.А., Ковалевская Е.В. и др. Экономика, разработка и использование программного обеспечения ЭВМ. – М.: «Финансы и статистика», 1995. – 228 с.
9. Бобровски С.М. Oracle 7 и вычисления клиент-сервер: Пер. с англ. – М.: «Лори», 1996. – 651 с.
10. Бойко В.В., Савинков В.М. Проектирование баз данных информационных систем. – М.: «Финансы и статистика», 1989. – 351 с.
11. Брукинг А., Джонс П., Кокс Ф. и др. Экспертные системы. Принципы работы и примеры. Пер. с англ. – М.: Радио и связь, 1987. – 224 с.
12. Бусленко Н.П. Моделирование сложных систем. – М.: «Наука», 1968. – 400 с.

13. Валькман Ю.Р. Интеллектуальные технологии исследовательского проектирования. Формальные системы и семиотические модели. – Киев: «Port-Royal», 1998. – 249 с.
14. Васильев Г.П., Горский В.Е., Щяуткулис В.И. и др. Программное обеспечение неоднородных распределенных систем: Анализ и реализация. – М.: «Финансы и статистика», 1986. – 160 с.
15. Васильев В.И., Ильясов Б.Г. Интеллектуальные системы управления. Теория и практика. – М.: «Радиотехника», 2009. – 392 с.
16. Векес А.Дж., Гандерлоу Майк. Access и SQL Server. Руководство разработчика: Пер. с англ. – М.: «Лори», 1997. – 362 с.
17. Вигерс К. Разработка требований к программному обеспечению. Пер. с англ. – М.: ИТД «Русская редакция», 2004. – 579 с.
18. Виттих В.А. Организация сложных систем. – Самара: Изд-во СНЦ РАН, 2010. – 64 с.
19. Волкова В.Н., Денисов А.А. Основы теории систем и системного анализа. – СПб.: Изд-во СПбГТУ, 2001. – 512 с.
20. Гаврилова Т.А. Онтологический подход к управлению знаниями при разработке корпоративных информационных систем // Новости искусственного интеллекта, №2, 2003. – С. 24-30.
21. Гаврилова Т.А., Кудрявцев Д.В., Тулугурова Е.В. Система управления знаниями // &.Strategy, №11, 2004. – С. 16-19.
22. Гаврилова Т.А., Хорошевский В.Ф. База знаний интеллектуальных систем. – СПб.: «Питер», 2000. – 384 с.
23. Гаврилова Т.А., Червинская К.Р. Извлечение и структурирование знаний для экспертных систем. – М.: «Радио и связь», 1992. – 200 с.
24. Гайсарян С.С. Объектно-ориентированные технологии проектирования прикладных программных систем. – М.: Центр информационных технологий, 1998. // URL: http://citforum.ru/programming/oop_rsis/index.shtml
25. Гермейер Ю.Б. Игры с противоположными интересами. – М.: Наука, 1976. – 327 с.
26. Гладун А.Я., Рогушина Ю.В. Онтологии в корпоративных системах, Часть I // Корпоративные системы. №1, 2006. – С. 41-47 // URL: <http://www.management.com.ua/ims/ims115.html>

27. Гладун А.Я., Рогушина Ю.В. Онтологии в корпоративных системах, Часть II // Корпоративные системы. №1, 2006. – С. 72-85 // URL: <http://www.management.com.ua/ims/ims116.html>
28. Голенко Д.И. Статистические модели в управлении производством. – М.: «Статистика», 1973. – 368 с.
29. ГОСТ 34.601-90 Автоматизированные системы. Стадии создания. – М.: Изд-во стандартов, 1990.
30. ГОСТ 34.602-89 Техническое задание на создание автоматизированной системы. – М.: Изд-во стандартов, 1989.
31. Грофф Дж.Р., Вайнберг П.Н. SQL: Полное руководство: Пер. с англ. – Киев: «ВНУ», 1998. – 608 с.
32. Гульяев А.К. MATLAB 5.3. Имитационное моделирование в среде Windows. – СПб.: «Корона Принт», 2001. – 400 с.
33. Дейт К.Дж. Введение в системы баз данных. Пер. с англ. – М.: ИД «Вильямс», 2005. – 1328 с.
34. Демина Е.Д., Иодко Е.К., Майофис Л.И. и др. Организация, планирование и управление предприятием связи. – М.: Радио и связь, 1990. – 352 с.
35. Джексон, Г. Проектирование реляционных баз данных для использования с микроЭВМ. Пер. с англ. – М.: «Мир», 1991. – 252 с.
36. Димов Э.М. Имитационное моделирование и алгоритмизация управления дискретно-непрерывным производством. Дисс. д.т.н. – Ташкент, АН УзССР, 1990. – 306 с.
37. Димов Э.М. Имитационное моделирование и оптимизация управления в сложных производственных системах. – Саратов: Изд-во СГУ, 1983. – 165 с.
38. Димов Э.М. Оптимальное управление комплексным производством дискретно-непрерывного типа на основе имитационного моделирования // Опыт применения прикладных методов математики и вычислительной техники в народном хозяйстве. – М.: Статистика, 1975. – С. 58-73.
39. Димов Э.М., Кондратович Д.А., Маслов О.Н., Новаковский В.Ф. Экспертная система для управления экономической деятельностью компании электросвязи (постановка задачи и принципы разработки) // Сб. материалов МНК и РНШ «Си-

- стемные проблемы надежности, качества, информационных и электронных технологий (ИННОВАТИКА-2004)». Москва-Сочи, 2004. Ч. 7. Т.1. – С. 61-68.
40. Димов Э.М., Маслов О.Н. О развитии математических принципов метода имитационного моделирования // Инфокоммуникационные технологии. Т.1, №2, 2003. – С. 5-11.
 41. Димов Э.М., Маслов О.Н. О точности и адекватности метода статистического имитационного моделирования // Инфокоммуникационные технологии. Т.5, №1, 2007. – С. 60-67.
 42. Димов Э.М., Маслов О.Н., Кондратович Д.А., Новаковский В.Ф. Имитационное моделирование и управление процессом ремонтных работ в телекоммуникационной компании // Телекоммуникации, №5, 2003. – С.46-48.
 43. Димов Э.М., Маслов О.Н., Новаковский В.Ф., Чаадаев В.К. Экспертные системы поддержки инновационных процессов в телекоммуникациях // Сб. материалов РНПК «Интеллектуальная подготовка инновационных процессов», Пенза, 2003. – С.27-30.
 44. Димов Э.М., Маслов О.Н., Пчеляков С.Н., Скворцов А.Б. Новые информационные технологии: подготовка кадров и обучение персонала. Ч. 2. Имитационное моделирование и управление бизнес-процессами в инфокоммуникациях. – Самара: Изд-во СНЦ РАН, 2008. – 350 с.
 45. Димов Э.М., Маслов О.Н., Раков А.С. Управление информационной безопасностью корпорации с применением критериев риска и ожидаемой полезности // Информационные технологии. Т.22, №8, 2016. – С. 620-627.
 46. Димов Э.М., Маслов О.Н., Скворцов А.Б. Новые информационные технологии: подготовка кадров и обучение персонала. Ч. 1. Реинжиниринг и управление бизнес-процессами в инфокоммуникациях. – М.: «ИРИАС», 2006. – 386 с.
 47. Димов Э.М., Маслов О.Н., Трошин Ю.В. Выбор средств программного обеспечения процесса статистического имитационного моделирования // Информационные технологии. Т.21, №2, 2015. – С. 132-139.
 48. Димов Э.М., Маслов О.Н., Трошин Ю.В. Снижение неопределенности выбора управленческих решений с помощью метода

- статистического имитационного моделирования // Информационные технологии. №6 (214), 2014. – С. 51-57.
49. Димов Э.М., Маслов О.Н., Трошин Ю.В., Халимов Р.Р. Динамика разработки имитационной модели бизнес-процесса // Инфокоммуникационные технологии. Т. 11, №1, 2013. – С. 63-78.
 50. Димов Э.М., Маслов О.Н., Швайкин С.К. Имитационное моделирование, реинжиниринг и управление в компании сотовой связи (новые информационные технологии). – М.: Радио и связь, 2001. – 256 с.
 51. Диязитдинова А.Р., Коныжева Н.В. Интегрированное информационное пространство корпорации. – Самара: Изд-во ПГУТИ, 2013. – 140 с.
 52. Диязитдинова А.Р., Матвеева Е.А., С.Г. Симагина С.Г. Информационные системы и технологии: история развития, проектирование, защита. – Самара: Изд-во «Офорт», 2006. – 193 с.
 53. Диязитдинова, А.Р. Модель представления знаний в экономической экспертной системе компании сотовой связи // Инфокоммуникационные технологии. Т.5, №2, 2003. – С. 16-22.
 54. Диязитдинова А.Р., Коныжева Н.В. Управление разработкой информационных систем. – Самара: Изд-во ПГУТИ, 2013. – 194 с.
 55. Друкер П. Задачи менеджмента в XXI веке. Пер. с англ.: – М.: ИД «Вильямс», 2004. – 272 с.
 56. Елиферов В.Г., Репин В.В. Процессный подход к управлению. Моделирование бизнес-процессов. – М.: «ИНФРА-М», 2005. – 319 с.
 57. Елиферов В.Г., Репин В.В. Бизнес-процессы: Регламентация и управление. – М.: РИА «Стандарты и качество», 2004. – 408 с.
 58. Ефимов В.В. Управление знаниями. – Ульяновск: Изд-во УлГТУ, 2005. – 111с.
 59. Информатика. Под ред. Н.В. Макаровой. – М.: Финансы и статистика, 1997. – 768 с.
 60. Искусственный интеллект. Кн. 1: Системы общения и экспертные системы. Справочник. Под ред. Э.В. Попова. – М.: Радио и связь, 1990. – 461 с.

61. Искусственный интеллект. Кн. 2. Модели и методы. Справочник. Под ред. Д.А. Поспелова. – М.: Радио и связь, 1990. – 304 с.
62. Искусственный интеллект. Кн. 3: Программные и аппаратные средства. Справочник. Под ред. В.Н. Захарова, В.Ф. Хорошевского. – М.: Радио и связь, 1990. – 320 с.
63. Калиниченко Л.А. Методы и средства интеграции неоднородных баз данных. – М.: «Наука», 1983. – 423 с.
64. Калянов, Г.Н. CASE-технологии. Консалтинг в автоматизации бизнес-процессов. – М.: Горячая линия – Телеком, 2002. – 320 с.
65. Карелин В.П. Интеллектуальные технологии и системы искусственного интеллекта для поддержки принятия решений // Вестник Таганрогского института управления и экономики. №2, 2011. – 79-84.
66. Красов А.В. Теория информационных процессов и систем // Краткий курс лекций. – СПб.: СПбГУТ, – 57 с.
67. Кузнецов О.П., Суховеров В.С., Шипилина Л.Б. Онтология как систематизация научных знаний: структура, семантика, задачи // Труды НТК «Технические и программные средства систем управления, контроля и измерения», Москва, 2010. – С. 762.
68. Ларичев О.И. Теория и методы принятия решений, а также Хроника событий в Волшебных странах. – М.: Логос, 2003. – 392 с.
69. Левин Р., Дранг Д. Практическое введение в технологию искусственного интеллекта и экспертных систем с иллюстрацией на Бейсике. – М.: Финансы и статистика, 1991. – 239 с.
70. Лорин Г. Распределенные вычислительные системы: Пер. с англ. – М.: «Радио и связь», 1984. – 296 с.
71. Лорьер Ж.-Л. Системы искусственного интеллекта. Пер. с франц. – М.: Мир, 1991. – 568 с.
72. Лотов А.В., Моисеев Н.Н., Петров А.А. Некоторые вопросы моделирования программного метода управления социально-экономической системой // Модели и алгоритмы программного метода планирования сложных систем. М.: Изд. ВЦ АН СССР, 1079. – С. 4-10.

73. Марселлус Д. Программирование экспертных систем на Турбо Прологе. Пер. с англ. – М.: Финансы и статистика, 1994. – 256 с.
74. Мартин Дж. Вычислительные сети и распределенная обработка данных: программное обеспечение, методы и архитектура: Пер. с англ. – М.: «Финансы и статистика», 2986. – 269 с.
75. Мартин Дж. Организация баз данных в вычислительных системах. – М.: «Мир». 1980. – 662 с.
76. Маслов О.Н. Безопасность корпорации: моделирование и прогнозирование внутренних угроз методом риска. – Самара: Изд-во ПГУТИ-АЭРОПРИНТ, 2013. – 170 с.
77. Маслов О.Н. Применение метода статистического имитационного моделирования для исследования случайных антенн и проектирования систем активной защиты информации // Успехи современной радиоэлектроники. №6, 2011. – С. 42-55.
78. Маслов О.Н., Фролова М.А. Функционал ожидаемой полезности в задачах управления сложными системами организационно-технического типа // Инфокоммуникационные технологии. Т.14, №2, 2016. – С. 168-178.
79. Маслов О.Н., Фролова М.А. Функционал ожидаемой полезности: принципы моделирования и практического применения // Инфокоммуникационные технологии. Т.13, №3, 2015. – С. 291-297.
80. Матвеев Л.А. Компьютерная поддержка решений. – СПб: «Специальная литература», 1998. – 472 с.
81. Моисеев Н.Н. Элементы теории оптимальных систем. – М.: Наука, 1975. – 528 с.
82. Назаров С.В., Перитков В.И., Тарищев В.А Компьютерные технологии обработки информации. – М.: «Финансы и статистика», 1995. – 248 с.
83. Нейлор, К. Как построить свою экспертную систему. Пер. с англ. – М.: Энергоатомиздат, 1991. – 286 с.
84. Нейман Дж. фон, Моргенштерн О. Теория игр и экономическое поведение. Пер. с англ. – М.: «Наука», 1970. – 708 с.

85. Новаковский В.Ф. Новые информационные технологии в управлении компанией – альтернативным оператором электросвязи. – М.: Радио и связь, 2004. – 225 с.
86. Одинцов Б.Е. Проектирование экономических экспертных систем. – М.: ЮНИТИ, 1996. – 168 с.
87. Осуга С. Обработка знаний. Пер. с англ. – М.: Мир, 1989. – 293 с.
88. Патент RU 2 469 370 С1. Способ реанимирования нефтяных скважин / Димов Э.М., Маслов О.Н., Халимов Р.Р. и др. Заявл. 16.05.2011, опубл. 10.12.2012, бюлл. №34.
89. Питерс Т., Уотерман Р. В поисках совершенства: Уроки самых успешных компаний Америки. Пер. с англ. – М.: «Альпина Паблишер», 2014. – 524 с.
90. Питер Дж. Введение в экспертные системы. Пер. с англ. – М.: ИД «Вильямс», 2001. – 624 с.
91. Попов Э.В., Фоминых И.Б., Кисель Е.Б. и др. Статические и динамические экспертные системы. – М.: «Финансы и статистика», 1996. – 320с.
92. Построение экспертных систем. Под ред. Ф. Хейес-Рота, Д. Уотермена, Д. Лената. Пер. с англ. – М.: Мир, 1987. – 441 с.
93. Приобретение знаний. Под ред. С. Осуги, Ю. Сазки. Пер. с англ. – М.: Мир, 1990. – 304 с.
94. Райли Дж., Кринер М. NGOSS: Построение эффективных систем поддержки и эксплуатации сетей для оператора связи. Пер. с англ. – М.: «Альпина Бизнес Букс», 2007. – 192 с.
95. РД 50-34.698-90 Автоматизированные системы. Требования к содержанию документов. – М.: Изд-во стандартов, 1990.
96. Ройс У Управление проектами по созданию программного обеспечения. Пер. с англ. – М.: «Лори», 2002. – 424 с.
97. Романов А.Н., Одинцов Б.Е. Советующие информационные системы в экономике. – М.: «ЮНИТИ-ДАНА», 2000. – 487 с.
98. Романов В.П. Интеллектуальные информационные системы в экономике. – М.: Изд-во «Экзамен», 2003. – 496 с.
99. Руководство к своду знаний по управлению проектами. 4-е издание (руководство РВВОК) Американский национальный стандарт.– РМІ, 2008. – 241 с.

100. Рутковская Д., Пилиньский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткие системы. Пер. с польск. – М.: Горячая линия – Телеком, 2004. – 452 с.
101. Ручкин В.Н., Фулин В.А. Универсальный искусственный интеллект и экспертные системы. – СПб.: «БХВ-Петербург», 2009. – 240 с.
102. Рыбина Г.В. Основы построения интеллектуальных систем. – М.: «Финансы и статистика»; «ИНФРА-М», 2010. – 432 с.
103. Самуйлов К.Е., Чукарин А.В., Яркина Н.В. Бизнес-процессы и информационные технологии в управлении телекоммуникационными компаниями. – М.: «Альпина Паблишер», 2009. – 442 с.
104. Таусенд К., Фохт Д. Проектирование и программная реализация экспертных систем на персональных ЭВМ. – М.: Финансы и статистика, 1990. – 320 с.
105. Телков А.Ю. Экспертные системы. – Воронеж: ИПЦ Воронежского ГУ, 2007. – 83с.
106. Тельнов Ю.Ф. Интеллектуальные информационные системы в экономике. – М.: «СИНТЕГ», 1998, – 216 с.
107. Уотермен Д. Руководство по экспертным системам. Пер. с англ. – М.: Мир, 1989. – 388 с.
108. Уткин В.Б., Балдин К.В. Информационные технологии управления. – М.: «Академия ИЦ», 2008. – 395 с.
109. Форрестер Дж. Основы кибернетики предприятия (индустриальная динамика). Пер. с англ. – М.: «Прогресс», 1971. – 340 с.
110. Чаадаев В.К. Бизнес-процессы в компаниях связи. – М.: «Эко-Трендз», 2004. – 176 с.
111. Частиков А.П., Гаврилова Т.А., Белов Д.Л. Разработка экспертных систем. Среда CLIPS. – СПб.: «БХВ-Петербург», 2003. – 608 с.
112. Шелдрейк Дж. Теория менеджмента: от тейлоризма до японизации. Пер. с англ. – СПб: Питер, 2001. – 352 с.
113. Шумейкер П. Модель ожидаемой полезности: разновидности, подходы, результаты и пределы возможностей. Пер.с англ. // THESIS. Вып. 5, 1994. – С. 29-80.

114. Ярушкина Н.Г. Основы теории нечетких и гибридных систем. – М.: «Финансы и статистика», 2004. – 320 с.
115. Chen P.P. The Entity-Relationship Model: A Basis for the Enterprise View of Data // ACM Transaction on Database Systems. Vol. 1, No. 1, 1976. – P. 9 – 36.
116. Ouchi W. Theory Z: How American business can meet the Japanese challenge // Business Horizons. Vol. 24, Issue 6, 1981. – P. 82-83.
117. Savage L.J. The Foundations of Statistics. – N.Y.: Wiley, 1954. – 310 p.
118. SQL–Windows/SAL – язык приложений баз данных с архитектурой клиент/сервер. – М.: «ДИАЛОГ-МИФИ», 1996. – 286 с.

Димов Э.М., Диязитдинова А.Р.,
Маслов О.Н., Новаковский В.Ф.

НОВЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ: ПОДГОТОВКА КАДРОВ И ОБУЧЕНИЕ ПЕРСОНАЛА

Часть 3. Интеллектуальные информационные системы и управление бизнес-процессами в инфокоммуникациях

Подписано в печать 07.04.2017г.
Формат 60x84 1/16. Бумага офсетная. Печать оперативная.
Усл. печ. л. 26,4
Тираж 300 экз. Заказ № 121.

Отпечатано в типографии АНО «Издательство СамНЦ РАН»
443001, Самарская область, г. Самара, Студенческий пер., 3А.
тел.: (846) 242-37-07, ИНН 6315944726.