

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
«Пензенский государственный университет
архитектуры и строительства»

Ю.В. Блинков

ОСНОВЫ ТЕОРИИ ИНФОРМАЦИОННЫХ ПРОЦЕССОВ И СИСТЕМ

Допущено Учебно-методическим объединением вузов
по университетскому политехническому образованию
в качестве учебного пособия
для студентов высших учебных заведений,
обучающихся по специальности 230201
«Информационные системы и технологии»

Пенза 2011

УДК 004 (075.8)
ББК 32.81 я73
Б69

Рецензенты: кафедра «Информатика» Пензенской государственной технологической академии (зав. кафедрой кандидат технических наук, доцент В.В. Пикулин);
кандидат технических наук, доцент
Л.Ю. Акулова (ПГТА)

Блинков Ю.В.

Б69 Основы теории информационных процессов и систем: учеб. пособие / Ю.В. Блинков. – Пенза: ПГУАС, 2011. – 184 с.
ISBN 978-5-9282-0725-0

Изложены основы общей теории систем. Большое внимание уделено классификации, решаемым задачам, принципам построения и функционирования информационных систем. Рассмотрены способы представления, преобразования, обработки и циркуляции непрерывной и цифровой информации в системах, особенности моделирования сигналов, используемых в системах для передачи информации, методы определения степени неопределенности и количества информации, а также методы кодирования информации в отсутствии и при наличии помех.

Представлены методики декомпозиции и агрегирования систем, используемых при их анализе и синтезе, а также особенности моделирования процессов функционирования агрегатов.

Значительное внимание уделено методикам проведения экспериментов при анализе систем.

Учебное пособие подготовлено на кафедре «Информационные системы и компьютерное моделирование» ПГУАС и предназначено для студентов специальности 230201 «Информационные системы и технологии» всех форм обучения.

ISBN 978-5-9282-0725-0

© Пензенский государственный университет
архитектуры и строительства, 2011
© Блинков Ю.В., 2011

ПРЕДИСЛОВИЕ

Учебное пособие, являющееся частью учебно-методического комплекса дисциплины «Теория информационных процессов и систем», содержит введение и пять разделов, посвященных основам теории информационных процессов и систем, примеры, поясняющие содержание материала, контрольные вопросы по изучаемым темам, а также приложение.

В первом разделе рассматриваются основы теории систем.

Второй раздел посвящен изучению основ построения и использования информационных систем.

В третьем разделе излагаются основы теории информации, а также ее преобразования и обработки в информационных системах.

Четвертый раздел содержит темы, посвященные изучению основных методов разработки и исследования информационных систем.

В пятом разделе приведены основы экспериментальных исследований систем, позволяющих создавать и уточнять их модели.

Приложение содержит краткое описание структуры одной из типовых информационных систем, а также некоторых ее фрагментов.

Следует отметить, что методы и средства дисциплины «Теория информационных процессов и систем» продолжают развиваться по мере роста сложности систем самого различного назначения, прежде всего информационных систем. В пособии предложены основы для дальнейшего самостоятельного повышения квалификации в данном направлении.

К сожалению, учебный план не позволяет более подробно рассмотреть в данном учебном пособии информационные системы. Планируя написание учебного пособия по такой обширной тематике, как современная теория информационных процессов и систем, приходится искать компромисс между многообразием сторон затрагиваемых вопросов, полнотой их описания и требований учебного плана. Поэтому данное учебное пособие далеко не исчерпывает всех задач, возникающих при проектировании и эксплуатации АИС. Например, в учебном пособии не затронуты вопросы, связанные с количественным подходом к решению многокритериальных задач оптимального управления (Подиновский В.В., Ногин В.Д. Парето-оптимальные решения многокритериальных задач. М.: Физматлит, 2007) методами расчёта надёжности систем, включая оценку надёжности программного обеспечения АИС (Черкесов Г.Н. Надёжность аппаратно-программных комплексов: учебное пособие. СПб.: Питер, 2005), что является неотъемлемой частью современной теории информационных процессов и систем.

ВВЕДЕНИЕ

Человек – активная часть природы. Добиваясь своих целей, человек использует природу, воздействует на нее, преобразует ее и себя. Без преувеличения можно сказать, что самыми важными и самыми интересными для человечества являются вопросы о возможностях человека в его отношениях с природой, о способах реализации этих возможностей, о факторах, способствующих и препятствующих расширению этих возможностей. Даже основной вопрос философии – о соотношении материи и сознания – принадлежит к этому кругу вопросов.

Отметим самые очевидные и обязательные из признаков системности: структурированность системы, взаимосвязанность составляющих ее частей, подчиненность организации всей системы определенной цели. По отношению к человеческой деятельности эти признаки и в самом деле очевидны, поскольку каждый из нас легко обнаружит их в своем собственном практическом опыте. Всякое наше осознанное действие преследует определенную цель. Во всяком действии легко увидеть его составные части, более мелкие действия. При этом легко убедиться, что эти составные части должны выполняться не в произвольном порядке, а в определенной последовательности. Это и есть та самая подчиненная определенной цели взаимосвязанность составных частей, которая и является признаком системности.

Всякая деятельность тем успешнее, чем выше уровень ее системности; неудачи вызваны недостаточной системностью. Выделяют три уровня системности труда: механизация, автоматизация, кибернетизация. Возможности механизации ограничены участием человека. Автоматизировать можно только алгоритмизируемые процессы. Кибернетизация включает в технологический процесс интеллект – естественный или искусственный – в тех случаях, когда алгоритмизация решаемой задачи затруднительна или невозможна.

Основная идея разрешения проблем, связанных со сложными системами, состоит в том, чтобы в тех случаях, когда автоматизация (т.е. формальная алгоритмизация) невозможна, использовать человеческую способность, которая именно в таких случаях проявляется и которая называется интеллектом: способность ориентироваться в незнакомых условиях и находить решение слабоформализованных задач. При этом человек выполняет именно те операции в общем алгоритме, которые не поддаются формализации (например, экспертная оценка или сравнение многомерных и неколичественных вариантов, принятие управленческих решений, взятие на себя ответственности). Именно на этом принципе строятся автоматизированные (в отличие от автоматичес-

ких) системы управления, в которых формализованные операции выполняют автоматы и ЭВМ, а неформализованные (и, возможно, неформализуемые) операции – человек.

Однако на этом возможности кибернетизации не кончаются, а, скорее, лишь начинаются. Вполне логично возникает вопрос: нельзя ли смоделировать интеллектуальные возможности человека – хотя бы в той части, которая необходима для выполнения конкретных, пусть частных, интеллектуальных операций? Здесь опыт науки и техники подсказывает два пути: «подглядеть» у природы алгоритмы интеллектуальной деятельности (т.е. изучить естественный интеллект) либо «изобрести» эвристически алгоритм предположительно с интеллектуальными свойствами и исследовать, что это даст (т.е. конструировать интеллект искусственно). Техническими средствами для этого служат автоматы и ЭВМ. Их совершенствование и использование для самых разнообразных нужд, создание алгоритмов и программных систем для ЭВМ привели к появлению науки информатики. Существенной частью информатики являются разделы, связанные с изучением, проектированием и использованием информационных процессов и систем, применение которых позволяет значительно повысить эффективность любой человеческой деятельности.

1. ОСНОВЫ ТЕОРИИ СИСТЕМ

Системные исследования получили в настоящее время широкое распространение в целом ряде наук. Важную роль они играют в технике, где сформировалось особое научное направление – системотехника.

Основные проблемы общей теории систем и системотехники во многом совпадают, хотя и имеют свою специфику, связанную с особенностями создания сложных технических комплексов.

Значительное место занимают вопросы оптимизации процессов функционирования и оценки эффективности систем.

Другой широкий круг вопросов посвящен проблемам проектирования. С вопросами проектирования, оптимизации, оценки эффективности тесно связаны вопросы моделирования сложных технических комплексов, особенно имитационного моделирования.

Общий анализ развития системотехники показывает, что помимо отмеченных проблем на первый план выдвигаются следующие:

- 1) выявление и описание наиболее общих системных характеристик и закономерностей, не зависящих от конкретного типа технических комплексов;

- 2) разработка экспериментальных методов, позволяющих с достаточной точностью при приемлемом объеме затрачиваемых ресурсов оценивать теоретические концепции;

- 3) разработка методов реализации принципов системотехники при создании и использовании конкретных систем.

Анализ опубликованных материалов показывает, что объектом системотехники являются сложные технические комплексы, называемые авторами по-разному: большими системами, сложными системами, интегральными системами, системотехническими комплексами.

Таким образом, предмет системотехники включает процессы проектирования, испытания, изготовления, установки, эксплуатации сложных технических комплексов.

Системотехника в силу своей ориентированности на наиболее общие системные характеристики носит междисциплинарный характер. Основой является общая теория систем.

Таким образом, системотехника – это научное направление, изучающее системные свойства системотехнических комплексов, процессы их создания, совершенствования, использования и ликвидации с целью получения максимального социального эффекта.

Основным методом системотехники является системный подход с его конкретными видами реализации: системным анализом, исследованием операций, системным синтезом, кибернетикой.

1.1. Системы и их основные свойства

Фундаментальным понятием системотехники является понятие «система» (гр. *systema* – это составленное из частей, соединение). В самом широком смысле под системой понимают множество закономерно взаимосвязанных элементов.

Среди систем выделяют простые и сложные.

Под *сложной системой* понимают совокупность объектов (простых и сложных элементов – компонент), взаимодействие которых обуславливает появление новых качеств, не свойственных объектам, входящим в систему. Этот принцип появления у системы новых свойств, не выводимых непосредственно из наблюдаемых свойств её частей и связей между ними, назван принципом эмерджентности – неожиданного появления новых свойств.

В отличие от простых систем в сложных системах *введение новых связей между ее элементами приводит к появлению новых свойств*. Четкую границу между простыми и сложными системами определить трудно. В них может быть множество элементов и множество связей.

Иерархичность системы подразумевает, что каждая система может быть разделена на подсистемы, которые, в свою очередь, могут быть разделены на более мелкие подсистемы следующего уровня иерархии, и т.д.

В принципе это деление можно производить до бесконечности. Сама система может входить в более крупную подсистему. Самый нижний уровень иерархии называют элементом, не подлежащим дальнейшему делению. Он определяется условиями конкретной задачи исследования.

Каждая сложная система обладает свойствами *целостности и членности*.

Сложная система обычно имеет многоуровневое управление и характеризуется большими потоками информации.

Сложная система обычно *функционирует в условиях воздействия большого числа случайных внешних и внутренних факторов*. Это значит, что сложную систему нельзя исследовать аналитически, так как в данном случае эти факторы являются доминирующими. Следовательно, сложную систему можно исследовать с помощью различных статистических методов, а самый лучший способ – создание их имитационных моделей в виде имитационных программ ЭВМ, с помощью которых можно многократно имитировать процесс функционирования системы в целом и все воздействия на неё, прежде всего случайные, на числах и статистически обработать полученные результаты. В связи с этим в теории сложных систем для создания их полноценных имитационных

моделей используется принцип абстрагирования от реальных физических свойств и процессов системы и её подсистем. При этом отдельные структурные блоки имитационных моделей рассматриваются как математические преобразователи информации.

Обязательными компонентами системы всегда являются элементы и связи между ними, а для сложных систем решающее значение имеют именно характер, способы, свойства связей между элементами системы, причем сами свойства связей между элементами, в свою очередь, существенно зависят от свойств элементов.

Любой элемент системы может быть расчленен на большое количество составляющих его элементов и т.д., вплоть до элементарных частей. Поэтому условимся называть элементом системы такую ее часть, которая является неделимой с точки зрения решаемой задачи.

Как видим, процесс расчленения системы на элементы и само понятие элемента условно. Тем не менее можно, исходя из практической целесообразности, достаточно удобно и четко выделять элементы так, что они будут обладать определенной типичной внутренней структурой и представлять собой образование, характеризующееся более высокой устойчивостью, чем система в целом. Это удобно для унификации таких элементов.

Элементы любых реальных систем являются некими физическими объектами, которые можно характеризовать различными параметрами. Однако с точки зрения их поведения в системе в большинстве случаев можно отвлечься от всех этих свойств и характеризовать их только возможностями образовывать различные виды связей – вещественные, энергетические и информационные – с другими элементами и с внешней средой.

Все три вида связей существуют практически всегда неотделимо друг от друга, но один из них является определяющим – по этому признаку относят связь к конкретному виду.

Основной характеристикой элемента в системе служит его способность к установлению связей, т.е. к порождению или к восприятию множества связей определенного вида. Общее количество связей, которое способен образовать элемент, называется его валентностью.

Вся совокупность возможных входов и выходов элемента представляет его контакты, т.е. валентность элемента равна числу его контактов, а установление связей – соединение соответствующих контактов.

Контакты и связи можно характеризовать их мощностью и направленностью. Под мощностью понимают пропускную способность контактов и связей в единицу времени.

Выходы элемента, формирующие исходную связь, являются его активными контактами, поглощающие связи – пассивными контактами; контакты, которые как генерируют, так и поглощают связи, называются нейтральными.

Все это позволяет при исследовании и синтезе систем пользоваться идеализированными моделями элементов и систем.

Идеализированный элемент – это некоторый абстрактный элемент, у которого отсутствуют любые физические свойства, кроме способности к регенерации связей с другими подобными идеализированными элементами, т.е. идеализированный элемент полностью определяется природой и направленностью связей.

Совокупность идеализированных элементов, объединенных необходимыми связями, образует идеализированную модель системы, которую удобно и наглядно представить в виде графа или матрицы. Такая формализация структуры системы делает более наглядным представление ее для исследователя и существенно упрощает ее анализ. Причем любой элемент системы всегда может, в свою очередь, рассматриваться как система, состоящая из ряда других элементов.

Любая система имеет, как правило, иерархическую структуру. При этом системы низшего уровня являются подсистемами систем более высокого уровня, которые, в свою очередь, представляют собой подсистемы систем еще более высокого уровня, и т.д.

Следует выделить следующие основные свойства систем:

1. Уникальность, целостность и членимость, разнородность подсистем и элементов. Система – это целостная совокупность элементов, т.е., с одной стороны, это целостная система, а с другой – в ее составе отчетливо иерархически выделяются целостные объекты (элементы), существующие только в системе. Вне системы это в лучшем случае объекты, обладающие «системозначимыми» свойствами.

Т.е. система рассматривается как единое целое, состоящее из взаимодействующих частей, часто разнокачественных, но одновременно совместимых.

2. Системы имеют иерархическую структуру. Отдельные части, называемые подсистемами, могут быть разделены на более мелкие части – более простые подсистемы, которые также могут быть разделены, и т.д., пока не получим элементы, под которыми понимают объекты, которые в условиях конкретной задачи не подлежат расчленению.

3. Свойства системы определяются не только свойствами отдельных ее элементов, но и характером связей и взаимодействия между ними. (Изучая только отдельные элементы, нельзя изучить и познать систему в целом.)

4. Сложная система имеет довольно сложную функцию, направленную на достижение заданной цели.

5. Сложная система имеет управление (часто с иерархической многоуровневой структурой), разветвленную информационную сеть и потоки информации.

6. Сложные системы взаимодействуют с внешней средой и функционируют в условиях воздействия множества случайных факторов различной природы.

1.1.1. Модель «черный ящик»

В определении системы, приведенном в подразд. 1.1, сделан акцент на назначении системы, а об ее устройстве говорится лишь косвенно. Для более определенной и точной характеристики конструкции системы следует развивать ее модель, преобразуя имеющиеся сведения так, чтобы в результате получить более удобную форму модели, включая в модель по мере необходимости дополнительные сведения.

Важную для человека роль играют наглядные, образные, визуальные модели. Перейдем к визуальному эквиваленту определения системы.

- Приведенное определение ничего не говорит о внутреннем устройстве системы. Поэтому ее можно изобразить в виде непрозрачного «ящика», выделенного из окружающей среды. Подчеркнем, что уже эта максимально простая модель по-своему отражает два следующих важных свойства системы: целостность и обособленность от среды.

- В определении системы косвенно говорится о том, что хотя «ящик» и обособлен, выделен из среды, но не является полностью от нее изолированным.

В самом деле, система связана со средой и с помощью этих связей воздействует на среду. Изобразим связи в виде стрелок, направленных от системы в среду. Эти связи называются выходами системы. Подчеркнем еще раз, что выходы системы в данной графической модели соответствуют слову «цель» в словесной модели (первом определении) системы.

Кроме того, в определении имеется указание и на наличие связей другого типа: система является средством; поэтому должны существовать и возможности ее использования, воздействия на нее, т.е. и такие связи со средой, которые направлены извне в систему. Изобразим эти связи также в виде соответствующих стрелок, направленных от среды в систему, и назовем их входами системы.

В результате мы построили модель системы, которая получила название «черного ящика» (рис. 1.1). Это название образно подчерки-

вадет полное отсутствие сведений о внутреннем содержании «ящика»: в этой модели задаются, фиксируются, перечисляются только входные и выходные связи системы со средой (даже «стенки ящика», т.е. границы между системой и средой, в этой модели обычно не описываются, а лишь подразумеваются, признаются существующими). Такая модель, несмотря на внешнюю простоту и на отсутствие сведений о внутренности системы, часто оказывается полезной.

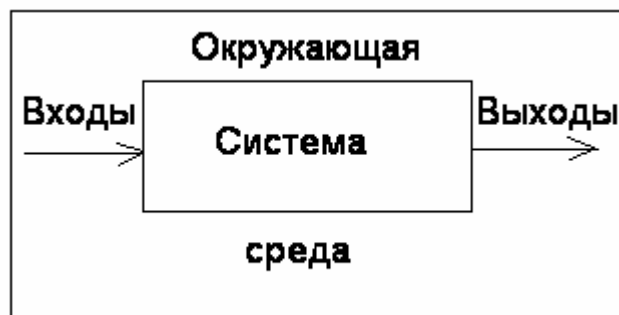


Рис. 1.1

Во многих случаях достаточно содержательного словесного описания входов и выходов; тогда модель «черного ящика» является просто их списком, или спецификацией. Например, бытовая модель телевизора такова: входы – шнур электропитания, антенна, ручки управления и настройки; выходы – экран кинескопа и динамики. В других случаях требуется количественное описание некоторых или всех входов и выходов. Пытаясь максимально формализовать модель «черного ящика», мы приходим к заданию двух множеств (X и Y) входных и выходных переменных, но никаких других отношений между этими множествами фиксировать нельзя (иначе это уже будет не «черный», а прозрачный ящик).

1.1.2. Сложности построения модели «черный ящик»

Теперь рассмотрим принципиально важный вопрос об обманчивой простоте модели «черного ящика». Казалось бы, так просто: перечислить входы и выходы системы – и модель готова. Но как только это потребуется сделать для конкретной реальной системы, мы столкнёмся с трудностями. Проиллюстрируем это на примерах.

Пример 1. Опишем выходы системы «наручные часы». Учитывая, что выходы соответствуют конкретизации цели, фиксируем в качестве выхода показание времени в произвольный момент. Затем принимаем во внимание, что сформулированная таким образом цель относится ко всем часам, а не только к нашим наручным часам. Для их различения

вносим добавление: удобство ношения часов на запястье; тогда появляется обязательность ремешка или браслета, а с ними еще один выход: удовлетворение требований санитарии и гигиены, т.к. не любое крепление часов на руке допустимо с этой точки зрения.

Далее, представив себе условия эксплуатации часов, можно добавить достаточную в бытовых условиях прочность; пылевлагонепроницаемость.

Затем, расширив понятие «условия эксплуатации часов», добавим еще два выхода: достаточную для бытовых нужд точность; легкость прочтения показаний часов при беглом взгляде на циферблат. Можно еще более расширить круг учитываемых требований к часам, что позволит добавить несколько выходов: соответствие моде и понятию красоты; соответствие цены часов покупательной способности потребителя. Очевидно, что список желаемых, т.е. включаемых в модель, выходов можно продолжать. Например, можно потребовать, чтобы имелась возможность прочтения показаний часов в полной темноте, и реализация этого выхода приведет к существенному изменению конструкции часов, в которой могут быть различные варианты самосвечения, подсветки, считывания на ощупь или подачи звуковых сигналов. А ведь мы в явной форме еще не говорили о габаритах, весе, многих других физических, химических, экономических и социальных аспектах использования наручных часов.

Пример 2. Попробуем перечислить входы системы «легковой автомобиль». Исходя из определения системы как средства достижения цели, мы связали понятие входа с управляющим воздействием на систему, воздействием, «подталкивающим» систему к цели. Поэтому сразу же выделим в автомобиле в качестве входов те его элементы, которые предназначены для управления во время движения: руль, педали сцепления, газа и тормоза, рычаг переключения коробки передач, переключатели сигнализации и освещения, ручка аварийного и стояночного тормоза.

Затем, учитывая, что регулирующие воздействия приходится осуществлять не только на ходу, в список входов автомобиля вносим регулировочные винты, гайки, эксцентрики.

Смазка и заправка – это также регулирующее и управляющее воздействия. Поэтому точки смазки и заправочные отверстия являются входами.

Нельзя не учитывать входы в буквальном смысле. Поэтому добавляем двери салона и (заодно) крышки багажника и капота.

И тут мы начинаем понимать, что входное воздействие на автомобиль оказывает не только водитель, но и пассажиры, а также окру-

жающая среда. Записываем в перечень входов окна и зеркала, с помощью которых поступает информация к водителю и пассажирам. Но тогда можно отметить, что свойства поверхности, по которой движется автомобиль, также оказывают входное воздействие: по-разному приходится действовать водителю при езде по асфальту, песку, гравии, в случае гололеда, грязи и т.д. Добавляем к списку входов механическое воздействие грунта на колеса.

Однако различие между песком и асфальтом для автомобиля существенно лишь потому, что существует поле тяготения Земли. Вместе с тем мы еще не упомянули другие способы воздействия среды на данную систему: ручки стеклоподъемников, аэродинамическое сопротивление воздуха, кнопки радиоприемника или кондиционера, а в последних моделях – входы вычислительных устройств. А разве не влияют на автомобиль и его пассажиров электрические и магнитные поля? Не зря же рекомендуют прикреплять к автомобилю проводящий ремень, который отводит накапливающиеся на кузове электрические заряды. Стали обязательными пристежные ремни, так как нельзя пренебрегать тем, что существует еще один вход – силы инерции, которые при авариях достигают опасных для здоровья и жизни величин. Очевидно, что список входов может быть еще продолжен.

Рассмотренные примеры свидетельствуют о том, что построение модели «черного ящика» не является тривиальной задачей, так как на вопрос о том, сколько и какие именно входы и выходы следует включать в модель, ответ не прост и не всегда однозначен. Установим причины этого факта.

1.1.3. Множественность входов и выходов

Главной причиной множественности входов и выходов в модели «черного ящика» является то, что всякая реальная система, как и любой объект, взаимодействует с объектами окружающей среды неограниченным числом способов. Строя модель системы, мы из этого бесчисленного множества связей отбираем конечное их число для включения в список входов и выходов. Критерием отбора при этом является целевое назначение модели, существенность той или иной связи по отношению к этой цели. То, что существенно, важно, включается в модель, то, что несущественно, неважно, – не включается. Именно здесь возможны ошибки. Тот факт, что мы не учитываем в модели, исключаем из рассмотрения остальные связи, не лишает их реальности, они все равно действуют независимо от нас. И нередко оказывается, что казавшееся несущественным или неизвестным для нас на самом деле является важным и должно быть учтено.

Особое значение этот момент имеет при задании цели системы, т.е. при определении ее выходов. Это относится и к описанию существующей системы по результатам ее обследования, и к проекту пока еще не существующей системы. Реальная система неизбежно вступает во взаимодействия со всеми объектами окружающей среды; поэтому необходимо как можно раньше, лучше всего еще на стадии построения (проектирования) модели, учесть все наиболее важное. В результате главную цель приходится сопровождать заданием дополнительных целей. К примеру 2 можно добавить пример пассажирского самолета; нужно не только чтобы он летал, но и чтобы при этом обеспечивались необходимый комфорт и безопасность пассажиров, не создавался сильный шум при полете над населенными пунктами, не требовались слишком длинные взлетно-посадочные полосы, соблюдались экономические выгоды в эксплуатации и многое другое. Важно подчеркнуть, что выполнения только основной цели недостаточно, что невыполнение дополнительных целей может сделать ненужным или даже вредным и опасным достижение основной цели. Этот момент заслуживает особого внимания, так как на практике часто обнаруживается незнание, непонимание или недооценка важности указанного положения. Между тем оно является одним из центральных во всей теории систем.

Модель «черного ящика» часто оказывается не только очень полезной, но в ряде случаев единственно применимой при изучении систем. Например, при исследовании психики человека или влияния лекарства на живой организм мы лишены возможности вмешательства в систему иначе, как только через ее входы, и выводы делаем только на основании наблюдения за ее выходами. Это вообще относится к таким исследованиям, в результате проведения которых нужно получить данные о системе в обычной для нее обстановке, где следует специально заботиться о том, чтобы измерения как можно меньше влияли на саму систему. Другая причина того, что приходится ограничиваться только моделью «черного ящика», – действительное отсутствие данных о внутреннем устройстве системы. Например, мы не знаем, как «устроен» электрон, но знаем, как он взаимодействует с электрическими и магнитными полями. Это и есть описание электрона на уровне модели «черного ящика».

1.1.4. Модель состава системы

Очевидно, что вопросы, касающиеся внутреннего устройства системы, невозможно решить только с помощью модели «черного ящика». Для этого необходимы более развитые, более детальные модели.

При рассмотрении любой системы прежде всего обнаруживается, что ее целостность и обособленность (отображенные в модели черного ящика) выступают как внешние свойства. Внутренность же «ящика» оказывается неоднородной, что позволяет различать составные части самой системы. При более детальном рассмотрении некоторые части системы могут быть, в свою очередь, разбиты на составные части и т.д. Те части системы, которые мы рассматриваем как неделимые, будем называть элементами. Части системы, состоящие более чем из одного элемента, назовем подсистемами. Можно ввести обозначения или термины, указывающие на иерархию частей (например, «подсистемы такого-то уровня»). В результате получается модель состава системы, описывающая, из каких подсистем и элементов она состоит (рис. 1.2).

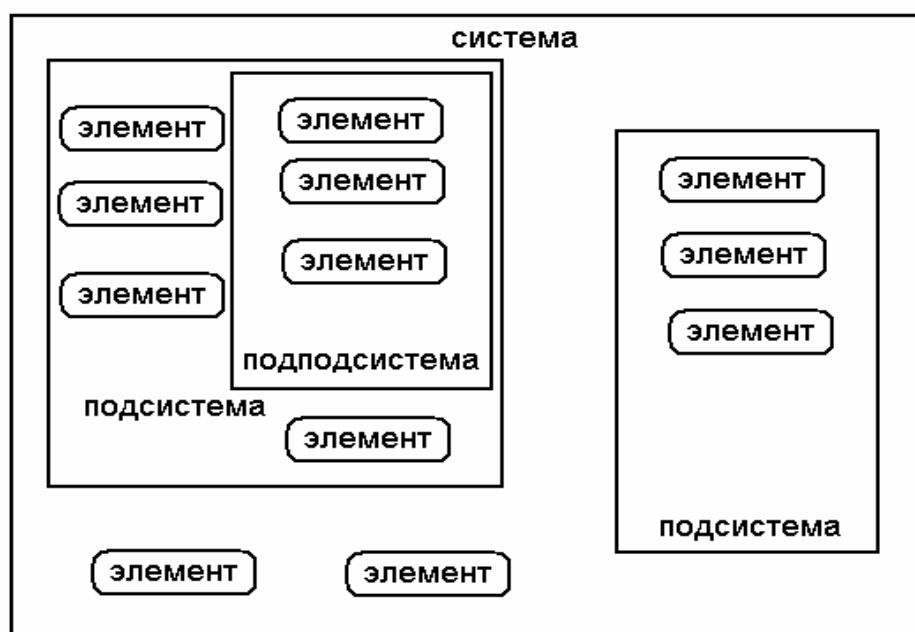


Рис. 1.2

Построение модели состава системы только на первый взгляд кажется простым делом. Если дать разным экспертам задание определить состав одной и той же системы, то результаты их работы будут различаться, и иногда довольно значительно. Причины этого состоят не только в том, что у них может быть различная степень знания системы: один и тот же эксперт при разных условиях также может дать разные модели. Существуют по крайней мере еще три важные причины этого факта. Во-первых, разные модели состава получаются вследствие того,

что понятие элементарности можно определить по-разному. То, что с одной точки зрения является элементом, с другой – оказывается подсистемой, подлежащей дальнейшему разделению.

Во-вторых, как и любые модели, модель состава является целевой, и для различных целей один и тот же объект потребуется разбить на разные части. Например, один и тот же завод для директора, главного бухгалтера, начальника пожарной охраны состоит из совершенно различных подсистем. Точно так же модели состава самолета с точек зрения летчика, стюардессы, пассажира и аэродромного диспетчера окажутся различными. То, что для одного обязательно войдет в модель, может совершенно не интересовать другого. В-третьих, модели состава различаются потому, что всякое разделение целого на части, всякое деление системы на подсистемы является относительным, в определенной степени условным. Например, тормозную систему автомобиля можно отнести либо к ходовой части, либо к подсистеме управления. Другими словами, границы между подсистемами условны, относительны.

Это относится и к границам между самой системой и окружающей средой; поэтому остановимся на этом моменте подробнее. В качестве примера рассмотрим систему «часы». Какую бы природу ни имели устройства, которые мы называем часами, в них можно выделить две подсистемы: датчик времени, т.е. процесс, ход которого изображает течение времени (это может быть равномерное раскручивание пружины, электрический ток с некоторым постоянным параметром, равномерное течение струйки песка, вращение Земли вокруг своей оси, колебания некоторой молекулы и т.д.); индикатор времени, т.е. устройство, преобразующее, отображающее состояние датчика в сигнал времени для пользователя. Модель состава часов можно считать полностью исчерпанной (если снова не разбивать эти две подсистемы). Однако, поскольку фактически каждые часы показывают состояние своего датчика, рано или поздно их показания разойдутся между собой. Выход из этого положения состоит в синхронизации всех часов с неким общим для всех эталоном времени, например с помощью сигналов «точного времени», передаваемых по радио. Здесь и возникает вопрос: включать эталон времени в состав часов как системы или рассматривать часы как подсистему в общей системе указания времени?

Модель состава ограничивается снизу тем, что считается элементом, а сверху – границей системы. Как эта граница, так и границы разбиения на подсистемы определяются целями построения модели и, следовательно, не имеют абсолютного характера. Это не означает, что сама система или ее состав нереальны, т.к. мы имеем дело с разными моделями системы.

1.1.5. Модель структуры системы

Еще раз подчеркнем, что для достижения ряда практических целей достаточно модели «черного ящика» или модели состава. Однако очевидно, что есть вопросы, решить которые с помощью этих моделей нельзя. Чтобы получить велосипед, недостаточно иметь «ящик» со всеми отдельными его деталями. Надо еще правильно соединить все детали между собой, или, говоря более общо, установить между элементами определенные связи – отношения. Совокупность необходимых и достаточных для достижения цели отношений между элементами называется структурой системы.

Перечень связей между элементами (т.е. структура системы) является отвлеченной, абстрактной моделью: установлены только отношения между элементами, но не рассмотрены сами элементы. Хотя на практике безотносительно к элементам говорить о связях можно лишь после того, как отдельно рассмотрены сами элементы (т.е. рассмотрена модель состава), теоретически модель структуры можно изучать отдельно.

Бесконечность природы проявляется и в том, что между реальными объектами, вовлеченными в систему, имеется невообразимое (может быть, бесчисленное) количество отношений. Однако когда мы рассматриваем некоторую совокупность объектов как систему, то из всех отношений важными, т.е. существенными для достижения цели, являются лишь некоторые. Точнее, в модель структуры (т.е. в список отношений) мы включаем только конечное число связей, которые, по нашему мнению, существенны по отношению к рассматриваемой цели.

Пример 1. При расчете механизма не учитываются силы взаимного притяжения его деталей, хотя, согласно закону всемирного тяготения, такие силы объективно существуют. Зато вес деталей (т.е. сила их притяжения к Земле) учитывается обязательно.

Пример 2. Между деталями молотка наряду с отношениями, существенными для его использования, имеются и несущественные. Например, если соприкасающиеся детали изготовлены из разных металлов, то между ними есть контактная разность потенциалов.

Пример 3. Рассмотрим систему «часы вообще». Считаем, что в состав такой системы входят три элемента: датчик, индикатор и эталон времени. Структура часов определяется следующими отношениями между парами элементов:

Пара элементов	Связь между ними
Датчик и индикатор	Однозначное соответствие
Эталон и датчик	Приблизительное соответствие
Индикатор и эталон	Периодическое сравнение и устранение расхождений

1.2. Классификация систем

В теории систем принято делить все системы на два больших класса:

- 1) большие системы (простые);
- 2) сложные системы.

Большие системы имеют строго иерархическую структуру или же иерархическую структуру с небольшим числом горизонтальных связей – сетевую структуру невысокой сложности. Такие системы удастся исследовать методом декомпозиции.

Сложные системы отличаются от больших числом горизонтальных связей – фактически это сложная сетевая структура, а исследовать ее методом чистой декомпозиции не удастся. Кроме того, сложные системы работают в условиях воздействия большого количества случайных факторов, что исключает чисто аналитический подход к анализу таких систем.

По характеру перехода из одного состояния в другое системы делят на **статические** и **динамические**.

Системы, в которых процесс перехода из одного состояния в другое не является рабочим – например, это всего лишь процесс настройки системы – называются статическими.

Системы, в которых переход из одного состояния в другое является рабочим режимом, называются динамическими.

Кроме того, динамическими являются системы, переход которых в новое состояние не может совершаться мгновенно, а происходит в результате некоторого процесса, растянутого во времени. Строго говоря, все реальные системы динамические. Однако практически можно пренебречь временем переходного процесса, если оно незначительно по сравнению с временем конкретного состояния системы.

Обилие возможных связей между подсистемами, наличие таких связей между системами и подсистемами как разных уровней (по вертикали), так и одного уровня (по горизонтали), а также наличие перекрестных связей сильно затрудняют анализ систем как единого целого. Введено понятие большой системы, т.е. такой, которую практически невозможно исследовать без выделения в ней более простых под-

систем. Такое ее расчленение на ряд более простых систем, рассматриваемых далее как относительно независимые, называется декомпозицией. Существуют системы, в которых такая декомпозиция в целях анализа несостоятельна, т.к. из-за тесных связей отдельных подсистем их изолированное исследование крайне затруднено и даже невозможно. Такие системы называют сложными.

Примеры – автомобиль (большая система), мозг человека (сложная система).

Большие системы переходят в сложные по мере усиления взаимодействия составляющих их компонентов, хотя четкую границу между ними провести невозможно.

С точки зрения взаимной обусловленности событий в системах их делят на детерминированные и стохастические.

В детерминированных системах всегда имеются однозначные взаимозависимости между входами и выходами; они могут быть описаны чисто аналитически. В данном случае считается, что случайные воздействия на исследуемый объект отсутствуют.

В стохастических системах событие X в элементе m_i может вызвать событие Y или Z , а может быть, Y , или Z , или U в некотором другом элементе m_j , причем возможность возникновения этих событий характеризуется соответственно различными вероятностями P_y, P_z, P_u и т.д. В таких системах связи между элементами и событиями в них носят вероятностный характер.

Строго детерминированных систем не существует, все – стохастические, такое деление является условным. В детерминированных системах вероятность ожидаемого события близка к 1.

С точки зрения общения с внешней средой системы делят на открытые и закрытые.

Открытые – это такие системы, в которых большинство входных воздействий формируется вне системы.

Закрытые – это такие системы, у которых все входные воздействия формируются внутри системы, хотя это понятие условное. Любая система является открытой, т.к. на нее всегда действуют различные внешние факторы.

Функционирование любой системы происходит в некоторой окружающей ее внешней среде, не входящей в систему, но взаимодействующей с ней.

Подобно связям между элементами системы, связи ее с внешней средой могут носить вещественный, энергетический и информационный характер. Количество таких связей велико; поэтому, исходя из

конкретных задач исследования, необходимо выделять лишь наиболее существенные связи, определяющие поведение системы в конкретных условиях.

1.3. Особенности функционирования систем

Состояние любой реальной системы в каждый момент времени можно описать с помощью некоторого множества характеризующих систему величин – параметров.

Для описания состояния и движения системы можно применять словесные описания, табличные или математические выражения, графические способы.

Словесное описание сводится к последовательному перечислению и характеристике параметров системы, тенденции их изменения, последовательности смены состояний системы – приближенно.

Таблицы и матрицы получили широкое распространение для количественной характеристики систем, выражаемой значениями их параметров в некоторые моменты времени. По этим таблицам строят диаграммы и графики, дающие визуальное представление о динамике системы.

Однако наиболее глубокой и наглядной является формализованная геометрическая интерпретация состояния и движения систем в так называемом пространстве состояний, или фазовом пространстве.

Процесс функционирования любой системы может быть представлен несколькими способами.

1. Входные и выходные воздействия представляются множеством чисел с помощью функций преобразования входов в выходы.

2. С помощью таблиц соответствия, в которых в одну графу заносятся значения входных воздействий, а в другую – значения выходных воздействий.

3. Графический способ является наиболее наглядным для описания динамических систем.

Для описания движения динамической системы широко применяется метод, основанный на использовании фазового пространства, так называемого n -мерного пространства, где n – число обобщенных координат системы. Для каждой системы с помощью n -мерного пространства формируют область допустимых состояний, т.е. в этой области может находиться точка, отображающая движение системы в случае ее нормального функционирования.

1.3.1. Пространство состояний системы

Пространством состояний системы называется пространство, каждой точке которого (изображающей точке) однозначно соответствует определенное состояние рассматриваемой динамической системы, а каждому процессу изменения состояний системы соответствует определенная траектория перемещения изображающей точки в пространстве.

Для описания движений динамических систем широко применяется метод, основанный на использовании фазового пространства (n -мерного Евклидова пространства), по осям которого откладываются значения всех n обобщенных координат рассматриваемой динамической системы. При этом взаимно однозначное соответствие между состояниями системы и точками фазового пространства достигается выбором числа измерений последнего, равного числу обобщенных координат рассматриваемой динамической системы.

Обозначим параметры некоторой системы Z_1, Z_2, \dots, Z_n , которые будем именовать в дальнейшем фазовыми координатами системы, а состояние (фазу) системы изобразим точкой $Z = (Z_1, Z_2, \dots, Z_n)$ в фазовом пространстве. Размерность этого пространства определяется, очевидно, числом фазовых координат, т.е. числом отобранных для описания системы и ее существенных параметров. Примеры:

- 1) одномерное фазовое пространство;
- 2) двухмерное фазовое пространство;
- 3) трехмерное фазовое пространство.

Геометрическое место точек при переходе системы из состояния Z_1 в Z_2 , представляющее собой некоторую траекторию в фазовом пространстве, отражает процесс движения системы.

В общем случае, когда число параметров, характеризующих систему, произвольно и больше трех, геометрическая интерпретация теряет наглядность. Однако геометрическая терминология в этих случаях остается удобной для описания состояния и движения систем в многомерном фазовом пространстве (гиперпространстве). Число независимых параметров системы называют числом степеней свободы.

Область фазового пространства, за пределы которого не должна выходить изображающая точка, называется областью допустимых состояний системы S_d . При исследовании и проектировании систем всегда исходят из того, что система остается в пределах области ее допустимых состояний.

Фактически в нормальных условиях эксплуатации параметры системы изменяются в более узких пределах, определяя рабочую область S_p . Геометрически это показано на рис. 1.3.

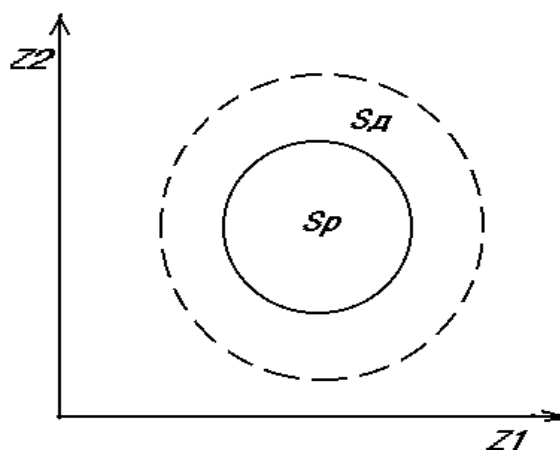


Рис. 1.3

Область допустимых состояний (или поле системы) включает в себя всевозможные фазовые траектории, т.е. «линии поведения системы».

Совокупность фазовых траекторий называется фазовым портретом рассматриваемой динамической системы.

Во всех случаях, когда параметры системы могут принимать в определенном интервале любые значения, т.е. изменяться плавно, изображающая точка может располагаться в любой точке области допустимых состояний – имеем дело с непрерывным пространством состояний.

Когда параметры системы могут принимать лишь дискретные значения, пространство состояний таких систем должно рассматриваться как дискретное. Точка, изображающая состояние системы, может находиться только в определенных фиксированных точках области допустимых состояний.

Изменение состояний такой системы интерпретируется скачками изображающей точки из одного состояния в другое и т.д. Траектория движения изображающей точки будет иметь дискретный, прерывистый характер. Однако деление характеристик систем на непрерывные и дискретные условно, т.к. любые непрерывные процессы дискретны (дискретна материя).

Любые процессы формирования системы могут условно рассматриваться в непрерывном и дискретном времени. Первое относится к случаю фиксации состояний системы последовательностями изображающих точек, которые соответствуют отсчетам, осуществляемым через сколь угодно малые интервалы времени Δt . Второе соответствует отсчетам, которые берутся через определенные (не бесконечно малые) интервалы времени Δt – например такты дискретного автомата.

1.3.2. Преобразования в системах

Движение любой системы представляет собой некоторую последовательность изменений ее состояний.

Характеризуя состояние системы в некоторый момент времени t_i вектором Z_i , а состояние в момент t_{i+1} – вектором Z_{i+1} , можно считать, что произошел переход (Z_i, t_i) в (Z_{i+1}, t_{i+1}) .

Закон, согласно которому каждому элементу некоторого множества соответствует элемент другого множества, называется оператором. При переходе системы в новое состояние оператором будет называться закон, в соответствии с которым происходит этот переход. Значение переменной величины, над которой совершается операция, вызывающая переход, называется операндом. Новое значение переменной, новое состояние, возникшее в результате воздействия оператора на операнд, называется образом.

Например, при реализации функции умножения $y = ax$ правило умножения есть оператор, x – операнд, y – образ.

Более сложны случаи, когда под воздействием некоторого оператора происходят переходы для некоторого множества операндов. Такое множество переходов называется преобразованием.

Преобразования, в которых каждому операнду соответствует только один образ, называются однозначными.

Преобразования, в которых каждому операнду могут соответствовать различные образы, называются неоднозначными.

Например: корень квадратный из $4 = 2 = -2$.

Любую систему, в которой происходят те или иные изменения (преобразования), в результате которых множество реакций системы $Y = (y_1, y_2, \dots, y_n)$ некоторым образом зависит от множества входных воздействий $X = (x_1, x_2, \dots, x_n)$, можно рассматривать как преобразователь, реализующий некоторую данную зависимость $y = f(x)$.

По особенностям преобразования все динамические системы можно разделить на два типа.

В системах 1-го типа (более простых) между вектором входных величин X и вектором выходных величин Y существует однозначная функциональная связь. Следовательно, вектор входных величин полностью определяет вектор выходных величин.

В системах 2-го типа (более сложных) имеют место две ступени функциональных связей. Одна из них описывает зависимость внутреннего состояния системы Z от вектора входных величин X , другая – зависимость вектора выходных величин Y от множества внутренних параметров системы и вектора входных величин.

Таким образом, во втором случае вместо зависимостей вида $y(t) = f[x(t)]$ рассматриваются зависимости вида $y(t) = f[x(t), z(t)]$.

По такому принципу работают конечные автоматы, являющиеся основой построения и функционирования ЭВМ.

1.3.3. Устойчивость систем

Под устойчивостью системы понимают ее свойство возвращаться в некоторое установившееся состояние или режим после нарушения последнего вследствие воздействия каких-либо внешних или внутренних факторов.

Основное содержание теории устойчивости составляет исследование влияния возмущающих воздействий на поведение системы. Под возмущающими факторами понимают силы, обычно неизвестные заранее и не учитываемые при описании движения системы. Система может иметь несколько состояний устойчивого и неустойчивого равновесия.

Одним из примеров устойчивого состояния системы есть ее цикличность – когда система при отсутствии возмущений периодически проходит одну и ту же последовательность состояний, например процесс незатухающих колебаний в электрическом колебательном контуре без потерь.

Если система возвращается в состояние равновесия или цикличности при любых возмущениях, то равновесие называется абсолютно устойчивым, а если она возвращается в состояние равновесия только из некоторой области, то равновесие называется устойчивым относительно этой области.

Если после воздействия на систему она сохраняет новое состояние, вызванное этим воздействием, то система называется безразлично устойчивой.

Устойчивость поведения систем является обычно положительным свойством, обеспечивающим их нормальное функционирование. Однако в ряде случаев устойчивость нежелательна, т.к. ограничивает возможности системы.

Характерные свойства кибернетических систем – большое число устойчивых состояний и способность самопроизвольно приходить в устойчивое состояние при нарушении равновесия.

Под устойчивостью функционирования сложной системы в более общем смысле понимают способность системы сохранять требуемые свойства в условиях действия возмущений.

Для системы, устойчивой относительно какого-либо возмущения или их комплекса, обычно указывают ограничения на эти возмущения, при которых система остается устойчивой.

Устойчивость кибернетических систем оценивается с помощью специальных критериев.

Подробно устойчивость систем рассматривается в теории управления.

1.3.4. Особенности управления сложными системами

В сложных системах важную роль играют процессы управления, представляющие собой процесс сбора, передачи и переработки информации, а также формирование на основе этой информации управляющих команд, изменяющих состояние и режимы функционирования системы.

В сложных системах выделяют специфические контуры управления, вдоль которых циркулируют потоки информации. Часто контуры управления являются замкнутыми и носят характер обратной связи: фактическое значение параметра сравнивают с его требуемым значением, и по величине отклонения вырабатывается корректирующий сигнал управления.

Управление сложной системой чаще всего осуществляется цифровым вычислительным устройством, работающим по специальному алгоритму.

Если управление сосредоточено в одном вычислительном устройстве, оно называется централизованным.

Серьезное значение имеет правильный выбор степени централизации управления в тех системах, в которых существуют большие массивы осведомительной информации, а качество управления существенно зависит от учета состояния всех элементов системы.

Децентрализация управления позволяет сократить объем передаваемой и перерабатываемой информации, но иногда снижает качество управления, что можно предотвратить, если использовать систему управления с иерархической структурой (рис. 1.4).

Особенность управления с иерархической структурой заключается в следующем: основная масса информации перерабатывается в соответствующих контурах низшего уровня, а на высшие уровни поступают лишь обобщенные данные, характеризующие не отдельные элементы, а целые подсистемы.

Многим сложным системам свойственны черты самоорганизации. Система называется самоорганизующейся, если она способна на основе оценки воздействий внешней среды путем последовательного изменения своих свойств прийти к некоторому устойчивому состоянию, когда воздействия внешней среды окажутся в допустимых пределах.

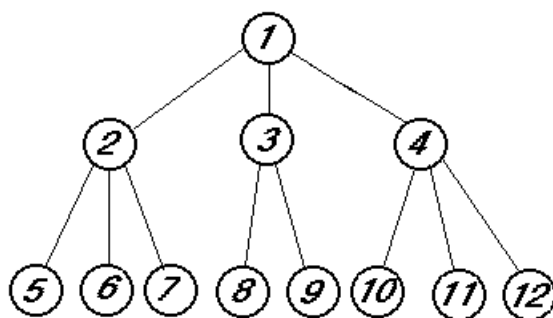


Рис. 1.4

Здесь 1 – центральное управляющее устройство; 2, 3, 4 – устройства управления низшего уровня; 5 .. 12 – объекты управления.

Реальные сложные системы функционируют в условиях действия большого числа случайных факторов, источниками которых являются воздействия внешней среды, ошибки, шумы, отклонения величин в самой системе.

Все случайные воздействия оказывают влияние на систему и могут существенно менять характер ее функционирования.

Рассеивание значений воздействий внешней среды или параметров системы приводит к рассеиванию результатов ее функционирования. Кроме того, доказано и практически подтверждено, что действие случайных факторов приводит к смещению средних значений результатов функционирования.

1.4. Критерии эффективности сложных систем

Качество функционирования сложных систем оценивают с помощью показателей, или критериев эффективности, под которыми понимают такую числовую характеристику, которая оценивает степень приспособленности системы к выполнению поставленной перед ней задачи.

Выбор показателя эффективности является заключительной стадией формулировки целей и задач системы.

Рассмотрим производственный процесс как сложную систему. Если принять критерием эффективности производительность, мы будем придавать максимальное значение факторам, связанным с производительностью, в ущерб другим – экономии сырья, фонду зарплаты, качеству изделия и т.д.

Если принимать за основу себестоимость продукции, на второй план может уйти производительность.

Для производственного процесса, как и для других систем, могут быть выбраны комплексные показатели, в частности для производственного процесса – это рентабельность.

Из примеров видно, что только выбор показателя эффективности делает описание целей и задач системы законченным.

Расчет показателей эффективности сложной системы – трудная задача, решаемая на ЦВМ с помощью математических и программных средств.

Чтобы показатель эффективности достаточно полно характеризовал качество работы системы, он должен учитывать все основные особенности и свойства системы, а также условия ее функционирования и взаимодействия с внешней средой, т.е. показатель эффективности определяется процессом ее функционирования. При этом можно представить себе множество процессов функционирования системы, элементы которого отличаются друг от друга вследствие различных условий и режимов работы системы. Каждому элементу этого множества можно поставить в соответствие элемент другого множества – значений показателя эффективности системы.

Следовательно, показатель эффективности можно считать функционалом, заданным на множестве процессов функционирования системы.

Функционал – это оператор, заданный на некотором множестве функций и принимающий значения из области действительных чисел.

Так как сложные системы работают в условиях действия случайных факторов, значения функционалов оказываются случайными числами, а при оценке показателей эффективности обычно пользуются средними значениями функционалов. Например, среднее количество изделий, выпускаемых за смену, среднее время ожидания в очереди.

Иногда в качестве показателей эффективности используют вероятность некоторых случайных событий, например вероятность застать абонентную линию занятой (телефон). Кажется, что здесь принципиально иная ситуация – элементам множества процессов функциональных систем ставится в соответствие множество случайных событий. Однако, если каждому событию поставить в соответствие функционал, принимающий два значения – 0 или 1 (событие не наступило, ... наступило), вероятность события равна среднему значению функционала. Могут быть построены совокупности функционалов, характеризующие и другие свойства сложных систем – надежность, помехозащищенность, качество управления и т.п. Как показывает опыт, наибольшей наглядностью и стройностью при постановке задачи отличаются совокупности функционалов, зависящие от показателей эффективности. Действительно, в большинстве случаев то или другое свойство системы имеет значение не само по себе, а лишь как фактор, влияющий на ее эффективность.

Рассмотрим показатель, характеризующий надежность сложной системы. Оценка надежности обычно производится с помощью специально выбранных функционалов, называемых показателями надежности системы. Следует заметить, что показатели надежности, заимствованные из теории надежности «простых» систем, учитывают лишь сам фактор появления или отсутствия отказов в элементах системы и не дают представления о влиянии отказов на конечный эффект функционирования систем. Очевидно, что для многих сложных систем выход некоторых элементов из рабочего состояния не только не приводит к потерям работоспособности, но иногда является планируемым событием (поломка автотранспорта и т.п.). Здесь может идти речь лишь о снижении качества работы системы, т.е. об изменении ее эффективности.

Постановка задачи об оценке надежности сложной системы сводится к следующему. Предполагаются известными характеристики, описывающие интенсивность отказов элементов сложной системы и определяемые экспериментально или другими методами оценки надежности.

Пусть в качестве показателя эффективности сложной системы выбран некоторый функционал Φ , значения которого зависят как от структуры и параметров системы, так и от значений характеристик надежности ее элементов, а Φ_1 – его значение, вычисленное в предположении, что отказы элементов имеют интенсивности, соответствующие заданным характеристикам, тогда в качестве показателя надежности сложной системы может быть выбрана разность $\Delta\Phi_{\text{надежн}} = \Delta\Phi - \Delta\Phi_1$, показывающая, насколько снижается эффективность системы из-за возможных отказов ее элементов по сравнению с эффективностью системы с абсолютно надежными элементами.

Если разность $\Delta\Phi_{\text{надежн}}$ мала, то отказы слабо влияют на эффективность системы; если $\Delta\Phi_{\text{надежн}}$ велика, то необходимо принять меры к повышению надежности; например, резервирование элементов, отказы которых оказывают наибольшее влияние на эффективность.

1.5. Основы разработки и исследования сложных систем

1.5.1. Основные этапы разработки сложных систем

Применительно к объектам большой сложности особое значение приобретает начальный этап проектирования, на котором осуществляются формулирование требований к системе и обоснование технического задания (ТЗ). Здесь должно быть получено следующее:

- 1) определение целей создания системы и круга возлагаемых на нее задач;

2) перечень действующих на систему факторов, подлежащих обязательному учету при проектировании, их числовые характеристики;

3) обоснованность показателей эффективности, надежности и т.п., по которым предполагается оценивать качество системы, и количественные требования к ним;

Эти сведения оформляются в виде ТЗ на проектирование, которое служит руководящим документом на всех этапах создания системы.

Проектирование сложных систем имеет две достаточно важные стадии. **Первая** относится к структурным вопросам и называется **макропроектированием** или внешним проектированием системы. На этой стадии производят выбор структуры системы, ее основных элементов, организации взаимодействия между ними, воздействия внешней среды, оценку показателей эффективности и соответствия рассматриваемого варианта системы общим требованиям ТЗ.

Вторая стадия – микропроектирование, или внутреннее проектирование, связанное с проектированием элементов сложных систем как физических единиц. Здесь осуществляются технические решения по основным элементам системы, их конструкции и параметрам, режиму эксплуатации, по организации производства.

В настоящее время накоплен большой опыт микропроектирования и создания элементов.

Внешнее проектирование – гораздо более сложная задача. Здесь нужны специалисты широкого профиля – инженеры-системотехники, с серьезной подготовкой по многим комплексным наукам.

Работа по внешнему проектированию начинается с изучения ТЗ и рассмотрения возможных вариантов построения системы. Далее производятся оценка вариантов системы и выбор одного из них для более детального исследования. Выбранный вариант подвергается всестороннему анализу с целью проверки его соответствия требованиям ТЗ. Для этого осуществляется оценка эффективности системы и ее других свойств.

Типичным методом исследования сложных систем на этом этапе является их моделирование на ЦВМ. Поскольку представление об элементах системы еще поверхностное, в моделях они обычно заменяются упрощенными эквивалентными схемами. Тем не менее модель должна удовлетворять следующим требованиям:

1) хорошо отражать структуру системы и быть чувствительной к таким ее изменениям, которые обычно производятся в процессе внешнего проектирования;

2) отражать специфику функционирования элементов системы с учетом условий внешней среды;

3) содержать все параметры системы, определяемые в результате микропроектирования.

В результате моделирования получаются зависимости, характеризующие влияние структуры и параметров системы на ее и другие свойства. Эти зависимости используются для определения оптимальной структуры и оптимальных значений параметров с точки зрения выбранных показателей эффективности.

Разработка технического проекта системы относится в основном к микропроектированию ее элементов и конкретной реализации их взаимодействия в процессе функционирования.

Разработка современной сложной системы сопровождается различными автономными и комплексными испытаниями. Автономные испытания относятся к элементам системы и имеют целью проверку правильности функционирования оборудования и получение экспериментальных данных для оценки параметров и режимов эксплуатации. Комплексные испытания (в целом или фрагментов) предназначены для отработки взаимодействия между элементами и подсистемами на различных уровнях иерархии и проверки соответствия ТЗ.

Подробное обследование элементов на стадии автономных испытаний позволяет освободить комплексные испытания от ненужных деталей и сократить объем регистрируемой информации.

На стадии комплексных испытаний приходится регистрировать не только общесистемные параметры, но и некоторые параметры элементов системы, так как в процессе автономных испытаний элементы функционируют часто не так, как в составе системы.

Для испытания сложных систем характерно некоторое использование разных видов имитации внешних воздействий. Любое внешнее воздействие на систему приводит к возникновению информации, которая становится составной частью потоков информации, циркулирующей в системе. Нередко информация, которая появилась в результате внешних воздействий, может быть заменена искусственной информацией, вырабатываемой имитатором.

1.5.2. Основные задачи исследования сложных систем

Среди задач, возникающих в связи с исследованием сложных систем, выделяют два основных класса:

- 1) задачи анализа, связанные с изучением свойств и поведения системы в зависимости от ее структуры и значений параметров;
- 2) задачи синтеза, сводящиеся к выбору структуры и значений параметров в зависимости от заданных свойств системы.

При анализе предполагаются заданными требуемые значения функциональных характеристик системы, а также область устойчивости системы. Необходимо выбрать структуру системы и значения параметров, обеспечивающие требуемые значения функциональных характеристик. Обычно синтез ставится как экспериментальная задача, т.е. речь идет о выборе такой структуры и значений параметров, при которых показатель эффективности имел бы \max или \min с учетом ограничений, налагаемых на остальные показатели (надежности, помехозащищенности и т.п.).

Следует заметить, что в настоящее время имеется достаточно много методов решения задач анализа сложных систем, среди которых больше всего распространены:

- 1) расчет показателей эффективности и других функциональных характеристик системы с помощью формул и уравнений, относящихся к данному узкому классу системы;

- 2) расчет показателей эффективности и других функциональных характеристик по результатам моделирования сложной системы на ЦВМ для систем общего вида.

Задача анализа может быть решена, если имеются все необходимые данные для расчета, чего нельзя сказать о задачах синтеза. В настоящее время почти нет методов, позволяющих строго формально решать задачи синтеза, за исключением нескольких узких классов систем, например конечных автоматов, для которых развиты формальные методы синтеза. Однако эти методы не могут быть непосредственно обобщены на другие классы систем, тем более на общий случай. Поэтому на практике пользуются различными неформальными методами синтеза сложных систем. В конечном счёте все они сводятся к перебору вариантов структур, или «синтезу через анализ», суть которого состоит в том, что, приступая к синтезу системы, исследователь намечает первоначальный вариант синтеза. Этот вариант известными методами исследуется – определяются параметры эффективности и другие характеристики.

Результаты анализа «первоначального» варианта системы сравниваются с заданными значениями. Как правило, «первоначальный» вариант не является удовлетворительным. Поэтому намечают следующий вариант синтеза системы с учетом недостатков «первоначального», который также исследуют, – и т.д. до получения системы с требуемыми характеристиками.

По мере обследования вариантов системы накапливаются сведения, ценные для синтеза. Зависимости функциональных характеристик от параметров системы исследуются для целенаправленного перехода к более удовлетворительным вариантам синтезированной системы.

Кроме того, эти сведения полезны и для разработки конкретных практических рекомендаций, например для оптимального планирования режимов наладки и замены инструментов на производстве, ремонта и профилактики оборудования и т.п.

Исключительно важное значение имеют исследования сложных систем в связи с вопросами управления. Для расчета систем управления необходимо уметь решать задачи, связанные с анализом процессов функционирования сложного оборудования, оценкой информационных потоков и законов управления, синтезом алгоритмов, выбором технических средств, обозначением параметров оборудования и т.п.

На основе анализа потоков информации необходимо решить самую важную задачу – задачу выделения информации, существенной для управления.

Чтобы построить оптимальные методы управления, следует выделить такое минимальное количество информации, которое еще обеспечивает заданное качество управления и не является избыточным.

Далее переходят к рассмотрению возможной структуры системного управления. Одним из вопросов, возникающих при этом, является оценка оптимальной централизации управления.

Анализ вариантов может быть использован для получения комплексной оценки эффективности разных вариантов автоматизации управления.

Контрольные вопросы

1. Каковы основные свойства систем?
2. Что такое сложная система?
3. Каковы основные свойства сложных систем?
4. Чем отличаются сложные системы от простых?
5. Что такое и для чего нужна модель «черный ящик»?
6. Что такое модель состава системы?
7. Что такое модель структуры системы?
8. Как можно представить процесс функционирования любой системы?
9. Что такое пространство состояний системы?
10. Что такое преобразования системы и какие они бывают?
11. Чем характеризуется устойчивость систем?
12. Каковы основные особенности управления сложными системами?
13. Что такое критерии эффективности сложных систем и каковы основные требования к ним?
14. Назовите и охарактеризуйте основные этапы разработки и основные задачи исследования сложных систем.

2. ИНФОРМАЦИОННЫЕ СИСТЕМЫ

2.1. Автоматизированные информационные системы

Процессы обработки информации всегда являлись основой человеческой деятельности, и объединение таких процессов с информационными ресурсами со временем стали называть информационными системами (ИС).

ИС – это комплекс, состоящий из информационной базы (хранилища информации) и процедур, позволяющих накапливать, хранить, корректировать, осуществлять поиск, обработку и выдачу информации. С появлением вычислительной техники ИС превратились в автоматизированные информационные системы (АИС), т.е. информационные системы, физической и функциональной компонентами которых являются программно-технический комплекс и средства связи.

Современные АИС представляют собой чрезвычайно сложные человеко-машинные комплексы, интегрированные в национальную и мировую информационные среды. Именно эта интеграция и создает эффективную научно-техническую базу информационного общества, так как изолированные АИС в настоящее время малоэффективны.

Эффективность АИС во многом определяется их качеством и доверием к ним пользователей. Качество изделий, процессов проектирования, производства и услуг является одной из узловых проблем, определяющей уровень жизни человека и состояние народного хозяйства, что полностью относится и к области информационных технологий. В АИС входят следующие основные компоненты:

- аппаратные средства вычислительной техники;
- аппаратные средства телекоммуникации (связи);
- программные средства реализации функций АИС;
- информационные базы данных (БД);
- документация, регламентирующая функции и применение компонент АИС;
- специалисты, обслуживающие и использующие программно-технические средства.

Аппаратные компоненты АИС имеют достаточно универсальный характер и относительно слабо зависят от функционального назначения конкретной информационной технологии. Остальные компоненты автоматизированных информационных систем составляют их интеллектуальную часть, определяющую назначение, функции и качество решения задач в конкретной области человеческой деятельности. Эти

компоненты могут отличаться новизной и большим разнообразием характеристик.

Архитектурная, техническая и программно-информационная совместимость различных АИС может быть обеспечена только путем стандартизации и сертификации программно-технических средств в соответствии с требованиями международных и государственных стандартов. Для этого необходимы стандартизация, сертификация и каталогизация средств, процессов и услуг, а также проведение единой технической политики при создании (приобретении) совместимых аппаратных и программных средств, организации взаимодействия и интеграции АИС различных уровней и назначения.

Любая реальная АИС действует в окружающей ее информационной среде, которую часто называют инфраструктурой.

Под инфраструктурой АИС обычно понимают телекоммуникационные сети и связываемые ими объекты: серверы, автоматизированные рабочие места, каталоги сетевых информационных ресурсов и т.п. Информационными ресурсами являются информационные базы (банки и базы данных) различного назначения и другие информационные структуры.

Таким образом, при создании сетей передачи данных нового поколения возникает необходимость решения комплекса задач, лежащих как в области телекоммуникаций, так и в области информационных технологий. Современное состояние телекоммуникационных технологий представляет собой весьма пеструю картину. С одной стороны, «всемирная паутина» (Internet) опутала весь земной шар и ближний космос, с другой – не иссякает поток критики со стороны пользователей данной «паутины».

Быстрое развитие и использование информационных технологий не только открывает новые возможности, но и создает новые проблемы:

- психобиологические, оказывающие отрицательное психологическое и физическое воздействие на пользователей;
- культурные, угрожающие национальной культурной самобытности пользователей;
- социально-экономические, увеличивающие неравенство возможностей получения доступа к качественным ИТ;
- политические, приводящие к разрушению гражданского общества в национальных государствах;
- бесконтрольное и несанкционированное использование чужой интеллектуальной собственности;
- технологические угрозы нанесения ущерба или разрушения самим АИС.

2.1.1. Содержание и структура теории ИС

Огромные средства затрачиваются во всем мире на разработку многочисленных прикладных систем, и недостаточное внимание уделяется теоретическим вопросам их исследования.

Необходимость концептуального, системного осмысления положения дел в некоторой предметной области возникает в силу того, что на определенном этапе развития этой области накапливается большое количество знаний, фактов, задач и интересов, слабо увязанных между собой. Такое положение дел периодически складывается во всех развивающихся областях человеческих знаний.

В частности, при разработке сложных перспективных АИС процесс создания включает большое разнообразие видов деятельности и требует тесного взаимодействия между представителями научно-технических профессий и лицами, принимающими политические и экономические решения. Возникает необходимость сведения воедино огромных объемов разнообразной информации, согласования большого числа различных и зачастую противоречивых целей и интересов.

Компоненты АИС принадлежат к различным предметным областям; поэтому теория АИС включают в себя положения из различных дисциплин.

При анализе и проектировании дискретных электронных схем для вычислительной техники и средств связи используется математический аппарат теории множеств, двоичной логики и теории кодирования. При разработке цифровых систем связи применяются методы математической статистики.

В разработках программного обеспечения и языков программирования используются методы автоматически-лингвистических моделей, аналитическая теория алгоритмов, модели исчисления предикатов, графовые модели и оптимизационные методы.

При разработке БД и информационных языков поиска в информационных БД обращаются к методам реляционной алгебры, моделям исчисления предикатов и математической лингвистики, а также графовым моделям.

При анализе эффективности, работоспособности и надежности АИС и их компонент необходимы методы теории массового обслуживания и математическая статистика.

Основой информационно-поисковых АИС являются документ – материальный объект с информацией, закрепленной созданным человеком способом для ее передачи во времени и пространстве, и классификатор – официальный документ, представляющий собой систематизированный свод наименований и кодов кодификационных группировок и (или) объектов классификации.

2.1.2. Предметная область АИС

Предметная область – это совокупность объектов реального или предполагаемого мира, рассматриваемых в пределах данного контекста, который понимается как отдельное рассуждение, фрагмент научной теории или теория в целом и ограничивается рамками данного контекста.

В научных и инженерных кругах, занимающихся вопросами теории и практики АИС, под предметной областью понимают систематизированную совокупность объектов, свойств объектов, связей между объектами, а также функции, выполняемые объектами, деятельность которых является предметом автоматизации с помощью АИС. Под простым объектом предметной области подразумевается носитель совокупности характеризующих его свойств, через которые проявляется сущность объекта и которые неотделимы от него.

Ядром (основой) любой АИС является информационная база (ИБ). ИБ представляет собой статичную информационную модель предметной области АИС, т.е. содержит систематизированное описание совокупности объектов, свойств объектов и связей между объектами в предметной области. Совокупность функций, выполняемых объектами, моделируется программным обеспечением АИС. Таким образом, АИС представляет собой динамическую модель конкретной предметной области.

В качестве примеров предметных областей АИС различного назначения можно назвать следующие:

- библиотека – для библиотечной АИС;
- цех или завод – для АИС управления предприятием;
- система целей – для АИС управления огнем средств ПВО;
- банк – для финансовой АИС;
- учебный курс или дисциплина – для автоматизированной системы обучения.

Разработка любой АИС начинается с системного анализа предметной области, в результате которого создается концептуальная схема предметной области (рис. 2.1.)

Концептуальное представление концентрирует внимание на смысле информации, именно концептуальная схема описывает это представление. На базе концептуальной схемы создаются внешние схемы, описания которых сосредоточены на том, как представлена информация для пользователя (т.е. описаны прототипы интерфейсов пользователя). Описание внутреннего физического представления данных в вычислительной системе содержится во внутренней схеме.

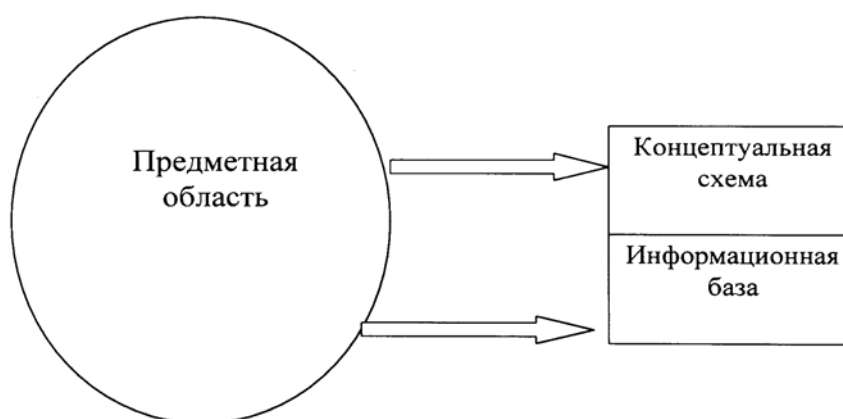


Рис. 2.1. Описание предметной области

Такая трехсхемная архитектура описания предметной области АИС признана достаточно широко на уровне международных стандартов. Концептуальная схема, помимо того что она является моделью предметной области на концептуальном уровне, полученной в результате системного анализа, играет также ключевую роль в проектировании БД.

Проектировщик ИС должен иметь возможность выразить концептуальную схему в терминах, соответствующих рассматриваемой предметной области и восприятию ее пользователем. В частности, не должно налагаться никаких ограничений на сущности, наличие которых допускается в предметной области, или на свойства, которыми они могут обладать. Проектировщик информационной системы ограничен только требованиями основных принципов логики, встроенных в информационный процессор, т.е. требованиями поддержания непротиворечивости.

Так как концептуальная схема должна быть простой в использовании и понимании для разнообразных пользователей, необходим механизм добавления лингвистических конструкций произвольной сложности (макроконструкций) в формальную систему, чтобы пользователи могли с ней взаимодействовать на любом требуемом уровне абстракции.

Более подробно особенности описания и использования предметной области рассматриваются в дисциплине «Управление данными».

2.1.3. Взаимодействие предметной области, пользователей и АИС

Концептуальная схема и информационная база абсолютно неизменны, пока нечто не произведет изменения в них. Это нечто называют информационным процессом. То есть, развивая концепцию АИС, вво-

дят понятие ИС, состоящей из концептуальной схемы, информационной базы и информационного процесса.

Информационный процесс производит изменения в информационной базе или концептуальной схеме только при получении сообщения. Сообщение содержит информацию и/или команды. Сообщения исходят из некоторой части реального мира, которую называют средой, которая может быть или не быть внешней по отношению к предметной области (рис. 2.2).

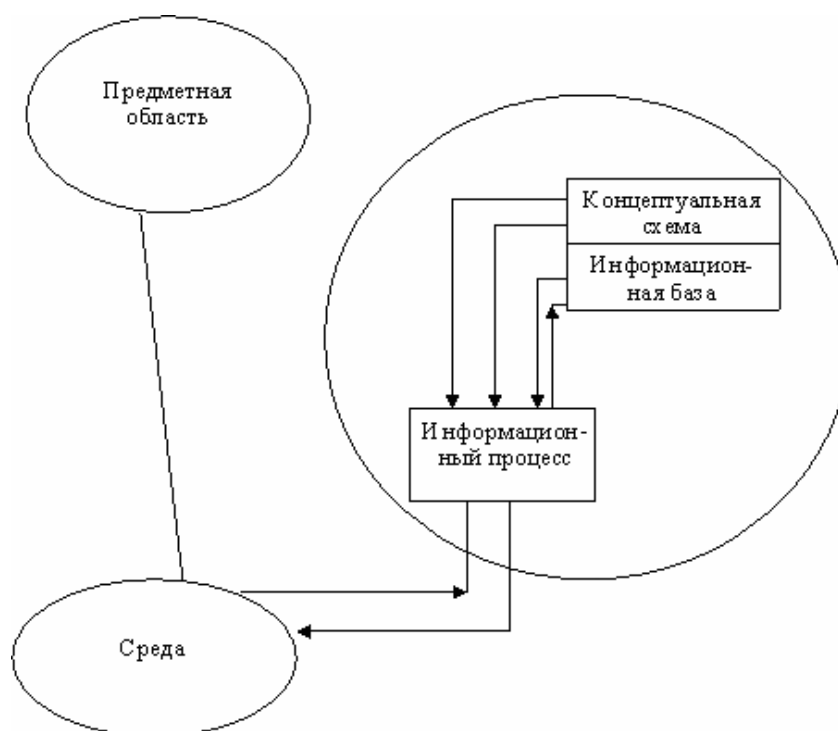


Рис. 2.2. Обобщенная структура ИС

При получении соответствующего сообщения, содержащего команду, информационный процесс может также выдавать информацию, имеющуюся в концептуальной схеме и информационной базе.

ИС отличается от среды следующим:

- ИС – это формальная система, среда в целом не является таковой;
- поведение ИС полностью определяется ограничениями и правилами поведения, которые установлены средой;
- ИС по своей инициативе никогда не устанавливает правила для среды;
- ИС не может отклоняться от предписанных правил или ограничений, а среда может отклоняться от своих правил.

В определенных случаях при системном анализе рассматривают ИС вместе со средой как части некоторой более сложной системы, но

последняя (сложная система) не может быть формальной или полностью предсказуемой. Поэтому далее термин «информационная система» используется с исключением из него среды и пользователей ИС, если другое не оговорено.

Пользователями ИС могут быть ЭВМ или другие системы, а также люди. Пользователь – это некто или нечто, посылающий команды и сообщения в ИС и получающий сообщения от ИС (он является частью среды). Некоторые пользователи также имеют полномочия устанавливать правила поведения или ограничения для ИС. При рассмотрении АИС управления (АИСУ) пользователи включаются в состав АИС, и тогда такая система называется организационно-технической или человеко-машинной.

Информационный процесс передает сообщения между средой и информационной базой или концептуальной схемой, как показано на рис. 2.2.

В АИС информационный процесс является вычислительной системой или какой-либо ее частью. В традиционных, не автоматизированных ИС роль информационного процесса играет человек при условии, что он не нарушает установленных правил. С другой стороны, вычислительные системы могут действовать как пользователи ИС. Пример – сеть автоматизированных информационных систем, обменивающихся сообщениями друг с другом. Если каждая имеет набор правил, независимый от других, тогда каждая играет роль пользователя других АИС.

2.1.4. Классификация АИС

АИС используются во всех областях человеческой деятельности. Далее приведена классификация АИС по признаку их применения:

- автоматизированная система управления (АСУ) – организационно-техническая система, созданная с применением автоматизированных информационных технологий для повышения эффективности процессов управления различными объектами;
- автоматизированная система научных исследований (АСНИ) – АИС, предназначенная для информационно-аналитического обеспечения научно-исследовательских работ;
- экспертная система – АИС, которая использует экспертные знания для обеспечения высокоэффективного решения задач в узкой предметной области;
- автоматизированная система контроля измерений (АСКИ) – АИС, предназначенная для сбора, анализа и хранения показаний контрольно-измерительных приборов;

- система автоматизированного проектирования (САПР) – организационно-техническая система, состоящая из программно-технического комплекса автоматизации проектирования, пользователями которого являются сотрудники подразделений проектной организации;
- автоматизированная система обучения – АИС, которая включает студентов, преподавателей, комплекс учебно-методических и дидактических материалов, автоматизированную систему обработки данных и предназначена для обеспечения процесса обучения с целью повышения его эффективности;
- автоматизированная справочная система – справочное руководство, содержание которого создается, хранится и доводится до пользователя с использованием автоматизированных информационных технологий;
- автоматизированная библиотечная система – АИС, обеспечивающая доступ к данным библиотечных каталогов и фондов, а также сбор, обработку и хранение соответствующей информации;
- автоматизированная система перевода – АИС, предназначенная для перевода текстов с одного языка на другой; составной частью такой системы является автоматизированный словарь;
- автоматизированная информационная юридическая система – АИС в предметной области юриспруденции;
- автоматизированные системы военного назначения – АИС, предназначенные для управления боевыми действиями, военными объектами, системами ПВО и т.д.

Можно привести еще много примеров АИС, но данный перечень имеет достаточно представительный характер.

Практически все АИС имеют в своем составе следующие компоненты:

- физическую – материальная основа носителя ИС;
- информационную – организованная определенным образом система записей данных (информационная база), характеризующаяся определенным языком, на котором выполнены образующие ее записи;
- функциональную – система процедур управления, обновления, поиска и завершающей обработки данных.

В локальной АИС компоненты сосредоточены в одном месте. Если же взаимодействие между компонентами АИС или их частями происходит посредством каналов связи, АИС называется распределенной.

2.2. Интегрированные корпоративные ИС

К сожалению, пока на многих российских предприятиях автоматизация управления носит эпизодический характер. Сосуществуют ИС, созданные на различных программно-аппаратных платформах и автоматизирующие решение отдельных функций управления на отдельных уровнях управления (локальные ИС). В то же время мировой опыт и опыт передовых отечественных компаний показывает, что решение проблемы автоматизации управления предприятием необходимо искать на путях внедрения интегрированных корпоративных информационных систем (ИКИС), охватывающих практически все стороны его деятельности и обладающих средствами поддержки корпоративного управления. Это управление персоналом, финансами и производством, логистика, бухгалтерский учет, ориентация на автоматизацию всех уровней управления – оперативного, тактического, стратегического. Главное преимущество таких систем – возможность работы с единой реляционной базой данных, что оптимизирует трудозатраты на ввод информации, уменьшает вероятность ошибок, облегчает решение проблемы защиты информации.

На российском рынке корпоративных систем присутствуют как западные продукты (например R/3 SAP, Oracle, Axapta/Navision, Frontstep), так и отечественные («Галактика», «Парус», «Фигаро», «КСТ М-3», 1С и др.) образцы.

По мнению экспертов, качество отечественных ИКИС по ряду параметров уступает международному уровню. Причина в том, что западные системы несут в себе в виде типовых решений отработанные многими десятилетиями бизнес-практики, для которых характерны ориентированность на аналитический менеджмент со строгой регламентацией процедур, высокой исполнительской дисциплиной, отлаженным механизмом принятия и контроля решений. Такие системы зачастую малопригодны в нашей стране, где условия бизнеса постоянно меняются, что требует быстрой адаптации решений.

Перечисленные причины и высокий уровень цен на услуги западных компаний делают постоянно развивающиеся российские ИКИС предпочтительными для значительной части отечественных предприятий.

По мнению экспертов, серьезного внимания заслуживают две российские корпоративные системы: «Галактика» и «Парус», которые близки по функциональному составу, по возможностям корпоративного управления и применения аналитической обработки данных для поддержки принятия решений. На стороне «Галактики» значительный опыт работы с крупными промышленными компаниями, ряд удачных

отраслевых решений, а также большое число внедрений (5600) на предприятиях России и других стран СНГ («Парус» внедрен приблизительно на 2500 предприятиях и организациях, из которых значительная часть – бюджетные). Заслуживает внимания образовательная политика корпорации «Галактика», которая в рамках программы «Галактика» и вузы» предоставляет учебным заведениям полнофункциональные демоверсии системы для внедрения в учебный процесс (об условиях участия в этой программе, к которой присоединились более 160 вузов, можно ознакомиться на сайте корпорации www.galaktika.ru).

2.2.1. Основы построения ИКИС

Простейшие ИС, реализующие отдельные функции управления на отдельных уровнях управления, получили название локальных. Предприятие, ограничивающееся лишь локальными ИС, не может быть конкурентоспособным в современных условиях. Необходим переход к применению полнофункциональных ИС, ориентированных на автоматизацию процессов управления на всех уровнях и обладающих средствами поддержки корпоративного управления.

Современная теория классифицирует экономические ИС по следующим признакам:

- по уровню функциональности и тесно связанной с ним степени интегрированности системы;
- по возможностям поддержки корпоративного управления;
- по степени реализации возможностей поддержки уровней управления – оперативного, тактического, стратегического.

Класс, к которому можно отнести экономическую ИС, во многом определяется масштабом предприятия, на котором она внедрена. Рассмотрим подробнее признаки классификации экономических ИС.

Уровень функциональности ИС. Наиболее простые ИС – **локальные**, реализующие отдельные функции управления (бухгалтерский учет, логистика и т.д.). Такие ИС применяются в настоящее время в основном на малых предприятиях, однако они вытесняются **многофункциональными** и **полнофункциональными** ИС, т.е. системами, в которых реализованы либо большинство, либо практически все функции управления.

Опыт показывает, что полнофункциональная ИС не может работать эффективно, не будучи интегрированной.

Интегрированная ИС (ИИС) основана на единой программно-аппаратной платформе и общей базе данных. В ИИС отдельные функциональные подсистемы (подсистемы управления персоналом, логистики, производства, бухгалтерского учета, управления финансами и

т.д.) взаимосвязаны на основе единого технологического процесса обработки информации.

Для функционирования ИИС необходимо организовать на предприятии локальную вычислительную сеть (ЛВС). Процессы обработки информации в ИИС базируются, как правило, на технологии клиент – сервер, т.е. распределены между сервером (программа, выполняющая функции управления и защиты базы данных) и клиентами.

Широкое распространение получили ИИС, ориентированные на автоматизацию процессов управления на малых предприятиях. Это **малые интегрированные информационные системы**. Характерные особенности ИИС малых предприятий – небольшое число рабочих мест (не более 5–10), отсутствие средств поддержки корпоративного управления. К достоинствам малых ИИС можно отнести их универсальность, обуславливающую небольшой цикл внедрения. В большинстве случаев разработчики снабжают такие системы простыми инструментальными средствами, позволяющими запрограммировать необходимые пользователю приложения. В результате получившаяся система иногда становится мало похожей на исходный прототип.

Наиболее известный представитель этого класса ИИС – «1С: Предприятие» (программные продукты компании «1С» внедрены на нескольких сотнях тысяч малых предприятий).

Возможность поддержки управления сложными структурами – корпорациями. Напомним, что корпорацией называют сложный хозяйствующий субъект, имеющий иерархическую структуру и включающий в себя предприятия самого различного масштаба (в том числе малые и средние) и профиля деятельности – производственные, транспортные, торговые, финансовые, учебные. Под это определение подпадают ИС большого диапазона: от ИС среднего предприятия, которое имеет находящиеся в пределах одного города цехи, склады, магазины и другие подразделения с той или иной степенью самостоятельности, до ИС транснациональных корпораций.

К **корпоративным** можно отнести **средние и крупные ИИС**. Таким образом, оба этих класса систем следует рассматривать как интегрированные корпоративные информационные системы (ИКИС). Такие системы должны быть, безусловно, полнофункциональными, но, кроме того, обладать средствами поддержки корпоративного управления. Средние ИКИС (из отечественных к ним можно отнести, например, систему «БЭСТ») имеют такие средства поддержки корпоративного управления, как возможность ведения консолидированной (совместной) базы данных корпорации, получения консолидированной отчетности по любому виду хозяйственной деятельности.

Крупные ИКИС (из отечественных наибольшую известность приобрели системы «Галактика», версии 7 и 8, и «Парус 8») помимо перечисленных средств поддержки корпоративного управления обладают большей глубиной поддержки процессов управления многофункциональными группами предприятий. В первую очередь здесь следует отметить средства корпоративного анализа с использованием технологии интерактивной аналитической обработки данных.

Отметим, что ИКИС – сравнительно новый класс систем; поэтому применяемая в этой сфере терминология пока еще не устоялась. Например, некоторыми авторами термины «корпоративная» и «интегрированная» используются фактически в качестве синонимов.

Поддержка управления корпорацией на различных уровнях. Выделяют следующие три основных уровня управления и соответствующие им системы:

- оперативный уровень (системы обработки данных/транзакций (СОД));
- тактический уровень (информационные системы управления (ИСУ));
- стратегический уровень (системы поддержки принятия решений (СППР)).

СОД предназначены для решения задач управления бизнес-процессами предприятия на оперативном уровне.

Главная функция СОД – регистрация в базе данных и обработка элементарных событий, сопутствующих протеканию бизнес-процессов: прием и увольнение сотрудников, приход и расход материальных ценностей на складах и в производственных подразделениях, оплата материальных ценностей и оказанных предприятию услуг через банк, ведение табеля учета рабочего времени и т.д. В результате функционирования СОД генерируются стандартные документы – платежные поручения, счета, расходные и приходные ордера и т.п. Задачи, решаемые СОД, носят повторяющийся, регулярный характер, а их временные рамки, как правило, не превышают одного дня.

Основная задача, стоящая перед ИС оперативного уровня, – обеспечить высокую скорость прохождения информационных потоков, связывающих участников бизнес-процессов. При необходимости СОД синхронизируют с ИС управления технологическими процессами на предприятии.

Задачи СОД решаются непосредственными исполнителями бизнес-процессов – кладовщиками, рабочими, кассирами, бухгалтерами, администраторами торговых залов – параллельно с выполнением ос-

новых обязанностей. От них требуется элементарная компьютерная грамотность.

СОД позволяет сделать прозрачным контроль за исполнением бизнес-процессов, поскольку в каждый момент времени можно оперативно выяснить состояние производственного процесса на всех рабочих местах.

На основе накопленной базы оперативных данных решаются задачи более высоких уровней управления – тактического и стратегического.

ИС управления служат для решения задач управления бизнес-процессами предприятия на тактическом уровне, к которому относятся процедуры среднесрочного (от нескольких дней до нескольких недель) планирования, анализа и организации работ. Если на оперативном уровне мы имеем дело с отдельным заказом и сопутствующими его выполнению транзакциями, то на тактическом уровне рассматриваются уже такие объекты, как, например, свод заказов для формирования производственной программы. Результаты решения подобных задач предназначены для менеджеров среднего звена – начальников производственных и транспортных цехов, руководителей служб снабжения и маркетинга, планово-финансовых отделов и т.д. Во многих случаях непосредственными исполнителями работ, связанных с функционированием ИСУ, являются высококвалифицированные сотрудники информационно-компьютерных подразделений предприятия.

Системы поддержки принятия решений предназначены для решения задач управления бизнес-процессами предприятия на стратегическом уровне, т.е. на уровне топ-менеджеров (руководства) фирм, предприятий, организаций, принимающих стратегические долгосрочные решения. На стратегическом уровне рассматриваются вопросы выпуска и продвижения на рынок новой продукции, поиска новых рынков сбыта, выбора источников финансирования, привлечения инвесторов, инжиниринга бизнес-процессов.

Эти решения определяют основные направления деятельности предприятий на длительные периоды (год и более). Таким образом, СППР являются рабочим инструментом для лиц, принимающих решения (ЛПР).

Для задач СППР характерны трудности формализации целей и ограничений, что обуславливает преобладание качественных оценок. Имеющаяся информация обычно неполна, нечетка и противоречива; поэтому для решения задач стратегического планирования применяются различные методы статистического анализа, экспертных систем, математического и имитационного моделирования, интеллектуального анализа данных.

СППР реализуются обычно в виде диалоговых человеко-машинных систем.

2.2.2. Базовая концепция ИКИС «Галактика»

Система «Галактика» модульная, каждый модуль взаимосвязан с другими компонентами, но самодостаточен, может работать автономно; поэтому продукт обычно внедряется постепенно, в связи с возникновением потребностей в функционале у заказчика. Модули объединены в контуры (рис. 2.3).

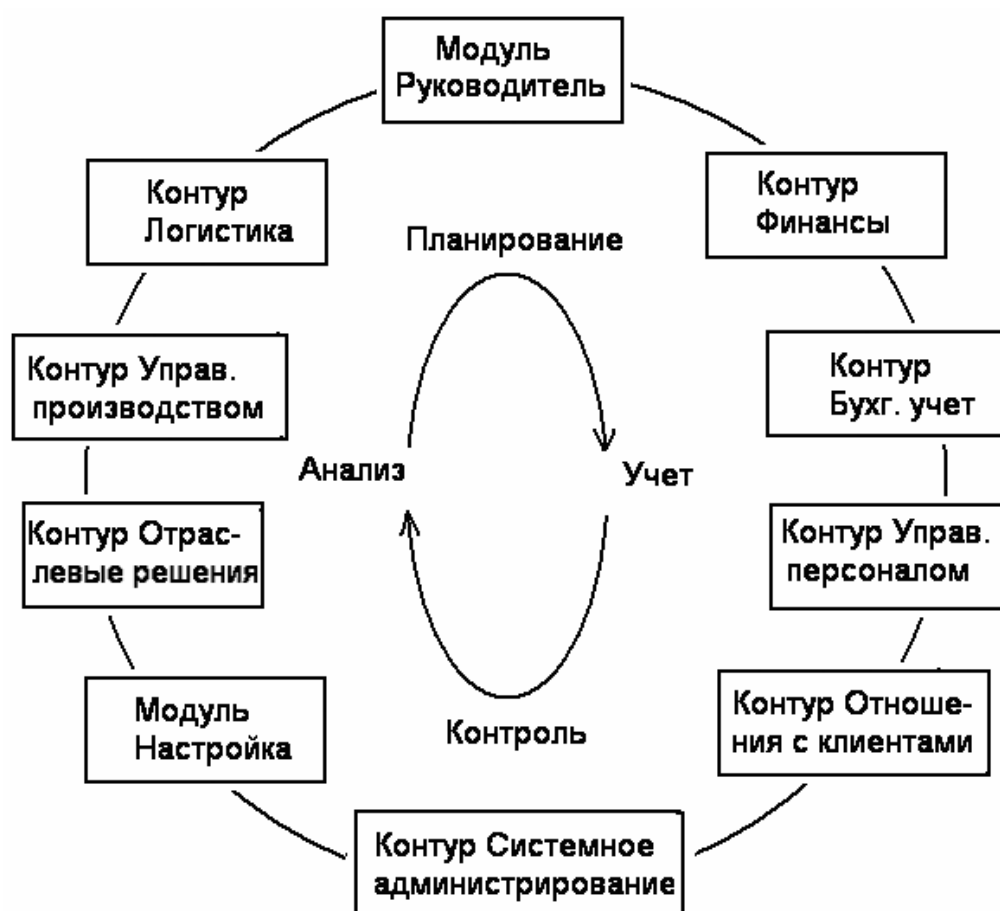


Рис. 2.3

Важнейшими принципами, лежащими в основе системы «Галактика» и определяющими ее высокую эффективность, являются:

- реализация «петли управления» (прогноз – планирование – контролируемая деятельность по выполнению планов – анализ результатов – коррекция прогнозов и планов) для всех сфер деятельности предприятия;
- использование технологии интерактивной аналитической обработки данных;

- обеспечение информационной поддержки принятия решений на разных уровнях вплоть до уровня топ-менеджеров предприятия;
- соответствие функциональности, технологичности и степени интеграции системы современным концепциям ERP, CSRP (Custom Synchronized Resource Planning – планирование ресурсов, синхронизированное с покупателем), SEM (Strategic Enterprise Management – стратегическое управление предприятием);
- сочетание высокой степени интеграции с гибкостью – система строится в виде набора взаимосвязанных, но относительно независимых компонентов, которые приобретаются в составе и порядке, удобном конкретному клиенту;
- обеспечение настройки на конкретную отрасль, регион, особенности бизнеса;
- отслеживание и оперативное включение в обновление системы всех изменений в российском законодательстве.

Основными объектами системы являются операционные документы, формируемые при проведении любой хозяйственной операции и образующие в совокупности документооборот предприятия.

Можно выделить четыре основных класса документов системы:

- документы-основания (ДО), регламентирующие операции как между юридическими лицами, так и внутри предприятия между его подразделениями, например договоры, счета, лимитно-заборные карты, требования;
- сопроводительные документы (товарные и финансовые), отражающие важнейшие стороны выполняемых операций;
- планы, отражающие намеченные показатели хозяйственной деятельности;
- отчеты, отражающие фактически достигнутые предприятием показатели хозяйственной деятельности.

Все перечисленные виды документов формируются в электронном виде и распечатываются по мере необходимости.

По всем первичным документам могут быть сформированы бухгалтерские проводки и финансовые операции с помощью механизма типовых хозяйственных операций (ТХО) и типовых финансовых операций (ТФО).

В системе интегрированы десятки (часто сотни) рабочих мест, размещенных во всех подразделениях предприятия. По мере создания операционных документов пользователями системы происходит накопление информации в единой базе данных предприятия.

Использование единой базы данных обеспечивает возможность однократного ввода информации, что устраняет дублирование работы поль-

зователей и снижает трудозатраты. Возможность однократного ввода информации в том месте, где она возникает, – важнейшее преимущество ИКИС, и оно реализовано в системе «Галактика» в полной мере.

Система позволяет осуществлять контроль корректности и целостности данных. Важно, что можно точно определить, кто из пользователей, когда и какие операции выполнял (персонификация действий пользователя).

Система «Галактика» – тиражно-заказной продукт. Это означает, что наиболее соответствующее потребностям пользователя отраслевое тиражное решение «доводится» в процессе совместной работы сотрудников предприятия-заказчика и работников одного из региональных представительств корпорации. Как правило, для осуществления такого рода «доводки» достаточно использовать настройки общесистемные и настройки модулей. Однако имеется возможность доработки программного кода модулей с помощью простых в освоении инструментальных средств.

Проведение настроек пользователя позволяет определить специфику работы каждого конкретного пользователя. Настройки прав доступа пользователей к документам устанавливаются администратором системы.

Все возможности, обеспечиваемые системой, независимы от программно-аппаратной платформы. Система «Галактика» может работать в среде самых различных операционных систем: Novell NetWare, Windows NT, Windows 2000. В процессе инсталляции системы можно выбрать в качестве рабочей одну из СУБД – Pervasive SQL (Btrieve), Oracle, MS SQL, обеспечивающих эффективное использование архитектуры клиент – сервер.

2.3. Основные виды обеспечения АИС

2.3.1. Информационное обеспечение АИС

Под информационным обеспечением АИС понимается система реализованных решений по объемам, размещению и формам организации информации, циркулирующей в АИС при ее функционировании. Специфическими формами организации информации в АИС являются:

- база данных (БД) – поименованная, целостная, единая система данных, организованная по определенным правилам, которые предусматривают общие принципы описания, хранения и обработки данных;
- база знаний (БЗ) – формализованная система сведений о некоторой предметной области, содержащая данные о свойствах объектов,

закономерностях процессов и явлений и правила использования в задаваемых ситуациях этих данных для принятия новых решений.

В БЗ центральным понятием является представление знаний в ИС, т.е. формализация метапроцедур, используемых человеком при решении интеллектуальных задач.

Базы данных. В современных АИС все средства обработки данных организовываются в виде системы управления базами данных (СУБД) – совокупности программных и языковых средств, предназначенных для управления данными в БД, ведения БД, обеспечения многопользовательского доступа к данным.

Основные сведения по БД изучаются в курсе «Управление данными», однако следует рассмотреть некоторые дополнения. Прежде всего отметим, что широко используется понятие автоматизированных банков данных (АБД) как совокупности СУБД и конкретных БД, находящихся под ее управлением, а также хранилищ данных, которые работают по принципу центрального склада. Хранилища данных отличаются от традиционных БД тем, что они проектируются для поддержки процессов принятия решений, а не просто для эффективного сбора и обработки данных. Как правило, хранилище содержит многолетние версии обычной БД, физически размещаемые в той же самой базе. Данные в хранилище не обновляются на основании отдельных запросов пользователей. Вместо этого вся БД периодически обновляется целиком. Хранилища данных могут быть внушительных размеров.

Проблемы накопления, хранения и использования информационных ресурсов резко обостряются при росте организации и усложнении ее структуры.

Ниже перечислены основные характеристики хранилищ данных.

1. Данные организованы по предметным областям.
2. Данные, поступающие в хранилище из оперативных приложений, приводятся к некоторому общему формату, т.е. интегрируются.
3. Информация в хранилище данных стабильна.
4. Информация отражает историю изменения данных.

Базы знаний. Происхождение понятия «база знаний» связано с системами искусственного интеллекта. В этих системах базы знаний генерируются для экспертов и систем, основанных на знаниях. В таких системах компьютеры используют правила вывода для получения ответов на вопросы пользователя.

Основные сведения по БЗ изучаются в соответствующем курсе, однако сделаем некоторые дополнения.

Ресурсы знаний различаются в зависимости от отраслей индустрии и приложений, но, как правило, они включают руководства, письма,

новости, информацию о заказчике, сведения о конкурентах и данные, накопившиеся в процессе работы. При этом используются разнообразные технологии: электронная почта; базы и хранилища данных; системы групповой поддержки; браузеры и системы поиска; корпоративные сети и Internet; экспертные системы и системы баз знаний; интеллектуальные системы.

Если хранилища данных содержат в основном количественные данные, то хранилища знаний ориентированы в большей степени на качественные данные. Знания генерируются из широкого диапазона БД, хранилищ данных, рабочих процессов, статей новостей, внешних баз, Web-страниц (как внешних, так и внутренних), и, конечно, люди представляют свою информацию. Хранилища знаний обычно распределены по большому количеству серверов. Базы знаний могут содержать руководства и правила проектирования, спецификации и требования. Другим распространенным приложением является база знаний кадровых ресурсов, содержащая данные о квалификации и профессиональных навыках сотрудников, об уровне их образования, перечень специальностей, сведения об опыте работы и т.д. В связи с бурным развитием Internet все чаще используют Web-технологии.

2.3.2. Средства обработки данных

В сфере розничной торговли, в банковской системе, на рынке телекоммуникаций, в инвестиционных и страховых компаниях, в государственных структурах и во многих других областях деятельности успех организации часто напрямую зависит от того, насколько полную информацию она имеет о своей клиентуре и состоянии рынка и насколько эффективно способна использовать эту информацию для принятия решений. Поэтому системы поддержки принятия решений (СППР) перестают быть роскошью.

Системы поддержки принятия решений представляют собой совокупность интеллектуальных информационных приложений и инструментальных средств, которые используются для манипулирования данными, их анализа и предоставления результатов такого анализа конечному пользователю.

Возможности и преимущества СППР можно показать на примере телекоммуникационного рынка, который традиционно считается одним из самых высокотехнологичных и динамичных.

СППР позволяет создавать центры по обслуживанию клиентов, которые будут реагировать на обращения заказчиков в режиме, приближенном к реальному времени. Здесь всё решают концентрация информации о клиенте в одном месте и обеспечение быстрого доступа ко

всем необходимым данным о сделанных вызовах за определенный период времени. Такая оперативная поддержка не только позволит добиться высокого уровня обслуживания, но и может принести компании реальный доход. Обслуживать старых клиентов всегда гораздо экономичнее, чем строить взаимоотношения с новыми. Текучесть заказчиков – одна из самых острых проблем телекоммуникационной индустрии. Как раз хранилище данных и аналитические приложения могут помочь ее решить. Современные средства моделирования анализируют поведение клиентов: подсоединение к сети оператора связи, отключение от нее, смена провайдера. Вкупе с программами анализа «текучести» эти средства позволяют выявлять группы высокого риска и предотвращать потерю клиентов.

Выделив определенные группы клиентов с теми или иными потребностями, компания может проводить целевой маркетинг новых услуг и тем самым добиваться быстрого получения прибыли. Например, оператор связи собирается ввести новые тарифы. С помощью маркетинга базы данных можно выбрать группу клиентов, для которой эти тарифы будут наиболее приемлемы. Система поддержки принятия решений позволит определять результаты использования новых тарифов для выбранной группы практически в режиме реального времени.

В конечном счёте, объединив в хранилище все данные о рынке, клиентах, сделанных вызовах и внутренних ресурсах и построив на его основе эффективную среду поддержки бизнеса, телекоммуникационная компания может по-новому организовать основные направления своей деятельности. Единое хранилище информации позволит полностью интегрировать все функции и создать модель работы компании: единый вызов, единый счет, единый взгляд на клиента.

2.3.3. Интерфейсы пользователя

АИС представляет собой человеко-машинную систему, и это обстоятельство влечет за собой постановку вопроса об исключительной важности проблем общения между человеком-пользователем и программно-техническим комплексом АИС, которые решаются с помощью пользовательского интерфейса. **Интерфейс пользователя – это программно-технические средства, которые обеспечивают взаимодействие пользователя с АИС.** В АИС рассматриваются и другие интерфейсы:

- межкомпонентные интерфейсы;
- интерфейсы между различными АИС;
- интерфейсы в телекоммуникационных сетях.

Но основное внимание уделяется пользовательскому интерфейсу. Интерес к пользовательскому интерфейсу стал возникать после того, как человек-пользователь основную часть своего рабочего времени стал проводить на АРМе – автоматизированном рабочем месте. Это порождает массу проблем: психологических, физиологических, социально-экономических и т.д.

Внедрение АИС может нарушить сложившиеся рабочие отношения в производственном подразделении. Эти социальные факторы могут усилить или ослабить опасения пользователя относительно системы. Поэтому одним из важнейших принципов создания АИС является привлечение пользователя к процессу разработки интерфейса на всех этапах. В этом случае пользователь относится к системе без предубеждений, так как основные характеристики реальной системы согласовываются с ним.

Конструктивные особенности оборудования и размещение его в рамках АРМа влияют на чувство физического комфорта пользователя при работе с системой. Необходимо обращать внимание на следующие эргономические характеристики:

- конструктивные особенности оборудования;
- качество разработки диалога;
- доступность и надежность компонент;
- чувствительность компонент.

Психологическая эргономика – это анализ соответствия функций системы психологическим процессам человека. Например, пользователь, который не может физически различить символы из-за бликов или плохого контраста, ощутит физический дискомфорт. Пользователь, который может прочесть текст, но не может понять его смысл из-за непонятных слов или непривычной формы представления текста, ощутит психологический дискомфорт. Нет смысла прилагать усилия для создания условий, когда у пользователя не будет болеть спина и не будут уставать глаза, если от сообщений системы его голова пойдет кругом.

С понятием психологической эргономики тесно связаны еще два фактора, хотя они важны и сами по себе – это доступность и чувствительность системы.

В принципе, в разумных пределах, пользователь должен получить доступ к системе в то время, когда это ему нужно. Число АРМов должно быть достаточным для обслуживания всех потенциальных пользователей. Обычно нужно иметь на каждом рабочем месте одну рабочую станцию.

Такой же отрицательный эффект, как невозможность получить доступ к системе, имеет и длительное ожидание ответа на запрос в течение 20 с и более. Еще хуже, если в какие-то дни ответ выводится через 2 с, а в другие приходится ждать 20 с – разное время реакции системы наводит пользователя на мысль, что система испортилась. Обеспечение приемлемого времени реакции системы – это один из самых технически важных и дорогостоящих аспектов разработки интерактивных систем.

Таким образом, для обеспечения эффективной работы пользователя нужно учитывать его эмоциональные, психологические и физиологические особенности. Если пользователь расстроен, раздражен или подавлен, он не сможет работать хорошо. Компонентом АИС, который может вызвать или, наоборот, снять стресс, является интерфейс человек – компьютер, т.е. среда, через которую пользователь взаимодействует с системой.

Пользователь работает в рамках определенных ограничений и допусков. Его глазам нужно, чтобы образы имели определенный размер, уровень яркости, контрастности и располагались на удобном расстоянии. Некоторые цвета воспринимаются легче других, наложение одних цветов помогает различать их, а наложение других – мешает. Суставы ограничены в движениях, подвижность рук тоже ограничена. Если нужно сохранять определенную позу длительное время, то нужна опора и т.д. Эти физические ограничения хорошо изучены и должны приниматься во внимание при разработке АИС.

Разработчик интерфейса, хорошо знакомый с организацией работ, должен уметь учитывать требования к удобству в работе совершенно незнакомых ему пользователей. Он должен не только удовлетворить их требования с точки зрения вычислительных задач, но и создать интерфейс, удобный для пользователя.

Интерфейс пользователя в содержательном плане является частью лингвистического обеспечения АИС – совокупности языковых средств для формализации естественного языка, построения и сочетания информационных символов при общении пользователей с АИС. При этом формализованные информационные языки (ФИЯ) – это языки, созданные на базе естественных языков путем наложения ограничений на их лексику и грамматику, а также путем применения специальных обозначений для элементов этих языков. ФИЯ предназначены для описания и обработки данных в АИС, в их число входят:

- входные языки АИС – языки, которые служат для составления предписаний на ввод, обновление, поиск, обобщение, редактирование и выдачу информации;

- языки описания данных (ЯОД) – формальные языки со специальными конструкциями, которые используются для описания схем баз данных;
- нормированные языки – строго определенное подмножество естественного языка, предложения которого, по крайней мере по отношению предметной области, семантически однозначны;
- язык манипулирования данными (ЯМД) – язык, предназначенный для формулирования запросов на поиск, обмен данными между прикладной программой и базой данных, а также для расширения языка программирования либо как самостоятельный язык;
- язык запросов (справочный язык) – язык, обеспечивающий взаимодействие конечного пользователя с информационной системой.

Основные требования к пользовательскому интерфейсу

1. Интерфейс должен выделяться в качестве отдельного компонента системы. Так же как структуры данных в БД АИС изолируют от прикладных программ обработки этих данных, интерфейс человек – компьютер – до определенной степени – необходимо отделить от прикладной задачи.

2. Разработчики интерфейса должны тщательно анализировать возможности и ограничения аппаратных и программных средств.

3. В процессе разработки требуется новизна, упрощающая работу пользователя.

4. Существует большое число общепринятых в эргономике рекомендаций, которые нужно использовать при разработке и организации АРМов в АИС. В то же время форма представления информации на экране не одинакова для различных систем: графический дизайн зависит от распределения информации на экране, словарного состава предложений, способа выделения ключевых элементов представления данных и т.д.

5. Разработчик интерфейса должен понимать и оценивать действия пользователя, направленные на достижение цели решения задачи, а также особенности потенциальных пользователей системы.

6. Немного найдется разработчиков, хорошо знакомых со всеми областями применения своих разработок или глубоко чувствующих психологические особенности потенциальных пользователей.

Существуют принципы, которым нужно следовать. Один из типичных принципов заключается в том, что аналитик получает сведения путем опроса будущих пользователей. Полезный способ подбора

нужных вопросов – это поставить себя на место пользователя, работающего с системой.

И все-таки лучший способ оценки доступности интерфейса – это посмотреть, как на самом деле пользователь взаимодействует с системой в нормальных рабочих условиях.

7. Необходимо предусматривать средства адаптации в рамках интерфейса. Хотя общие принципы определяют основу создания интерфейса, они не могут удовлетворять любого пользователя. Даже если условия задачи остаются постоянными, потребности пользователей, как и они сами, меняются.

Правильно спроектированный интерфейс должен быть настраиваемым на нужды разных пользователей, а также на одного пользователя в разные периоды его работы.

Критерии оценки интерфейса охватывают три основных направления:

- простота освоения и запоминания операций системы;
- быстрота достижения целей задачи, решаемой с помощью системы;
- субъективная удовлетворенность при эксплуатации системы.

2.3.4. Классификация программного обеспечения АИС

Под программным обеспечением АИС – функциональной компонентой – понимают совокупность общесистемных и функциональных программ, а также программ системы обработки данных и программных документов, необходимых для эксплуатации этих программ. Программы, обеспечивающие возможность выполнения АИС основных функций, практически не зависящих от специфики конкретных задач, входят в состав программного обеспечения общесистемного назначения. Тогда прикладное программное обеспечение АИС – это программное обеспечение, предназначенное для решения определенной задачи в предметной области или для предоставления пользователю определенных услуг.

Иногда используется термин «программная система» – программное средство (ПС), т.е. программная продукция, представляющая собой совокупность программ и (или) подсистем, имеющих общее целевое назначение. ПС, оформленное в соответствии с определенными требованиями, называется программным изделием – программой на носителе данных, являющейся продуктом промышленного производства.

На рис. 2.4 представлен состав программного обеспечения АИС; компоненты которого классифицированы согласно функциям, выполняемым ими на различных этапах полного жизненного цикла АИС. В

целом, при изложении данного курса мы последовательно придерживаемся концепции полного жизненного цикла АИС; поэтому и классификация ПО АИС основывается на этой концепции.

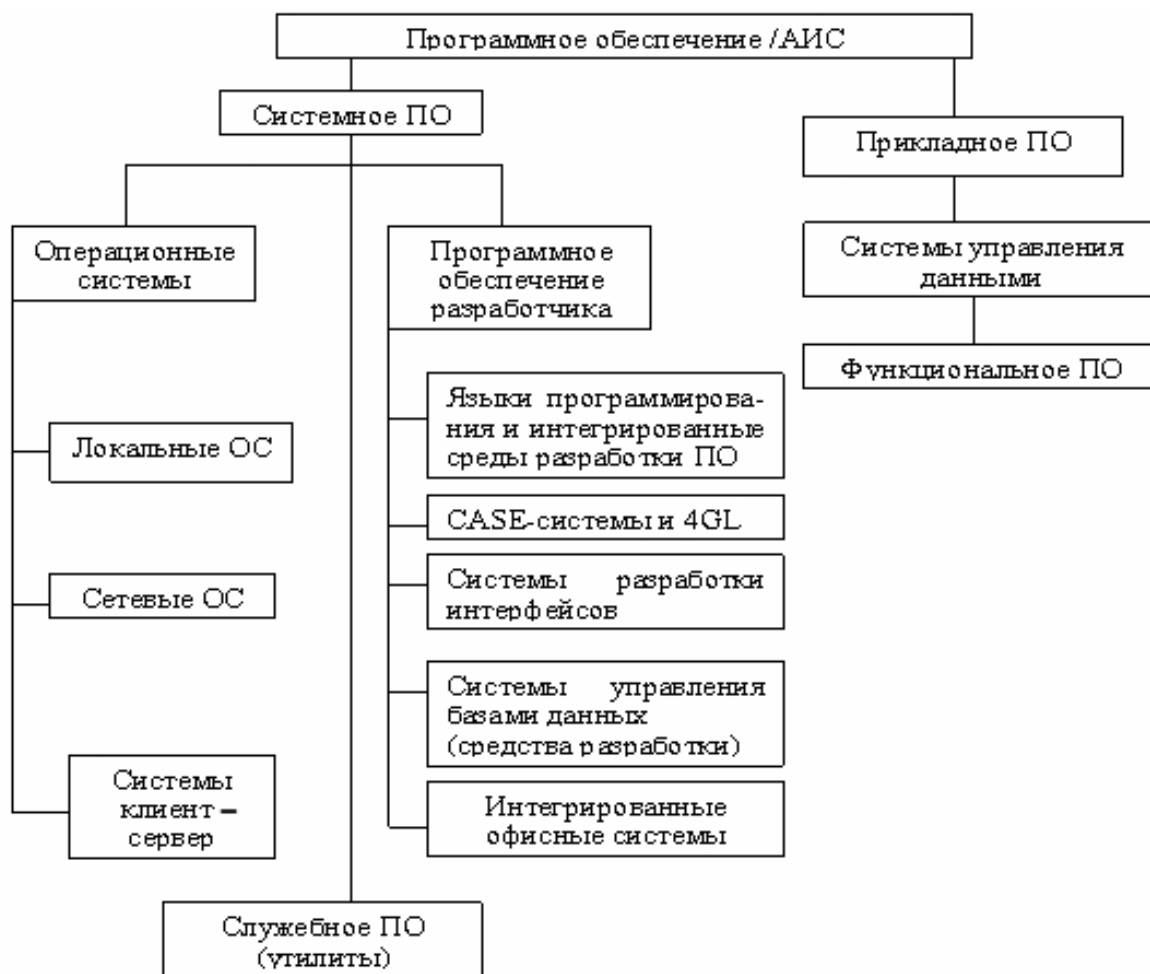


Рис. 2.4

Базой ПО АИС, услугами которой пользуются все остальные компоненты, является операционная система (ОС) – система программ, предназначенная для обеспечения определенного уровня эффективности вычислительной системы за счет автоматизированного управления ее работой и предоставляемых пользователям определенного набора услуг. В группу базового ПО кроме ОС входят программные средства служебного назначения, которые выполняют частные общесистемные функции:

- системы управления базами данных (СУБД) – совокупность программ и языковых средств, предназначенных для управления данными в БД, ведения БД и обеспечения взаимодействия ее с прикладными программами;

- служебные программы (утилиты) – набор специальных программ, предназначенных для повышения эффективности рутинных работ по обслуживанию информационно-вычислительной среды, например копирования и сжатия данных, дефрагментация жесткого диска и т.д.

На этапах концептуального, технического и рабочего проектирования наиболее интенсивно используются программные средства разработки, которые в настоящее время оформляются в виде интегрированных сред – программных систем, включающих все необходимые пользователю средства и обеспечивающие их единообразие.

Можно выделить интегрированные среды следующего назначения:

- средства автоматизации проектирования и переноса АИС (CASE-технологии) – системы программного обеспечения, которые основываются на методологиях коллективной разработки и сопровождения АИС и обеспечивают автоматизацию всех этапов жизненного цикла АИС;

- интегрированная среда разработки программ (ИСП) – система программ, которая упрощает процесс программирования и делает его более эффективным. ИСП содержит весь комплекс средств, необходимых для написания, редактирования, компиляции, компоновки и отладки программ;

- интегрированная система пользователя – несколько взаимосвязанных пакетов прикладных программ, в том числе текстовый редактор, электронная таблица, база данных, деловая графика, средства коммуникации;

- инструментальная система – программно-техническая система, позволяющая пользователю, не владеющему языками программирования, создавать личные приложения и (или) их модели.

С учетом быстрого развития в современных АИС средств телеобработки данных – совокупности методов, обеспечивающих пользователям дистанционный доступ к ресурсам систем обработки данных и средств связи – целесообразно выделить в составе ПО АИС сетевые программные средства. Это средства разработки и реализации методов, обеспечивающих установление связи, обработку и передачу данных между различными ЭВМ или (и) удаленными или локальными терминалами и абонентскими пунктами.

В состав прикладного ПО АИС входят:

- функциональные прикладные программы, предназначенные для решения задачи или класса задач в определенной области применения систем обработки данных;

- интерфейсы пользователя – совокупность методов, правил, программных и аппаратных средств, обеспечивающих взаимодействие пользователей с программами или другими видами ресурсов АИС.

Работы по всему жизненному циклу выполняются на автоматизированных рабочих местах (АРМах) – программно-технических системах, обеспечивающих возможности доступа пользователя к средствам разработки и ресурсам автоматизированной информационной системы.

2.3.5. Операционные системы

Рассмотрим основные особенности ОС применительно к АИС.

В современных АИС используются следующие виды ОС:

- локальные ОС, обеспечивающие управление ресурсами в пределах конкретного вычислительного комплекса, без использования каналов связи;
- сетевые ОС, осуществляющие управление ресурсами в локальных и глобальных сетях передачи данных.

Могут использоваться специальные ОС, например:

- ОС автоматизированного проектирования – часть программного обеспечения автоматизированного проектирования, предназначенная для управления проектированием;
- ОС реального времени – ОС, обеспечивающая режим работы ЭВМ в реальном времени, и др.

Система реального времени – это система, осуществляющая информационный обмен с другими системами, периферийными устройствами или датчиками при таких временных характеристиках, которые позволяют немедленно обрабатывать всю поступающую информацию и информацию для вывода. Эти системы обычно используются в управлении техническими объектами (АСУ ТП).

Значительная доля современных АИС имеет распределенный характер (рис. 2.5), то есть их компоненты распределены по различным вычислительным комплексам, которые связаны между собой каналами передачи данных.

Распределенные АИС базируются на сетях ЭВМ, предназначенных для организации распределенной обработки информации.

Для построения распределенных АИС особое значение имеет концепция открытых информационных сетей. Открытая информационная сеть – это сеть, в которой взаимодействие всех реальных систем подчиняется требованиям стандартов Международной организации по стандартизации.

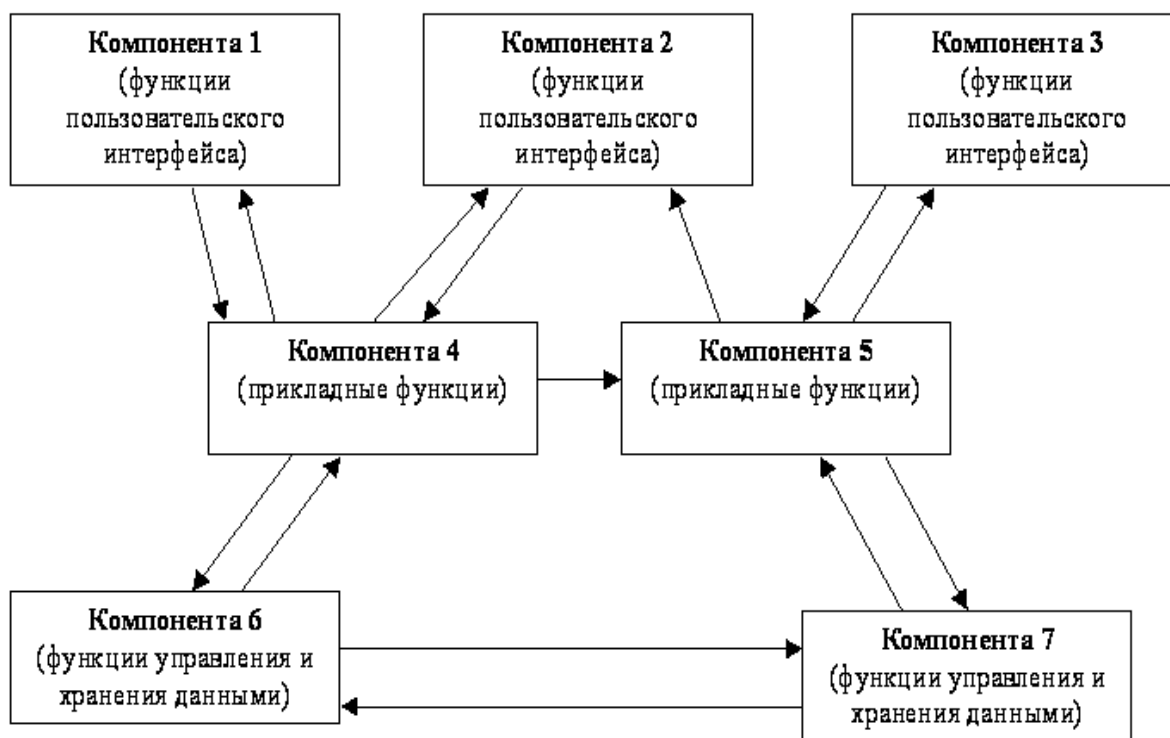


Рис. 2.5

Разновидностью открытых систем являются АИС, реализованные на базе корпоративных сетей типа «intranet», созданных на базе технологий Internet. В отличие от «Internet» «intranet» является чисто искусственным словообразованием, которое принято писать с маленькой буквы. Internet – имя собственное вполне конкретной, глобальной сети, а intranet – обозначение абстрактного понятия. В русском языке используется его перевод – «интрасеть».

Логика развития корпоративных компьютерных сетей привела к объединению локальных сетей, реализованных на уровне подразделения, в глобальные. В качестве примера можно представить себе сеть корпораций, имеющих сотни подразделений по всему миру. Или банк, имеющий несколько десятков отделений, в каждом из которых есть свои собственные локальные сети (а иногда и несколько). Использование идей Internet для создания достаточно больших корпоративных сетей является довольно очевидным: здесь то же самое разнообразие типов компьютеров, линий связи, а также довольно хаотичное изменение топологии. Причем актуальность данной проблемы проявляется не только на уровне крупных корпораций или банков, но и при создании общей сети даже не самого крупного завода или вуза.

Однако создание интрасетей на уровне прямого взаимодействия компьютеров через Internet является невозможным. Во-первых, это может быть просто нецелесообразно для относительно небольших сетей, в которых объединение компьютеров выполняется за счет чисто

внутренних линий (например, на территории завода). Во-вторых (и это самое главное), из-за необходимости обеспечения информационной защиты при взаимодействии с внешним миром. Ведь открытые протоколы Internet делают прозрачными все включенные в него узлы. Узел в принципе не в состоянии сам определить, с каким абонентом – внутренним или внешним – он имеет дело. Решением данной проблемы является физическое отделение внутренней корпоративной сети от внешнего мира, а общение между ними выполняется через специальные шлюзы, получившие название прокси серверов.

Специфика программных продуктов и технологий для интрасетей заключается в том, что число узлов в них является ограниченным и изменение топологии сети и содержимого информации носит некоторый предсказуемый характер. А в этом случае организация таких процедур, как маршрутизация и поиск информации, в отличие от хаотичного Internet может быть оптимизирована совершенно другими методами.

При создании АИС на базе интрасетей нужно решить следующие задачи:

- интегрировать внутренние локальные сети с Internet для расширения возможностей коммуникации предприятий с клиентами и партнерами;
- реализовать новые методы навигации, применяемые в Internet, во всех продуктах, с тем чтобы облегчить операции поиска и анализа информации, а также взаимодействия с партнерами;
- упростить разработку, внедрение и администрирование прикладных программ с целью оптимизации деловых процессов на предприятиях и сокращения циклов разработки;
- интегрировать новые продукты и Internet-технологии с существующими инфраструктурами, чтобы пользователи могли на основе уже имеющихся решений развивать свои информационные системы.

Современные сетевые ОС – это системы, обеспечивающие управление ресурсами и координацию функционирования компонентов АИС, распределенных в корпоративных или глобальных сетях передачи данных. В основе их работы лежит технология клиент – сервер (рис. 2.6). Клиент – это компьютер или программа, обеспечивающие доступ к ресурсам другого компьютера или программы, предоставленным в совместное использование, а сервер – компьютер или программа, предоставляющие свои ресурсы для совместного использования в сети.

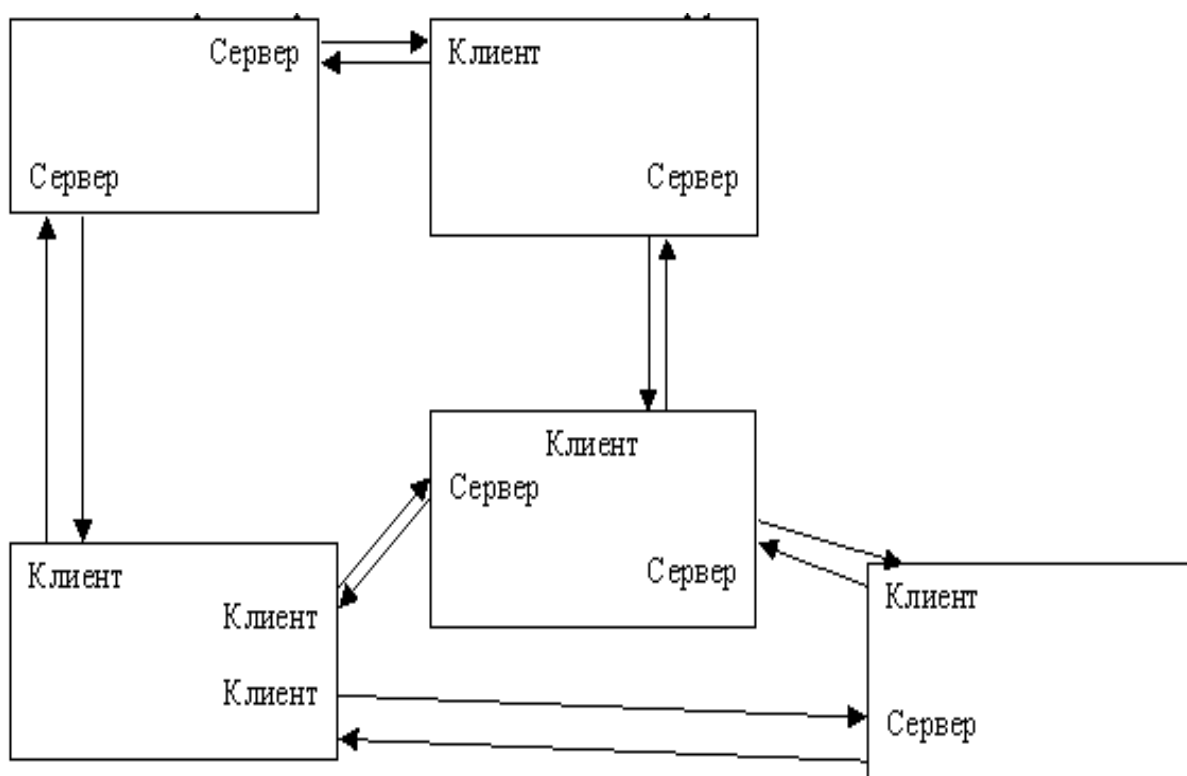


Рис. 2.6

Следует отметить, что многие современные сетевые ОС (ОС семейств Windows Server, Unix) могут выступать как:

- файл-сервер;
- сервер печати;
- сервер приложений;
- контроллер домена;
- сервер удаленного доступа;
- сервер Internet;
- сервер обеспечения безопасности данных;
- сервер резервирования данных;
- сервер связи;
- сервер вспомогательных служб.

В режиме файла-сервера ОС обеспечивает работу компьютера как централизованного хранилища большого количества файлов (баз данных), коллективно используемых в АИС, в том случае, когда хранение этой информации на локальном компьютере нецелесообразно. Все файловые ресурсы независимо от того, на каком диске они расположены (жестком или CD-ROM), сразу могут быть предоставлены для совместного использования.

Как сервер печати ОС позволяет работать и предоставлять в совместное пользование неограниченное число принтеров.

Режим сервера приложений используется в системах, способных исполнять основное («тяжелое») приложение на мощном высокопроизводительном сервере, а результаты передавать на маломощные клиентские АРМы.

В режиме сервера резервирования данных ОС обеспечивает возможность резервного копирования файлов на магнитную ленту. Администратор системы определяет пользователя, ответственного за эту операцию, и только он регулярно выполняет копирование данных. При необходимости эту операцию можно автоматизировать.

Служба удаленного доступа (Remote Access Service – RAS) состоит из двух частей: серверной, устанавливаемой на ЭВМ с сетевой ОС, и клиентской, устанавливаемой на АРМах с MS-DOS, Windows 98 или Windows 2000. Пользователь АРМа, связанного с сетью через сервер удаленного доступа, чувствует себя работающим непосредственно в сети: он может осуществлять доступ к файлам и данным, печатать документы, подключаться к другим серверам и обмениваться сообщениями по электронной почте.

Такой прозрачный доступ к сети удобен тем, кто постоянно бывает в разъездах, командировках, а также администраторам системы. Он широко применяется и для связи территориально удаленных филиалов предприятий; при этом одновременно поддерживается до 256 сессий удаленного доступа.

Режим сервера связи сетей подразумевает возможность соединения между собой различных сегментов сети, в том числе сопряжения разнородных сетей.

Диспетчер заданий позволяет просмотреть список исполняемых задач и все процессы, работающие в системе; в нем можно в наглядной графической форме следить за загруженностью процессора и использованием памяти. Эти возможности позволяют быстро обнаружить некорректно работающую задачу или процесс и, при необходимости, прервать его.

Программы-мастера администрирования позволяют:

- создавать новые учетные записи пользователей и групп;
- управлять предоставлением доступа к ресурсам сервера;
- создавать новые принтеры в системе и предоставлять их в совместное использование;
- конфигурировать модемы;
- устанавливать клиентское программное обеспечение;
- добавлять новые приложения или удалять существующие;
- отслеживать правильность соблюдения лицензионной политики.

Информационные сети, как и ОС, изучаются в специальных курсах.

2.3.6. Средства автоматизации проектирования АИС

В настоящее время чисто «ручные» методы проектирования АИС из-за их высокой трудоемкости и себестоимости вытесняются методами и средствами автоматизации проектирования.

Основная цель развития современных технологий проектирования АИС состоит в повышении экономической эффективности всего жизненного цикла АИС в различных проблемно-ориентированных областях. Это достигается снижением трудоемкости, ускорением и упрощением проектирования всей совокупности компонент, проведением комплексной автоматизации технологий обеспечения всего жизненного цикла АИС.

Совокупность методов и инструментальных средств автоматизации технологического процесса разработки сложных АИС объединяется под названием CASE (Computer Aided Software Engineering – автоматизированное проектирование программных средств).

CASE-технология регламентирует (рис. 2.7) порядок организации и проведения работ, неавтоматизированного и автоматизированного выполнения технологических операций, направленных на получение в имеющихся организационно-технических условиях готовой АИС с заданными функциями и качеством. Для этого разработан и активно используется ряд принципов:

- максимально возможное повышение уровня абстракции при описании компонент и действий над ними на различных этапах проектирования;
- сокрытие всей информации, избыточной для данного этапа или объекта проектирования;
- модульность и строгая иерархия в структурном построении программных и информационных компонент;
- унификация правил проектирования, структурного построения и взаимодействия компонент между собой и с внешней средой;
- поэтапный контроль полноты и качества решения функциональных задач.

Особенности предметной области отражаются в специальных алгоритмах прикладной части системы. Эффективность создания и функционирования прикладного ПО АИС зависит от качества ОС, СУБД и CASE-средств. Возможность использования готовых решений, отработанных в ранних версиях продукта или разработках других фирм, определяется степенью поддержки протоколов и интерфейсов с пользователями, с прикладными программами и стандартной документацией. Интегрированные CASE-средства служат для извлечения знаний из заказчика на этапе проведения обследования, а также для проектиро-

вания концептуальной и логической структур баз данных прикладной системы; кроме того, с их помощью осуществляются сопровождение и развитие проекта. Организация, внедряющая у себя CASE-технологию, должна произвести определенные затраты на внедрение соответствующей методологии и средств на обучение персонала.



Рис. 2.7

Основными функциями CASE-средств являются:

- объектно-ориентированное системное и логическое проектирование программных средств и баз данных;
- планирование и оценка затрат ресурсов на разработку программных средств и баз данных;
- стратегическое планирование и управление проектами на всем жизненном цикле АИС;
- анализ требований, структурное проектирование ПС и БД, разработка и применение спецификаций требований;
- организация и управление базами данных и хранилищами проектов;
- повторное использование отработанных программных компонент, а также перенос их на иные операционные и аппаратные платформы.

В приложении приведены фрагменты примера Автоматизированной информационной системы (АИС) учета материальных ценностей в строительной фирме, созданные в средах Case_Analyst, ERWin, Access 2002.

Языки четвертого поколения (4GL) являются средством разработки приложений. Они легки в изучении и уменьшают размер программного кода. В результате достигаются значительно меньшие затраты на программирование и возросшая более чем в 10 раз производительность, по сравнению с языками низкого уровня. Языки 4GL также облегчают программистам следующие задачи:

- поддержку и модификацию программных приложений;
- написание более строгой программной документации;
- осуществление быстрого макетирования приложений;
- управление проектированием программных средств.

Многие 4GL предоставляют возможность решать самостоятельно некоторые задачи, например, написание запросов и отчетов БД лицам, не являющимся профессиональными программистами.

Операция, требующая сотни команд кода 3GL, может быть выполнена с помощью нескольких простых предложений на 4GL. Эти предложения затем транслируются в соответствующий код низкого уровня. Под термином 4GL подразумеваются как конкретные языки программирования, так и средства для разработки программного обеспечения, поставляемые вместе с языками. К этим средствам относятся:

- отладчики;
- редакторы текстов и баз данных;
- генераторы приложений и меню;
- средства создания форм и отчетов;
- средства форматирования экрана;
- системы презентации.

Компонентная структура CASE-системы представлена на рис. 2.8. Здесь 1 – навигатор объектов, 2 – матричный диаграммер, 3 – средства администрирования, 4 – средства моделирования процессов, 5 – диаграммер ER моделей, 6 – диаграммер иерархии функций, 7 – диаграммер потоков данных, 8 – диаграммер структуры приложения, 9 – навигатор параметров, 10 – навигатор процедурной логики, 11 – диаграммер баз данных, 12 – диаграммер модулей, 13 – генератор сервера, 14 – генератор форм, 15 – генератор отчетов. Под диаграммерами здесь понимаются средства построения диаграмм, а под навигаторами – средства поиска.

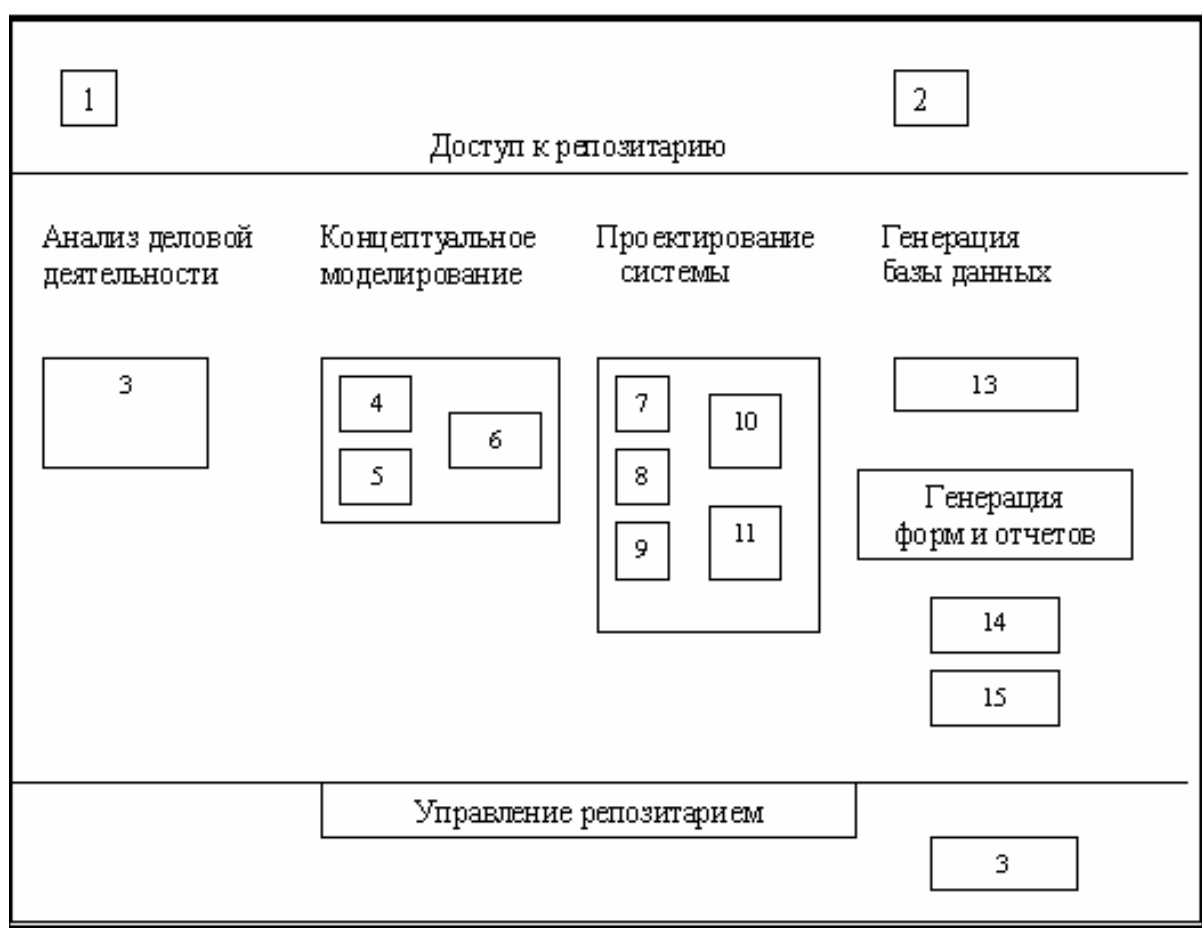


Рис 2.8

2.3.7. Программное обеспечение интерфейсов АИС

По мере роста мощности компьютеров увеличиваются затраты на диалоговую компоненту программного обеспечения. Термин «эффективность» постепенно изменил свое значение; если раньше он отражал такие характеристики, как процессорное время и объем занимаемой памяти, то теперь под ним понимают простоту разработки, легкость сопровождения и удобство работы с программой. Поэтому затраты на разработку пользовательского интерфейса являются оправданными. Большинство современных пользовательских интерфейсов основывается на аналогичных идеях:

- активное использование «мыши»;
- ориентированность на объекты;
- графика и имитация процессов и явлений;
- возможность использования алгоритмов, знакомых человеку из его опыта.

Следовательно, необходимо вспомогательное программное обеспечение, предназначенное для создания стандартных интерфейсов. Одна-

ко разнообразие аппаратных и системных платформ, на которых будет работать это программное обеспечение, требует его переносимости на уровне исходного кода. Рассмотрим пример такой реализации – широко известную X Window.

X Window (далее X) – это система для создания графического пользовательского интерфейса, изначально – на компьютерах, работающих под управлением ОС UNIX. Она построена по схеме клиент – сервер. Взаимодействие X-клиента и X-сервера происходит в рамках соответствующего протокола прикладного уровня. В качестве транспортного протокола может служить как локальный, так и любой сетевой, например TCP. Это означает, что различия в архитектуре X-клиента и X-сервера не играют никакой роли и их совместимость обеспечивается стандартом X-протокола.

Система обеспечивает графический вывод на экран, воспринимает сигналы от устройств ввода (клавиатура и мышь) и передает их программам. Устройство вывода может иметь более одного экрана и X обеспечивает вывод на любой из них. Всё это: экран (экраны), устройства ввода – называется в терминах X Window дисплеем.

Благодаря своей архитектуре X Window свободно используется в распределенных вычислительных системах, например в сетях TCP/IP (Internet), и позволяет пользователю общаться со многими программами одновременно. Система создаёт на экране дисплея «виртуальные подэкраны» – окна. Каждое приложение (как правило) рисует только в своём окне (или своих окнах). X предоставляет набор средств для создания окон и работы с ними.

X функционирует согласно идеологии управляемости событиями – она организует общение между самими программами и между программами и внешней средой посредством событий. Событие есть единица информации, идентифицирующая происходящие в системе изменения или действия. По идентификатору события можно получить информацию о нём: вид события, его характеристики, где оно произошло и т.п.

X Window представляет собой совокупность программ и библиотек. Ядром ее является специальная программа – X-сервер. Именно сервер знает особенности конкретной аппаратуры, знает, что надо предпринять, чтобы вывести какой-либо графический объект на экран. Он умеет воспринимать и обрабатывать сигналы, приходящие от клавиатуры и мыши. Сервер общается с программами-клиентами, посылая им или принимая от них пакеты данных.

Если сервер и клиент находятся на разных машинах, то данные посылаются по сети, а если на одном, – используется внутренний канал. Например, если сервер обнаруживает, что пользователь нажал кнопку

мыши, то он подготавливает соответствующий пакет (событие) и посылает его тому клиенту, в чьём окне находился курсор мыши в момент нажатия кнопки. И наоборот, если программе надо вывести что-либо на экран дисплея, она создаёт необходимый пакет данных и посылает его серверу. Описание этого взаимодействия, форматов пакетов и т.п. и составляет спецификацию вышеупомянутого X-протокола.

Чтобы программировать для X, не нужно знать детали реализации сервера и X-протокола. Система предоставляет стандартную библиотеку процедур, осуществляющих доступ программ к услугам X на высоком уровне. Так, для того чтобы вывести на экран точку, достаточно вызвать соответствующую стандартную процедуру, передав ей требуемые параметры. Эта процедура выполнит всю работу по формированию пакетов данных и передаче их серверу.

Окно – это базовое понятие в X. Это прямоугольная область на экране, предоставляемая системой программе-клиенту. Последняя использует окно для вывода графической информации. Окно имеет внутренность и край. Основными атрибутами окна являются ширина и высота внутренности, а также ширина (толщина) края. Эти параметры называются геометрией окна.

С каждым окном связывается система координат, начало которой находится в левом верхнем углу окна (точнее – его внутренности). Ось X направлена вправо, а ось Y – вниз. Единица измерения по обеим осям – пиксел.

X Window позволяет программе создавать несколько окон одновременно. Они связаны в иерархию, в которой одни являются «родителями», а другие – «потомками». Сам сервер на каждом экране создает одно основное окно, являющееся самым верхним «родителем» всех остальных окон. Это окно называется «корневым» (root).

Окна могут располагаться на экране произвольным образом и даже перекрывать друг друга. X Window имеет набор средств, пользуясь которыми, программа-клиент может изменять размеры окон и их положение на экране. Особенностью системы является то, что она не имеет встроенной возможности управлять окнами с помощью мыши или клавиатуры. Чтобы это можно было осуществить, нужен специальный клиент – менеджер окон. Однако менеджер не может корректно управлять окнами, ничего о них не зная. Окна могут обладать различными свойствами, которые должен обеспечивать именно менеджер окон. Например, во многих случаях удобно иметь заголовки окон, в других – желательно, чтобы окно нельзя было сделать меньше или, наоборот, больше определённого размера. Окно может быть «схлопнуто» в пиктограмму («иконку») – в этом случае менеджер должен знать, какую

пиктограмму использовать и как её назвать. Клиенты могут сообщать менеджеру свои пожелания относительно окон двумя способами:

- при создании окна системе X могут быть переданы «рекомендации» о начальном положении окна, его геометрии, размерах и т.д.;
- можно использовать встроенный в X способ общения между программами – механизм «свойств».

2.3.8. Техническое обеспечение АИС

В развитии технического обеспечения АИС наблюдаются процессы, схожие с процессами развития программного обеспечения:

- интеграция компонентов и систем;
- масштабируемость;
- сотрудничество различных производителей по сближению характеристик продукции на базе стандартов.

Под техническим обеспечением АИС понимается комплекс применяемых для создания АИС технических средств, в который входят: средства вычислительной техники, периферийное оборудование, средства связи и др.

Ряд фирм-производителей, которые преобладают на рынке технических средств АИС, стараются выпускать весь спектр средств вычислительной техники, который способен полностью решить проблему интеграции технического обеспечения корпоративных и других АИС. Масштабируемые компьютерные системы призваны удовлетворить потребности как разработчиков, так и предприятий, предъявляющих высокие требования по гибкости и надежности приложений. Для этого разрабатываются, производятся и поставляются на рынок рабочие станции, серверы и системы хранения информации уровня предприятия, а также сетевые компьютеры.

Сосредоточивая усилия на ключевых областях информационных технологий, во всех других областях эти фирмы основываются на стратегическом распределении усилий, полагаясь на дополняющих их партнеров и поставщиков. По мере того как развивались новые направления деятельности, ключевые технологии претерпевали эволюцию. Например, когда обнаруживались ограничения в существующих микропроцессорных технологиях, увеличивались капиталовложения в разработку процессоров и заключались соглашения о стратегическом партнерстве с производителями полупроводников.

Некоторые производители средств ВТ проникают (диверсифицируют) в смежные области и начинают разрабатывать и производить для своих систем программное обеспечение, которое содержит как среды

программирования, так и операционные системы, а также соответствующие утилиты.

2.3.9. Нормативно-техническое обеспечение качества, эффективности и безопасности АИС

Методической основой технологии, регламентирующей деятельность специалистов в процессе разработки АИС, является её типовой жизненный цикл и технологический процесс, который отражает этапы и систему операций в последовательности их выполнения и взаимосвязи, обеспечивающих ведение разработки от подготовки технического задания до завершения испытаний и ввода в эксплуатацию.

Типовой жизненный цикл АИС отображается в системе нормативно-технических документов, которые содержат правила, общие принципы или характеристики, касающиеся различных видов деятельности, стандарты, технические условия, инструкции и регламенты по применению. Правила оформления, содержание, структура и привязка к конкретному этапу типового жизненного цикла АИС (ЖЦ АИС) нормативно-технических документов содержатся в международных и национальных стандартах. Стандартизация типового жизненного цикла АИС проходила постепенно в течение многих лет, и к настоящему времени ее еще нельзя считать завершенной. Тем не менее создание эффективных сложных АИС было бы сегодня невозможно без следования требованиям стандартов.

Качество АИС описывается совокупностью показателей – критериев, для каждого из которых должны быть определены метрики и методы их измерения. Эти критерии и метрики АИС позволяют описывать свойства АИС как конечного продукта независимо от способа достижения этих свойств.

Основными показателями и критериями качества АИС являются следующие:

- научно-технический уровень – степень использования технических решений, отвечающих современным научно-техническим достижениям;
- уровень автоматизации – степень использования автоматизированных информационных технологий при выполнении функций АИС;
- эффективность – свойство системы, заключающееся в выполнении предписанных функций, с учетом соотношения затрат с результатами;
- экономичность – количество и степень занятости ресурсов: реализующей ЭВМ, АРМов пользователей, каналов сети и др.;

- функциональные критерии качества – отражают специфику областей применения и степень соответствия АИС их основному целевому назначению;
- конструктивные критерии качества – отражают эффективность использования ресурсов вычислительных средств, а также надежность и другие общие характеристики функционирования АИС.

В существующих стандартах описывается до 30 показателей и приводятся их определения. Однако значения многих показателей могут определяться только экспертно. В процессе разработки контроль всей совокупности показателей нецелесообразен, так как конструктивное влияние на процессы разработки и применения оказывают только некоторые из них. В зависимости от класса, функций и особенностей АИС целесообразно выбирать различные наборы критериев, адекватные свойствам конкретной АИС. Критериев, наиболее полно отражающих назначение и функциональные характеристики АИС, обычно не более 10. Каждый из основных критериев может быть описан еще несколькими частными показателями, характеризующими основные свойства объекта или процесса в разных аспектах.

Концепция и основные понятия безопасности АИС

Безопасность – это понятие, характеризующее способность системы обработки данных обеспечивать защиту, достоверность и надежность хранения информации. В случае непредумышленных угроз вводится понятие технологической безопасности АИС, т.е. системы методов и средств предотвращения отказов функционирования системы при случайных, дестабилизирующих воздействиях, а также снижения воздействия этих угроз до допустимого уровня, который определяется соответствующими нормативно-техническими документами.

Проблемы защиты (средства для ограничения доступа или использования всей АИС или ее части) имущественных прав на информацию возникли в связи с превращением информации в объект собственности. В этом случае формируются следующие цели защиты информации в АИС:

- предотвращение утечки, хищения, утраты, несанкционированного уничтожения, искажения, модификации (подделки), несанкционированного копирования, блокирования информации;
- сохранение полноты, достоверности, целостности информации и программ обработки, установленных собственником или уполномоченным им лицом;

- обеспечение конституционных прав граждан на сохранение личной тайны и конфиденциальности персональной информации, накапливаемой в базах данных;
- сохранение секретности, конфиденциальности информации в соответствии с действующим законодательством;
- соблюдение прав авторов программной и информационной продукции, используемой в информационных системах.

В деятельности по применению мер и средств защиты информации выделяются следующие самостоятельные направления:

- защита информации от несанкционированного доступа;
- юридическая защита электронных документов;
- защита от компьютерных вирусов и других опасных воздействий по каналам распространения программ;
- защита от несанкционированного копирования и распространения программ и ценной компьютерной информации.

Защита конфиденциальной и ценной информации от несанкционированного доступа и модификации технически сводится к задаче разграничения функциональных полномочий и доступа к АИС. При этом под словом «доступ» следует понимать не только возможность потенциального нарушителя «читать» хранящуюся в компьютере информацию, но и возможность модифицировать ее. Требования по защите в различных приложениях могут существенно различаться, однако они всегда направлены на достижение трех основных свойств защищаемой информации:

- конфиденциальность – засекреченная информация должна быть доступна только тому, кому она предназначена;
- целостность – информация, на основе которой принимаются важные решения, должна быть достоверной и точной, а также защищена от возможных непреднамеренных и злоумышленных искажений;
- готовность – информация и соответствующие автоматизированные службы должны быть доступны, готовы к обслуживанию всегда, когда в них возникает необходимость.

Защита юридической значимости электронных документов оказывается необходимой при обработке, хранении и передаче сообщений, файлов, баз данных, содержащих в себе приказы, платежные поручения, контракты и другие распорядительные, договорные, финансовые документы. Их общая особенность заключается в том, что в случае возникновения споров (в том числе и судебных) должна быть обеспечена возможность доказательства истинности факта того, что автор действительно фиксировал акт своего волеизъявления в отчуждаемом

электронном документе, а не был фальсифицирован каким-либо третьим лицом.

Для решения данной проблемы используются современные криптографические методы проверки подлинности информационных объектов, связанные с применением так называемых «цифровых подписей». Эти методы основаны на включении в документы специальных меток, логически неразрывно связанных с его текстом, для порождения которых используется индивидуальный секретный криптографический ключ. При наличии индивидуальных криптографических ключей исключается возможность «подделки» таких меток со стороны других лиц. Цифровая подпись может неопровержимо свидетельствовать об авторстве того или иного конкретного лица, и этот факт может быть проверен получателем.

Защита информации и ПС от компьютерных вирусов и других опасных воздействий по каналам распространения программ приобрела за последнее время особую актуальность. Масштабы реальных проявлений «вирусных эпидемий» оцениваются сотнями тысяч случаев «заражения» программ во всех странах. Хотя некоторые из вирусных программ оказываются вполне безвредными, многие из них имеют разрушительный характер. Особенно опасны вирусы для компьютеров, входящих в состав информационно-вычислительных сетей. Некоторые особенности современных АИС создают благоприятные условия для распространения вирусов. К ним, в частности, относятся:

- необходимость совместного использования ПС и БД многими пользователями;
- трудность ограничения в применении программ и данных;
- ненадежность существующих механизмов защиты и разграничения доступа к информации с точки зрения противодействия вирусу.

Защита от несанкционированного копирования и распространения программ и ценной информации является самостоятельным видом защиты имущественных прав, ориентированным на охрану интеллектуальной собственности. Эта защита обычно осуществляется с помощью специальных программ, подвергающих защищаемые программы и данные предварительной обработке (вставка парольной защиты, проверок по обращению к устройствам хранения ключа и ключевым дискетам, блокировка отладочных прерываний, проверка рабочей ЭВМ по ее уникальным характеристикам и т.п.), которая приводит объектный код защищаемой программы в состояние, препятствующее его выполнению на «чужих» ЭВМ. Применяются дополнительные аппаратные блоки (ключи), подключаемые к разъему прин-

тера или к системной шине ЭВМ, а также производится шифрование файлов, содержащих объектный код программы.

Общим свойством средств защиты программ от несанкционированного копирования является ограниченная стойкость такой защиты, так как в конечном счёте объектный код программы поступает на выполнение в центральный процессор в открытом виде и может быть прослежен с помощью аппаратных отладчиков. Однако это обстоятельство не снижает потребительские свойства средств защиты до нуля, так как основной целью их применения является в максимальной степени затруднить, хотя бы временно, возможность массового тиражирования новых версий ПС до появления последующих версий.

Контрольные вопросы

1. Что такое информационные системы (ИС) и какие основные функции они выполняют?
2. Дайте обобщенную структуру ИС и охарактеризуйте общий алгоритм ее функционирования.
3. Назовите и охарактеризуйте основные типы ИС по назначению и их основные функции.
4. Что такое интегрированные корпоративные ИС?
5. Чем отличаются локальные ИС от полнофункциональных?
6. Как классифицируют экономические ИС по уровню функциональности и степени интегрированности?
7. Как классифицируют экономические ИС по возможностям поддержки корпоративного управления?
8. Как классифицируют экономические ИС по возможностям поддержки уровней управления?
9. Какие основные виды программного обеспечения ИС вы знаете?
10. Что является базовым программным обеспечением ИС?
11. Какие программные средства разработки ИС вы знаете?
12. Какие программные средства прикладного программного обеспечения ИС вы знаете?
13. Какие операционные системы используются для поддержки ИС?
14. Какие функции выполняют ОС в распределенных ИС?
15. Какие средства автоматизации проектирования ИС вы знаете?
16. Какие основные функции CASE-средств вы знаете?
17. Какое программное обеспечение интерфейсов АИС вы знаете?
18. Какие основные виды технического обеспечения ИС вы знаете?
19. Какие основные виды нормативно-технического обеспечения качества, эффективности и безопасности ИС вы знаете?

3. ИНФОРМАЦИЯ В СИСТЕМАХ

Современное понимание того, что такое информация и какова ее роль в искусственных и естественных системах, сложилось не сразу; оно представляет собой совокупность знаний, полученных разными науками.

Хотя физика старается изучать явления природы в максимально объективированной форме, ей не удалось полностью исключить «человеческий фактор». Во-первых, при экспериментальном исследовании физических явлений невозможно обойтись без измерения ряда величин, и наличие этих величин в теоретических моделях требует специального рассмотрения того, как именно проводятся измерения, – а без влияния человека не проводятся даже автоматические эксперименты. Во-вторых, физика не могла не заняться изучением технических (т.е. созданных человеком) устройств и тут-то и «попала в ловушку»: полностью объективистский подход к машинам все равно обнаружил в их поведении следы человеческой деятельности. Впервые это произошло в термодинамике – науке, изучающей процессы в тепловых машинах. Оказалось, что без введения понятия энтропии невозможно дать исчерпывающего описания действия таких машин. Скачок в понимании природы этой величины произошел, когда Л. Больцман дал ей статистическую интерпретацию (1877); уже сам Больцман обронил фразу о том, что энтропия характеризует недостающую информацию, но тогда этой фразы никто не понял. После построения К. Шэнноном теории информации (1948), когда оказалось, что формула Шэннона для информационной энтропии и формула Больцмана для термодинамической энтропии оказались тождественными, разгорелись споры о том, является ли это совпадение чисто формальным или оно выражает глубокую связь термодинамики и теории информации. Дискуссии привели к современному пониманию этой неразрывной связи. Совсем с другой стороны к этой проблеме подошла философская теория познания. Изначальный смысл слова «информация» как «знания, сведения, сообщения, уведомления, известия, ведомости», т.е. нечто присущее только человеческому сознанию и общению, начал расширяться и обобщаться. Признав, что наше знание есть отражение реального мира, материалистическая теория познания установила, что отражение является всеобщим свойством материи. Сознание человека есть высшая, специфическая форма отражения. И теперь, как только состояния одного объекта находятся в соответствии с состояниями другого объекта, мы говорим, что один объект отражает другой, содержит информацию о другом.

В настоящее время информация рассматривается как фундаментальное свойство материи.

Таким образом, роль информации в самом существовании систем, искусственных и естественных, огромна. Понятие информации, обладая всеобщностью, приобрело смысл философской категории. Для кибернетики и системологии понятие информации столь же фундаментально, как понятие энергии для физики. Недаром признается определение кибернетики, предложенное А.Н. Колмогоровым: кибернетика – это наука, изучающая системы любой природы, способные воспринимать, хранить и перерабатывать информацию и использовать ее для управления и регулирования.

3.1. Кодирование информации и алфавиты

Любое сообщение, подлежащее передаче по каналу связи, записи в память или переработке в ИС, должно быть предварительно закодировано. Кодирование определяют как процесс представления информации в виде некоторых символов или их последовательностей (кодовых комбинаций), причем эти символы, в свою очередь, могут быть представлены (перекодированы) в виде совокупностей физических сигналов той или иной природы – акустических, оптических, электрических и т.д. Так, примером естественного кодирования является представление в виде слов-символов информации, возникающей в процессе восприятия человеком явлений окружающего мира – его отражения мозгом человека. Однако в дальнейшем для фиксации этой информации на бумаге необходимо ее перекодирование в виде букв и их сочетаний, для передачи по электрическому каналу связи – в виде электрических сигналов и т.д. Все эти многочисленные этапы преобразования форм представления информации будем называть, как это принято, кодированием. Так, при передаче словесно-буквенного текста по телеграфу осуществляется кодирование его одним из телеграфных кодов, представляющих собой организованные определенным образом сочетания импульсов электрического тока, например кодом Морзе или кодом Бодо. При приеме этого сообщения необходимо восстановить передававшийся первоначальный словесно-буквенный текст – осуществить обратное перекодирование, которое носит название декодирования.

Согласно строгой математической терминологии «кодированием называется отображение произвольного множества A в множество конечных последовательностей (слов) в некотором алфавите B , а декодированием – обратное отображение».

Это отображающее множество В, включающее в себя множество знаков (символов) и слов (кодowych комбинаций), составленных из этих знаков по определенным правилам и предназначенных для однозначного отображения элементов множества А, называется кодом.

Конечное множество (список) попарно различных знаков, букв, цифр или любых других символов, применяемых в той или иной области (языке), носит название алфавита. Соответственно, рассматривая проблемы кодирования, можно говорить об алфавите данного кода. Количество различных букв, входящих в алфавит, называют объемом алфавита.

К любой системе кодирования предъявляются следующие основные требования:

а) взаимная однозначность преобразований отображаемого множества в отображающее множество при кодировании и обратного преобразования при декодировании, что составляет необходимое условие отсутствия ошибок в интерпретации исходной информации;

б) экономичность кодирования, обеспечиваемая прежде всего минимизацией средней длины кодовой комбинации, а значит, и длины информационных текстов, что, в свою очередь, обеспечивает сокращение времени, необходимого для передачи и обработки информации, и экономию носителей информации;

в) помехоустойчивость, т.е. возможность обнаружения и исправления ошибок в кодовых комбинациях под влиянием тех или иных помех и сбоев в процессе передачи и обработки информации, повышающая достоверность работы кибернетических систем.

Необходимо отметить, что второе и третье требования взаимно противоречат друг другу, так как повышение помехоустойчивости кодов достигается увеличением длины кодовых комбинаций и, следовательно, снижает экономичность систем кодирования. В технике связи и обработки информации предложено большое количество различных способов кодирования, обеспечивающих более или менее успешный компромисс в выполнении этих требований в различных кибернетических системах и условиях их функционирования.

Не вдаваясь в детали классификации различных кодов, приведем сведения об их основных классах. При этом будем говорить только о цифровых кодах, которые получили практически повсеместное распространение, в том числе и при обработке буквенных текстов, также кодируемых числами, т.е. цифровыми комбинациями. В зависимости от длины кодовой комбинации (количества образующих ее символов) коды делятся на равномерные и неравномерные. В равномерных кодах все комбинации имеют одинаковую длину, в неравномерных – различную. С точки зрения осуществления процессов кодирования и после-

дующей передачи и обработки закодированной информации (особенно в условиях помех) удобнее и в конечном счете выгоднее применять равномерные коды. Далее коды делятся на неизбыточные и избыточные. Неизбыточные – это коды минимальной длины, определяемой только возможностью их различения. Избыточные коды содержат еще и дополнительные символы или группы символов, специально предназначенные для обнаружения ошибок (обнаруживающие коды) или исправления их (корректирующие коды). И те, и другие коды могут быть систематическими (разделимыми) и несистематическими. В систематических имеет место четкое разделение кодовой комбинации на две части: одну, несущую основную информацию, и другую, контрольную, предназначенную специально для обнаружения и исправления ошибок.

3.2. Сигналы в системах

Для того чтобы два объекта содержали информацию друг о друге, необходимо, чтобы между их состояниями существовало соответствие, только при этом условии по состоянию одного объекта можно судить о состоянии другого. Такое соответствие может возникнуть только в результате физического взаимодействия между этими объектами. Соответствие между состояниями двух объектов может устанавливаться и с помощью взаимодействия с промежуточными объектами, часто даже целой совокупностью промежуточных объектов.

3.2.1. Понятие сигнала

Сигнал есть материальный носитель информации, средство перенесения информации в пространстве и времени.

Утверждая, что объекты выступают в качестве сигналов, следует сделать уточнение. Один и тот же объект может выступать в качестве разных сигналов: колебания воздуха могут нести звуки музыки, речь лектора, пение птиц или шум самолета; с магнитной ленты можно стереть одну запись и сделать другую и т.д. Следовательно, в качестве сигналов используются не сами по себе объекты, а их состояния.

Не всякое состояние имеет сигнальные свойства. Точнее говоря, данный объект взаимодействует не только с тем объектом, информацию о котором мы хотели бы получить, но и с другими, не интересующими нас объектами. В результате соответствие состояний ослабевает, разрушается. Условия, обеспечивающие установление и способствующие сохранению сигнального соответствия состояний, называются кодом. В искусственных системах кодом называют комплекс правил об-

разования сигнала. При более подробном рассмотрении этого процесса в технических системах кодом называют условные, варьируемые правила, а диктуемые техникой условия – модуляцией. Будем пока использовать самое общее употребление слова «код». Посторонние воздействия, нарушающие это соответствие, называются помехами или шумами. Нарушение соответствия может происходить не только вследствие помех, но и из-за рассогласования кодов взаимодействующих объектов. В искусственных системах, где такое согласование организуется специально, это явно видно на примере криптографии, основанной на засекречивании кодов. В природных системах согласование кодов происходит в самой структуре систем через естественный отбор различных вариантов.

Информация есть свойство материи, состоящее в том, что в результате взаимодействия объектов между их состояниями устанавливается определенное соответствие. Чем сильнее выражено это соответствие, тем полнее состояние одного объекта отражает состояние другого, тем больше информации один объект содержит о другом.

Таким образом, сигнал есть материальный носитель информации. В качестве сигналов используются состояния физических объектов или полей. Соответствие между сигналом и несомой им информацией устанавливается по специальным правилам, называемым кодом.

3.2.2. Типы сигналов

Поскольку сигналы служат для переноса информации в пространстве и времени, для образования сигналов могут использоваться только объекты, состояния которых достаточно устойчивы по отношению к течению времени или к изменению положения в пространстве. С этой точки зрения сигналы делятся на два типа:

1. Сигналы, являющиеся стабильными состояниями физических объектов (например, книга, фотография, магнитофонная запись, состояние памяти ЭВМ, положение триангуляционной вышки и т.д.). Такие сигналы называются статическими.

2. Сигналы, в качестве которых используются динамические состояния силовых полей. Такие поля характеризуются тем, что изменение их состояния не может быть локализовано в (неизолированной) части поля и приводит к распространению возмущения. Конфигурация этого возмущения во время распространения обладает определенной устойчивостью, что обеспечивает сохранение сигнальных свойств. Примерами таких сигналов могут служить звуки (изменение состояния поля сил упругости в газе, жидкости или твердом теле), световые и

радиосигналы (изменения состояния электромагнитного поля). Сигналы указанного типа называются динамическими.

Динамические сигналы используются для передачи, а статические – для хранения информации, но можно найти и противоположные примеры (динамические запоминающие устройства, письма, газеты).

Сигналы играют в системах особую, очень важную роль. Если энергетические и вещественные потоки, образно говоря, питают систему, то потоки информации, переносимые сигналами, организуют все ее функционирование, управляют ею. Н. Винер, например, подчеркивал, что общество простирается до тех пределов, до каких распространяется информация. Пожалуй, это следует отнести к любой системе.

Таким образом, главное отличие подхода к изучению любого объекта как системы, а не как просто объекта, состоит в том, что мы ограничиваемся не только рассмотрением и описанием вещественной и энергетической его сторон, но прежде всего проводим исследование его информационных аспектов: целей, сигналов, информационных потоков, управления, организации и т.п.

3.3. Математическая модель сигналов

Казалось бы, после того как мы установили, что сигналами служат состояния физических объектов, никаких проблем с их математическим описанием не должно быть: ведь физика имеет богатый опыт построения математических моделей физических процессов и объектов. Например, можно зафиксировать звуковые колебания, соответствующие конкретному сигналу, в виде зависимости давления x от времени t и изобразить этот сигнал функцией $x(t)$; такой же функцией можно изобразить и статический сигнал, например запись этого звука на магнитной ленте или на грампластинке, поставив параметру t в соответствие протяженность (длину) записи.

3.3.1. Непредсказуемость — основное свойство сигналов

Однако имеется существенное различие между просто состоянием $x(t)$ объекта и сигналом $x(t)$. Оно состоит в том, что единственная функция $x(t)$ не исчерпывает всех важных свойств сигналов. Ведь понятие функции предполагает, что нам известно значение x (либо правило его вычисления) для каждого t . Если же это известно получателю сигнала, то отпадает необходимость в передаче: функция $x(t)$ может быть и без этого воспроизведена на приемном конце.

Следовательно, единственная однозначная функция вещественного аргумента не может служить моделью сигнала. Такая функция приобретает сигнальные свойства только тогда, когда она является одной из возможных функций. Другими словами, моделью сигнала может быть набор (или, как еще говорят, ансамбль) функций параметра t , причем до передачи неизвестно, какая из них будет отправлена; это становится известным получателю только после передачи. Каждая такая конкретная функция называется реализацией. Если теперь еще ввести вероятностную меру на множество реализаций, то мы получим математическую модель, называемую случайным процессом.

Случайный процесс хорошо отображает главное свойство реальных сигналов – их неизвестность до момента приема. Однако некоторые модели обладают и «лишними», не присущими реальным сигналам свойствами. Например, описание реализаций случайного процесса аналитическими функциями предполагает бесконечную точность, чего в реальности не бывает.

Прежде чем перейти к некоторым подробностям этой модели, отметим, что у нее все-таки имеются качества, которых нет у реальных сигналов (еще раз обратим внимание на нетождественность модели и оригинала). Дело в том, что реальные системы всегда оперируют только с конечным объемом данных, а понятие аналитической функции предполагает ее точное значение для каждого значения аргумента, т.е. такое, которое может быть представлено только с помощью бесконечного ряда цифр. Иначе говоря, в «более правильной» модели сигнала любая реализация не должна определяться с бесконечной точностью. Но пока не существует подходящего математического аппарата, столь же удобного, как математический анализ, и мы вынуждены пользоваться традиционным понятием функции. Это сопряжено с тем, что в выводах могут появиться (и появляются) парадоксы, природа которых связана не с реальными сигналами, а с их моделью. Во всяком случае, там, где конечная точность реальных сигналов существенна для самой постановки задачи, ее вводят в модель либо как добавочный «шум», либо как «квантование» непрерывного сигнала (наподобие шкалы с делениями у измерительных приборов).

Вернемся к рассмотрению случайных процессов как моделей сигналов. Имеется несколько различных подходов к тому, как вводить вероятностную меру на множестве реализаций. Для инженерных приложений оказывается удобным определение случайного процесса как такой функции времени $x(t)$, значение которой в каждый данный момент является случайной величиной. Случайная величина полностью характеризуется распределением вероятностей, например плотностью $P_1(x_1|t_1)$;

однако, чтобы охарактеризовать случайный процесс, нужно описать, связаны ли (и если да, то как) значения реализации, разделенные некоторыми интервалами времени. Так как связь только двух таких значений, описываемая распределением второго порядка $P_2(x_1, x_2 | t_1, t_2)$, может неполно характеризовать процесс в целом, вводят распределения третьего, четвертого, ..., n -го порядков: $P_n(x_1, \dots, x_n | t_1, \dots, t_n)$. В конкретных задачах обычно ясно, до какого порядка следует доходить в описании процесса.

3.3.2. Классы случайных процессов

Необходимость моделирования самых разнообразных сигналов приводит к построению частных моделей случайных процессов наложением дополнительных ограничений на параметры распределений и на сами распределения. Перечислим наиболее важные классы случайных процессов.

Непрерывные и дискретные по времени процессы. Случайный процесс с непрерывным временем характеризуется тем, что его реализации определяются для всех моментов из некоторого (конечного или бесконечного) интервала T параметра t . Дискретный по времени процесс задается на дискретном ряде точек временной оси (обычно равноотстоящих).

Непрерывные и дискретные по информативному параметру процессы. Они различаются в зависимости от того, из какого (непрерывного, дискретного) множества принимает значение реализация x случайной величины X .

Стационарные и нестационарные процессы. Так называются процессы в зависимости от постоянства или изменчивости их статистических характеристик. Случайный процесс называется стационарным в узком смысле, если для любого n конечномерные распределения вероятностей не изменяются со временем, т.е. при любом τ выполняется условие

$$P_n(x_1, \dots, x_n | t_1 + \tau, t_2 + \tau, \dots, t_n + \tau) = P_n(x_1, \dots, x_n | t_1, t_2, \dots, t_n).$$

Если же условие независимости от времени выполняется только для первых двух моментов – среднего и функции автокорреляции, то процесс называется стационарным в широком смысле.

Эргодические и неэргодические процессы. На практике при описании случайных величин вместо рассмотрения их распределений часто ограничиваются только их числовыми характеристиками, обычно

моментами. В тех случаях, когда распределение неизвестно, моменты (и другие числовые характеристики) можно оценить статистически.

Перенос такой практики на произвольные случайные процессы требует не только учета зависимости отстоящих друг от друга («разнесенных») во времени значений, но и наложения дополнительных требований. Требование совпадения величин, получающихся при усреднении по ансамблю (т.е. при фиксированном времени) и при усреднении по времени (точнее, по одной реализации), и называется условием эргодичности. Это требование можно толковать и как совпадение результатов усреднения по любой реализации. Как и для стационарности, различают эргодичность в узком и широком смысле.

Можно продолжать классификацию случайных процессов и дальше, но мы будем делать это при рассмотрении конкретных вопросов.

Таким образом, случайный процесс может служить математической моделью сигнала. Необходимо только следить за тем, чтобы конкретные особенности изучаемых сигналов были корректно отображены в свойствах случайного процесса.

3.4. Математические модели реализаций случайных процессов

Для рассмотрения конкретных свойств систем бывает необходимо учесть особенности сигналов, циркулирующих по каналам связи этих систем. Такие особенности можно описать по-разному: просто перечислить возможные реализации (если число их конечно) либо задать в той или иной форме общие свойства реализаций, входящих в ансамбль. О дискретных процессах будем говорить отдельно, а сейчас рассмотрим математические модели реализаций непрерывных сигналов.

Приведем примеры, с которыми часто имеют дело в теории сигналов.

3.4.1. Моделирование конкретных реализаций

Гармонические сигналы. Обозначим через S_c множество всех синусоидальных сигналов:

$$\begin{aligned} S_c = \{x : x(t) = S \cos(\omega t + \varphi); \\ -\infty < t < \infty; \quad S, \omega \in R^+, \varphi \in [0, 2\pi]\}, \end{aligned} \quad (1)$$

где R^+ — множество всех положительных действительных чисел;
 S — амплитуда;
 ω — круговая частота (частота);
 φ — фаза гармонического колебания.

Модулированные сигналы. В технических системах полезная информация может переноситься каким-нибудь одним параметром «гармонического» колебания. Конечно, колебание при изменении этого параметра во времени перестает быть гармоническим. Процесс изменения параметра синусоиды называется модуляцией, а выделение этого изменения в чистом виде, как бы «снятие» модуляции, называется демодуляцией. Само колебание называется несущим. Различают амплитудную (2), частотную (3) и фазовую (4) модуляцию в зависимости от того, на какой из параметров несущего колебания «накладывают» полезную информацию (параметры с индексом 0 считаются постоянными):

$$S_{\text{ам}} = \{x : x(t) = S(t) \cos(\omega_0 t + \varphi_0)\}; \quad (2)$$

$$S_{\text{чм}} = \{x : x(t) = S_0 \cos(\omega(t) \cdot t + \varphi_0)\}; \quad (3)$$

$$S_{\text{фм}} = \{x : x(t) = S_0 \cos(\omega_0 t + \varphi(t))\}. \quad (4)$$

Необходимо отметить, что «физический смысл» модуляции сохраняется лишь в том случае, когда модулирующий сигнал является «медленно меняющимся» по сравнению с немодулированным несущим колебанием: только при этом условии имеется возможность нестрого, «по инженерному», но все-таки не без смысла говорить о «синусоиде с переменной амплитудой (фазой, частотой)». Однако удобнее все-таки говорить не о «переменной амплитуде», а об «огibaющей», как это принято в радиотехнике.

Периодические сигналы. Сигналы называются периодическими, а временной интервал τ – периодом, если

$$S_\tau = \{x : x(t) = x(t + \tau); \quad -\infty < t < \infty\}. \quad (5)$$

Сигналы с ограниченной энергией. О сигналах из множества

$$S_\varepsilon = \left\{ x : \int_{-\infty}^{\infty} x^2(t) dt \leq K < \infty \right\} \quad (6)$$

говорят, что их энергия ограничена величиной K . Происхождение этого названия связано с тем, что если $x(t)$ есть напряжение, то интеграл в формуле (6) представляет собой энергию, выделенную сигналом $x(t)$ на единичном сопротивлении. Конечно, если $x(t)$ есть, например, глубина бороздки на грампластинке, то интеграл связан с энергией лишь косвенно.

Сигналы ограниченной длительности. Интервал T называется длительностью сигнала $x(t)$, если

$$S_T = \{x : x(t) \equiv 0, \quad t \notin T\}; \quad (7)$$

конечно, предполагается, что внутри этого интервала сигнал $x(t)$ не везде обращается в нуль.

Особую роль среди сигналов с ограниченной длительностью играют импульсные сигналы, характеризующиеся «кратковременностью» T , которую трудно формализовать, но которая проявляется в практике: звуки типа «щелчок», «взрыв», «хлопок»; световые «вспышки»; тактильные сигналы «укол», «щипок», «удар» и т.п. В таких случаях сигнал $x(t)$ обычно называется «формой импульса». На практике широко распространены периодические последовательности импульсов (радиолокация, электрокардиография, ультразвуковая гидролокация и т.п.); не менее важны непериодические импульсные последовательности (телеграфия, цифровая телеметрия); в реальности все импульсные последовательности, в свою очередь, имеют ограниченную длительность.

Сигналы с ограниченной полосой частот. Фурье-преобразование $X(f)$ временной реализации $x(t)$ сигнала называется ее спектром:

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{i2\pi ft} dt. \quad (8)$$

Физический смысл спектра состоит в том, что колебание $x(t)$ представляется в виде суммы (в общем случае в виде интеграла) составляющих его гармонических колебаний с определенными амплитудами $|x(f)|$, частотами f и соответствующими фазами. Между $x(t)$ и $X(f)$ имеется взаимно однозначное соответствие, так как

$$x(t) = \int_{-\infty}^{\infty} X(f) e^{-i2\pi ft} df. \quad (9)$$

Условием существования и обратимости Фурье-преобразования является ограниченность энергии сигнала (интегрируемость в квадрате функций $x(t)$ и $X(f)$ и его непрерывность).

Если функция $X(f)$ на оси f имеет ограниченную «длительность» F (в смысле соотношения (7), но в частотной области), то говорят, что сигнал $x(t)$ имеет ограниченную полосу частот шириной F :

$$S_B = \{x : X(f) = \int_{-\infty}^{\infty} x(t) e^{i2\pi ft} dt = 0 \quad \text{для всех } f > |F|\}. \quad (10)$$

3.4.2. Некоторые модели ансамбля реализации

Нормальный шум. Удобной моделью помех и некоторых полезных сигналов является стационарный нормальный случайный процесс. Такое название вызвано тем, что случайные мгновенные значения величины $x(t)$ предполагаются подчиненными нормальному закону (с нулевым средним), т.е. плотность распределения первого порядка выражается формулой

$$p_1(x_1 | t_1) = \frac{1}{\sqrt{2\pi N}} \exp\left\{-\frac{1}{2N}\right\}. \quad (11)$$

Для плотности распределения второго порядка справедлива формула

$$\begin{aligned} p_2(x_1, x_2 | t_1, t_2) &= p_2(x_1, x_2 | \tau = t_2 - t_1) = \\ &= -\frac{1}{2\pi N \sqrt{1 - p^2(\tau)}} \exp\left\{-\frac{1}{2(1 - p^2(\tau))N} [x_1^2 - 2p(\tau)x_1x_2 + x_2^2]\right\}. \end{aligned}$$

Использованные здесь обозначения таковы:

$N = E(X^2)$ – мощность шума (дисперсия мгновенных значений);

$p(\tau) = N^{-1}E(X_t X_{t+\tau})$ – нормированная функция автокорреляции.

Сигнал как колебание со случайными огибающей и фазой. Понятия амплитуды и фазы, введенные первоначально для гармонических сигналов, с помощью модуляции были обобщены на сигналы, которые уже не являются гармоническими. Легко обобщить их на произвольные сигналы.

Негармонический сигнал изобразится как вектор, совершающий вращения или колебания (изменяется угол) и изменяющий свою длину R .

При всей простоте такая модель позволяет легко описывать различные преобразования сигналов. Например, сумма двух колебаний с частотой ω_0 изобразится как сумма двух соответствующих векторов (при этом автоматически учитываются фазовые соотношения).

Вектор шума случаен, обе его компоненты нормальны с нулевым средним и дисперсией σ^2 . Сложение сигнала и шума приводит к вектору с нормальными компонентами, у одной из которых среднее равно S , а у другой – нулю:

$$p(x, y) dx dy = \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{1}{2\sigma^2} [(x - S)^2 + y^2]\right\} dx dy. \quad (12)$$

Поскольку существуют радиотехнические устройства, чувствительные только к огибающей или только к фазе принимаемого сигнала, имеет смысл рассмотреть их статистические свойства.

Можно показать, что огибающая распределена по закону Рэлея, фаза – равномерно в интервале $[0, 2\pi]$, R и ϕ независимы (так как их совместное распределение факторизовалось).

3.5. О некоторых свойствах непрерывных сигналов

Хотя мы излагаем лишь элементы теории сигналов, преследуя только ознакомительные цели, представляется интересным рассмотреть два основных ее аспекта, относящихся к свойствам непрерывных сигналов.

3.5.1. Частотно-временное представление сигналов

Известно, что некоторая функция $x(t)$ и ее спектр $X(f)$ однозначно выражаются друг через друга. Следовательно, сигнал можно рассматривать в любом из этих эквивалентных представлений – временном или частотном. При этом масштабные параметры этих представлений связаны обратно пропорциональной зависимостью. Пусть $x(t)$ имеет спектр $X(f)$. Изменим масштаб по оси времени в a раз (например, воспроизведем запись $x(t)$ с другой скоростью) и найдем спектр функции $X(at)$:

$$X_a(f) = \int_{-\infty}^{\infty} x(at) e^{i2\pi ft} dt = \frac{1}{a} X\left(\frac{f}{a}\right). \quad (13)$$

Как видим, масштаб по частотной оси изменился в $1/a$ раз. Более того, из свойств преобразования Фурье следует, что сигналы с ограниченной длительностью имеют спектры неограниченной ширины, а сигналы с ограниченной полосой частот длятся бесконечно долго.

Этот математический результат находится в противоречии с практикой: в реальности все сигналы конечны по длительности, а все чувствительные к сигналам устройства не могут воспринимать и воспроизводить абсолютно все частоты. Например, диапазон частот, к которым чувствителен слух человека, простирается от нескольких герц до 20–30 кГц, а все различимые звуки человеческой речи длятся доли секунды.

Тот факт, что аналитическая функция времени не может быть одновременно ограниченной и по длительности, и по ширине спектра, является, как видим, не свойством реальных сигналов, а свойством данной модели сигналов. Мы уже отмечали, что если не отказываться

от достоинств аппарата аналитических функций, то выход состоит в том, чтобы как-то иначе ввести в рассмотрение конечную точность реализаций функций времени. Правда, пока эта конечная точность не будет свойством самих реализаций, ее искусственное введение в модель можно проводить на разных этапах, что придает результатам некоторую относительность.

Например, говорить об одновременной ограниченности сигналов и по времени, и по спектру оказывается возможным при использовании энергетического критерия точности: сигнал считается имеющим конечную длительность ΔT , если в этом интервале времени сосредоточена основная часть всей энергии функции $x(t)$. В то же время и ширина спектра ΔF сигнала определяется как область частот, содержащая эту же часть всей энергии спектра $X(f)$:

$$\int_{\Delta T} x^2(t) dt = \int_{\Delta F} X^2(f) df = \mu \int_{-\infty}^{\infty} x^2(t) dt = \mu \int_{-\infty}^{\infty} X^2(f) df \quad (14)$$

Здесь величина μ меньше единицы, но достаточно близка к ней, а величина $(1 - \mu)$ характеризует косвенным образом точность, о которой шла речь.

Теперь можно говорить о том, какую «площадь» на плоскости «частота – время» занимает тот или иной сигнал. Если строго следовать теории Фурье-преобразований, то получим, что эта площадь для всех сигналов бесконечна, но для большинства из них энергетический критерий позволит ограничить ее естественным образом. Меняя форму сигнала $s(t)$, можно менять и занимаемую им площадь. Оказывается, что уменьшать эту площадь можно лишь до некоторого предела. Этот предел достигается на кривой, являющейся гармоническим колебанием, которое модулировано по амплитуде гауссовым импульсом; интересно, что спектр этой кривой имеет такую же форму:

$$s(t) = e^{-a(t-t_0)^2} e^{i2\pi f_0 t}; \quad (15)$$

$$S(f) = e^{-\left(\frac{\pi}{a}\right)^2 (f-f_0)^2} e^{-i2\pi f t_0}. \quad (16)$$

Существование предела, ниже которого нельзя сжать площадь сигнала, занимаемую им на плоскости «частота – время», и называется (по аналогии с принципом неопределенности в квантовой механике) принципом частотно-временной неопределенности сигналов:

$$\Delta F \cdot \Delta T \geq \text{const} > 0. \quad (17)$$

Все реальные сигналы длятся конечное время, и все реальные системы не могут реагировать на абсолютно все частоты. Следовательно, как длительность, так и ширина спектра сигнала ограничена. Аналитическая модель сигнала позволяет говорить об этих характеристиках приближенно. В этом мы видим конкретный пример проявления расхождений между моделью и оригиналом. Практически всякий сигнал можно представить состоящим из более простых компонент. В математике этой идее соответствует представление функций их разложениями в ряды и интегралы. В теории сигналов особое внимание уделяется условиям, при которых непрерывные функции можно однозначно представить в виде дискретного набора элементарных функций.

3.6. Цифровое представление непрерывных сигналов

Вторым важным аспектом теории сигналов является проблема цифрового представления непрерывных сигналов. Вопрос формулируется так: существуют ли условия (и если да, то какие), при которых любой непрерывной функции $x(t)$ можно поставить во взаимно однозначное соответствие дискретное множество чисел $\{C_k(x)\}$, $k = \dots -2, -1, 0, 1, 2, \dots$. Положительный ответ на этот вопрос имеет как теоретическое, так и практическое значение, так как рассмотрение случайных величин вместо реализаций непрерывных случайных процессов существенно упрощает решение многих задач, вся теория становится проще и может быть продвинута дальше, а соответствие $x(t) \leftrightarrow \{C_k(x)\}$ можно использовать и в технических устройствах, работающих с непрерывными сигналами (например, технически проще хранить или передавать $\{C_k(x)\}$ вместо $x(t)$).

Ограничимся более конкретной формулировкой поставленной задачи и рассмотрим условия выполнения равенства

$$x(t) = \sum_k C_k(x) \varphi_k(t). \quad (18)$$

Функции $\{\varphi_k(t)\}$ называются координатными функциями, они не должны зависеть от $x(t)$; более того, они заранее известны. Ряд в правой части равенства называется разложением $x(t)$ по координатным функциям. Числовые коэффициенты $\{C_k(x)\}$ содержат всю информацию об $x(t)$, необходимую для восстановления этой функции по фор-

муле (6); следовательно, множества $\{C_k(x)\}$ являются функционалами от функции $x(t)$.

Функционал – это отображение множества функций на множество чисел.

Известны разложения по системе ортогональных и нормированных функций. Это означает, что функции $\{\varphi_k(t)\}$ удовлетворяют условиям

$$\int \varphi_i(t) \varphi_k(t) dt = \delta_{ik} = \begin{cases} 1 & \text{при } i = k \\ 0 & \text{при } i \neq k \end{cases} \quad (19)$$

Умножим обе части равенства (6) на $\varphi_i(t)$ и проинтегрируем (опуская тонкости, будем считать, что все операции обоснованы):

$$\int \varphi_i(t) x(t) dt = \sum_k C_k(x) \int \varphi_i(t) \varphi_k(t) dt = C_i. \quad (20)$$

Такое представление называют рядом Фурье, а $C_k(x)$ – коэффициентами Фурье. Условия сходимости ряда Фурье к функции $x(t)$ подробно исследованы и, кратко говоря, сводятся к тому, чтобы были оправданы все необходимые математические операции, а коэффициенты Фурье убывали достаточно быстро (точнее, $\sum_k C_k^2 \log^2 k < \infty$). Это

не очень жесткое ограничение, но все же оно связывает свойства системы координатных функций и самих функций $x(t)$. Например, если $\varphi_k(t)$ – гармонические функции кратных частот, то $x(t)$ должна быть периодической функцией с периодом T , равным периоду самой низкочастотной гармоники:

$$x(t) = \sum_{k=-\infty}^{\infty} C_k e^{\frac{i\pi kt}{T}}; \quad (21)$$

$$C_k = \frac{1}{2F} \int_{-F}^F x(t) e^{-i \frac{2\pi kt}{T}} dt. \quad (22)$$

Значительный интерес привлекли разложения реализаций случайного процесса с ограниченной полосой частот. Для таких сигналов В.А. Котельников доказал следующую теорему (теорему отсчетов): *любая функция со спектром, находящимся в интервале $[0, F]$, полностью определяется последовательностью ее значений в точках, отстоящих друг от друга на $1/(2F)$ единиц времени.*

Пусть $x(t)$ имеет спектр $X(f)$; причем $X(f)$ отлично от нуля только в интервале $|f| \leq F$. В этом интервале применимо разложение:

$$X(f) = \sum_{k=-\infty}^{\infty} C_k(X) e^{i \frac{\pi k f}{F}} = \sum_{k=-\infty}^{\infty} C_k(X) e^{i \frac{2\pi k f}{2F}},$$

а коэффициенты Фурье этого разложения таковы:

$$C_k(X) = \frac{1}{2F} \int_{-F}^F X(f) e^{i \frac{2\pi k f}{2F}} df = \frac{1}{2F} x\left(\frac{k}{2F}\right).$$

Следовательно,

$$X(f) = \frac{1}{2F} \sum_{k=-\infty}^{\infty} x\left(\frac{k}{2F}\right) e^{i \frac{2\pi k f}{2F}}, \quad (23)$$

и это соотношение доказывает теорему отсчетов в силу однозначной связи $X(f)$ с $x(t)$.

Можно показать, что восстанавливать $x(t)$ для значений t между точками отсчетов следует согласно формуле

$$\begin{aligned} x(t) &= \int_{-F}^F X(f) e^{i 2\pi f t} df = \frac{1}{2F} \sum_{k=-\infty}^{\infty} x\left(\frac{k}{2F}\right) \int_{-F}^F X(f) e^{i 2\pi f \left(\frac{k}{2F} - t\right)} df = \\ &= \sum_{k=-\infty}^{\infty} x\left(\frac{k}{2F}\right) \frac{\sin(2\pi F t - k\pi)}{2\pi F t - k\pi}. \end{aligned} \quad (24)$$

Итак, мы получили разложение реализации, координатными функциями которого являются функции вида $(\sin u)/u$, сдвинутые друг относительно друга на интервалы времени $1/(2F)$, а коэффициентами — значения («отсчеты») самой реализации, взятые в моменты $k/(2F)$.

Иногда говорят, что эта теорема является теоретическим обоснованием возможности на практике восстанавливать $x(t)$ по отсчетам $x\left(\frac{k}{2F}\right)$. Однако дело в том, что координатные функции имеют неограниченную длительность и, следовательно, физически нереализуемы. Кроме того, ряд (18) имеет неограниченное число членов. Все это снова возвращает нас к проблеме точности фиксации сигналов, пока не получившей полного освещения.

Следовательно, из многочисленных результатов теории сигналов мы выделяем два, как существенно поясняющие природу непрерывных сигналов. Первый состоит в том, что сигналы обнаруживают своеобразную «упругость» занимаемой ими площади на плоскости «частота —

время». Это явление называется частотно-временной неопределенностью сигналов. Вторым результатом является то, что определенный класс непрерывных сигналов допускает взаимно однозначное соответствие между любой реализацией из этого класса и дискретным набором отсчетов данной реализации.

3.6.1. Особенности цифрового представления непрерывных сигналов. Решетчатые функции

Входные и выходные сигналы непрерывных систем есть функции непрерывного времени t . Если переменная t принимает непрерывную последовательность значений, то сигнал, или функция, называется непрерывным. Сигнал, или функция, называется дискретным, если переменная t принимает только конечное множество значений t_k , где $k = 0; \pm 1; \pm 2; \dots$

Квантование – процедура преобразования данных из непрерывной формы в цифровую. В цифровых системах используются различные виды квантования: квантование по уровню (релейные системы), квантование по времени (импульсные системы), квантование по уровню и по времени (цифровые системы). Сигналы, дискретные только во временной области, представляют собой последовательности импульсов, появляющихся в определенные моменты времени. Обычно дискретный сигнал получается в результате дискретизации – периодического прерывания непрерывного сигнала с постоянным тактом. При дискретизации промежуток времени между передними фронтами соседних импульсов называется периодом или тактом дискретизации.

Существуют различные способы модуляции (формирования) отдельных импульсов, входящих в последовательность. Они отличаются допустимыми значениями амплитуд, шириной импульсов и модулирующей частотой. В цифровых системах широкое распространение получила амплитудно-импульсная модуляция, представляющая собой такое преобразование непрерывного сигнала в последовательность импульсов, при котором высота импульса пропорциональна амплитуде текущего значения непрерывного сигнала, ширина постоянна, а интервалы между импульсами одинаковы и равны такту дискретизации. Устройство, преобразующее непрерывный сигнал в дискретный, называется дискретизатором.

На рис. 3.1 представлен дискретизатор как эквивалентный амплитудно-импульсный модулятор. Входной сигнал $f(t)$ умножается на несущий сигнал $p(t)$, который является последовательностью периодических сигналов, например импульсов единичного веса.

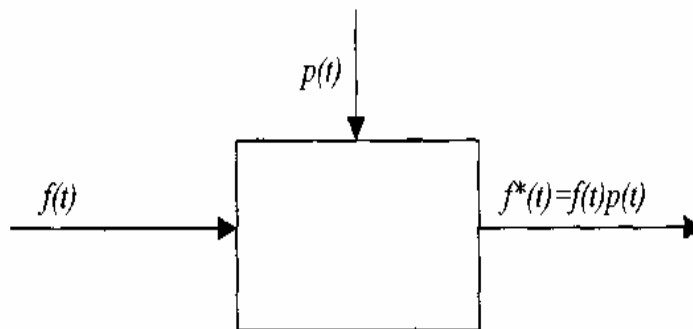


Рис. 3.1

Для входного сигнала $f(t)$, который является функцией непрерывно изменяющегося параметра t , сигнал на выходе дискретизатора $f^*(t)$ представляет собой последовательность импульсов конечной длительности, амплитуда которых промодулирована входным сигналом. Длительность импульсов равна h , период квантования T_0 и соответствующая ему частота квантования $F_0 = 1/T_0$ (Гц). На рис. 3.2 представлены типичные формы входного (а), несущего (б) и выходного (в) сигналов. Период квантования может быть непостоянным или изменяться циклически. Во всех случаях использование модели амплитудно-импульсного модулятора для представления дискретизатора очень удобно, так как при различных законах дискретизации меняется только несущий сигнал $p(t)$.

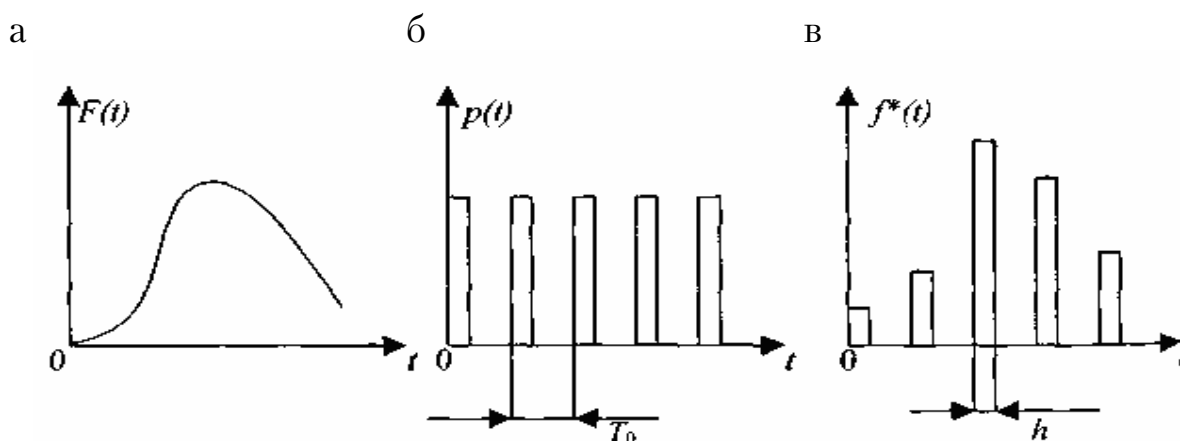


Рис. 3.2

Кроме того, на практике применяется широтно-импульсная модуляция. В этом случае выходной сигнал представляет собой последовательность импульсов, длительность которых является функцией амплитуды входного сигнала в момент выборки. На рис. 3.3 показаны входной (а) и выходной (б) сигналы широтно-импульсного модулятора.

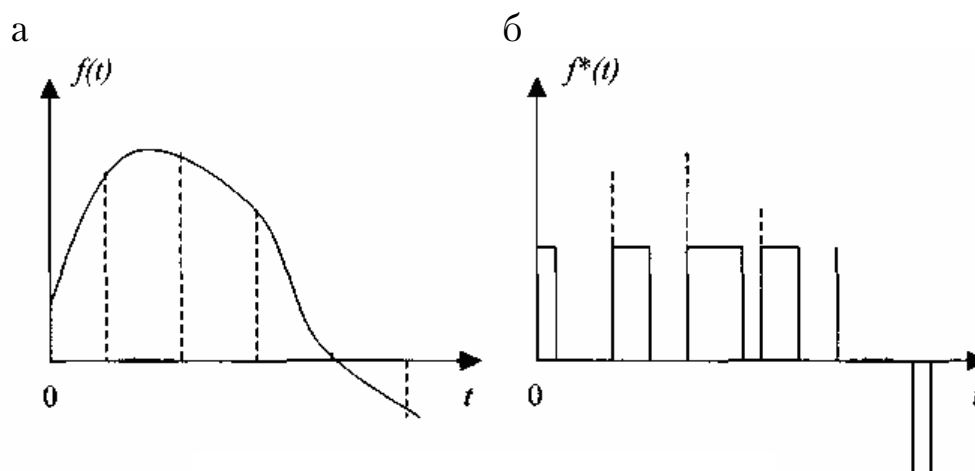


Рис. 3.3

Рассмотрим амплитудно-импульсную модуляцию с помощью дискретизатора (рис. 3.4) с постоянным периодом и конечным временем выборки, представленного в виде ключа, замыкающегося и размыкающегося мгновенно.

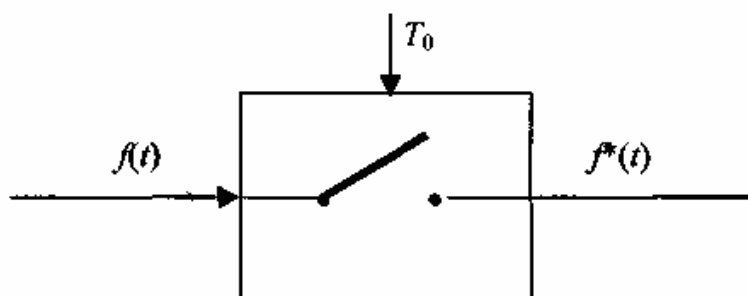


Рис. 3.4

Если длительность импульса h ($h > 0$) существенно меньше такта дискретизации T_0 , а за ключом стоит линейное звено с постоянной времени $T > h$, то последовательность импульсов $f^*(t)$ на выходе квантователя можно приближенно рассматривать как дискретный сигнал $f_T(kT_0)$, показанный на рис. 3.5 (а – непрерывный, б – дискретный). В этом случае ключ действует как идеальный дискретизатор, и величины $f_T(kT_0)$ равны мгновенным значениям амплитуд.

Модулированная по амплитуде дискретная функция $f^*(t)$, получаемая дискретизацией непрерывного сигнала $f(t)$ с постоянным тактом T_0 , математически описывается следующими выражениями:

$$\begin{aligned} f_T(t) &= f(kT_0) && \text{при } t = kT_0 \\ f_T(t) &= 0 && \text{при } kT_0 < t < (k+1)T_0 \quad (k=0, 1, 2, \dots) \end{aligned} \quad (25)$$

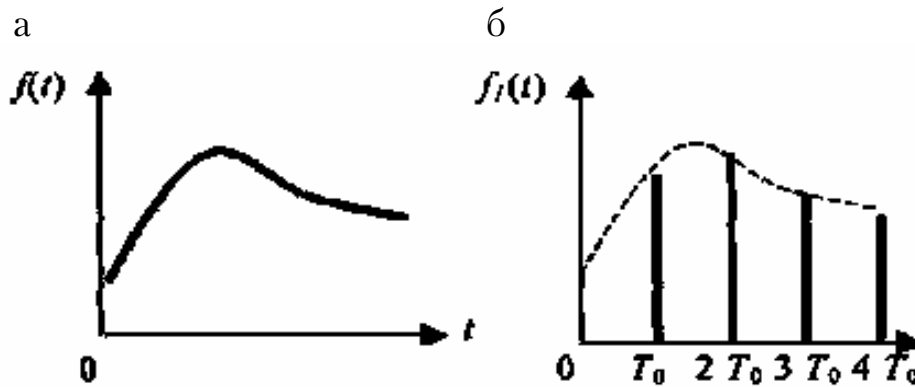


Рис. 3.5

Таким образом, идеальный дискретизатор может быть представлен в виде так называемой решетчатой функции времени $f_T(kT_0)$, или в сокращенной записи – $f(k)$, значения которой определены в дискретные моменты времени $t = kT_0$, где k – целое число, T_0 – период квантования. В этом случае решетчатая функция представляет собой модулированную последовательность δ -функций. Идеальную δ -функцию можно записать в виде:

$$\delta(t) = \begin{cases} 0 & \text{при } t \neq 0 \\ \infty & \text{при } t = 0 \end{cases}. \quad (26)$$

Выходная величина простейшего импульсного элемента, или идеального дискретизатора, представляющая собой модулированную последовательность δ -функций, определяется по формуле

$$f_T(t) = f_T \delta_T(t), \quad (27)$$

где $\delta_T(t)$ – немодулированная последовательность импульсов единичной площади или δ функций, равноотстоящих по времени (рис. 3.6):

$$\delta_T(t) = \sum_{k=0}^{\infty} \delta(t - kT_0).$$

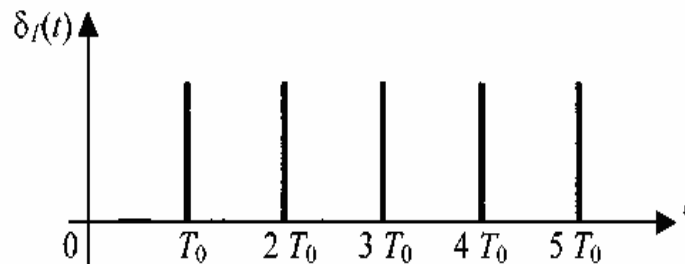


Рис. 3.6

С учетом того, что последовательность идеальных импульсов определена только в моменты времени $f = kT_0$ ($k = 0, 1, 2, \dots$), функция $f_T(t)$ описывается следующим выражением:

$$f_T(t) = \sum_{k=0}^{\infty} f(kT_0) \delta(t - kT_0).$$

Обратная задача – формирование непрерывной функции из решетчатой – не может быть решена однозначно, так как функции, заданной в дискретные моменты времени, может соответствовать бесконечное множество непрерывных функций. Непрерывные функции, совпадающие с заданными дискретными, называются огибающими решетчатой функции.

3.6.2. Особенности прохождения непрерывного сигнала в цифровых системах

В настоящее время применяются преимущественно цифровые ИС, что вызвано как постоянным повышением требований к таким системам, так и стремительным ростом возможностей цифровой вычислительной техники.

Несмотря на то что в цифровых системах происходит некоторая потеря информации ввиду прерывистого характера ее передачи, они имеют ряд преимуществ перед непрерывными. Особенно характерно это для цифровых систем управления, в которых контролируемые и управляемые процессы имеют в основном аналоговый характер. В большинстве таких систем циркулируют как аналоговые, так и цифровые сигналы. Поэтому необходимо преобразование сигналов, обеспечивающее взаимодействие цифровых и аналоговых элементов системы.

При этом информационная или управляющая система содержит процессор, аналогово-цифровой (АЦП) и цифроаналоговый (ЦАП) преобразователи. ЦАП, или декодер, – это устройство, преобразующее цифровой код в аналоговый сигнал (операция декодирования цифровых входных данных). АЦП – это устройство, преобразующее аналоговый сигнал в цифровой код. Если характеризовать эти устройства только функцией передачи информации, их можно описать с помощью структурных схем и передаточных функций. На рис. 3.7 показана укрупненная схема такого типа. Если в схеме управления входным сигналом является цифровой код, то она состоит из цифровой ЭВМ, ЦАП, управляемого объекта (процесса) и АЦП.

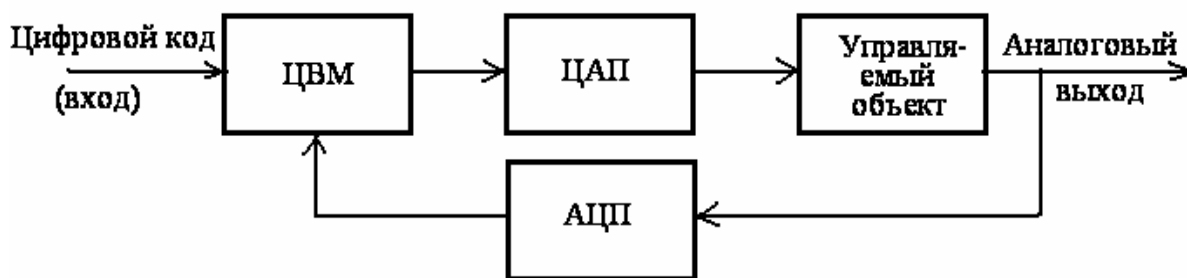


Рис. 3.7

ЭВМ преобразует входной сигнал в цифровой управляющий сигнал в соответствии с алгоритмом, заложенным в программу. ЦАП преобразует цифровой код ЭВМ в аналоговый сигнал, который воздействует на управляемый процесс. АЦП преобразует выходной сигнал управляемого процесса в последовательные импульсы.

Если применяются аналоговый вход и аналоговый выход, то используется схема, приведенная на рис. 3.8.

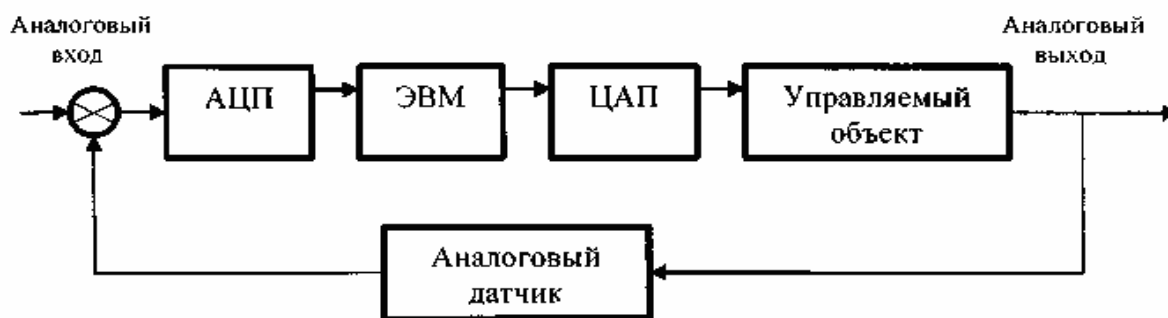


Рис. 3.8

В современной технике управления и обработки информации цифровые системы обладают следующими главными преимуществами:

1. Повышенной чувствительностью, высокой надежностью, отсутствием дрейфа, высокой помехоустойчивостью.
2. Небольшими габаритами и массой.
3. Низкой стоимостью, удобством при программировании.
4. Высокой гибкостью по сравнению с аналоговыми системами, которая заключается в возможности изменения алгоритма преобразования информации без каких-либо изменений в аппаратных средствах системы.
5. Возможностью работы в режиме разделения времени.

ЭВМ в цифровой системе управления работает в режиме реального времени. Систему обработки данных в режиме реального времени можно определить как систему, получающую исходные данные и выдающую результаты с такой скоростью, которая обеспечивает своевременную реакцию системы на изменения, происходящие во внешней

среде. Время ответа управляющей ЭВМ должно соответствовать времени регулирования управляемого процесса.

На рис. 3.9 представлена схема, имитирующая прохождение непрерывного сигнала через ЦВМ, включенную в информационную систему.

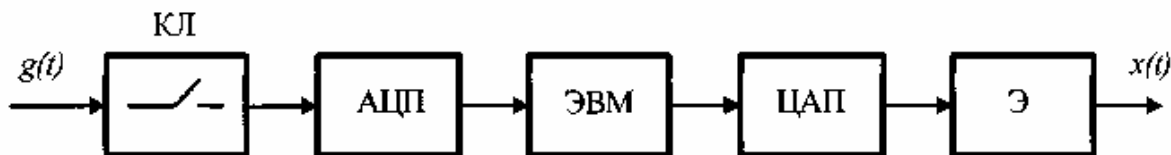


Рис. 3.9

Составные элементы этой схемы:

- ключ (КЛ), или импульсный элемент, преобразующий непрерывный сигнал в дискретный (дискретизатор);
- аналогово-цифровой преобразователь (АЦП), преобразующий амплитуду дискретного сигнала в цифровой код;
- цифроаналоговый преобразователь (ЦАП);
- экстраполятор (Э).

Дискретный сигнал на выходе ключа может принимать любое значение в заданном амплитудном диапазоне. Далее сигнал поступает в ЭВМ в цифровой бинарной форме со скоростью, соответствующей интервалу дискретизации по времени, для обработки по заданному алгоритму. После ЭВМ цифровой сигнал трансформируется преобразователем цифра-аналог в дискретно-аналоговую форму. Наконец, экстраполятор приводит сигнал к аналоговому непрерывному виду, форма которого определяется характеристиками экстраполятора.

Влияние на динамику оказывают лишь три из перечисленных элементов: ключ, ЭВМ и экстраполятор, а характеристики АЦП и ЦАП обычно не влияют на математическое описание системы; поэтому схему, изображённую рис. 3.10, можно упростить.

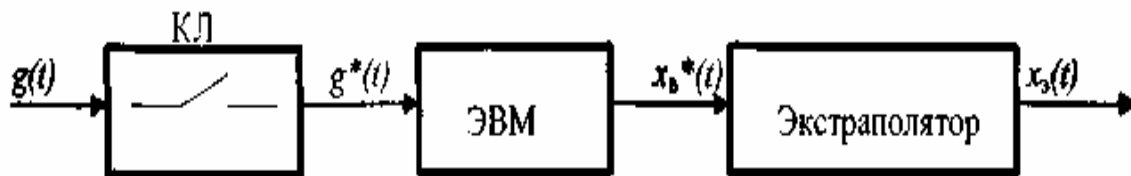


Рис. 3.10

В дальнейшем предполагается, что анализ схемы с ЦВМ справедлив при следующих ограничениях:

- Шаг дискретности T_0 постоянен.
- Запоздыванием, создаваемым процессом вычисления в ЭВМ, можно пренебречь.

- ЭВМ выполняет любую линейную операцию (дифференцирование, интегрирование, упреждение, запаздывание, решение дифференциальных и интегральных уравнений и т.д.).

- ЭВМ работает в реальном времени.

- ЭВМ может использовать настоящую и прошлую, но не будущую информацию (принцип физической осуществимости).

ИС с помощью ЦАП дискретизирует сигнал по времени и квантует по уровню. Можно показать, что квантование по уровню создает на выходе ошибку второго порядка малости по сравнению с эффектом от дискретизации по времени (ошибку квантования по уровню можно считать некоторым возмущающим воздействием). В связи с этим при дальнейшем рассмотрении динамики системы квантованием по уровню можно пренебречь.

3.7. Вейвлеты и вейвлетный анализ сигналов

3.7.1. Основы вейвлет-преобразования сигналов

Вейвлет-преобразование сигналов является обобщением спектрального анализа, типичный представитель которого – классическое преобразование Фурье. Термин «вейвлет» (wavelet) в переводе с английского означает «маленькая (короткая) волна». Вейвлеты – это обобщенное название семейств математических функций определенной формы, которые локальны во времени и по частоте и в которых все функции получаются из одной базовой (порождающей) посредством ее сдвигов и растяжений по оси времени. Вейвлет-преобразования рассматривают анализируемые временные функции в терминах колебаний, локализованных по времени и частоте. Как правило, вейвлет-преобразование (WT) подразделяют на дискретное (DWT) и непрерывное (CWT).

DWT используется для преобразований и кодирования сигналов, CWT – для анализа сигналов. Вейвлет-преобразования в настоящее время, принимаемые на вооружение для огромного числа разнообразных применений, нередко заменяют обычное преобразование Фурье. Это наблюдается во многих областях, включая молекулярную динамику, квантовую механику, астрофизику, геофизику, оптику, компьютерную графику и обработку изображений, анализ ДНК, исследования белков, исследования климата, общую обработку сигналов и распознавание речи.

Вейвлетный анализ представляет собой особый тип линейного преобразования сигналов и отображаемых этими сигналами физических данных о процессах и физических свойствах природных сред и объектов. Базис собственных функций, по которому проводится вейвлетное разложение сигналов, обладает многими специфическими свойствами и возможностями. Вейвлетные функции базиса позволяют

сконцентрировать внимание на тех или иных локальных особенностях анализируемых процессов, которые не могут быть выявлены с помощью традиционных преобразований Фурье и Лапласа. К таким процессам в геофизике относятся поля различных физических параметров природных сред. В первую очередь это касается полей температуры, давления, профилей сейсмических трасс и других физических величин. Принципиальное значение имеет возможность вейвлетов анализировать нестационарные сигналы с изменением компонентного содержания во времени или в пространстве.

Вейвлеты имеют вид коротких волновых пакетов с нулевым интегральным значением, локализованных по оси аргументов (независимых переменных), инвариантных к сдвигу и линейных к операции масштабирования (сжатия/растяжения). По локализации во временном и частотном представлении вейвлеты занимают промежуточное положение между гармоническими (синусоидальными) функциями, локализованными по частоте, и функцией Дирака, локализованной во времени.

Теория вейвлетов не является фундаментальной физической теорией, но она дает удобный и эффективный инструмент для решения многих практических задач. Основная область применения вейвлетных преобразований – анализ и обработка сигналов и функций, нестационарных во времени или неоднородных в пространстве, когда результаты анализа должны содержать не только общую частотную характеристику сигнала (распределение энергии сигнала по частотным составляющим), но и сведения об определенных локальных координатах, на которых проявляют себя те или иные группы частотных составляющих или происходят быстрые изменения частотных составляющих сигнала. По сравнению с разложением сигналов на ряды Фурье, вейвлеты способны с гораздо более высокой точностью представлять локальные особенности сигналов, вплоть до разрывов 1-го рода (скачков). В отличие от преобразований Фурье, вейвлет-преобразование одномерных сигналов обеспечивает двумерную развертку, при этом частота и координата рассматриваются как независимые переменные, что дает возможность анализа сигналов сразу в двух пространствах.

Одна из главных и особенно плодотворных идей вейвлетного представления сигналов на различных уровнях декомпозиции (разложения) заключается в разделении функций приближения к сигналу на две группы: аппроксимирующую – грубую, с достаточно медленной временной динамикой изменений, и детализирующую – с локальной и быстрой динамикой изменений на фоне плавной динамики, с последующим дроблением функций приближения и детализацией на других уровнях декомпозиции сигналов. Это возможно как во временной, так и в частотной областях представления сигналов вейвлетами.

3.7.2. Основные свойства вейвлетных преобразований сигналов

Аналитика вейвлетных преобразований сигналов определяется математической базой разложения сигналов, которая аналогична преобразованиям Фурье. Основной отличительной особенностью вейвлет-преобразований является новый базис разложения сигналов – вейвлетные функции. Свойства вейвлетов принципиально важны как для самой возможности разложения сигналов по единичным вейвлетным функциям, так и для целенаправленных действий над вейвлетными спектрами сигналов, в том числе с последующей реконструкцией сигналов по обработанным вейвлетным спектрам.

Вейвлеты могут быть ортогональными, полуортогональными, биортогональными. Эти функции могут быть симметричными, асимметричными и несимметричными, с компактной областью определения и без таковой. Некоторые функции имеют аналитическое выражение, другие – быстрый алгоритм вычисления вейвлет-преобразования. Вейвлеты различаются также степенью гладкости. Для практики желательно было бы иметь ортогональные симметричные и асимметричные вейвлеты. К сожалению, такими вейвлетами являются лишь вейвлеты Хаара, которые не обладают достаточной гладкостью и не подходят для большинства приложений. Наибольшее применение находят биортогональные вейвлеты.

3.7.3. Вейвлетный кратномасштабный анализ сигналов

В практике передачи информации часто требуется представить сигнал в виде совокупности его последовательных приближений. Например, при просмотре и передаче изображений с выборкой из некоторой базы данных можно сначала передать грубую его версию, а затем (при необходимости) последовательно ее уточнять. При сжатии изображений без потери качества можно убирать из изображения незначимые мелкомасштабные детали.

В общем случае произвольный информационный сигнал в теории цифровых временных рядов обычно рассматривается в виде суммы разнотипных составляющих: региональной функции тренда – средних значений по большим интервалам усреднения, циклических компонент с определенным периодом повторения, как правило, достаточно гладких по форме, локальных особенностей (аномалий) разного порядка, вплоть до интервенций – резких изменений в определенные редкие моменты, и флуктуаций значений более высокого порядка (шумов) вокруг всех вышеперечисленных составляющих сигнала. Инструментом разделения (декомпозиции) сигналов на такие составляющие с учетом разрешения по времени и по частоте, анализа их порядка и

реконструкции сигналов из определенных составляющих (или с исключением определенных составляющих, например шумов или малозначимых деталей) является кратномасштабный (многомасштабный) анализ (КМА).

КМА позволяет получать хорошее разрешение по времени на высоких частотах и хорошее разрешение по частоте на низких частотах. Этот подход становится особенно эффективным, когда сигнал имеет высокочастотные компоненты короткой длительности и протяженные низкочастотные компоненты. Именно такие сигналы и встречаются чаще всего на практике.

Идея кратномасштабного анализа заключается в том, что разложение сигнала производится по ортогональному базису, образованному сдвигами и кратномасштабными копиями вейвлетной функции. Свертка сигнала с вейвлетами позволяет выделять характерные особенности сигнала в области локализации этих вейвлетов, причем, чем больший масштаб имеет вейвлет, тем более широкая область сигнала будет оказывать влияние на результат свертки. Теория кратномасштабного анализа базируется на теории функциональных пространств. Понятие КМА (Multiresolution analyses) является фундаментальным в теории вейвлетов. Это определяется тем, что для КМА возможен каскадный алгоритм вычислений, подобный быстрому преобразованию Фурье.

3.8. Фрактальные стохастические процессы

3.8.1. Что такое фракталы и хаос

Когда-то людям казалось, что геометрия в природе ограничивается такими простыми фигурами, как линия, круг, коническое сечение, многоугольник, сфера, квадратичная поверхность, а также их комбинациями. К примеру, что может быть красивее утверждения о том, что планеты в нашей солнечной системе движутся вокруг солнца по эллиптическим орбитам? Этот замечательный закон – один из трех постулатов планетарного движения, сформулированных И. Кеплером на основе наблюдений и измерений, сделанных Тихо Браге. Позднее Исаак Ньютон вывел закон обратных квадратов для гравитационного притяжения как решение некоторого дифференциального уравнения, причем законы Кеплера следовали из его решения. Как в этом, так и в других случаях, когда применение простых геометрических моделей оказалось удачным, это привело к огромным научным достижениям.

Однако многие природные системы настолько сложны и нерегулярны, что использование только знакомых объектов классической геометрии для их моделирования представляется безнадежным. Как, к примеру, построить модель горного хребта или кроны дерева в терми-

нах геометрии? Как описать то многообразие биологических конфигураций, которое мы наблюдаем в мире растений и животных? Представьте себе всю сложность системы кровообращения, состоящей из множества капилляров и сосудов и доставляющей кровь к каждой клетке человеческого тела. Представьте, как хитроумно устроены легкие и почки, напоминающие по структуре деревья с ветвистой кроной.

Столь же сложной и нерегулярной может быть и динамика реальных природных систем. Как подступиться к моделированию каскадных водопадов или турбулентных процессов, определяющих погоду? Какая математика отвечает за ритмы сердца и головного мозга, наблюдаемые на электрокардиограмме и энцефалограмме, в особенности за те внезапные приступы аритмии, которые могут вызвать сбой в работе сердца? Можно ли математически описать внезапное возникновение волны паники на финансовых рынках или даже построить математическую модель социального поведения?

Фракталы и математический хаос – подходящие средства для исследования поставленных вопросов. Термин «фрактал» относится к некоторой статичной геометрической конфигурации, такой, как мгновенный снимок водопада. Хаос – термин динамики, используемый для описания явлений, подобных турбулентному поведению погоды.

Нередко то, что мы наблюдаем в природе, интригует нас бесконечным повторением одного и того же узора, увеличенного или уменьшенного во сколько угодно раз. Например, у дерева есть ветви, а на них есть ветки поменьше и т.д. Теоретически элемент «разветвление» повторяется бесконечно много раз, становясь все меньше и меньше. То же самое можно заметить, разглядывая фотографию горного рельефа. Попробуйте немного приблизить изображение горной гряды – вы снова увидите горы. Приблизьте картинку еще – вы по-прежнему будете различать нечто, напоминающее горы, благодаря вашей способности (статистической по сути) различать тип объекта на рисунке. Так проявляется характерное для фракталов свойство самоподобия.

Во многих работах по фракталам самоподобие используется в качестве определяющего свойства, согласно которому фракталы должны определяться в терминах фрактальной (дробной) размерности. Отсюда и происхождение слова «фрактал». Понятие дробной размерности представляет собой весьма сложную концепцию. Прямая – это одномерный объект, а плоскость – двумерный. Хорошенько перекрутив прямую или плоскость, можно повысить размерность полученной конфигурации, при этом новая размерность обычно будет дробной в некотором смысле, который нам предстоит уточнить. Связь дробной размерности и самоподобия состоит в том, что с помощью самоподобия

можно сконструировать множество дробной размерности наиболее простым образом. Даже в случае гораздо более сложных фракталов, когда чистое самоподобие отсутствует, имеется почти полное повторение базовой формы во все более в более уменьшенном виде.

В английском языке хаос обычно определяется как состояние полного беспорядка или неразберихи. Некоторые словари прибегают к понятию состояния, в котором правит случай. Термин «хаос» в математике используется в узком смысле.

Хотя универсального определения математического хаоса не существует, имеется полное согласие в том, что любой вид хаоса обладает свойством непредсказуемости. Как ни странно, оно не эквивалентно случайному поведению. По сути дела, математический хаос – это характерная черта именно детерминированных динамических систем. Поэтому наблюдаемые в состоянии хаоса флуктуации только кажутся случайными – их значения полностью предопределены входными параметрами. Но на практике мы никогда не располагаем абсолютно точной информацией о начальных условиях. Ошибки, пусть и ничтожные, всегда имеют место при измерении входных параметров. То, что кажется нам случайным результатом на выходе динамической системы, обусловлено большими ошибками, которые могут появиться, когда система ведет себя хаотично.

Когда-то считалось, что в детерминированной системе, при наличии достаточного объема вычислительных ресурсов, мы всегда в состоянии сделать значимое предсказание (например, дать надежный прогноз погоды), несмотря на маленькие ошибки измерения текущего состояния. В присутствии хаоса это не так. Никакой самый мощный компьютер не позволит нам сделать точный прогноз на основе математической системы с существенной зависимостью от начальных условий.

Наиболее интересный вопрос теории фракталов и хаоса состоит в том, как связать эти понятия воедино. Многие важные фракталы, включая снежинку Коха, ковер Серпинского и классическое множество Кантора, могут быть получены как аттракторы систем итерированных функций. Анализ этих систем итерированных функций указывает путь к построению хаотических операторов, ассоциированных с упомянутыми фракталами.

Как видим, изучение и использование фракталов и хаоса открывают замечательные возможности как в исследовании бесконечного числа приложений, так и в области чистой математики.

3.8.2. Теория фрактальных процессов в телекоммуникациях

В отличие от детерминированных фракталов, стохастические фрактальные процессы не обладают четким сходством составных частей в мельчайших деталях. Несмотря на это, стохастическая самоподобность является свойством, которое может быть проиллюстрировано наглядно и оценено математически.

Для количественной оценки и описания пульсирующей структуры телекоммуникационного трафика в большинстве случаев достаточно использовать статистические характеристики 2-го порядка. Корреляционная функция самоподобного процесса играет весьма важную роль, являясь основным критерием, относительно которого успешно определяется масштабная инвариантность подобных процессов. Существование корреляции «на расстоянии» называется долговременной зависимостью (ДВЗ). Отличием корреляционной функции самоподобных процессов от функции, которая обычно имеет место, является то, что корреляция как функция временной задержки предполагает полиномиальное, а не экспоненциальное убывание.

В телекоммуникационных приложениях свойствам стохастической самоподобности (фрактальности) удовлетворяют измеренные трафиковые трассы. Здесь предполагается, что мерой схожести является вид трафика с соответствующей амплитудной нормировкой. Для измеренных трафиковых трасс сложно наблюдать четкую структуру, однако самоподобность позволяет учитывать стохастическую природу многих сетевых устройств и событий, которые совместно влияют на сетевой трафик. Будем считать, что трафик телекоммуникационной сети является выборкой реализации стохастического процесса. Тогда, ослабив степень схожести обоих процессов и выбрав некоторую статистическую характеристику перемасштабированного временного ряда, можно получить точную подобность математических объектов и асимптотическую подобность их конкретных выборок относительно этих ослабленных критериев схожести.

Последние исследования локального и глобального трафика показали, что сетевой трафик проявляет изменчивость в широком диапазоне масштабов времени. Поразительна повсеместность этого явления, наблюдаемого в различных сетевых технологиях, от Ethernet до АТМ, LAN и WAN, сжатом видео и WWW-трафике, основанном на HTTP. Такая масштабно-инвариантная изменчивость не совместима с традиционными моделями сетевого трафика, которые проявляют пульсирующий характер на коротких масштабах времени, но сильно сглажены на больших масштабах времени, поэтому в них отсутствует ДВЗ. Поскольку инвариантная к масштабу пульсирующая структура тра-

фика может оказывать сильное влияние на производительность сети, то анализ причин и последствий самоподобности в трафике является очень важной задачей. Многочисленные измерения сделали очевидным, что инвариантная к масштабу пульсирующая структура является не отдельным, побочным явлением, а скорее характерной особенностью, сложившейся в пределах сетевых окружений.

Известны работы по физическому моделированию фрактальных процессов, в которых делаются попытки объяснить физические причины самоподобности в сетевом трафике. Эти исследования основаны на сетевых механизмах и опытно установленных свойствах распределенных систем, которые совместно генерируют самоподобную пульсирующую структуру, мультиплексируя процессы на сетевом уровне.

Ведутся работы по разработке математических моделей долговременно зависимого трафика с позиции облегчения анализа эффективности при построении очередей. Эти работы чрезвычайно важны, поскольку позволяют очертить границы сетевой производительности, исследуя характер построения очередей с долговременно зависимым входом. В частности, показывается, что распределение длины очереди в бесконечном буфере системы при долговременно зависимом процессе на входе затухает медленнее, чем при экспоненциальном (или субэкспоненциальном). Напротив, для кратковременно зависимых процессов на входе затухание имеет экспоненциальный характер.

Когда трафик на входе является самоподобным, буферизация оказывается неэффективной, поскольку возникают несоразмерные задержки в очереди.

Самоподобность подразумевает существование корреляционной структуры на промежутке, который может использоваться с целью управления трафиком. Для этих целей на больших временных масштабах извлекают информацию о корреляционных свойствах анализируемых процессов, корректируя выходной характер структуры управления перегрузками.

Стохастический процесс называется фрактальным, когда некоторые из его важных статистических характеристик проявляют свойства масштабирования с соответствующими масштабными показателями.

Поскольку масштабирование математически приводит к степенным соотношениям в масштабируемых величинах (т.е. степенной зависимости статистических характеристик), можно сделать вывод, что, например, телекоммуникационный трафик проявляет фрактальные свойства, когда некоторые из его оцененных статистических характеристик имеют степенную зависимость в широком временном или частотном диапазоне.

3.9. Энтропия

Установив, что случайные процессы являются адекватной моделью сигналов, мы получим возможность воспользоваться результатами и мощным аппаратом теории случайных процессов. Кроме того, обнаружив, что некоторые типы непрерывных сигналов допускают дискретное представление, мы упрощаем задачу, сводя все к рассмотрению случайных величин.

Это не означает, что теория вероятностей и теория случайных процессов дают готовые ответы на все вопросы о сигналах: подход с новых позиций выдвигает такие вопросы, которые раньше просто не возникали. Так и родилась теория информации, специально рассматривающая сигнальную специфику случайных процессов. При этом были введены принципиально новые понятия, которые мы рассмотрим далее, и получены новые результаты, имеющие характер научных открытий.

3.9.1. Понятие неопределенности

Первым специфическим понятием теории информации является понятие неопределенности случайного объекта, для которой удалось ввести количественную меру, названную энтропией. Начнем с простейшего варианта – со случайного события. Пусть, например, некоторое событие может произойти с вероятностью 0,99 и не произойти с вероятностью 0,01, а другое событие имеет вероятности соответственно 0,5 и 0,5. Очевидно, что в первом случае результатом опыта «почти наверняка» является наступление события, во втором же случае неопределенность исхода так велика, что от прогноза разумнее воздержаться.

Для характеристики размытости распределений широко используется второй центральный момент (дисперсия) или доверительный интервал. Однако эти величины имеют смысл лишь для случайных числовых величин и не могут применяться к случайным объектам, состояния которых различаются качественно, хотя и в этом случае можно говорить о большей или меньшей неопределенности исхода опыта. Следовательно, мера неопределенности, связанной с распределением, должна быть некоторой его числовой характеристикой, функционалом от распределения, никак не связанным с тем, в какой шкале измеряются реализации случайного объекта.

3.9.2. Энтропия и ее свойства

Примем (пока без обоснования) в качестве меры неопределенности случайного объекта A с конечным множеством возможных состояний A_1, \dots, A_n с соответствующими вероятностями p_1, \dots, p_n величину

$$H(A) = H(\{p_i\}) = -\sum_{k=1}^n p_i \log p_i, \quad (28)$$

которую и называют энтропией случайного объекта A (или распределения $\{p_i\}$). Этот функционал обладает следующими свойствами, которые вполне естественны для меры неопределенности.

1. $H(p_1, \dots, p_n) = 0$ в том и только в том случае, когда какое-нибудь одно значение из множества $\{p_i\}$ равно единице (а остальные – нули). Это соответствует случаю, когда исход опыта может быть предсказан с полной достоверностью, т.е. когда отсутствует всякая неопределенность. Во всех других случаях энтропия положительна. Это свойство проверяется непосредственно.

2. $H(p_1, \dots, p_n)$ достигает наибольшего значения при $p_1 = p_2 = \dots = p_n = 1/n$.

Действительно, вариация H по p_i при условии $\sum p_i = 1$ дает $p_i = \text{const} = \frac{1}{n}$.

3. Если A и B – независимые случайные объекты, то

$$H(A \cap B) = H(\{\pi_{ik} = p_i q_k\}) = H(\{p_i\}) + H(\{q_k\}) = H(A) + H(B).$$

Это свойство проверяется непосредственно.

4. Если A и B – зависимые случайные объекты, то

$$H(A \cap B) = H(A) + H(B|A) = H(B) + H(A|B),$$

где условная энтропия $H(B|A)$ определяется как математическое ожидание энтропии условного распределения.

Это свойство проверяется непосредственно.

5. Имеет место неравенство $H(A) \geq H(A|B)$, что согласуется с интуитивным представлением о том, что знание состояния объекта B может только уменьшить неопределенность объекта A , а если они независимы, то оставит ее неизменной.

Это свойство доказывается с помощью тождественного неравенства

$$\sum_k \lambda_k f(x_k) \geq f\left(\sum_k \lambda_k x_k\right),$$

справедливого для любой выпуклой функции $f(x)$, если в этом неравенстве положить $f(x) = \log x$, $\lambda_k = p_k$, $x_k = q_{kl}$.

Как видим, свойства функционала H позволяют использовать его в качестве меры неопределенности. Интересно отметить, что если пойти в обратном направлении, т.е. задать желаемые свойства меры неопределенности и искать обладающий указанными свойствами функционал, то уже только условия 2 и 4 позволят найти этот функционал, причём единственным образом (с точностью до постоянного множителя).

3.9.3. Дифференциальная энтропия

Обобщение столь полезной меры неопределенности на непрерывные случайные величины наталкивается на ряд сложностей. Можно по-разному преодолеть эти сложности; выберем кратчайший путь. Прямая аналогия

$$-\sum p_k \log p_k \rightarrow \int_X p(x) \log p(x) dx$$

не приводит к нужному результату; плотность $p(x)$ является размерной величиной (размерность плотности $p(x)$ обратна размерности x , так как элемент вероятности $dP(x) = p(x)dx$ безразмерен), а логарифм размерной величины не имеет смысла. Однако положение можно исправить, умножив $p(x)$ под знаком логарифма на величину ϵ , имеющую ту же размерность, что и x :

$$-\sum p_k \log p_k \rightarrow \int_X p(x) \log[\epsilon \cdot p(x)] dx.$$

Теперь величину ϵ можно принять равной единице измерения x , что приводит к функционалу

$$h(x) = -\int_X p(x) \log p(x) dx, \quad (29)$$

который получил название дифференциальной энтропии. Это аналог энтропии дискретной величины, но аналог условный, относительный: ведь единица измерения произвольна. Запись (29) означает, что мы как бы сравниваем неопределенность случайной величины, имеющей плот-

ность $p(x)$, с неопределенностью случайной величины, равномерно распределенной в единичном интервале. Поэтому величина $h(X)$ в отличие от $H(X)$ может быть не только положительной. Кроме того, $h(X)$ изменяется при нелинейных преобразованиях шкалы x , что в дискретном случае не играет роли. Остальные свойства $h(X)$ аналогичны свойствам $H(X)$, что делает дифференциальную энтропию очень полезной мерой.

Пусть, например, задача состоит в том, чтобы, зная лишь некоторые ограничения на случайную величину (моменты, пределы сверху и снизу области возможных значений и т.п.), задать для дальнейшего – каких-то расчетов или моделирования – конкретное распределение. Один из подходов к решению этой задачи дает принцип максимума энтропии: из всех распределений, отвечающих данным ограничениям, следует выбирать то, которое обладает максимальной дифференциальной энтропией. Смысл этого критерия состоит в том, что, выбирая экстремальное по энтропии распределение, мы гарантируем наибольшую неопределенность, связанную с ним, т.е. имеем дело с наихудшим случаем при данных условиях.

Важным шагом в построении теории информации является введение количественной меры неопределенности – энтропии. Оказывается, что функционал (28) обладает качествами, которые логично ожидать от меры неопределенности, и, наоборот, единственным функционалом с такими свойствами является именно функционал энтропии. Обобщение понятия энтропии на непрерывные случайные величины приводит к выводу, что такое обобщение – дифференциальная энтропия – возможно лишь как относительная мера.

Оказывается, что энтропия связана с глубокими свойствами случайных процессов. Например, для дискретных процессов имеет место свойство асимптотической равновероятности реализаций из высоковероятной группы.

3.9.4. Фундаментальное свойство энтропии случайного процесса

Особое значение энтропия приобретает в связи с тем, что она связана с очень глубокими, фундаментальными свойствами случайных процессов. Покажем это на примере процесса с дискретным временем и дискретным конечным множеством возможных состояний.

Назовем каждое такое состояние символом, множество возможных состояний – алфавитом, их число m – объемом алфавита. Число всевозможных последовательностей длины n , очевидно, равно m^n . Появление конкретной последовательности можно рассматривать как реа-

лизацию одного из m^n возможных событий. Зная вероятности символов и условные вероятности появления следующего символа, если известен предыдущий (в случае их зависимости), можно вычислить вероятность $P(C)$ для каждой последовательности C . Тогда энтропия множества $\{C\}$, по определению, равна

$$H_n = - \sum_{C \in \{C\}} P(C) \log P(C). \quad (30)$$

Определим энтропию процесса H – среднюю неопределенность, приходящуюся на один символ, следующим образом:

$$H = \lim_{n \rightarrow \infty} \frac{H_n}{n}. \quad (31)$$

Существование такого предела для любого стационарного процесса можно строго доказать.

На множестве $\{C\}$ можно задать любую числовую функцию $f_n(C)$, которая, очевидно, является случайной величиной. Определим $f_n(C)$ с помощью соотношения $f_n(C) = -\frac{1}{n} \log P(C)$. Математическое ожидание этой функции $Mf_n(C) = \sum_C P(C) f_n(C) = -\frac{1}{n} \sum_C P(C) \log P(C)$, откуда следует, что

$$M \left[-\frac{1}{n} \log P(C) \right] = \frac{H_n}{n} \quad \text{и} \quad \lim_{n \rightarrow \infty} M \left[-\frac{1}{n} \log P(C) \right] = H. \quad (32)$$

Соотношение (32), весьма интересное уже само по себе, является, однако, лишь одним из проявлений гораздо более общего свойства дискретных эргодических процессов. Оказывается, что не только математическое ожидание величины $Mf_n(C)$ при $n \rightarrow \infty$ имеет своим пределом H , но и сама эта величина $Mf_n(C)$ стремится к H при $n \rightarrow \infty$. Другими словами, как бы малы ни были $\varepsilon > 0$ и $\delta > 0$, при достаточно большом n справедливо неравенство

$$P \left[\left| \frac{1}{n} \log P(C) + H \right| > \varepsilon \right] < \delta, \quad (33)$$

т.е. близость $Mf_n(C)$ к H при больших n является почти достоверным событием.

Для большей наглядности сформулированное фундаментальное свойство случайных процессов обычно излагают следующим образом.

Для любых заданных $\varepsilon > 0$ и $\delta > 0$ можно найти такое n_0 , что реализации любой длины $n > n_0$ распадаются на два класса:

1) группа реализаций, вероятности $P(C)$ которых удовлетворяют неравенству

$$\left| \frac{1}{n} \log P(C) + H \right| > \varepsilon; \quad (34)$$

2) группа реализаций, вероятности которых этому неравенству не удовлетворяют.

Так как согласно неравенству (33) суммарные вероятности этих групп равны соответственно $(1 - \delta)$ и δ , то первая группа называется высоковероятной, а вторая – маловероятной.

Это свойство эргодических процессов приводит к ряду важных следствий, из которых три заслуживают особого внимания.

1. Независимо от того, каковы вероятности символов и каковы статистические связи между ними, все реализации высоковероятной группы приблизительно равновероятны (неравенство (34)).

В связи с этим фундаментальное свойство иногда называют «свойством асимптотической равномерности». Это следствие, в частности, означает, что по известной вероятности $P(C)$ одной из реализаций высоковероятной группы можно оценить число N_1 реализаций в этой группе:

$$N_1 = \frac{1}{P(C)}.$$

2. Энтропия H_n с высокой точностью равна логарифму числа реализаций в высоковероятной группе:

$$H_n = nH = \log N_1. \quad (35)$$

3. При больших n высоковероятная группа обычно охватывает лишь ничтожную долю всех возможных реализаций (за исключением случая равновероятных и независимых символов, когда все реализации равновероятны и $H = \log m$).

Действительно, из соотношения (26) имеем $N_1 = a^{nH}$, где a – основание логарифма. Число всех возможных реализаций есть $N = m^n = a^{n \log m}$. Доля реализаций высоковероятной группы в общем числе реализаций выражается формулой

$$N_1/N = a^{-n(\log m - H)}, \quad (36)$$

и при $H < \log m$ эта доля неограниченно убывает с ростом n . Например, если $a = 2, n = 100, H = 2,75, m = 8$, то $N_1/N = a^{-n(\log m - H)}$, т.е. к высоковероятной группе относится лишь одна тридцатимиллионная доля всех реализаций!

Строгое доказательство фундаментального свойства эргодических процессов сложно и здесь не приводится. Однако следует отметить, что в простейшем случае независимости символов это свойство является следствием закона больших чисел. Действительно, закон больших чисел утверждает, что с вероятностью, близкой к 1, в длинной реализации i -й символ, имеющий вероятность p_i , встретится примерно np_i раз. Следовательно, вероятность реализации высоковероятной группы есть $P(C) = \prod_{i=1}^m p_i^{np_i}$, откуда

$$-\log P(C) = -n \sum_{i=1}^m p_i \log p_i = nH,$$

что и доказывает справедливость фундаментального свойства в этом случае.

Таким образом, связав понятие неопределенности дискретной величины с распределением вероятности по возможным состояниям и потребовав некоторых естественных свойств от количественной меры неопределенности, приходим к выводу, что такой мерой может служить только функционал, названный энтропией. С некоторыми трудностями энтропийный подход удалось обобщить на непрерывные случайные величины (введением дифференциальной энтропии) и на дискретные случайные процессы.

3.10. Количество информации

В основе всей теории информации лежит открытие, что информация допускает количественную оценку. В простейшей форме эта идея была выдвинута в 1928 г. Хартли, но заверченный и общий вид придал ей Шэннон в 1948 г. Не останавливаясь на том, как развивалось и обобщалось понятие количества информации, дадим сразу его современное толкование.

3.10.1. Количество информации как мера снятой неопределенности

Процесс получения информации можно интерпретировать как изменение неопределенности в результате приема сигнала. Проиллюстрируем это на примере достаточно простого случая, когда передача сигнала происходит при следующих условиях: 1) полезный (отправляемый) сигнал x_i является последовательностью статистически независимых символов с вероятностями $p(x_i)$, $i = 1, \dots, m$; 2) принимаемый сигнал является последовательностью символов y_k того же алфавита; 3) если шумы (искажения) отсутствуют, то принимаемый сигнал совпадает с отправляемым $y_k = x_i$; 4) если шум имеется, то его действие приводит к тому, что данный символ может либо остаться прежним (i -м), либо быть подмененным любым другим (k -м) символом, вероятность этого равна $p(y_k|x_i)$; 5) искажение очередного символа является событием, статистически независимым от того, что произошло с предыдущими символами.

Итак, до получения очередного символа ситуация характеризуется неопределенностью того, какой символ будет отправлен, т.е. априорной энтропией $H(X)$. После получения символа y_k неопределенность относительно того, какой символ был отправлен, меняется: в случае отсутствия шума она вообще исчезает (апостериорная энтропия равна нулю, поскольку точно известно, что был передан символ $x_k = y_k$), а при наличии шума мы не можем быть уверены, что полученный нами символ и есть отправленный, и возникает неопределенность, характеризуемая апостериорной энтропией

$$H(X|y_k) = H(\{p(x_i|y_k)\}) > 0.$$

Определим теперь количество информации как меру снятой неопределенности: числовое значение количества информации о некотором объекте равно разности априорной и апостериорной энтропии этого объекта:

$$I(X,Y) = H(X) - H(X|Y). \quad (37)$$

Можно показать, что

$$I(X,Y) = H(Y) - H(Y|X). \quad (38)$$

В явной форме равенство (1) запишется так:

$$\begin{aligned}
 I(X, Y) &= H(X) - H(X|Y) = -\sum_{i=1}^m p(x_i) \log p(x_i) + \\
 &+ \sum_{k=1}^m p(y_k) \sum_{i=1}^m p(x_i|y_k) \log p(x_i|y_k) = \\
 &= -\sum_k \sum_i p(x_i, y_k) \log p(x_i) + \sum_i \sum_k p(x_i, y_k) \log p(x_i|y_k) = \\
 &= \sum_{i=1}^m \sum_{k=1}^m p(x_i, y_k) \log \frac{p(x_i|y_k)}{p(x_i)},
 \end{aligned} \tag{39}$$

а для равенства (38) имеем

$$I(X, Y) = \sum_{i=1}^m \sum_{k=1}^m p(x_i, y_k) \log \frac{p(y_k|x_i)}{p(y_k)} \tag{40}$$

3.10.2. Количество информации как мера соответствия случайных объектов

Формулам (39) и (40) легко придать полную симметричность: умножив и разделив логарифмируемое выражение в (39) на $p(y_k)$, а в (40) на $p(x_i)$, сразу получим:

$$I(X, Y) = \sum_{i=1}^m \sum_{k=1}^m p(x_i, y_k) \log \frac{p(x_i, y_k)}{p(x_i) p(y_k)}. \tag{41}$$

Эту симметрию можно интерпретировать так: количество информации в объекте X об объекте Y равно количеству информации в объекте Y об объекте X . Таким образом, количество информации является не характеристикой одного из объектов, а характеристикой их связи, соответствия между их состояниями. Следовательно, среднее количество информации, вычисляемое по формуле (37), есть мера соответствия двух случайных объектов.

Это определение позволяет прояснить связь понятий информации и количества информации. Информация есть отражение одного объекта другим, проявляющееся в соответствии их состояний. Один объект может быть отражен с помощью нескольких других, часто какими-то лучше, чем остальными. Среднее количество информации и есть числовая характеристика степени отражения, степени соответствия. Подчеркнем, что при таком описании как отражаемый, так и отражающий объекты выступают совершенно равноправно. С одной стороны, это

подчеркивает обоюдность отражения: каждый из них содержит информацию друг о друге. Это представляется естественным, поскольку отражение есть результат взаимодействия, т.е. взаимного, обоюдного изменения состояний. С другой стороны, фактически одно явление (или объект) всегда выступает как причина, другой – как следствие; это никак не учитывается при введенном количественном описании информации.

Формула (41) обобщается на непрерывные случайные величины, если в соотношения (37) и (38) вместо H подставить дифференциальную энтропию h :

$$I(X, Y) = \int \int_{XY} p(x, y) \log \frac{p(x, y)}{p(x) p(y)}, \quad (42)$$

где $p(x)$, $p(y)$ и $p(x, y)$ – соответствующие плотности вероятностей.

Количество информации можно определить как меру уменьшения неопределенности в результате получения сигнала. Это соответствует разности энтропии до и после приема сигнала.

Среди свойств количества информации выделяются следующие: 1) количество информации (в отличие от энтропии) имеет одинаковый смысл как для дискретных, так и для непрерывных случайных объектов; 2) при обработке данных содержащееся в них количество информации не может быть увеличено. Следовательно, обработка делается лишь для представления информации в более удобном, компактном виде и – в лучшем случае – без потери полезной информации.

3.10.3. Свойства количества информации

Отметим некоторые важные свойства количества информации.

1. Количество информации в случайном объекте X относительно объекта Y равно количеству информации в Y относительно X :

$$I(X, Y) = I(Y, X). \quad (43)$$

2. Количество информации неотрицательно:

$$I(X, Y) \geq 0. \quad (44)$$

Это можно доказать по-разному. Например, варьированием $p(x, y)$ при фиксированных $p(x)$ и $p(y)$ можно показать, что минимум I , равный нулю, достигается при $p(x, y) = p(x)p(y)$.

3. Для дискретных X справедливо равенство $I(X, X) = H(X)$.

4. Преобразование $\varphi\left(\begin{smallmatrix} \bullet \\ \bullet \end{smallmatrix}\right)$ одной случайной величины не может увеличить содержание в ней информации о другой, связанной с ней, величине:

$$I(\varphi(X), Y) \leq I(X, Y). \quad (45)$$

5. Для независимых пар величин количество информации аддитивно:

$$I(\{X_i, Y_i\}) = \sum_i I(X_i, Y_i). \quad (46)$$

3.10.4. Единицы измерения энтропии и количества информации

Рассмотрим вопрос о единицах измерения количества информации и энтропии. Из определений I и H следует их безразмерность, а из линейности их связи – одинаковость их единиц. Поэтому для определенности будем говорить об энтропии. Начнем с дискретного случая. За единицу энтропии примем неопределенность случайного объекта, такого, что

$$H(X) = -\sum_{i=1}^m p_i \log p_i = 1. \quad (47)$$

Легко установить, что для однозначного определения единицы измерения энтропии необходимо конкретизировать число m состояний объекта и основание логарифма. Возьмем для определенности наименьшее число возможных состояний, при котором объект еще остается случайным, т.е. $m = 2$, и в качестве основания логарифма также возьмем число 2. Тогда из равенства $-p_1 \log_2 p_1 - p_2 \log_2 p_2 = 1$ вытекает, что $p_1 = p_2 = 1/2$. Следовательно, единицей неопределенности служит энтропия объекта с двумя равновероятными состояниями. Эта единица получила название «бит». Бросание монеты дает количество информации в один бит. Другая единица («нит») получается, если использовать натуральные логарифмы; обычно она употребляется для непрерывных величин.

3.11. Основные результаты теории информации

Теория информации вводит ряд новых понятий, описывающих информационные процессы, происходящие в любых системах, введение новых количественных параметров, позволяющих производить изме-

рения и расчеты. Часть этих понятий и величин рассмотрена выше. Однако главная ценность теории информации заключается в получении новых результатов, в открытии ранее неизвестных свойств систем.

Чтобы познакомиться с основными из этих результатов, введем еще несколько понятий и параметров информационных процессов и систем.

3.11.1. Избыточность

Одной из важнейших характеристик сигнала является содержащееся в нем количество информации. Однако по ряду причин количество информации, которое несет сигнал, обычно меньше, чем то, которое он мог бы нести по своей физической природе, т.е. информационная нагрузка на каждый элемент сигнала меньше той, которую он способен нести. Для описания этого свойства сигналов введено понятие избыточности и определена ее количественная мера.

Пусть сигнал длиной в n символов содержит количество информации I . Если это представление информации обладает избыточностью, то такое же количество информации I может быть представлено с помощью меньшего числа символов. Обозначим через n_0 наименьшее число символов, необходимое для представления I без потерь. На каждый символ в первом случае приходится $I_1 = I/n$ бит информации, во втором — $I_{1\max} = I/n_0$ бит. Очевидно, $nI_1 = n_0I_{1\max} = I$. В качестве меры избыточности R принимается относительное удлинение сигнала, соответствующее данной избыточности:

$$R = \frac{n - n_0}{n} = 1 - \frac{n_0}{n} = 1 - \frac{I_1}{I_{1\max}}. \quad (48)$$

В дискретном случае имеются две причины избыточности: неравновероятность символов и наличие статистической связи между символами. В непрерывном случае — это неэкстремальность распределений (т.е. отклонение от распределений, обладающих максимальной энтропией), что в широком смысле сводится к отклонениям от экстремальности распределения первого порядка и от минимальности связи во времени (от равномерности спектра при его ограниченности).

Не следует думать, что избыточность — явление всегда отрицательное. При искажениях, выпадениях и вставках символов именно избыточность позволяет обнаружить и исправить ошибки.

3.11.2. Скорость передачи и пропускная способность

Следующим важнейшим понятием является скорость передачи информации. Так называется количество информации, передаваемое в единицу времени. Эта величина определяется по формуле

$$R = H(X) - H(X|Y), \quad (49)$$

где указанные энтропии исчисляются на единицу времени. В дискретном случае единицей времени удобно считать время передачи одного символа, тогда в формуле (49) фигурируют априорная и апостериорная энтропии на один символ. Для непрерывных каналов единицей времени может служить либо обычная единица (например секунда), либо интервал между отсчетами; тогда в формулу (49) входят соответствующие дифференциальные энтропии. Для более наглядного представления о величине R укажем, что темп обычной речи соответствует скорости порядка 20 бит/с.

Скорость передачи информации по каналу связи зависит от многих факторов – от энергии сигнала, числа символов в алфавите избыточности, полосы частот, способа кодирования и декодирования. Если имеется возможность изменять некоторые из них, то, естественно, следует делать это так, чтобы максимально увеличить скорость передачи информации. Однако обычно существует предел, выше которого увеличение скорости невозможно. Этот предел называется пропускной способностью канала.

Скорость передачи информации зависит от условий A , в которых функционирует канал связи A , где $\{A\}$ – множество вариантов условий. Так как множество $\{A\}$ можно определить по-разному, имеет смысл говорить о нескольких типах пропускных способностей. Наиболее важным является случай, когда мощность сигнала (объем алфавита) фиксирована, а варьировать можно только способ кодирования. Именно таким образом пропускную способность определил К. Шэннон. С другой стороны, В.И. Сифоров показал, что целесообразно рассмотреть предел, к которому стремится шэнноновская пропускная способность C при стремлении мощности полезного сигнала к бесконечности. Оказалось, что все каналы связи разбиваются на два класса: каналы первого рода (терминология Сифорова), для которых указанный предел бесконечен, и каналы второго рода, имеющие конечную пропускную способность даже при бесконечной мощности передатчика. Этот предел называют собственной пропускной способностью. При других требованиях, предъявляемых к множеству $\{A\}$, мы придем к тем или иным условным пропускным способностям.

Для представления о порядках величин C приведем примеры. Прямыми измерениями установлено, что пропускные способности зрительного, слухового и тактильного каналов связи человека имеют порядок 50 бит/с. Возможно, ограничивающим фактором являются не сами рецепторы, а нервные волокна, передающие возбуждения. Если включить в канал и «исполнительные» органы человека (например, предложить ему нажимать педаль или кнопку в темпе получения сигналов), то пропускная способность близка к 10 бит/с. Интересно отметить, что многие бытовые технические устройства слабо согласованы с органами чувств человека. Например, канал телевидения имеет пропускную способность в десятки миллионов бит/с.

3.11.3. Кодирование в отсутствие шумов

С помощью введенных понятий можно рассмотреть многие информационные процессы. Начнем с дискретных систем без шумов. Здесь главное внимание привлекает проблема эффективности: важно, чтобы данная информация заняла в запоминающем устройстве как можно меньше ячеек, при передаче желательно занимать канал связи на максимально короткий срок. В такой постановке задачи легко распознается проблема устранения всякой избыточности. Однако эта проблема не тривиальна.

В теории информации доказываются теоремы об условиях достижимости этих требований. Мы не будем приводить эти теоремы; укажем лишь, что речь идет не только о принципиальной возможности, но и о разработанных процедурах построения кодов, обеспечивающих безызыточное кодирование (либо в случае невозможности этого – сколь угодно близкое к нему).

3.11.4. Кодирование при наличии шумов

Наиболее интересные и важные результаты были получены при рассмотрении передачи информации по каналам связи с шумами. В этом случае безызыточное кодирование приведет к безвозвратным потерям информации: искаженный символ нельзя ни обнаружить, ни исправить. Для борьбы с влиянием помех необходимо ввести избыточность в сигнал. Основываясь на интуитивных соображениях (например, на опыте многократного повторения), легко прийти к выводу, что при неограниченном повышении требований к малости вероятности ошибки избыточность при любом способе кодирования должна неограниченно возрастать, а скорость передачи – стремиться к нулю. Здесь имеем пример того, как сильно интуиция может привести к за-

блуждению. Шэннон показал, что существуют такие способы введения избыточности, при которых обеспечиваются одновременно и сколь угодно малая вероятность ошибки, и конечная (отличная от нуля) скорость передачи информации, причем эта скорость может быть сколь угодно близкой к пропускной способности канала. Это замечательное открытие и привлекло столь большое внимание к теории информации.

Воспроизведем упрощенное доказательство указанного утверждения. Рассмотрим схему передачи по каналу с шумом (рис. 3.11). Будем считать, что на вход кодера сигналы поступают закодированными безизбыточно. Кодер вносит в сигналы избыточность, увеличивая длительность кодовых слов. Число возможных последовательностей сразу резко увеличивается, но избыточность и состоит в том, что к отправке предназначаются не все из них, а лишь разрешенные (отмеченные на рис. 3.11 черными кружками). Согласно фундаментальному свойству энтропии, число всевозможных последовательностей длины n равно $2^{nH(X)}$, а число разрешенных к отправке равно $2^{nH} < 2^{nH(X)}$ (считаем, что энтропия исчисляется в битах); H – энтропия на символ во множестве разрешенных к отправке последовательностей («энтропия источника», или «скорость создания информации»), $H(X)$ – энтропия на символ во множестве всевозможных последовательностей. В результате воздействия шумов какие-то из символов отправленной последовательности подменяются другими и на приемный конец поступает другая, отличная от отправленной, последовательность. Поскольку $p(x|y)$ считается известным, каждой принятой последовательности соответствует $2^{nH(X|Y)}$, возможно, отправленных последовательностей. Декодирование (т.е. принятие решения о том, какая последовательность была отправлена) можно выразить как разбиение всего множества Y принимаемых последовательностей на 2^{nH} подмножеств, сопоставляемых с разрешенными к отправке: если, например, принят сигнал i -й группы, то считается, что был послан i -й разрешенный сигнал, который тут же выдается в «чистом» виде получателю.

Итак, в построенной модели проблема выбора кода (способа передачи) состоит в размещении разрешенных последовательностей среди множества всевозможных на передающем конце и в разбиении этого множества из $2^{nH(X)}$ последовательностей на 2^{nH} подмножеств на приемном конце. Идея Шэннона состоит не в том, чтобы указать некоторый регулярный способ кодирования со сколь угодно малой вероятностью ошибки, а в том, чтобы показать, что такой код вообще существует.

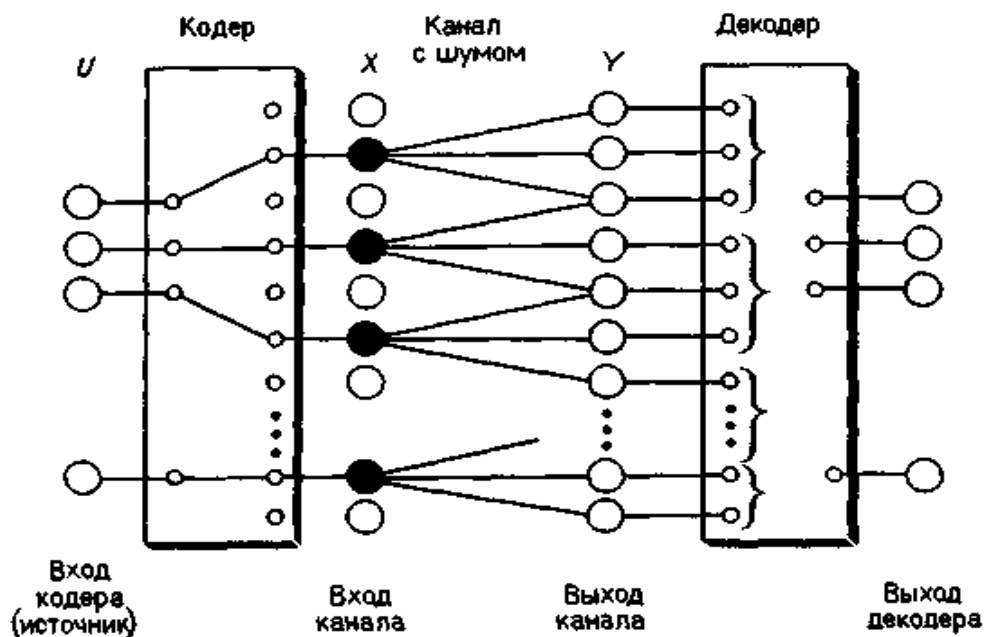


Рис. 3.11

Само знание потенциальных возможностей имеет огромное значение: сравнение характеристик реальных систем с теоретическими пределами позволяет судить о достигнутом уровне и о целесообразности дальнейших затрат на его повышение. Прикладные же вопросы рассматриваются в специальном разделе теории информации – теории кодирования, которая изучает способы построения конкретных кодов и их свойства, в частности точные или граничные зависимости вероятностей ошибок от параметров кода.

Для напоминания об основных понятиях теории информации приведем вариант блок-схемы системы передачи информации (рис. 3.12), на котором приведено большинство рассмотренных ранее понятий и обозначений.

Вместе с тем некоторые специалисты различных отраслей науки некритически распространяли методы и конкретные результаты теории информации на любые информационные процессы в реальном мире. Уже сам Шэннон заметил эту тенденцию и настоятельно предостерегал от того, чтобы ей безоглядно следовать. Дело в том, что шэнноновское количество информации является характеристикой, лишь с одной стороны описывающей информационные отношения. Именно эта сторона – соответствие состояний – играет главную роль в технических устройствах. Однако в этом соответствии не проявляют себя (и, следовательно, могут не учитываться) такие свойства информации, как смысл, доброкачественность (степень истинности), ценность, полезность, временное старение, причинно-следственная направленность и

т.д. – свойства, весьма существенные для информационных отношений с участием живых организмов, людей, коллективов.

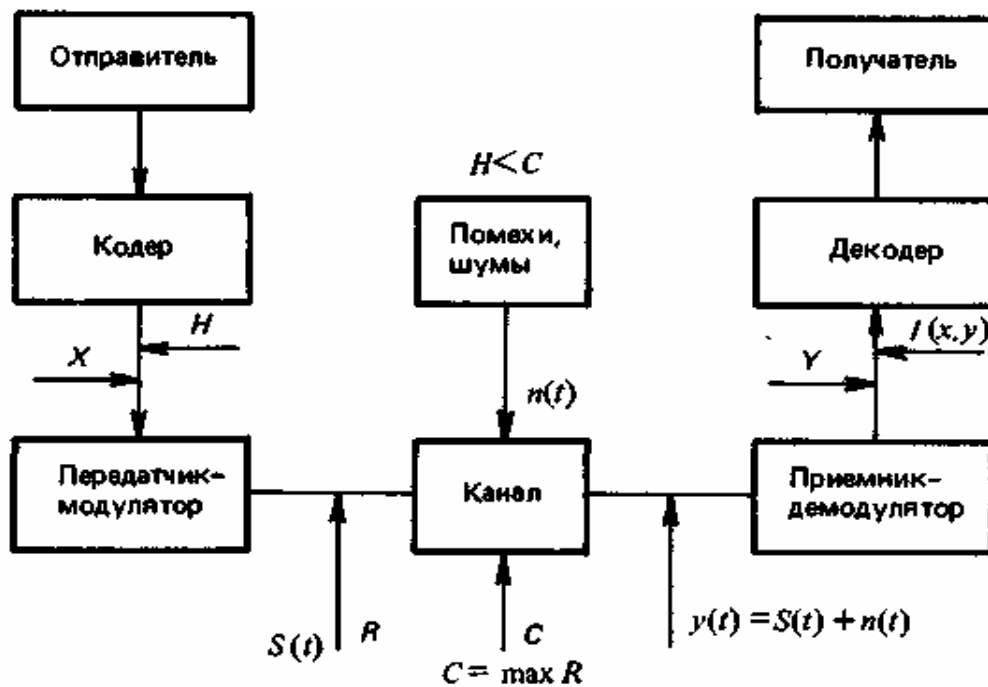


Рис. 3.12

Для системного анализа теория информации имеет двоякое значение. Во-первых, ее конкретные методы и результаты позволяют провести ряд количественных исследований информационных потоков в изучаемой или проектируемой системе (если в этом возникнет необходимость). Однако более важным является эвристическое значение основных понятий теории информации – неопределенности, энтропии, количества информации, избыточности, пропускной способности и пр. Их применение столь же важно для понимания системных процессов, как и использование понятий, связанных с временными, энергетическими процессами. Системный анализ неизбежно выходит на исследование ресурсов (наличных и необходимых), которые потребуются для решения анализируемой проблемы. Информационные ресурсы играют ведущую роль наряду с остальными ресурсами – материальными, энергетическими, временными, кадровыми.

Контрольные вопросы

1. Что такое кодирование информации и для чего оно используется?
2. Охарактеризуйте основные принципы кодирования.
3. Что такое алфавиты и как они используются?
4. Какие основные требования предъявляют к кодированию?

5. Что такое сигналы в системах и как они используются?
6. Какие основные типы сигналов вы знаете?
7. Что является основным свойством сигналов?
8. Какие классы случайных процессов вы знаете?
9. Какие математические модели реализаций случайных процессов вы знаете?
10. Что такое гармонические сигналы?
11. Что такое модулированные сигналы?
12. Что такое периодические сигналы?
13. Что такое сигналы с ограниченной энергией?
14. Что такое сигналы ограниченной длительности?
15. Что такое сигналы с ограниченной полосой частот?
16. Что такое частотно-временное представление сигналов?
17. Как осуществляется цифровое представление непрерывных сигналов?
18. Что такое решетчатые функции?
19. Охарактеризуйте основные особенности прохождения непрерывного сигнала в цифровых системах.
20. Что такое вейвлеты и вейвлетный анализ сигналов?
21. Что такое фрактальные стохастические процессы?
22. Что такое энтропия?
23. Назовите основные свойства энтропии
24. Что такое дифференциальная энтропия?
25. В чем заключается фундаментальное свойство энтропии случайного процесса?
26. Что такое количество информации и как оно определяется?
27. Охарактеризуйте основные свойства количества информации.
28. Назовите единицы измерения энтропии и количества информации.
29. Что такое избыточность информации и как она используется?
30. Что такое кодирование в отсутствие шумов?
31. Что такое кодирование при наличии шумов?

4. ДЕКОМПОЗИЦИЯ И АГРЕГИРОВАНИЕ СИСТЕМ

4.1. Модели систем как основание декомпозиции

Основной операцией анализа является представление целого в виде частей. Задача распадается на подзадачи, система – на подсистемы, цели – на подцели и т.д. При необходимости этот процесс повторяется, что приводит к иерархическим древовидным структурам. Обычно (если задача не носит чисто учебного характера) объект анализа сложен, слабоструктурирован, плохо формализован; поэтому операцию декомпозиции выполняет эксперт. Если поручить анализ одного и того же объекта разным экспертам, то полученные древовидные списки будут различаться. Качество «деревьев», построенных экспертами, зависит как от их компетентности в данной области знаний, так и от применяемой методики декомпозиции.

Обычно эксперт легко разделяет целое на части, но испытывает затруднения, если требуется привести доказательство полноты и безызыбыточности предлагаемого набора частей. Стремясь перейти от чисто эвристического, интуитивного подхода к более осознанному, алгоритмическому выполнению декомпозиции, следует понять, почему эксперт разделяет целое именно так, а не иначе, и именно на данное, а не на большее или меньшее число частей. Объяснение состоит в том, что основанием всякой декомпозиции является модель рассматриваемой системы.

Остановимся на этом важном положении подробнее. Операция декомпозиции представляется как сопоставление объекта анализа с некоторой моделью, как выделение в нем того, что соответствует элементам взятой модели. Поэтому на вопрос, сколько частей должно получиться в результате декомпозиции, можно дать следующий ответ: столько, сколько элементов содержит модель, взятая в качестве основания. Вопрос о полноте декомпозиции – это вопрос завершенности модели. Декомпозиция проводится по модели входов организационной системы (рис. 4.1), которая включает входы от «нижестоящих» систем.

Пример 1. В начале 70-х годов прошлого века проводились работы по системному анализу целей развития морского флота. Первый уровень дерева целей выглядел в виде схемы, изображенной на рис. 4.2.

Здесь клиентуры – подцель 1 и подцель 2 от «вышестоящих» систем (т.е. от народного хозяйства в целом); от «существенной среды» – подцель 3 (в данном случае от флотов капиталистических государств) и подцель 4 (имеются в виду социалистические государства). Очевидно, что такая декомпозиция неполна, поскольку отсутствует подцель,

связанная с собственными интересами морского флота. Это, по-видимому, не столько ошибка экспертов-аналитиков (абстрактно упоминается дерево целей социально-экономического развития), сколько результат преобладавшего тогда «остаточного» подхода к проблемам быта, присущего тогдашнему стилю руководства. Позднее пришлось говорить о том, что отсутствие учета компоненты соцкультбыта создало серьезные проблемы в работе флота в целом.

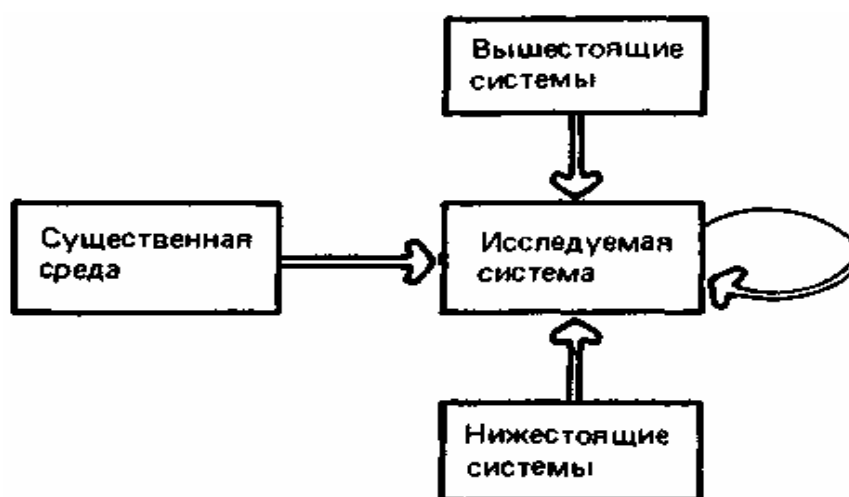


Рис. 4.1

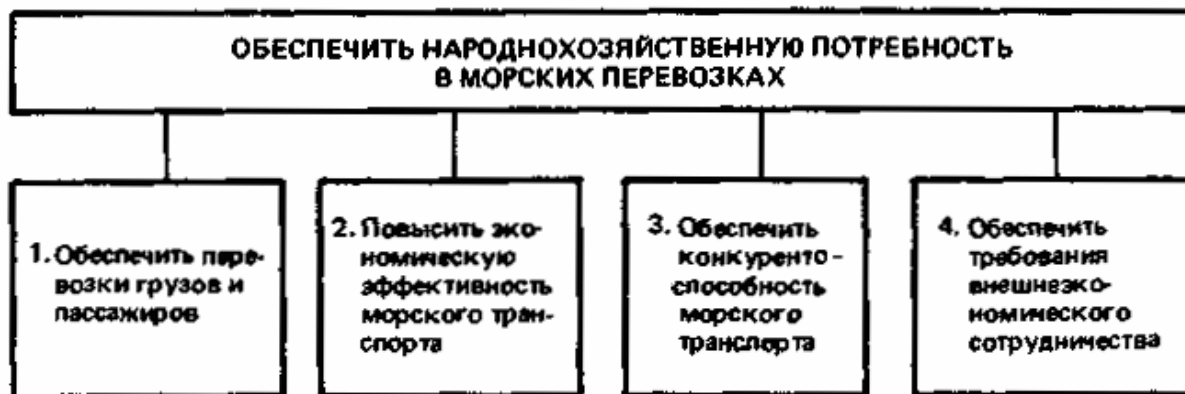


Рис. 4.2

Итак, объект декомпозиции должен сопоставляться с каждым элементом модели-основания. Однако и сама модель-основание может с разной степенью детализации отображать исследуемый объект. Например, в системном анализе часто приходится использовать модель типа «жизненный цикл», позволяющую декомпозировать анализируемый период времени на последовательные этапы от его возникновения до окончания. С помощью такой декомпозиции шахматную партию можно разбить на дебют, миттельшпиль и эндшпиль; в жизни человека принято различать молодость, зрелость и старость, но можно выделять

и более мелкие этапы, например детство, отрочество и юность. Такое же разнообразие может иметь место и при декомпозиции жизненного цикла любой проблемы. Разбиение на этапы дает представление о последовательности действий, начиная с обнаружения проблемы и кончая ее ликвидацией (иногда такую последовательность рассматривают как «алгоритм системного анализа»).

Перейдем к рассмотрению вопроса о том, какие модели брать за основания декомпозиции. Прежде всего напомним, что при всем практически необозримом многообразии моделей формальных типов моделей немного: это модели «черного ящика», состава, структуры, конструкции (структурной схемы) – каждая в статическом или динамическом варианте. Это позволяет организовать нужный перебор типов моделей, полный или сокращенный, в зависимости от необходимости.

Основанием для декомпозиции может служить только конкретная, содержательная модель рассматриваемой системы. Выбор формальной модели лишь подсказывает, какого типа должна быть модель-основание; формальную модель следует наполнить содержанием, чтобы она стала основанием для декомпозиции. Это позволяет несколько прояснить вопрос о полноте анализа, который всегда возникает в явной или неявной форме.

Полнота декомпозиции обеспечивается полнотой модели-основания, а это означает, что предварительно следует позаботиться о полноте формальной модели. Благодаря формальности, абстрактности такой модели часто удается добиться ее абсолютной полноты.

Пример 2. Схема входов организационной системы, представленная на рис. 4.1, является полной: к ней нечего добавить (перечислено все, что воздействует на систему), а изъятие любого элемента лишит ее полноты.

Пример 3. Формальный перечень типов ресурсов состоит из энергии, материи, времени, информации (для социальных систем добавляются кадры и финансы). При анализе ресурсного обеспечения любой конкретной системы этот перечень не позволит пропустить что-то важное.

Итак, полнота формальной модели должна быть предметом особого внимания. Поэтому одна из важных задач информационного обеспечения системного анализа и состоит в накоплении наборов полных формальных моделей (в искусственном интеллекте такие модели носят название фреймов).

Таким образом, основанием для декомпозиции является содержательная модель проблеморазрешающей системы. Это означает, что в разделяемом целом мы должны найти часть, соответствующую каждому из элементов модели-основания. Ориентиром для построения содержательной модели (т.е. основания декомпозиции) служат формальные модели известных типов. Предметом особого внимания является полнота модели, и существует несколько приемов, помогающих добиться полноты в формальных моделях и повысить полноту содержательных моделей.

4.2. Алгоритмизация процесса декомпозиции

В подразд. 4.1 рассмотрены некоторые аспекты того, каким образом эксперт осуществляет единичный акт разложения целого на части. Теперь можно дать дальнейшие рекомендации по осуществлению всего многоступенчатого процесса декомпозиции.

4.2.1. Компромиссы между полнотой и простотой модели

Начнем с обсуждения требований к древовидной структуре, которая получится как итог работы по всему алгоритму. С количественной стороны эти требования сводятся к двум противоречивым принципам: полноты (проблема должна быть рассмотрена максимально всесторонне и подробно) и простоты (все дерево должно быть максимально компактным – «вширь» и «вглубь»). Эти принципы относятся к количественным характеристикам (размерам) «дерева». Компромиссы между ними вытекают из качественного требования – главной цели: свести сложный объект анализа к конечной совокупности простых подобъектов либо (если это не удастся) выяснить конкретную причину неустранимой сложности.

Принцип простоты требует сокращать размеры дерева. Мы уже знаем, что размеры «вширь» определяются числом элементов модели, служащей основанием декомпозиции. Поэтому принцип простоты вынуждает брать как можно более компактные модели-основания. Наоборот, принцип полноты предусматривает использование как можно более развитых, подробных моделей. Компромисс достигается с помощью понятия существенности: в модель-основание включаются только компоненты, существенные по отношению к цели анализа (релевантные). Как видим, это понятие неформальное, поэтому решение вопроса о том, что же является в данной модели существенным, а что – нет, возлагается на эксперта. Чтобы облегчить работу эксперта, в алгоритме должны быть предусмотрены возможности внесения (в случае

необходимости) поправок и дополнений в модель-основание. Одна из таких возможностей заключается в дополнении элементов, которые эксперт счел существенными, еще одним элементом «все остальное»; он может не использоваться экспертом для декомпозиции, но будет постоянно пробуждать у эксперта сомнение в полноте предложенной им модели. Другая возможность состоит в разукрупнении, разбиении отдельных элементов модели-основания в случае необходимости, которая может возникнуть на последующих стадиях анализа.

Перейдем теперь к вопросу о размерах дерева «вглубь», т.е. о числе «этажей» дерева, числе уровней декомпозиции. Конечно, желательно, чтобы оно было небольшим (принцип простоты), но принцип полноты требует, чтобы в случае необходимости можно было продолжать декомпозицию как угодно долго до принятия решения о ее прекращении по данной ветви (разные ветви иногда могут иметь различную длину). Такое решение принимается в нескольких случаях. Первый, к которому мы обычно стремимся, наступает, когда декомпозиция привела к получению (подцели, подфункции, подзадачи и т.п.) простого, понятного, реализуемого, обеспеченного, заведомо выполнимого результата. Будем называть его элементарным. Для некоторых задач понятие элементарности может быть конкретизировано до формального признака, в других задачах анализа оно неизбежно остается неформальным, и проверка фрагментов декомпозиции на элементарность поручается экспертам.

Неэлементарный фрагмент подлежит дальнейшей декомпозиции по другой (не использовавшейся ранее) модели-основанию. Очевидно, что эффективность работы эксперта, размеры получающегося дерева и в конечном счете качество анализа в определенной мере зависят от последовательности, в которой эксперт использует имеющиеся модели. Например, алгоритм декомпозиции, встроенный в компьютерную диалоговую систему, должен, ради удобства эксперта, допускать предъявление моделей в том порядке, который определит сам эксперт. Вместе с тем должен быть предусмотрен и режим совета эксперту, рекомендуемый определенный порядок взятия оснований, упрощающий дело.

Если эксперт перебрал все фреймы, но не достиг элементарности на какой-то ветви дерева, то прежде всего выдвигается предположение, что дальнейшая декомпозиция может все-таки довести анализ до получения элементарных фрагментов и следует дать эксперту возможность продолжить декомпозицию. Такая возможность состоит во введении новых элементов в модель-основание и продолжении декомпозиции по ним. Поскольку новые существенные элементы могут быть получены только расщеплением уже имеющихся, в алгоритме декомпозиции

должна быть заложена возможность возврата к использованным ранее основаниям. При этом нет необходимости рассматривать заново все элементы модели, так как обрабатываемый фрагмент находится на ветви, соответствующей только одному элементу каждого основания. Тогда следует рассмотреть возможность расщепления именно этого элемента (например, при рассмотрении системы «вуз» вход «абитуриенты» можно разделить на абитуриентов со стажем и без него, выход «научная информация» – на выходы «монографии», «статьи», «отчеты по НИР» и т.п.). На этой же стадии можно рекомендовать эксперту решить, не настала ли пора выделить из «всего остального» и включить в число существенных еще один элемент. Пройдя таким образом всю предысторию неэлементарного фрагмента, получим новые основания для его декомпозиции и возможность продолжить анализ, надеясь достичь элементарности по всем ветвям.

Итак, итеративность алгоритма декомпозиции придает ему вариативность, возможность пользоваться моделями различной детальности на разных ветвях, углублять детализацию сколько угодно (если это потребуется).

4.2.2. Типы сложности моделей

Несмотря на возможности, предоставляемые сменой моделей и итеративностью, может наступить момент, когда эксперт признает, что его компетентности недостаточно для дальнейшего анализа данного фрагмента и что следует обратиться к эксперту другой квалификации. По существу, сложность такого типа есть сложность из-за неинформированности, которую можно преодолеть с помощью информации, рассредоточенной по разным экспертам и источникам. Однако случай, когда декомпозиция заканчивается элементарными фрагментами на всех ветвях дерева, является простейшим. Не имеет значения, один или несколько экспертов довели анализ до конца, а важно, что это оказалось возможным и, следовательно, первоначальная сложность была вызвана не столько недостатком информации, сколько большой размерностью проблемы. В действительно сложных случаях получение вполне завершенной декомпозиции должно не только радовать, но и настораживать: не связана ли реальная сложность с пропущенной ветвью дерева, сочтенной экспертами несущественной? Опасность неполноты анализа следует иметь в виду всегда (примеры – проблема поворота северных рек, проблемы Байкала и Ладожского озера и т.д.). Один из приемов (не дающий полной гарантии, но иногда полезный) – не только предлагать экспертам выявлять доводы в пользу рассматриваемого проекта, но и сопровождать его обязательным указанием возможных отрица-

тельных последствий. В частности, в классификатор выходов – конечных продуктов – любой системы помимо полезных продуктов обязательно должны быть включены отходы. Невозможность доведения декомпозиции до получения элементарного фрагмента, которая либо эвристически констатируется экспертом на ранних стадиях анализа, либо обнаруживается в виде «затягивания» анализа по данной ветви, является не только отрицательным, но и положительным результатом. Хотя при этом сложность не ликвидируется полностью, ее сфера сужается, обнаруживается и локализуется истинная причина этой сложности. Знание о том, что именно мы не знаем, быть может, не менее важно, чем само позитивное знание. Правда, вокруг таких результатов часто возникает атмосфера неприятия. Даже физики, говоря «отрицательный результат – тоже результат», чаще желают просто утешить коллегу-неудачника, а сам отрицательный результат стараются обходить стороной. Так было в начале XX века с «ультрафиолетовой катастрофой» до возникновения квантовой механики, похожая ситуация сложилась с объяснением природы шаровой молнии. Однако если в науке сложность из-за непонимания расценивается как временно неустраняемое и терпимое явление, то в управлении (т.е. в деловых, административных, политических вопросах) она часто воспринимается как неприемлемый вариант, ведущий к недопустимой отсрочке решения. Не потому ли именно в управлении нередко прибегают к интуитивным и волевым решениям? И не из-за отрицательного ли (в целом) опыта таких решений в последнее время наблюдаются быстрое сближение образа мышления управленцев и ученых, повышение роли научных методов в управлении?

Итак, если рассматривать анализ как способ преодоления сложности, то полное сведение сложного к простому возможно лишь в случае сложности из-за неинформированности; в случае сложности из-за непонимания анализ не ликвидирует сложность, но локализует ее, позволяет определить, каких именно сведений нам не хватает. Поэтому (с некоторой натяжкой) можно сказать, что метод декомпозиции не дает новых знаний, а лишь «вытягивает» знания из экспертов, структурирует и организует их, обнажая возможную нехватку знаний в виде «дыр» в этой структуре. Дело в том, что в действительности не только обнаружение нехватки конкретных знаний все-таки является новым знанием, но и по-иному скомбинированные фрагменты старых знаний также обладают новыми качествами.

4.3. Алгоритм декомпозиции

Сам алгоритм декомпозиции представлен в виде блок-схемы (рис. 4.3). К тому, что было уже сказано об изображаемых блоками операциях алгоритма, добавим следующее.

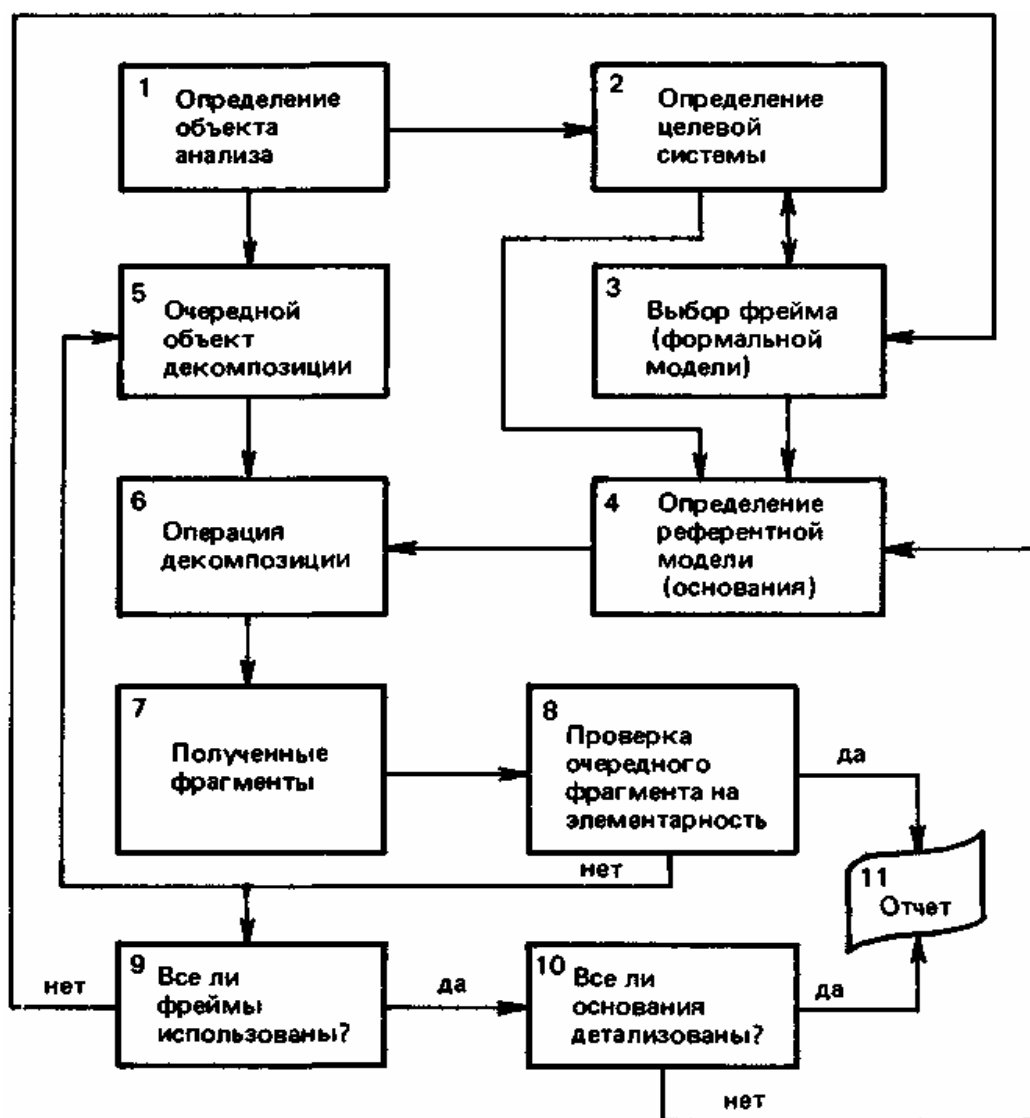


Рис. 4.3

Блок 1. Объектом анализа может стать все, что угодно, – любое высказывание, раскрытие смысла которого требует его структурирования. На определение объекта анализа иногда затрачиваются весьма значительные усилия. Когда речь идет о действительно сложной проблеме, ее сложность проявляется и в том, что сразу трудно правильно сформулировать объект анализа. Даже в таком основательно регламентированном документами случае, как работа министерства, формулировка глобальной цели возглавляемой им отрасли требует неоднократного уточнения и согласования, прежде чем она станет объектом

анализа. Это относится не только к формулировкам цели, но и к определению любого высказывания, подлежащего анализу. От правильности выбора объекта анализа зависит, действительно ли будет делаться нужное дело.

Блок 2. Этот блок определяет, зачем нужно то, что мы будем делать. В качестве целевой системы выступает система, в интересах которой осуществляется весь анализ. Снова подчеркнем, что более формального определения целевой системы дать нельзя, что многое зависит от конкретных условий. Например, опыт построения деревьев целей для Минобразования показал, что, хотя результаты анализа будет использовать министерство, целевой системой для верхних уровней дерева должна быть отрасль в целом и лишь на нижних уровнях требуются модели самого министерства.

Блок 3. Этот блок содержит набор фреймовых моделей и рекомендуемые правила их перебора либо обращение к эксперту с просьбой самому определить очередной фрейм.

Блок 4. Содержательная модель, по которой будет произведена декомпозиция, строится экспертом на основании изучения целевой системы. Хорошим подспорьем ему могут служить различные классификаторы, построенные в различных областях знаний, а также собранные в справочниках и специальных энциклопедиях.

Блоки 5–10 интуитивно понятны.

Блок 11. Окончательный результат анализа оформляется в виде дерева, конечными фрагментами ветвей которого являются либо элементарные фрагменты, либо фрагменты, признанные экспертом сложными, но не поддающимися дальнейшему разложению. Причины такой сложности могут состоять либо в ограниченности знаний данного эксперта или данной группы экспертов (сложность из-за неинформированности), либо в том, что нужные знания существуют, но еще не объединены в объясняющие модели (сложность из-за непонимания), либо в принципиальном отсутствии нужных знаний (сложность из-за незнания).

Блок-схема, изображенная на рис. 4.3, является слишком укрупненной; она предназначена для разъяснения основных идей алгоритма декомпозиции.

Таким образом, один из способов упрощения сложного – метод декомпозиции – состоит в разложении сложного целого на все более мелкие (и простые) части. Компромиссы между требованием не упустить важного (принцип полноты) и требованием не включать в модель

лишнего (принцип простоты) достигаются с помощью понятий существенного (необходимого), элементарного (достаточного), а также постепенной нарастающей детализации базовых моделей и итеративности алгоритма декомпозиции.

4.4. Агрегирование, эмерджентность и внутренняя целостность систем

Операцией, противоположной декомпозиции, является операция агрегирования, т.е. объединения нескольких элементов в единое целое. Необходимость агрегирования может вызываться различными целями и сопровождаться разными обстоятельствами, что приводит к различным (иногда принципиально различным) способам агрегирования. Однако у всех агрегатов (так называют результат агрегирования) есть одно общее свойство, получившее название эмерджентности. Это свойство присуще всем системам, и ввиду его важности остановимся на нем подробнее.

4.4.1. Эмерджентность как проявление внутренней целостности систем

Будучи объединенными, взаимодействующие элементы образуют систему, обладающую не только внешней целостностью, обособленностью от окружающей среды, но и внутренней целостностью, природным единством. Если внешняя целостность отображается моделью «черного ящика», то внутренняя целостность связана со структурой системы. Наиболее яркое проявление внутренней целостности системы состоит в том, что свойства системы не являются только суммой свойств ее составных частей. Система есть нечто большее, система в целом обладает такими свойствами, которых нет ни у одной из ее частей, взятой в отдельности. Модель структуры подчеркивает главным образом связанность элементов, их взаимодействие. Мы же стремимся сейчас сделать акцент на том, что при объединении частей в целое возникают новые качества, такие, чего не было и не могло быть без этого объединения.

Пусть имеется некоторый цифровой автомат S , преобразующий целое число на его входе в число, на единицу большее входного (рис. 4.4,а). Если соединить два таких автомата последовательно в кольцо (рис. 4.4,б), то в полученной системе обнаружится новое свойство: она генерирует возрастающие последовательности на выходах A к B , причем одна из этих последовательностей состоит только из четных, другая – из нечетных чисел.

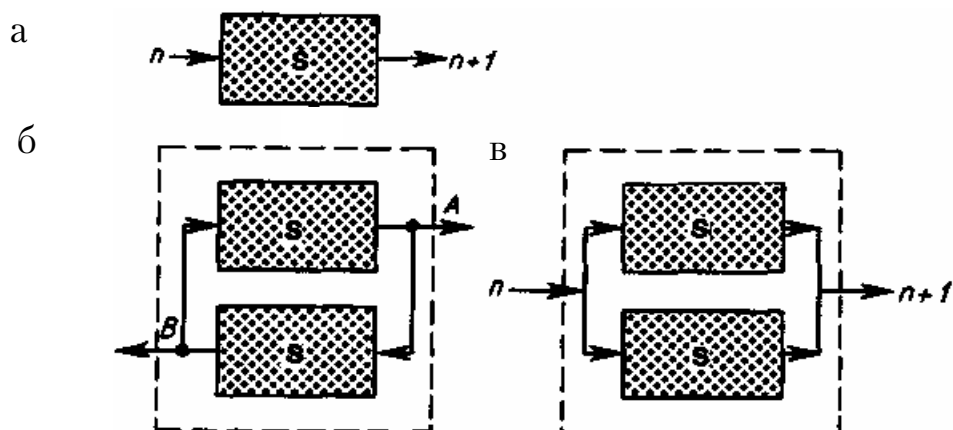


Рис. 4.4

4.4.2. Эмерджентность как результат агрегирования

Такое «внезапное» появление новых качеств у систем и дало основание присвоить этому их свойству название эмерджентности. Английский термин *emergence* означает возникновение из ничего, внезапное появление, неожиданную случайность. В специальной литературе на русском языке не делалось попыток найти эквивалентный русский термин. Однако сам термин имеет обманчивый смысл. Какие бы удивительные свойства ни возникали при объединении элементов в систему, ничего мистического, взявшегося «ниоткуда», здесь нет: новые свойства появляются благодаря конкретным связям между конкретными элементами. Другие связи дадут другие свойства, не обязательно столь же очевидные. Например, параллельное соединение тех же автоматов (рис. 4.4, в) ничего не изменяет в арифметическом отношении, но увеличивает надежность вычислений, если на выход поступает сигнал только от исправного автомата.

Свойство эмерджентности признано и официально: при государственной экспертизе изобретений патентоспособным признается новое, ранее не известное соединение хорошо известных элементов, если при этом возникают новые полезные свойства.

Итак, агрегирование частей в единое целое приводит к появлению новых качеств, не сводящихся к качествам частей в отдельности. Это свойство и является проявлением внутренней целостности систем, или – иначе – системообразующим фактором. Новые качества систем определяются в сильной степени характером связей между частями и могут варьироваться в весьма широком диапазоне.

4.5. Виды агрегирования

Как и в случае декомпозиции, техника агрегирования основана на использовании определенных моделей исследуемой или проектируемой системы. Именно избранные нами модели жестко определяют, какие части должны войти в состав системы (модель состава) и как они должны быть связаны между собой (модель структуры). Разные условия и цели агрегирования приводят к необходимости использования разных моделей, что определяет тип окончательного агрегата и технику его построения.

В самом общем виде агрегирование можно определить как установление отношений на заданном множестве элементов. Благодаря значительной свободе выбора множества элементов и отношений на этом множестве, получается весьма обширное количественно и разнообразное качественно множество задач агрегирования. Отметим здесь лишь основные агрегаты, типичные для системного анализа: конфигуратор, агрегаты-операторы и агрегаты-структуры.

4.5.1. Конфигуратор

Всякое действительно сложное явление требует разностороннего, многопланового описания, рассмотрения с различных точек зрения. Только совместное (агрегированное) описание в терминах нескольких качественно различающихся языков позволяет охарактеризовать явление с достаточной полнотой. Например, автомобильная катастрофа должна рассматриваться не только как физическое явление, вызванное механическими причинами – техническим состоянием автомобиля, дорожного покрытия, силами инерции, трения, ударов и т.д., но и как явление медицинского, социального, экономического, юридического характера. Даже движение планет имеет не только механические аспекты, но и социальные, – вспомним, какие потрясения вызвал переход от геоцентрического к гелиоцентрическому описанию этого движения. В реальной жизни не бывает проблем чисто физических, экономических, общественных или даже системных – эти термины обозначают не саму проблему, а выбранную точку зрения на нее.

Эта многоплановость реальной жизни имеет важные последствия для системного анализа. С одной стороны, системный анализ отличается междисциплинарным характером. Системный аналитик готов использовать при исследовании системы данные из любой отрасли знаний, привлечь эксперта любой специальности, если этого потребуют интересы дела. С другой стороны, перед ним встает неизбежный вопрос

о допустимой минимизации описания явления. Однако если при декомпозиции этот вопрос решался компромиссно с помощью понятия существенности, что давало некоторую свободу выбора, сопровождаемую риском недостаточной полноты или излишней подробности, то при агрегировании этот вопрос обостряется: риск неполноты становится почти недопустимым, поскольку при неполноте речь может идти вообще не о том, что мы имеем в виду; напротив, риск переопределения связан с большими излишними затратами. Рассмотренные соображения привели к введению понятия агрегата, состоящего из качественно различных языков описания системы, причём число этих языков минимально, но необходимо для заданной цели. Такой агрегат называют конфигуратором. Продемонстрируем смысл этого понятия на примерах.

Пример 1. Конфигуратором для задания любой точки n -мерного пространства является совокупность ее координат. Обратим внимание на эквивалентность разных систем координат (разных конфигураторов) и на предпочтительность ортогональных систем, дающих независимое описание на каждом «языке» конфигуратора.

Пример 2. Конфигуратором для описания поверхности любого трехмерного тела на «плоскостных» языках является совокупность трех ортогональных проекций, принятая в техническом черчении и начертательной геометрии. Обратим внимание на невозможность уменьшения числа проекций и на избыточность большего числа «точек зрения».

Пример 3. В электронике для одного и того же прибора используется конфигуратор: структурная схема, принципиальная схема, монтажная схема. Структурная схема может определяться теми технологическими единицами, которые выпускаются промышленностью, и тогда прибор членится на такие единицы. Принципиальная схема предполагает совершенно иное расчленение: она должна объяснить функционирование этого прибора. На ней выделены функциональные единицы, которые могут не иметь пространственно локализованных аналогов. Приборы могут иметь различные структурные схемы и одинаковые принципиальные схемы, и наоборот. Наконец, монтажная схема является результатом расчленения прибора в зависимости от геометрии объема, в пределах которого производится его монтаж.

Подчеркнем, что главное в конфигураторе не то, что анализ объекта должен производиться на каждом языке конфигуратора отдельно (это разумеется само собой), а то, что синтез, проектирование, производство и эксплуатация прибора возможны только при наличии всех трех его

описаний. Однако этот пример дает возможность подчеркнуть еще и зависимость конфигуратора от поставленных целей. Например, если кроме целей производства будем преследовать и цели сбыта, то в конфигуратор радиоаппаратуры придется включить и язык рекламы, позволяющий описать внешний вид и другие потребительские качества прибора.

Заметим, что конфигуратор является содержательной моделью высшего возможного уровня. Перечислив языки, на которых будем говорить о системе, мы тем самым определим, синтезируем тип системы, зафиксируем наше понимание природы системы. Как всякая модель, конфигуратор имеет целевой характер и при смене цели может утратить свойства конфигуратора. Как каждая модель, конфигуратор в простых случаях, очевидно, адекватен, адекватность других подтверждает практика, в полноте третьих мы лишь более или менее уверены и готовы пойти на их дополнение новыми языками.

4.5.2. Агрегаты-операторы

Одна из наиболее частых ситуаций, требующих агрегирования, состоит в том, что совокупность данных, с которыми приходится иметь дело, слишком многочисленна, плохо обозрима, с этими данными трудно «работать». Именно интересы работы с многочисленной совокупностью данных приводят к необходимости агрегирования. В данном случае на первый план выступает такая особенность агрегирования, как уменьшение размерности: агрегат объединяет части в нечто целое, единое, отдельное.

Простейший способ агрегирования состоит в установлении отношения эквивалентности между агрегируемыми элементами, т.е. в образовании классов. Это позволяет говорить не только о классе в целом, но и о каждом его элементе в отдельности.

Можно рассматривать различные задачи, связанные с классификацией элементов и ее использованием. Приведем примеры таких задач: классификация как самая простая, самая первая, «атомарная» форма моделирования (это проблематика философии и психологии); элемент как представитель класса (совсем не простой вопрос; так, в теории случайных процессов это приводит к проблеме эргодичности); искусственная классификация и природная кластеризация (т.е. образование классов «для удобства» и естественная общность определенных явлений, возможные соотношения между ними; например, созвездия – первое или второе или отчасти то и другое?); иерархическая и сетевая классификации и т.д. Классификация является очень важным и многофункциональным, многосторонним явлением в человеческой практике

вообще и в системном анализе в частности. С практической точки зрения одной из важнейших является проблема определения, к какому классу относится данный конкретный элемент. Обсудим ее подробнее.

4.5.3. Классификация как агрегирование

Если признак принадлежности к классу является непосредственно наблюдаемым, то кажется, что особых трудностей классификации нет. Однако и в этих случаях на практике возникает вопрос о надежности, правильности классификации. Например, разложить окрашенные куски картона по цветам – трудная задача даже для ученых-психологов: отнести ли оранжевый кусок к «красным» или «желтым», если между ними нет других классов? По одежде военных можно определить не только, к какому государству они принадлежат, но и в каком роде войск служат, в каком чине состоят и т.д. Но эту четкость и однозначность враг может использовать, переодевая своих диверсантов в форму противника, и тогда распознавание «своих» от «чужих» осуществляется с помощью других признаков. Если же непосредственно наблюдаемый признак принадлежности к классу формулируется на естественном языке, то, как известно, некоторая неопределенность (расплывчатость) становится неизбежной (например, кого отнести в класс «высоких людей?»).

Сложности классификации резко возрастают, если признак классификации не наблюдается непосредственно, а сам является агрегатом косвенных признаков. Типичным примером является диагностика заболевания по результатам анамнеза: диагноз болезни (ее название есть имя класса) представляет собой агрегат большой совокупности ее симптомов и характеристик состояния организма. Если классификация имеет природный характер, то агрегирование косвенных признаков может рассматриваться как обнаружение закономерностей в таблицах экспериментальных данных, т.е. как поиск устойчивых, достаточно часто повторяющихся в обучающей выборке «сцеплений» признаков. При этом приходится перебирать все возможные комбинации признаков с целью проверки их повторяемости в обучающей выборке (например, в таблице, строки которой соответствуют данным предварительных опроса и анализов для каждого пациента). Вообще, метод перебора вариантов – самый очевидный, простой и надежный способ поиска решения. Несмотря на трудоемкость, его нередко с успехом применяют. Однако уже при совсем небольшом количестве признаков полный перебор становится нереальным даже при использовании ЭВМ. Успех в значительной степени зависит от того, удастся ли найти метод со-

кращения перебора, приводящий к «хорошим» решениям, и разработке таких методов посвящено значительное количество исследований.

Все сказанное свидетельствует о том, что агрегирование в классы является эффективной, но далеко не тривиальной процедурой. Если представлять класс как результат действия агрегата-оператора, то такой оператор имеет вид «ЕСЛИ <условия на агрегируемые признаки>, ТО <имя класса>». Как было отмечено, иногда класс непосредственно задается совокупностью признаков, а в ряде случаев, наоборот, требуется доопределить оператор, выявив экспериментально, при каких условиях объект будет принадлежать заданному классу.

4.5.4. Статистики как агрегаты

Важный пример агрегирования данных дает статистический анализ. Среди различных агрегатов (называемых в этом случае статистиками, т.е. функциями выборочных значений) особое место занимают достаточные статистики, т.е. такие агрегаты, которые извлекают всю полезную информацию об интересующем нас параметре из совокупности наблюдений. Однако при агрегировании обычно потери информации неизбежны, и достаточные статистики сводят эти потери к минимуму. В таких условиях становятся важными оптимальные статистики, т.е. позволяющие свести неизбежные в этих условиях потери к минимуму в некотором заданном смысле. Наглядный пример статистического агрегирования представляет собой факторный анализ, в котором несколько переменных сводятся в один фактор.

Наконец, подчеркнем, что с созданием агрегата-оператора связан не только выигрыш, ради которого он и создается, но и риск попасть в «ловушки». Отметим основные из них:

- потеря полезной информации. Агрегирование является необратимым преобразованием (например, по сумме нельзя восстановить слагаемые), что в общем случае и приводит к потерям; достаточные статистики – лишь счастливое исключение (если сумма есть достаточная статистика, то информация об отдельных слагаемых и не нужна);
- агрегирование представляет собой выбор определенной модели системы, причем с этим выбором связаны непростые проблемы адекватности;
- некоторым агрегатам-операторам присуща внутренняя противоречивость, сопряженная с отрицательными (по отношению к целям агрегирования) последствиями. Наиболее ярким примером этого является теорема о невозможности, но не присуще ли это свойство (хотя и выраженное в разной степени) всем агрегатам?

4.5.5. Агрегаты-структуры

Важной (а на этапе синтеза – важнейшей) формой агрегирования является образование структур. Как любой вид агрегата, структура является моделью системы и, следовательно, определяется тройственной совокупностью – объекта, цели и средств (в том числе среды) – моделирования. Это и объясняет многообразие типов структур (сети, матрицы, деревья и т.д.), возникающих при выявлении, описании структур (познавательные модели).

При синтезе мы создаем, определяем, навязываем структуру будущей, проектируемой системе. Если это не абстрактная, а реальная система, то в ней вполне реально (т.е. независимо от нашего желания) возникнут, установятся и начнут «работать» не только те связи, которые мы спроектировали, но и множество других, не предусмотренных нами, вытекающих из самой природы сведенных в одну систему элементов. Поэтому при проектировании системы важно задать ее структуры во всех существенных отношениях, так как в остальных отношениях структуры сложатся сами, стихийным образом (конечно, не совсем независимо от установленных и поддерживаемых проектных структур). Совокупность всех существенных отношений определяется конфигуратором системы, и отсюда вытекает, что проект любой системы должен содержать разработку столько структур, сколько языков включено в ее конфигуратор. Например, проект организационной системы должен содержать структуры распределения власти, распределения ответственности и распределения информации. Подчеркнем, что, хотя эти структуры могут весьма сильно отличаться топологически (например, структура подчиненности иерархична, а функционирование организовано по матричной структуре), они лишь с разных сторон описывают одну и ту же систему и, следовательно, не могут быть не связанными между собой.

В современных системных науках все большее внимание уделяется одному из специфических видов структур – так называемым семантическим сетям.

4.6. Обобщенная модель агрегата

Серьезную помощь при разработке моделей агрегатов-операторов может оказать владение типовыми математическими схемами, широко используемыми в арсенале моделирования и многократно проверенными опытом. Среди них наиболее простой является схема конечного автомата. Конечный автомат характеризуется конечными множествами состояний Z , входных сигналов X , выходных сигналов Y . В каждый

момент времени в него поступает входной сигнал $x(t)$, под действием которого автомат переходит в новое состояние в соответствии с функцией переходов $z(t) = f_1[z(t-1), x(t)]$ и выдает выходной сигнал, определяемый функцией выходов $y(t) = f_2[z(t-1), x(t)]$. Автомат можно задать также таблицами переходов и выходов или графической схемой переходов и выходов.

Математические модели широкого класса детерминированных объектов (т.е. без учета случайных факторов), функционирующих в дискретном времени, приводятся к различным типам конечных автоматов.

Детерминированные объекты, функционирующие в непрерывном времени, обычно описывают дифференциальными уравнениями.

Стохастические объекты, функционирующие в дискретном времени, можно представить вероятностными автоматами. Функция переходов вероятностного автомата определяет не одно конкретное состояние, а лишь распределение вероятностей на множестве состояний, т.е. автомат со случайными переходами, а функция выходов – распределение вероятностей на множестве выходных сигналов, т.е. автомат со случайными выходами. Функции этих автоматов изучаются с помощью теории цепей Маркова.

Математическими моделями стохастических объектов с непрерывным временем служат системы массового обслуживания или представители марковских случайных процессов.

Остановимся на некоторых общих свойствах динамических систем, используемых в качестве математических моделей элементов сложных систем:

1. Элемент функционирует во времени; в каждый момент времени t он находится в одном из возможных состояний z .
2. С течением времени под действием внутренних и внешних причин элемент переходит из одного состояния в другое.
3. В процессе функционирования элемент взаимодействует с другими элементами системы и объектами внешней среды.

Наличие общих свойств моделей элементов сложных систем позволяет дать им общее математическое описание, полезное для решения ряда теоретических и практических вопросов системного анализа и называемое агрегатом.

В каждый момент времени $t = (0, T)$ агрегат находится в одном из возможных состояний, являющихся элементами некоторого множества Z , т.е. состояние z изменяется как функция времени $z(t)$, причем, состояние агрегата $z(t)$ зависит как от предыдущих состояний, так и от внешних входных и управляющих воздействий.

С точки зрения моделирования агрегат выступает как универсальный преобразователь информации – он воспринимает входные и управляющие сигналы и выдает выходные сигналы.

Совокупность входных, управляющих и выходных сигналов, расположенных в порядке их поступления или выдачи, называют входными, управляющими или выходными сообщениями.

У агрегата выделяют так называемые особые состояния, под которыми понимают его состояния в моменты получения входного, управляющего или выдачи выходного сигнала.

Математическая модель агрегата приведена на рис 4.5. Состояние агрегата $z(t)$ для произвольного момента времени $t > t_0$ определяется по предыдущим состояниям оператором H (в общем случае имеющим случайный характер): $z(t) = H[z(t_0), t]$, т.е. данному $z(t_0)$ ставится в соответствие в общем случае не одно конкретное $z(t)$, а множество значений $z(t)$ с некоторым законом распределения и статистическими характеристиками, зависящими от вида оператора H .

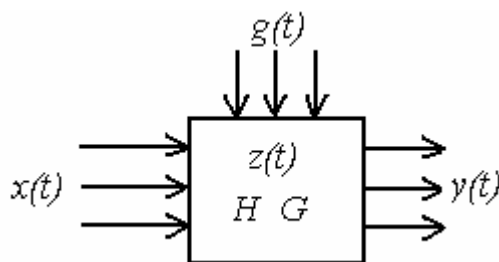


Рис. 4.5

Конкретное значение $z(t)$ определяется как реализация в соответствии с этим законом распределения.

Выходной сигнал $y(t)$ является элементом некоторого множества Y и определяется по состояниям агрегата $z(t)$ при помощи оператора G .

Оператор H называется оператором переходов; G – оператором выходов.

Агрегат представляет собой математическую схему общего вида, частными случаями которой являются функции алгебры логики, конечные и вероятностные автоматы и т.п. С точки зрения моделирования он выступает как достаточно универсальный преобразователь информации – воспринимает входные, управляющие и выдает выходные сигналы, обрабатывает их.

4.7. Некоторые особенности моделирования процесса функционирования агрегата

Целью моделирования в любом случае является получение характеристик, определяемых состояниями моделируемой системы. Для этого в процессе моделирования необходимо фиксировать достаточно полную информацию о состояниях системы в соответствующие моменты модельного времени.

Моменты поступления входных и управляющих сигналов играют значительную роль при построении моделирующего алгоритма. Будем рассматривать моделирование как последовательную цепь переходов из одного особого состояния агрегата в другое, начиная с начального состояния $z(0)$.

Вид моделирующего алгоритма существенно зависит от того, известны ли заранее моменты поступления входных и управляющих сигналов и вообще моменты последующих особых состояний.

При этом возможны две ситуации:

1. Законы поступления входных и управляющих сигналов заданы (обычно это сигналы от внешней среды и объектов относительно системы).

2. Входные и управляющие сигналы вырабатываются в процессе моделирования как выходные сигналы других агрегатов системы.

Рассмотрим некоторые варианты первого случая. Входные и управляющие сигналы, поступающие в агрегат в течение интервала времени моделирования $(0, T)$ можно записать в память машины в виде таблицы, содержащей для всех моментов значения параметров. Однако, если таких сигналов много и каждый из них характеризуется большим числом координат, этот метод неудобен.

Одним из распространенных способов является генерирование сигналов с помощью чисел, найденных по заданному детерминированному закону, или с помощью случайных чисел, распределенных по заранее заданному закону. При этом моделирующий алгоритм в общем должен содержать:

- а) группу операторов, формирующих моменты поступления сигналов;

- б) группу операторов, моделирующих процесс функционирования агрегата между последовательными моментами поступления внешнего сигнала, в том числе и процессы выдачи выходных сигналов. Алгоритмы этого типа называют алгоритмами по принципу особых состояний.

Во втором случае, когда входные и управляющие сигналы неизвестны заранее и не могут быть описаны математически до начала

моделирования, факт поступления внешнего сигнала может быть выяснен по ходу моделирования. Для этого в моделирующем алгоритме должны быть предусмотрены процедуры проверки факта поступления внешних сигналов за некоторый интервал модельного времени Δt . Поскольку момент следующего состояния, при котором поступает внешний сигнал, часто неизвестен, обычно приходится выбирать тип моделирования Δt . Алгоритмы этого типа называют алгоритмами по принципу Δt . Они более трудоемки, так как требуют гораздо больше памяти и машинного времени. В любом случае анализируются состояния системы и выходные сигналы.

На практике часто применяют комбинированные алгоритмы.

4.8. Агрегативные системы

Во многих случаях процессы, объекты и системы представляют в виде агрегативных систем (А-систем), под которыми понимают сложные системы, расчленяемые на элементы (в общем случае неоднозначно), каждый из которых представляет собой агрегат. Каждый элемент может быть агрегатом с полным комплектом свойств, в том числе А-системой, или его частным случаем.

На рис. 4.6 представлен пример А-системы. Здесь А1–А7 – агрегаты, x и q – соответственно входные и управляющие внешние относительно системы воздействия.

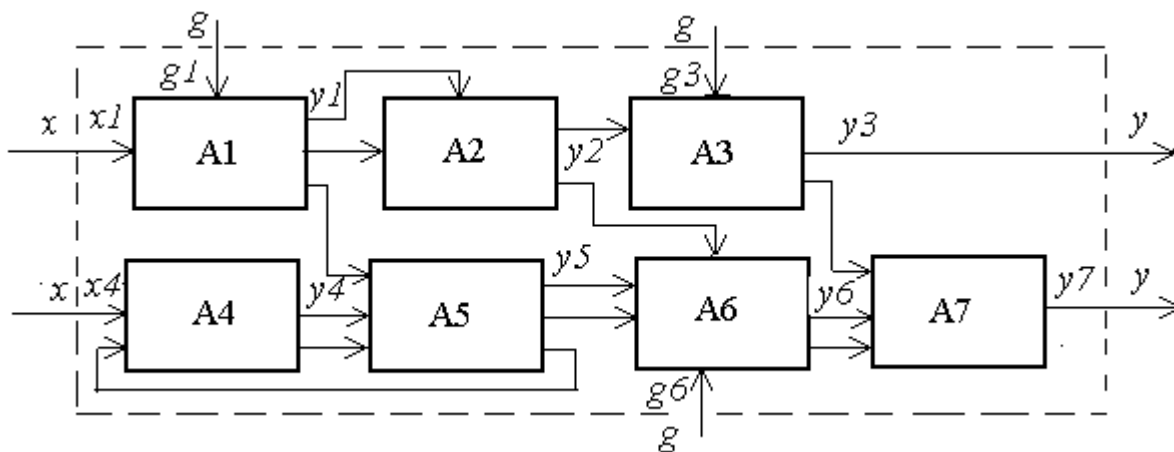


Рис. 4.6

Вся информация А-системы делится на внешнюю и внутреннюю, вырабатываемую агрегатами системы. Обмен информацией между А-системой и внешней средой происходит через агрегаты, называемые полюсами системы. Имеются входные (А1, А4), управляющие (А1, А3, А6) и выходные (А3, А7) полюсы.

Агрегаты, не являющиеся полюсами системы, называются внутренними агрегатами А-системы.

Передача информации в А-системах происходит мгновенно и без искажений, т.е. реальные каналы связи следует представить в виде простейших агрегатов, осуществляющих задержку и искажение информации, если это необходимо.

А-система называется комплексом, если любой её агрегат связан хотя бы с одним агрегатом этой А-системы (комплексом может быть А-система, состоящая из одного агрегата).

Агрегат А2 непосредственно подчинен агрегату А1 (агрегат А1 непосредственно управляет агрегатом А2), так как часть управляющей информации агрегата А2 есть часть выходной информации агрегата А1.

Агрегат А6 подчинен агрегату А1, так как он непосредственно подчинен агрегату А2, который непосредственно подчинен агрегату А1.

Комплекс А подчинен комплексу В, если часть выходной информации комплекса В есть часть управляющей информации комплекса А.

А-система называется иерархической, если она состоит из некоторого количества комплексов, подчиненных одному (управляющему) комплексу.

А-система может состоять из нескольких иерархических А-систем, подчиненных одному управляющему комплексу. Такая система называется иерархической с двумя уровнями управления. Аналогично может быть образовано произвольное число уровней управления.

Следует заметить, что создание имитационных моделей особо сложных систем обычными методами требует много времени и средств; кроме того, повышение трудоемкости имитационного моделирования постоянно возрастает по мере роста сложности систем. Радикальным способом снижения трудоемкости при увеличении сложности моделируемых объектов является автоматизация процедур построения и реализации моделей. Одним из таких способов является применение языков моделирования, однако их использование эффективно лишь при разработке программного обеспечения.

Пользователь применяет более простые средства, в основу которых заложен принцип стандартной формы описания сложной системы. В качестве стандартной формы обычно используют одноуровневую А-систему с элементами кусочно-линейных агрегатов общего или специального вида.

Стандартная форма является связующим звеном между математическими моделями сложных систем, поступающими от пользователя, и комплексом программ, осуществляющих имитацию процесса функционирования системы на ЦВМ. Модель объекта в общем случае представ-

ляется в виде многоуровневой системы, элементы которой описываются, как правило, различными типичными схемами – конечными и вероятностными автоматами, системами массового обслуживания и т.п.

Поскольку такие системы являются частными случаями агрегата, можно построить алгоритмы достаточно точного преобразования моделей элементов к стандартной форме.

Контрольные вопросы

1. Что такое декомпозиция и для чего она применяется?
2. Как производится декомпозиция?
3. Охарактеризуйте основные этапы декомпозиции.
4. Что такое агрегирование и для чего оно применяется?
5. Что такое эмерджентность?
6. Чем отличается внутренняя целостность систем от внешней?
7. Назовите и кратко охарактеризуйте основные виды агрегирования.
8. Что такое конфигуратор?
9. Что такое агрегаты-операторы?
10. Что такое классификация?
11. Что такое агрегаты-статистики?
12. Что такое агрегаты-структуры?
13. Охарактеризуйте обобщенную модель агрегата.
14. Охарактеризуйте основные особенности моделирования процесса функционирования агрегата
15. Охарактеризуйте обобщенную модель агрегативной системы.

5. ЭКСПЕРИМЕНТ В АНАЛИЗЕ СИСТЕМ

В изначальном смысле отношение между экспериментом и информацией такое же, как и между курицей и яйцом: они находятся в одном цикле, и крайне трудно определить, что было «в самом начале». Эксперимент с некоторым объектом проводится, чтобы уточнить модель этого объекта; поэтому постановка эксперимента определяется имеющейся до опыта моделью. Это полностью относится и к экспериментальному исследованию систем.

Отношение между экспериментом и теоретической моделью двоякое. С одной стороны, эксперимент позволяет проверить и при необходимости уточнить модель, т.е. эксперимент является источником информации для моделирования. С другой стороны, модель диктует, какой именно эксперимент следует проводить, т.е. модель является источником информации для организации эксперимента.

Современное понимание эксперимента предполагает, что:

- 1) измерения кроме количественного могут носить качественный характер;
- 2) измерение может не снимать неопределенность, если она имеет расплывчатую природу;
- 3) измерение обычно сопровождается неизбежными погрешностями;
- 4) интересующая нас величина часто не наблюдаема и поддается лишь косвенным измерениям.

Природу эксперимента хорошо понимали выдающиеся естествоиспытатели древности и понимают современные исследователи. Приведем некоторые высказывания крупных ученых по этому поводу.

Леонардо да Винчи: «... мне кажется, что пусты и полны заблуждений те науки, которые не порождены опытом, отцом всякой достоверности, и не завершаются в наглядном опыте ... Опыт никогда не ошибается, ошибаются ваши суждения, ожидая от него такого действия, которое не является следствием ваших экспериментов ... Мудрость есть дочь опыта».

Н. Винер: «Любой эксперимент – всегда некий вопрос. Если вопрос неточен, получить точный ответ на него трудно. Глупые ответы, т.е. противоречивые, расходящиеся друг с другом или не относящиеся к делу результаты экспериментов, обычно указывают на то, что сам вопрос был поставлен глупо».

Общая мысль этих высказываний ясна. Мы вернулись (правда, уже с другой стороны) к проблеме соотношения реальности и созданных нами ее моделей. Отличие от сказанного по этому поводу ранее состоит

в том, что не только опыт является критерием истинности модели, но и сама постановка эксперимента диктуется моделью, так как вытекает из необходимости ее проверки или уточнения.

Рассмотрим теперь возможности опытов с системами. Начнем обсуждение с модели «черного ящика», т.е. с информации о входах и выходах системы. Выбор именно этих входов и выходов и есть построение модели, которая и будет определять организацию опыта. Если мы только регистрируем события на выбранных входах и выходах, то опыт называется пассивным экспериментом (или наблюдением). Если же мы не только фиксируем происходящее на входах и выходах, но и воздействуем на некоторые из них, одни намеренно поддерживая неизменными, другие изменяя нужным образом, то опыт называется активным (или управляемым) экспериментом.

Результаты опыта регистрируются с помощью измерений, т.е. изображения результатов опыта в виде символов, номеров или чисел. Способы осуществления такого отображения различны. Важно, что современное понимание измерения существенно шире только количественного измерения.

Если ваша догадка сформулирована плохо или достаточно неопределенно и если метод, которым вы пользуетесь для оценки последствий, достаточно расплывчат, то ваша теория всем хороша – ведь ее нельзя опровергнуть. Кроме того, если ваш метод расчетов последствий достаточно нечеток, при некоторой ловкости всегда можно сделать так, чтобы результаты экспериментов были похожи на предполагаемые последствия.

Оставив неизменным принцип проверки адекватности модели на опыте, современный подход позволил расширить понятие измерений по крайней мере в четырех отношениях.

1. Стало ясно, что есть наблюдаемые явления, которые в принципе не допускают числовой меры (например, «количество материнской любви»), но которые можно фиксировать в «слабых», «качественных» шкалах и эти результаты учитывать в моделях, получая качественные, но вполне научные выводы.

2. Расплывчатость некоторых наблюдений также признана их неотъемлемым природным свойством, которому придана строгая математическая форма, и разработан формальный аппарат работы с такими наблюдениями.

3. Хотя по-прежнему считается, что чем точнее измерения, тем лучше, теперь пришло осознание, что погрешности измерений являются не только чем-то побочным, чуждым для измерений (сторонние помехи, результат небрежности или ошибок оператора и т.п.), но и не-

отъемлемым, естественным и неизбежным свойством самого процесса измерения («шумы квантования», соотношения неопределенности, собственные шумы аппаратуры). Проверяемые на практике модели должны быть не только гипотезами об исследуемом объекте, но и гипотезами об ошибках измерения.

4. Широкое распространение получили статистические измерения, т.е. оценивание функционалов распределений вероятностей по реализации случайного процесса; этот класс измерений важен потому, что большинство временных зависимостей входов и выходов носит сигнальный характер. Для таких измерений требуется специфическая методика и техника.

Далее будет затронута еще одна важная тема. Хотя для проведения эксперимента необходима модель объекта, с которым мы экспериментируем, а для уточнения модели объекта требуется проведение эксперимента, здесь нет порочного круга: после завершения очередного цикла следующий начинается с новой, измененной модели. Ситуация напоминает не вращающееся колесо, а катящийся снежный ком: с каждым оборотом он становится больше, весомее. Рассмотрим, как происходит переход от модели «черного ящика» к модели «белого ящика» при использовании результатов измерений.

5.1. Измерительные шкалы

Измерение – это алгоритмическая операция, которая данному наблюдаемому состоянию объекта, процесса, явления ставит в соответствие определенное обозначение: число, номер или символ. Такое соответствие обеспечивает то, что результаты измерений содержат информацию о наблюдавшемся объекте; количество же информации зависит от степени полноты этого соответствия и разнообразия вариантов. Нужная нам информация получается из результатов измерения с помощью их преобразований или обработки.

Совершенно ясно, что чем теснее соответствие между состояниями и их обозначениями, тем больше информации можно извлечь в результате обработки данных – однако степень этого соответствия зависит не только от организации измерений (т.е. от экспериментатора), но и от природы исследуемого явления – и что сама степень соответствия, в свою очередь, определяет допустимые (и недопустимые) способы обработки данных.

Будем рассматривать только такие объекты, про любые два состояния которых можно сказать, различимы они или нет, и только такие алгоритмы измерения, которые различным состояниям ставят в

соответствие разные обозначения, а неразличимым состояниям – одинаковые обозначения. Это означает, что как состояния объекта, так и их обозначения удовлетворяют следующим аксиомам тождества:

1. Либо $A = B$, либо $A \neq B$.
2. Если $A = B$, то $B = A$.
3. Если $A = B$ и $B = C$, то $A = C$.

Здесь символ « $=$ » обозначает отношение эквивалентности; в том случае, когда A и B – числа, он означает их равенство.

5.1.1. Шкалы наименования

Предположим, что число различных состояний (математический термин – число классов эквивалентности) конечно. Каждому классу эквивалентности поставим в соответствие обозначение, отличное от обозначений других классов. Теперь измерение будет состоять в том, чтобы, проведя эксперимент над объектом, определить принадлежность результата к тому или иному классу эквивалентности и записать это с помощью символа, обозначающего данный класс. Такое измерение называется измерением в шкале наименований (иногда эту шкалу называют также номинальной или классификационной); указанное множество символов и образует шкалу.

Особенности шкалы наименований рассмотрим на примерах. Естественнее всего использовать шкалу наименований в тех случаях, когда классифицируются дискретные по своей природе явления (например, различные объекты). Для обозначения классов могут применяться слова естественного языка (географические названия, собственные имена людей и т.д.), произвольные символы (гербы и флаги государств, эмблемы родов войск, всевозможные значки и т.д.), номера (регистрационные номера автомобилей, официальных документов, номера на майках спортсменов), а также их различные комбинации (например, почтовые адреса, экслибрисы личных библиотек, печати и пр.) Все эти обозначения эквивалентны простой нумерации (в некоторых странах человек при рождении получает номер, под которым он фигурирует в государственных информационных системах всю жизнь), но на практике часто предпочтение отдают другим обозначениям (вообразите, что вместо имен и фамилий ваших друзей и знакомых вы должны будете использовать номера).

Поскольку присваиваемое классу объектов обозначение в принципе произвольно (хотя после присвоения и однозначно), эту свободу в выборе можно использовать для удобства. Так, при большом и/или нефиксированном числе классов их конкретизация упрощается и облегчается, если обозначения вводятся иерархически. Примером могут слу-

жить почтовые адреса: страна – территориальная административная единица (республика, штат, кантон, графство, область) – населенный пункт – улица – дом – квартира – адресат. Другой пример – автомобильные номера: в их символике есть обозначение как территории, так и принадлежности машины (государственная или личная).

Необходимость классификации возникает и в тех случаях, когда классифицируемые состояния образуют непрерывное множество. Задача сводится к предыдущей, если все множество разбить на конечное число подмножеств, искусственно образуя тем самым классы эквивалентности. Теперь принадлежность состояния к какому-либо классу снова можно регистрировать в шкале наименований. Однако условность введенных классов (не их шкальных обозначений, а самих классов) рано или поздно проявится на практике. Например, возникают трудности точного перевода с одного языка на другой при описании цветовых оттенков: в английском языке голубой, лазоревый и синий цвета не различаются; не исключено, что англичане иначе видят мир (например, в одном английском толковом словаре слово «синий» объясняется как «цвет чистого неба, древесного дыма, снятого молока, свинца», а в другом – как «цвет неба или моря, а также вещей namного бледнее или темнее, как дым, удаленные холмы, лунный свет, синяк”).

Перейдем теперь к вопросу о допустимых операциях над данными, выраженными в номинальной шкале. Подчеркнем еще раз, что обозначения классов – это только символы, даже если для этого использованы номера. Номера лишь внешне выглядят как числа, но на самом деле числами не являются. Если у одного спортсмена на спине номер 4, а другого – 8, то никаких других выводов, кроме того, что это разные участники соревнований, делать нельзя: так, нельзя сказать, что второй «в два раза лучше» или что у одного из них форма новее. С номерами нельзя обращаться как с числами, за исключением определения их равенства или неравенства: только эти отношения определены между элементами номинальной шкалы.

Поэтому при обработке экспериментальных данных, зафиксированных в номинальной шкале, непосредственно с самими данными можно выполнять только операцию проверки их совпадения или несовпадения. Изобразим эту операцию с помощью символа Кронекера: $\delta_{ij} = \{1: x_i = x_j; 0: x_i \neq x_j\}$, где x_i и x_j – записи разных измерений.

С результатами этой операции можно выполнять более сложные преобразования: считать количества совпадений (например, число наблюдений k -го класса равно $n_k = \sum_{j=1}^n \delta_{kj}$, где n – общее число наблюдений), вычислять относительные частоты классов (например, частота

k -го класса есть $p_k = n_k/n$), сравнивать эти частоты между собой, выполнять различные статистические процедуры, строго следя, однако, чтобы в этих процедурах с исходными данными не выполнялось ничего, кроме операции проверки их на совпадение (например, можно использовать χ^2 -тест, другие тесты на относительных частотах, коэффициент согласия и т.д.).

Измерение – операция, ставящая наблюдаемому явлению в соответствие один из элементов подходящей измерительной шкалы. Измерительная шкала может иметь разную силу, в зависимости от того, являются ли ее элементы символами, номерами или числами. Измерительную шкалу следует выбирать максимально сильной, однако сила шкалы должна соответствовать природе наблюдаемого явления и не быть завышенной.

5.1.2. Порядковые шкалы

В тех случаях, когда наблюдаемый (измеряемый) признак состояния имеет природу, позволяющую не только отождествлять состояния с одним из классов эквивалентности, но и в каком-то отношении сравнивать разные классы, для измерений можно выбрать более сильную шкалу, чем номинальная. Если же не воспользоваться этим, то мы откажемся от части полезной информации. Однако усиление измерительной шкалы зависит от того, какие именно отношения между классами существуют в действительности. Это и явилось причиной появления измерительных шкал разной силы.

Следующей по силе за номинальной шкалой является порядковая шкала (используется также название «ранговая шкала»). Этот класс шкал появляется, если кроме аксиом тождества 1–3 классы удовлетворяют следующим аксиомам упорядоченности:

4. Если $A > B$, то $B < A$.

5. Если $A > B$ и $B > C$, то $A > C$.

Обозначив такие классы символами и установив между этими символами те же отношения порядка, мы получим шкалу простого порядка. Примерами применения такой шкалы являются нумерация очередности, воинские звания, призовые места в конкурсе.

Иногда оказывается, что не каждую пару классов можно упорядочить по предпочтению: некоторые пары считаются равными. В таком случае аксиомы 4 и 5 видоизменяются:

4'. Либо $A < B$, либо $A > B$.

5'. Если $A > B$ и $B > C$, то $A > C$.

Шкала, соответствующая аксиомам 4' и 5', называется шкалой слабого порядка. Примером такой шкалы служит упорядочение по сте-

пени родства с конкретным лицом (мать = отец > сын = дочь, дядя = тетья < брат = сестра и т.п.).

Иная ситуация возникает, когда имеются пары классов, несравнимые между собой, т.е. ни $A \leq B$, ни $B \leq A$ (это отличается от условия слабого порядка, когда одновременно $A \geq B$ и $B \geq A$, т.е. $A = B$). В таком случае говорят о шкале частичного порядка. Шкалы частичного порядка часто возникают в социологических исследованиях субъективных предпочтений. Например, при изучении покупательского спроса субъект часто не в состоянии оценить, какой именно из двух разнородных товаров ему больше нравится (например, клетчатые носки или фруктовые консервы, велосипед или магнитофон и т.д.); затрудняется человек и упорядочить по предпочтению любимые занятия (чтение литературы, плавание, вкусная еда, слушание музыки и т.д.).

Характерной особенностью порядковых (в строгом смысле) шкал является то, что отношение порядка ничего не говорит о дистанции между сравниваемыми классами. Поэтому порядковые экспериментальные данные, даже если они изображены цифрами, нельзя рассматривать как числа, над ними нельзя выполнять действия, которые приводят к получению разных результатов при преобразовании шкалы, не нарушающем порядка.

Однако допустима операция, позволяющая установить, какое из двух наблюдений – x_i или x_j – предпочтительнее, хотя формально эту операцию мы можем выразить через разность ($x_i - x_j$).

Введем индикатор положительных чисел – функцию

$$C(t) = \{1: t \geq 0; \quad 0: t < 0\}.$$

Тогда если $x_i \geq x_j$ и мы ввели цифровую шкалу порядка, то $C(x_i - x_j) = 1$, а $C(x_j - x_i) = 0$, что и позволяет установить предпочтительность x_i перед x_j . Число $R_i = \sum_{j=1}^n C(x_i - x_j)$, где n – число сравни-

ваемых объектов ($1 \leq R_i \leq n$), называется рангом i -го объекта. (Отсюда происходит другое название порядковых шкал – ранговые.) Если имеет место слабый порядок, то часть наблюдений совпадает (в статистике такая группа наблюдений называется связкой) и все члены связки получают одинаковый (старший для них) ранг. Когда это неудобно, прибегают либо к присвоению ранга, среднего для данной связки (мидранга), либо присваивают ранги от младшего до старшего случайным образом.

Итак, при измерениях в порядковых (в строгом смысле) шкалах обработка данных должна основываться только на допустимых для этих шкал операциях – вычислении δ_{ij} и R_i . С этими числами можно «работать» дальше уже произвольным образом: кроме нахождения час-

тот и мод (как и для порядковой шкалы) появляется возможность определить выборочную медиану (т.е. наблюдение с рангом R_i , ближайшим к числу $n/2$); можно разбить всю выборку на части в любой пропорции, находя выборочные квантили любого уровня p , $0 < p < 1$ (т.е. наблюдения с рангом R_i , ближайшим к величине n_p); можно определить коэффициенты ранговой корреляции между двумя сериями порядковых наблюдений; строить с помощью полученных величин другие статистические процедуры.

Подчеркнем еще раз, что даже в тех случаях, когда состояния, которые допускают только порядковые сравнения, в эксперименте измеряются через величины, связанные с ними косвенно, но фиксируемые в числовых шкалах, эти измерения все равно остаются измерениями в порядковой шкале.

В качестве примера рассматривается испытание умственных способностей, при котором измеряется время, затрачиваемое испытуемым на решение тестовой задачи. В таких экспериментах время хотя и измеряется в числовой шкале, но, как мера интеллекта, принадлежит порядковой шкале.

5.1.3. Модифицированные порядковые шкалы

По-видимому, опыт работы с сильными числовыми шкалами и желание уменьшить относительность порядковых шкал, придать им хотя бы внешнюю независимость от измеряемых величин побуждают исследователей к различным модификациям, придающим порядковым шкалам некоторое (чаще всего кажущееся) усиление. Другая важная причина попыток усиления шкалы состоит в том, что многие измеряемые в порядковых (принципиально дискретных) шкалах величины имеют действительный или мыслимый непрерывный характер: сила ветра или землетрясения, твердость вещества, глубина и прочность знаний, овладение навыками и т.п. Сама возможность введения между любыми двумя шкальными значениями третьего способствует тому, чтобы попытаться усилить шкалу.

Все это вместе взятое привело к появлению и использованию на практике ряда порядковых шкал, но не в таком «строгом смысле», как те, о которых мы говорили выше. При этом иногда с полученными данными начинают обращаться как с числами, даже если произведенная модификация не выводит шкалу из класса порядковых. Это сопряжено с ошибками и неправильными решениями. Рассмотрим некоторые из известных модификаций.

Балльные шкалы оценки знаний учащихся. Слушая ответы учащихся или сравнивая их письменные работы, опытный преподаватель

может обнаружить разницу между ними и установить, чьи ответы лучше; это типичное отношение порядка. Методом сравнения можно определить, кто в классе лучше других знает данный предмет; сложнее, но иногда возможно (это зависит от состава класса) определить лучшего ученика в классе. Сравнение старшеклассника с учеником младшего класса по степени овладения знаниями проблематично.

Потребность общества в официальном определении степени квалифицированности проходящих обучение независимо от того, где, когда и как они получают образование, способствовала введению общепринятых шкал для оценивания знаний учащихся в виде баллов (такие шкалы введены повсеместно). Все испытывают, в том числе и на собственном опыте, неточность, приблизительность этой шкалы. Одна из попыток «улучшить» шкалу баллов состоит в увеличении числа градаций. В наших школах принята 5-балльная, в вузах – 2-балльная (для зачетов) и 4-балльная (для экзаменов) системы оценок, в некоторых европейских странах – 10-балльная, а в англоязычных странах – 100-балльная система. Это не спасает положения, и преподаватели неофициально («для себя») вводят дополнительные градации – присоединяют к баллам плюсы, минусы, точки. Примечательно, что и при 100-балльной шкале некоторые преподаватели используют дробные баллы. Все это происходит потому, что не существует ни абсолютного стандарта, единого для всех людей, ни даже условного общедоступного стандарта, наподобие эталонов твердости или высоты волн, и знания могут оцениваться только в порядковой шкале. Тем не менее мало кто (не только учащиеся, но и преподаватели) понимает, что балльная шкала принадлежит к классу порядковых. Дело доходит до того, что даже в официальных вопросах, влияющих на судьбы людей, учитывают среднеарифметический балл – величину, не имеющую смысла в порядковой шкале! Некоторый оттенок объективности и количественности балльной шкале пытаются придать директивным определением того, каким требованиям должен удовлетворять учащийся, чтобы иметь право на тот или иной балл, т.е. ввести независимые стандарты. Однако преподаватели неизбежно по-разному понимают и выполняют инструкции, и оценки все равно получаются относительными: известно, что уровни знаний отличников разных школ или вузов заметно различаются. Именно поэтому в ответственных случаях устраивают не конкурсы документов об успеваемости, а конкурсы самих претендентов, т.е. возвращаются к порядковому измерению, непосредственному сравнению обладателей знаний.

Порядковая шкала Черчмена и Акоффа. В социологических исследованиях часто оказывается полезным и возможным предложить

опрашиваемому не только упорядочить заданный перечень альтернатив, но и указать, хотя бы грубо, силу предпочтения. Это уже существенная модификация упорядочения, и, как будет показано в дальнейшем, при достаточно жестких требованиях к весовым коэффициентам измерения могут быть переведены в разряд более сильных шкал, нежели шкала порядка. Мы же, обсуждая пока именно порядковые шкалы, рассмотрим случай, когда и взвешивание упорядоченных классов не выводит шкалу из разряда порядковых, хотя разницу между весами классов можно интерпретировать как «расстояние» между ними. Метод измерения проиллюстрируем на примере.

Пусть имеются четыре предмета. Сначала опрашиваемый упорядочивает их в порядке предпочтения: $A \geq B \geq C \geq D$. Затем его просят поставить в соответствие (приписать) предметам любые числа между нулем и единицей, выразив грубо «силу» предпочтения. Пусть результат таков:

A	B	C	D
1,00	0,85	0,75	0,20

Целью является уточнение с помощью дальнейших вопросов действительной силы предпочтений опрашиваемого. Например, что он предпочитает – A или B , C и D вместе взятые. Результат необходимо как-то отразить в весовых коэффициентах. Делается предположение, что весовой коэффициент совокупности альтернатив равен сумме их весовых коэффициентов. Если, например, $A > B \cap C \cap D$, приписывают новые коэффициенты:

A	B	C	D
1,00	0,65	0,20	0,10

Далее спрашивают, как упорядочиваются B и $C \cap D$. Если, по мнению опрашиваемого, $C \cap D > B$, то уменьшают вес B так, чтобы он был меньше суммы весов C и D ;

A	B	C	D
1,00	0,25	0,20	0,10

Другие начальные веса при указанных вопросах и ответах могут оставаться неизменными, если они сразу отвечали указанным требованиям. Например:

A	B	C	D	A	B	C	D
1,00	0,33	0,33	0,33	1,00	0,04	0,03	0,03

Чтобы уменьшить количество перебираемых комбинаций при уточнении шкалы, авторы метода предлагают наиболее предпочтительной

альтернативе приписывать единичный вес, а остальные группировать по три и действовать по указанной методике. Если и при этом количество перебираемых комбинаций окажется большим (что неизбежно при большом числе упорядочиваемых объектов), то можно прибегнуть к неполному перебору, применив случайный механизм выбора троек и установив критерий прекращения пересчета весов.

Основным предметом критики порядковой шкалы Черчмена и Акоффа является тот факт, что предположение об аддитивности весов предпочтения в психологии нередко не выполняется: скажем, опрашиваемый может оценивать смесь меда с дегтем иначе, чем суммой весов меда и дегтя в отдельности; то же может относиться и к оценке хлеба с маслом и хлеба и масла в отдельности.

Данные, зафиксированные в протоколе эксперимента, принадлежат определенной измерительной шкале. При обработке данных важно следить за тем, чтобы над ними выполнялись только такие действия, которые допустимы для использованной шкалы. Нарушение этого правила может привести к неправомерной интерпретации результатов опыта.

5.1.4. Шкалы интервалов

Если упорядочивание объектов можно выполнить настолько точно, что известны расстояния между любыми двумя из них, то измерение окажется заметно сильнее, чем в шкале порядка. Естественно выражать все расстояния в единицах, хотя и произвольных, но одинаковых по всей длине шкалы. Это означает, что объективно равные интервалы измеряются одинаковыми по длине отрезками шкалы, где бы они на ней ни располагались. Следствием такой равномерности шкал этого класса является независимость отношения двух интервалов от того, в какой из шкал эти интервалы измерены (т.е. какова единица длины интервала и какое значение принято за начало отсчета). В самом деле, если два интервала в одной шкале выражаются числами Δ_1x и Δ_2x , а при другом выборе нуля и единицы – числами Δ_1y и Δ_2y , то, поскольку это объективно те же самые интервалы, имеем $\Delta_1x/\Delta_2x = \Delta_1y/\Delta_2y$; откуда следует, что введенные шкалы могут иметь произвольные начала отсчета и единицы длины, а связь между показаниями в таких шкалах является линейной: $y = ax + b$, $a > 0$, $-\infty < b < \infty$. Это соотношение можно выразить словами: «шкала интервалов единственна с точностью до линейных преобразований». Построенные таким образом шкалы называются интервальными.

Примерами величин, которые по физической природе либо не имеют абсолютного нуля, либо допускают свободу выбора в установлении

начала отсчета и поэтому измеряются в интервальных шкалах, являются температура, время, высота местности.

Название «шкала интервалов» подчеркивает, что в этой шкале только интервалы имеют смысл настоящих чисел и только над интервалами следует выполнять арифметические операции: если произвести арифметические операции над самими отсчетами по шкале, забыв об их относительности, то имеется риск получить бессмысленные результаты. Например, если сказать, что температура воды увеличилась в два раза при ее нагреве от 9 до 18° по шкале Цельсия, то для тех, кто привык пользоваться шкалой Фаренгейта, это будет звучать весьма странно, так как в этой шкале температура воды в том же опыте изменится от 37 до 42°.

Подобно тому как определение значения символа Кронекера является единственной допустимой операцией над наблюдениями в номинальной шкале, а вычисление ранга наблюдения – в порядковой шкале, в интервальной шкале единственной новой допустимой операцией над наблюдениями является определение интервала между ними. Над интервалами же можно выполнять любые арифметические операции, а вместе с ними использовать подходящие способы статистической и иной обработки данных. Например, центральные моменты (в том числе дисперсия) имеют объективный физический смысл, а начальные моменты (в том числе среднее значение) являются относительными наряду с началом отсчета. Поэтому понятие относительной погрешности – коэффициента вариации, т.е. отношения стандартного отклонения к математическому ожиданию, – не имеет смысла для интервальной шкалы. Это не означает, что вообще нельзя суммировать показания в шкале интервалов, например вычислять выборочное среднее $\bar{x} = \frac{1}{n} \sum x_i$. Однако с такой величиной нужно обращаться так же, как и с другими исходными наблюдениями: она остается интервальной величиной и приобретает числовой смысл только в процессе определения интервалов. Поэтому выборочная дисперсия имеет объективный смысл, хотя и определяется через \bar{x} по формуле

$$S^2 = \overline{M(X - \bar{x})^2}.$$

Дело в том, что $X - \bar{x}$ является интервалом.

5.1.5. Шкалы отношений

Пусть наблюдаемые величины удовлетворяют не только аксиомам 4 и 5, но и аксиомам аддитивности:

6. Если $A = P$ и $B > 0$, то $A + B > P$.

$$7. A + B = B + A.$$

$$8. \text{ Если } A = P \text{ и } B = Q, \text{ то } A + B = P + Q.$$

$$9. (A + B) + C = A + (B + C).$$

Это существенное усиление шкалы: измерения в такой шкале являются «полноправными» числами, с ними можно выполнять любые арифметические действия, так как вычитание, умножение и деление – лишь частные случаи сложения. Введенная таким образом шкала называется шкалой отношений. Этот класс шкал обладает следующей особенностью: отношение двух наблюдаемых значений измеряемой величины не зависит от того, в какой из таких шкал произведены измерения: $x_1/x_2 = y_1/y_2$. Этому требованию удовлетворяет соотношение вида $y = ax$ ($a \neq 0$). Таким образом, величины, измеряемые в шкале отношений, имеют естественный, абсолютный нуль, хотя остается свобода в выборе единиц.

Примерами величин, природа которых соответствует шкале отношений, являются длина, вес, электрическое сопротивление, деньги.

5.1.6. Шкалы разностей

К числу шкал, единственных с точностью до линейных преобразований, относятся шкала интервалов ($y = ax + b$, $a > 0$ и b произвольны) и шкала отношений $y = ax$, $a > 0$ – преобразование растяжения). Рассмотрим особенности шкал, инвариантных к сдвигу: $y = x + b$.

Повторно применяя сдвиг к y ($z = y + b = x + 2b$), затем к z и т.д., обнаруживаем, что в такой шкале значение не изменяется при любом числе сдвигов: $y = x + nb$, $n = 0, 1, 2, \dots$. Постоянная b является характерным параметром шкалы и называется ее периодом. Полученную шкалу будем называть шкалой разностей (иногда ее также называют циклической или периодической). В таких шкалах измеряется направление из одной точки (шкала компаса, роза ветров и т.д.), время суток (циферблат часов), фаза колебаний (в градусах или радианах).

Циклические шкалы являются частным случаем интервальных шкал. Однако соглашение о хотя и произвольном, но едином для нас начале отсчета шкалы позволяет использовать показания в этой шкале как числа, применять к ним арифметические действия (до тех пор, пока кто-нибудь не забудет об условности нуля, например, при переходе на летнее время или обратно).

5.1.7. Абсолютная шкала

Рассмотрим такую шкалу, которая имеет и абсолютный нуль, и абсолютную единицу. Эта шкала не единственна с точностью до какого-либо преобразования, а просто уникальна. Именно такими качествами

обладает числовая ось, которую естественно назвать абсолютной шкалой. Важной особенностью абсолютной шкалы по сравнению со всеми остальными является отвлеченность (безразмерность) и абсолютность ее единицы. Указанная особенность позволяет производить над показаниями абсолютной шкалы такие операции, которые недопустимы для показаний других шкал: употреблять эти показания в качестве показателя степени и аргумента логарифма. Числовая ось используется как измерительная шкала в явной форме при счете предметов, а как вспомогательное средство присутствует во всех остальных шкалах. Внутренние свойства числовой оси при всей кажущейся ее простоте оказываются чрезвычайно разнообразными, и теория чисел до сих пор не исчерпала их до конца. А некоторые безразмерные числовые отношения, обнаруживаемые в природе, вызывают восхищение и изумление (явления резонанса; гармонические отношения размеров, звуков; законы теории подобия и размерности; квантование энергии элементарных частиц и т.п.).

В таблице приведены основные сведения о всех рассмотренных в данном подразделе измерительных шкалах. Можно сказать, что чем сильнее шкала, в которой производятся измерения, тем больше сведений об изучаемом объекте, явлении, процессе дают измерения. Поэтому так естественно стремление каждого исследователя провести измерения в возможно более сильной шкале. Однако важно иметь в виду, что выбор шкалы измерения должен ориентироваться на объективные отношения, которым подчинена наблюдаемая величина, и лучше всего производить измерения в той шкале, которая максимально согласована с этими отношениями. Можно измерять и в шкале, более слабой, чем согласованная (это приведет к потере части полезной информации), но применять более сильную шкалу опасно: полученные данные реально не будут иметь той силы, на которую ориентируется их обработка.

Аналогичная ситуация имеет место и после того, как проведены измерения. У исследователя могут быть причины, побуждающие его выполнять преобразования протокола наблюдений, переводя их из одной шкалы в другую.

Если при этом данные переводятся в более слабую шкалу, то обычно исследователь отдает себе отчет в том, что в результате происходит некоторое ухудшение качества выводов. Иногда же исследователи усиливают шкалы; типичный случай – «оцифровка» качественных шкал: классам в номинальной или порядковой шкале присваиваются номера, с которыми дальше «работают» как с числами. Если в этой обработке не выходят за пределы допустимых преобразований, то «оцифровка» – это просто перекодировка в более удобную (например,

для ЭВМ) форму. Однако применение других операций сопряжено с заблуждениями и ошибками, так как свойства, навязываемые подобным образом, на самом деле не имеют места.

Измерительные шкалы

Шкала	Номинальная	Порядковая	Интервальная
Характеристики			
Определяющие отношения	Эквивалентность	Эквивалентность. Предпочтение	Эквивалентность; предпочтение; сохранение отношения интервалов
Эквивалентное преобразование шкал	Перестановки наименований	Не изменяющее порядка (монотонное)	Линейное преобразование $y = ax + b, a > 0, b \in R$
Допустимые операции над данными (первичная обработка)	Вычисление символа Кронекера δ_{ij}	Вычисление δ_{ij} и рангов R_i	Вычисление δ_{ij} , рангов R_i и интервалов – разностей между наблюдениями
Вторичная обработка данных	Вычисление относительных частот и операций над ними	Вычисление относительных частот и выборочных квантилей, операции над ними	Арифметические действия над интервалами
Определяющие отношения	Эквивалентность; предпочтение; сохранение отношения интервалов; периодичность	Эквивалентность; предпочтение; сохранение отношения интервалов; сохранение отношения двух значений	Эквивалентность; предпочтение; сохранение отношения интервалов; сохранение отношения двух значений; абсолютная и безразмерная единица; абсолютный нуль
Эквивалентное преобразование шкал	Сдвиг $y = x + nb, b = \text{const}, n = 0, 1, 2, \dots$	Растяжение $y = ax, a > 0$	Шкала уникальна
Допустимые операции над данными (первичная обработка)	То же, что и для интервальной шкалы	Все арифметические операции	Все арифметические операции; использование в качестве показателя степени, основания и аргумента логарифма
Вторичная обработка данных	То же, что и для интервальной шкалы	Любая подходящая обработка	Любая необходимая обработка

Стоит упомянуть и об еще одной особенности преобразований протоколов наблюдений: некоторые из преобразований могут ненамеренно изменить уровень шкалы. Например, в акустике и радиотехнике часто

отношение мощностей сигналов представляется в децибелах: $N = 10 \lg (P_2/P_1)$. Мощности P_1 и P_2 измеряются в шкале отношений; следовательно, все необходимые операции допустимы. Но величина N принадлежит шкале интервалов, что следует учитывать при дальнейшем оперировании с нею (например, нельзя говорить, что мощность данного сигнала равна такому-то количеству децибел, и не указать, в сравнении с чем).

5.2. Расплывчатое описание ситуаций

Обратим внимание на то, что все, рассмотренные измерительные шкалы имеют одно общее свойство: они основаны на справедливости отношения эквивалентности (таблицу). Это отношение имеет силу как отдельно на множестве состояний наблюдаемого объекта и множестве наблюдений, зафиксированных в любой из шкал (два состояния или два измерения либо тождественны, либо различны), так и на их совокупности (состояния и соответствующие им измерения находятся во взаимно однозначном соответствии). Использование рассогласованной (т.е. более слабой, чем можно) шкалы приводит к образованию на множестве состояний новых классов эквивалентности, внутри которых состояния неразличимы в данной шкале (хотя их и можно различить в более сильной шкале). Однако и в этом случае отношение эквивалентности соблюдается.

Встречаются (и гораздо чаще, чем кажется) случаи, когда тождество или различие двух состояний и/или наблюдений нельзя утверждать с полной уверенностью. Наиболее явно это видно на примере шкал, в которых классы обозначаются конструкциями естественного языка. «В комнату вошел высокий молодой человек» – класс, к которому принадлежит человек, назван (т.е. измерение состоялось), но какого он роста и сколько ему лет? «В руках он держал довольно тяжелый сверток» – какого веса была его ноша? Если разобраться, то почти каждое наше слово обозначает некоторое не вполне определенное множество. («Почти» – какой процент? «Наше» – чье именно? «Некоторое» – какое же? «Не вполне» – насколько? «Определенное» – кем и как?» и т.д.) Это свойство естественного языка, природное и неотъемлемое, безусловно, полезное (иначе бы оно не закрепилось в процессе развития языка), но приводящее к затруднениям, когда сопровождающая его неопределенность мешает. Древние логики дискутировали вопрос о том, сколько песчинок должно быть собрано вместе, чтобы получилась куча песка; сегодня мы просто говорим, что слово «куча» – это лишь метка нечетко определенного множества. Спор о том, сколько песчинок в «куче»,

эквивалентен спору о том, в каком возрасте человек становится «старым» или сколько волосинок должно у него выпасть, чтобы он был «лысым».

Эта неопределенность смысла языковых конструкций является одной из основных трудностей автоматизации анализа и синтеза речи, автоматического (и не только автоматического) перевода с одного языка на другой. Например, одному английскому предложению, состоящему из пяти слов, можно дать пять разных (!) смысловых интерпретаций:

TIME FLIES LIKE AN ARROW
ВРЕМЯ ЛЕТИТ СТРЕЛОЙ
ВРЕМЯ ЛЕТИТ В НАПРАВЛЕНИИ СТРЕЛЫ
МУХАМ ВРЕМЕНИ НРАВИТСЯ СТРЕЛА
ИЗМЕРЯЙ СКОРОСТЬ МУХ ТАК ЖЕ, КАК СКОРОСТЬ СТРЕЛЫ
ИЗМЕРЯЙ СКОРОСТЬ МУХ, ПОХОЖИХ НА СТРЕЛУ

Все сказанное выше мотивирует введение понятия лингвистической переменной как переменной, значение которой расплывчато по своей природе, как метки размытого, расплывчатого множества. Хотя теория размытых множеств прекрасно иллюстрируется языковыми примерами и имеет интересные приложения в области искусственного интеллекта, размытость оказывается свойством не только естественного языка. Например, в математике с успехом применяются понятия «значительно больше» (символ \gg) и «приблизительно равно» (символ \approx или \simeq), являющиеся типично расплывчатыми.

Расплывчатость – это такое свойство явлений, при котором не выполняется отношение эквивалентности: явление одновременно может принадлежать данному классу и не принадлежать ему. Неопределенность такого типа описывается с помощью функции принадлежности; значение этой функции выражает степень уверенности, с которой мы относим данный объект к указанному классу. Сам класс в итоге становится не определяемым однозначно и называется расплывчатым множеством.

5.3. Вероятностное описание ситуаций. Статистические измерения

Говоря о наблюдениях над изучаемым объектом и о фиксации результатов этих наблюдений (измерений), напомним, что сама необходимость обращения к эксперименту вытекает из того, что нужно устранить некоторую неопределенность, свойственную нашим знаниям об

объекте до проведения этого эксперимента. В некоторых случаях эксперимент устраняет неопределенность полностью (как при бросании монеты или контрольном замере уровня масла в двигателе); в других случаях неопределенность лишь уменьшается до некоторого предела, относительного (т.е. в принципе преодолимого) или абсолютного (не уменьшаемого). Ясно, что и организация эксперимента, и обработка экспериментальных данных, определяющие степень уменьшения неопределенности, должны исходить из природы, существа, причины неопределенности.

5.3.1. Понятие случайной неопределенности

Оказывается, что неопределенность бывает разного происхождения. Один из ее видов – неизвестность – рассматривается теорией познания и философией; такого типа неопределенность характеризует ситуацию, когда мы задаемся вопросом «есть ли жизнь на Марсе?» (посадка советской автоматической станции на эту планету уменьшила неопределенность, но не сняла ее совсем) или «существуют ли внеземные цивилизации?» (поиск возможных искусственных радиосигналов, пока, к сожалению, безуспешен). Другой вид неопределенности – расплывчатость – был обсужден в подразд. 5.2; для нее характерно, что эксперимент в принципе не снимает ее полностью. Третий вид неопределенности – случайность – мы кратко рассмотрим сейчас; при этом будем исходить из элементов теории вероятностей.

Говоря о случайных явлениях, прежде всего обращают внимание на их непредсказуемость, противопоставляют случайность детерминированности, хаотичность – упорядоченности. Имеющее определенный смысл, такое противопоставление является односторонним, так как оставляет в тени тот факт, что под случайностью понимается вид неопределенности, подчиняющийся строгой закономерности, которая выражается распределением вероятностей. Зная распределение (например, плотность $p(x)$) вероятностей, можно ответить на любой вопрос о случайной величине; в каком интервале находятся ее возможные значения (определим носитель распределения X – множество элементов x , для которых $p(x) > 0$); около какого значения рассеиваются ее реализующие значения (найдем параметр положения распределения, например среднее, моду или медиану); насколько сильно разбросаны эти значения (найдем масштабный параметр – дисперсию или стандартное отклонение, средний модуль разности, энтропию); какова связь между разными реализациями (вычислим заданную меру зависимости) и т.д.

5.3.2. О природе случайности

Самая полная информация, которой можно располагать о случайном объекте, содержится в распределении вероятностей по возможным состояниям этого объекта. Важно, что само конкретное распределение уже есть закономерность: неоднозначная (но и не вполне произвольная) для того, какое именно состояние реализуется, и вполне однозначная для многих важных характеристик, выражаемых функционалами от распределения.

Различные определения вероятности рассматриваются в соответствующих курсах; мы же остановимся на природе случайности. Существует несколько точек зрения на этот счет, причем каждая из них имеет достаточные основания.

1. Согласно первой точке зрения, случайным нам представляется нечто такое, в чем мы пока не уловили закономерности. По мере познания явления в последнем остается всё меньше и меньше случайного. Ярким выразителем такой позиции был Лаплас, считавший, что случайность не присуща самим объектам, а связана только с незнанием, в принципе устранимым.

2. Противоположная точка зрения состоит в том, что случайность является объективным свойством всех явлений. Более ста лет назад О. Курно писал, что «случайность вмешивается во все, что творится на свете», что «миром управляет случай, или, говоря точнее, случай имеет свою часть, и притом весьма значительную, в управлении вселенной».

3. Промежуточная позиция признает как существование вполне детерминированных явлений, так и в принципе случайных, описываемых статистическими закономерностями (статистические законы физики, химии, термодинамики; законы квантовой механики и т.д.). Таким образом, случайность признается объективным свойством лишь некоторых явлений.

5.3.3. Статистические измерения

При всем различии этих точек зрения они не столь несовместимы, как это представляется на первый взгляд. Рассмотрим наглядную простейшую задачу обнаружения постоянного сигнала в нормальном шуме. Сигнал может принимать одно из двух (известных заранее) значений, S или O , с вероятностями P и Q соответственно (это и есть случайность незнания). В любом из этих случаев мы можем наблюдать только аддитивную смесь сигнала с гауссовым шумом, т.е. иметь выборку x_1, \dots, x_N либо из распределения $N_x(O, \sigma^2)$, либо из распределения $N_x(S, \sigma^2)$. Здесь шум представляет собой объективную и неустранимую

случайность, подчиненную закономерности нормального распределения. Зная априори величины P , Q , S , O , функцию $N_x(a, \sigma^2)$ при любых a и σ^2 , а также используя наблюдения x_1, \dots, x_N , мы можем уменьшить неопределенность того, какое же из возможных значений, S или O , имеет место, т.е. уменьшить случайность незнания. Однако объективная случайность шума не позволяет сделать это безошибочно; даже при оптимальных методах обработки измерений x_1, \dots, x_N вероятности ошибок отличны от нуля, хотя при неограниченном увеличении N они стремятся к нулю.

Как видим, в практических задачах объективные и субъективные случайности неразделимо переплетены. Такое слияние может быть еще более тесным: например, в непараметрической статистике распределения, характеризующие объективную случайность, считаются лишь существующими, но функционально неизвестными, т.е. субъективное незнание распространяется и на описание объективной случайности.

Итак, как и любые эксперименты, измерения случайных величин и процессов выполняются для уточнения их моделей, снятия или уменьшения неопределенности незнания. Обычно достаточно знать не всё распределение, а лишь какой-то из его параметров, и тогда задача сводится к оценке этого параметра по наблюдаемой выборке. Хотя это уже «вторичная» обработка данных, измерение выборочных значений и вычисление оценки в совокупности можно трактовать как «измерение параметра». То же относится и к определению по выборке более сложных характеристик – самих распределений, регрессий, корреляций, спектров и т.д. Такое совместное рассмотрение непосредственных измерений и их обработки оказывается полезным еще и потому, что можно проводить общую оптимизацию этого процесса, и она далеко не всегда совпадает с оптимизацией компонент в отдельности.

Все эти соображения и дают основания ввести понятие статистических измерений, рассматривать эту проблематику как самостоятельный раздел метрологии со своей теорией и измерительной техникой.

В заключение подчеркнем еще раз, что статистический, вероятностный подход относится к неопределенности, описываемой распределениями вероятностей. На то, что методы статистики надо применять осторожно, что многие экспериментальные ситуации могут быть хотя и хаотическими, но не иметь вероятностного характера, обращали внимание многие исследователи.

Еще один важный момент состоит в том, чтобы по возможности ослабить или хотя бы учесть влияние измерений на наблюдаемый объект. Особенно это существенно при социальных исследованиях, наблюдениях за людьми: сам факт осознания, что они являются объектом вни-

мания, заметно меняет их поведение. Воздействие измерительного устройства на измеряемый объект должно также учитываться при физических и химических экспериментах.

5.3.4. Регистрация экспериментальных данных и ее связь с их последующей обработкой

Результаты любого эксперимента фиксируют в той или иной форме, а затем используют для той цели, ради которой и проводился эксперимент. Иногда эти операции практически совмещены во времени, например при автоматическом управлении производственным процессом, при автоматизации экспериментов в реальном масштабе времени и т.п. В некоторых же видах человеческой практики (научные исследования; системный анализ; контрольная, ревизионная, следственная и другие виды административной деятельности; учебные эксперименты и пр.) обработка экспериментальных данных является отдельным, самостоятельным этапом, промежуточным между этапами получения информации (измерения) и ее использования (принятия решений и их выполнения). В таких случаях исходной информацией для обработки являются протоколы наблюдений (называемые также матрицами данных, экспериментальными таблицами).

Характер самих протоколов наблюдений и методы их обработки зависят от того, какова модель, для уточнения которой ставится эксперимент: фактически обработка данных – это просто преобразование информации к виду, удобному для использования, перевод ответов природы с языка измерений на язык уточняемой модели. Наши знания могут быть как первоначальными, грубыми, так и далеко продвинутыми, хорошо структурированными, хотя и требующими уточнения. Соответствующие два типа моделей разные авторы называют по-разному, в зависимости от того, какой их аспект они хотят подчеркнуть: дескриптивные и конструктивные, качественные и количественные, декларативные и процедурные, классификационные и числовые. Будем пользоваться последними терминами; отметим, что в них отражено и то различие, что классификационные модели описывают множество различных объектов, а числовые – один объект (или множество схожих объектов). Отметим также, что в классификационной модели могут участвовать количественные переменные, и это не меняет ее качественного характера (например, диагноз больному ставится с учетом количественных анализов); аналогично, в числовых моделях часть переменных может измеряться в слабых шкалах. Рассмотрим кратко особенности экспериментальных данных и их обработки для обоих типов моделей.

5.4. Классификационные модели

Всякая статистическая задача состоит в том, чтобы, несмотря на неустранимую неопределенность наблюдений «внутри» распределения, снять (или точнее, уменьшить до приемлемого предела) неопределенность интересующей нас характеристики распределения.

Классификационные модели являются основополагающими, первичными, исходными формами знания. Узнавание окружающих предметов – типичный пример классификационных процессов в мыслительной деятельности человека (и животных). И в науке познание начинается с соотнесения изучаемого объекта с другими, выявления сходства и различия между ними. Поэтому протокол наблюдений на классификационном уровне эксперимента содержит результаты измерения ряда признаков X для подмножества A объектов, выбранных из множества Γ : каждый объект $a_i \in A \subseteq \Gamma$ обладает значениями признаков $x_i = (x_{i0}, x_{i1}, \dots, x_{in}) \in \{X_0, X_1, \dots, X_n\} = X$, $i = \overline{1, N}$, n – число признаков, N – число объектов в A . Признак характеризует конкретное свойство объекта; поэтому и тогда такой протокол называют таблицей «объект – свойство».

Как уже отмечалось, способ обработки протокола зависит от цели обработки. Часто оказывается, что задача может быть сформулирована как определение по наблюдавшимся значениям признаков $x = (x_1, \dots, x_n)$ значений ненаблюдаемого («целевого») признака x_0 . Как правило, целевыми признаками являются те параметры модели, которые требуется уточнить по экспериментальным данным.

Рассмотрим различные типы задач для классификационных моделей.

Кластеризация (поиск «естественной» группировки объектов). Не заданы ни границы классов в пространстве признаков, ни число классов! Требуется их определить исходя из «близости», «похожести» или «различия» описаний объектов $x_i = (x_{i1}, \dots, x_{in})$. Компоненты вектора X_0 – признаки кластера, значения которых подлежат определению.

Классификация (распознавание образов). Число классов задано. Если также заданы границы между классами, то имеем априорную классификацию; если границы требуется найти, оценить по классифицированным примерам, то задача называется распознаванием образов по обучающей выборке. Целевой признак X_0 имеет значения в номинальной шкале (имена классов).

Упорядочивание объектов. Требуется установить отношения порядка между признаками $x_{10}, x_{20}, \dots, x_{N0}$ (или некоторой их частью) по определенному критерию предпочтения.

Уменьшение размерности модели. Классификационные модели, как первоначальные «сырые», учитывают множество предположений, которые еще надо проверять. Так, сам список признаков X формируется эвристически, часто «с запасом», и оказывается довольно длинным, а главное, избыточным, содержащим «дублирующие» и «шумящие» признаки. Поэтому одна из важных задач совершенствования классификационных моделей состоит в уменьшении размерности модели с помощью отбора наиболее информативных признаков, «склеивания» нескольких признаков в один и т.п. Как следует из практики, информативные признаки могут оказаться различными для разных классов.

5.5. Числовые модели

Числовые модели отличаются от классификационных тем, что: 1) целевые признаки x_0 измеряются в числовых шкалах; 2) числа x_0 представляют собой функционалы или функции признаков переменных (которые не обязательно все являются числовыми); 3) в этих моделях гораздо чаще учитываются связи переменных во времени (в классификационных задачах время иногда даже называют «забытой» переменной). В связи с этим и протоколы наблюдений могут не обязательно относиться к множеству объектов: модель можно уточнять и по экспериментам с одним объектом в разные моменты времени.

Отметим, что числовые модели могут задавать связь между переменными как в параметризованной форме (т.е. в виде функции с конечным числом параметров), так и в непараметризованной форме (в виде функционала). Например, зависимость между входом $x = \{x_i\}$ и выходом y некоторой системы может задаваться в виде параметризованной линейной регрессии $y = \sum a_i x_i + \xi$ либо в непараметризованной форме как функционал линии регрессии $y(x) = \int y \cdot p(y|x) dy$, где $p(y|x)$ – неизвестная плотность условного распределения вероятностей.

Приведем типичные задачи для числовых моделей.

Косвенные измерения (оценка параметра). Требуется определить значение x_0 по заданному множеству $\{x_{ij}\}$. В отличие от классификации x_0 измеряется не в номинальной, а в числовой шкале. Если значения $\{x_{ij}\}$ определены до некоторого момента t_0 , а x_0 требуется оценить для $t > t_0$, то задача называется прогнозированием. (Прогнозирование имеет смысл и в задаче классификации; например, ранняя диагностика заболевания.)

Поиск экстремума (планирование эксперимента). Считается, что имеется возможность пошагового изменения величин $\{x_{ij}(t_k)\}$, $t_k = t_0 + k\Delta t$, $k = 0, 1, 2, \dots$. Требуется изменять их так, чтобы в конце концов получить экстремальное значение целевого признака x_0 .

5.6. Особенности протоколов наблюдений

Отметим встречающиеся на практике особенности реальных протоколов наблюдений, которые следует учитывать при их обработке.

Большая размерность. Во многих исследованиях число объектов N и число признаков n велики, так что произведение $n \times N$ достигает нескольких десятичных порядков. Учет времени приводит к еще большему увеличению размерности блока данных. В настоящее время применение ЭВМ существенно расширяет количественные возможности обработки данных, но «проклятие размерности» остается в силе и для ЭВМ.

Разнотипность данных. Разные признаки могут измеряться в различных шкалах. Многие алгоритмы построены для обработки однотипных переменных, что часто вызывает необходимость приводить разнотипные данные к одной шкале. Ясно, что более правильной стратегией поведения является разработка алгоритмов, специально построенных так, чтобы имелась возможность обрабатывать разнотипные данные, не внося в протокол никаких изменений, не связанных с экспериментом.

Пропущенные значения. Незаполненная ячейка таблицы данных – не такой уж редкий случай, особенно если эксперимент производится не в лабораторных, а в естественных условиях. Исключить из таблицы строку и столбец, на пересечении которых находится пустая ячейка, – выход далеко не всегда приемлемый. Можно, используя избыточность таблицы, некоторым образом «восстановить» пропущенные значения, а затем обрабатывать таблицу так, будто их и не было. Однако критерий «восстановления» и цель обработки должны быть согласованы, поэтому не может быть универсального способа «восстановления» пропусков. Хотя этот путь в ряде случаев вполне допустим, перспективным представляется конструирование алгоритмов обработки, позволяющих использовать таблицы с пробелами без их предварительного заполнения.

Зашумленность. Довольно часто измерение, занесенное в протокол, на самом деле отличается от измеряемого значения на некоторую случайную величину. Статистические свойства этой добавочной поме-

хи могут не зависеть от измеряемой величины, и тогда мы говорим об аддитивном шуме. В противном случае имеет место неаддитивная, или зависящая, помеха. Все эти варианты должны по-разному учитываться при обработке экспериментальных данных.

Искажения, отклонения от предположений. Приступая к обработке протокола наблюдений, мы всегда исходим из определенных предположений о природе величин, занесенных в протокол. Любой способ обработки дает результаты ожидаемого качества только в том случае, если данные отвечают определенным предположениям. Далеко не всегда в ходе обработки данных обращают внимание на то, действительно ли данные отвечают предположениям, заложенным в алгоритм обработки.

Например, данные могут выглядеть как не размытые, но быть на самом деле расплывчатыми. Цифры в действительности могут быть символами, а мы можем считать, что они числа. Числовые шкалы предполагают одинаковость единиц измерения вдоль всей шкалы, а измерительный прибор может обладать нелинейной характеристикой; и если это не отражено в протоколе, то мы будем обрабатывать искаженные данные. Измеряемая величина может быть непрерывной, но в протоколе она неизбежно приводится с округлением, и это также является искажением.

Контрольные вопросы

1. Охарактеризуйте основные цели экспериментов при получении новых знаний.
2. Что такое измерительные шкалы?
3. Для чего используются шкалы наименования?
4. Охарактеризуйте основные свойства шкал наименования.
5. Для чего используются порядковые шкалы?
6. Охарактеризуйте основные свойства порядковых шкал.
7. Для чего используются модифицированные порядковые шкалы?
8. Охарактеризуйте основные свойства модифицированных порядковых шкал.
9. Для чего используются шкалы интервалов?
10. Охарактеризуйте основные свойства шкал интервалов.
11. Для чего используются шкалы отношений?
12. Охарактеризуйте основные свойства шкал отношений.
13. Для чего используются шкалы разностей?
14. Охарактеризуйте основные свойства шкал разностей.
15. Для чего используется абсолютная шкала?
16. Охарактеризуйте основные свойства абсолютной шкалы.

17. Что такое расплывчатое описание ситуаций?
18. Что такое вероятностное описание ситуаций?
19. Охарактеризуйте основные особенности статистических измерений.
20. Охарактеризуйте основные особенности классификационных моделей.
21. Охарактеризуйте основные особенности числовых моделей.
22. Охарактеризуйте основные особенности протоколов наблюдений.

ПРИЛОЖЕНИЕ

Пример структуры АИС

В примере рассмотрены фрагменты Автоматизированной информационной системы учета материальных ценностей в строительной фирме.

Назначение. АИС обеспечивает:

- Повышение надежности и точности процесса сбора, обработки и передачи информации.
- Автоматическое формирование отчетных документов.
- Сокращение затрат времени на обработку экономической информации.
- Повышение производительности труда сотрудников.
- Улучшение культуры производства – технологии обработки экономической информации – и качества подготовки документов.

Краткая характеристика АИС:

1. Администрация принимает заказы на выполнение строительных работ, поручает производственно-техническому отделу (ПТО) подготовить смету на объект строительства; заключает с заказчиком договор на выполнение строительных работ; выполняет анализ расходов строительных материалов.

2. ПТО составляет смету на строительство в соответствии с нормами расхода материалов, комплектует ведомость; ежемесячно принимает от прорабов отчеты о фактически выполненных работах и израсходованных материалах, сопоставляет данные с планами и нормами.

3. На склад или непосредственно на объекты строительства поступают строительные материалы от поставщиков; прорабы и заведующий складом передают после оформления документы на полученные материалы и изделия в бухгалтерию. Часть материалов на строительство передается со склада по накладным на внутреннее перемещение с указанием объектов строительства. На складе ведется картотека материалов (поступило, израсходовано, осталось).

4. На основании первичных документов о приходе, расходовании, перемещении материальных ценностей (МЦ) бухгалтерия составляет отчетные документы (ведомости прихода, расхода, движения МЦ и др.); передает данные о стоимости МЦ в ПТО для расчета финансовых затрат на строительство объекта за месяц.

Примеры форм табличных документов приведены в таблице.

175

Продолжение приложения

1	2	3	4
Склад (три склада в фирме)	Оформление документов при приемке продукции от поставщика и внутреннем перемещении Ведение картотеки МЦ	Прораб Бухгалтерия	Накладная на внутреннее перемещение Ведомости прихода и расхода; результаты инвентаризации
Поставщик	Оформление сопроводительной документации на поставляемую продукцию	Склад, прораб	Счет-фактура.
Администрация	Прием заказов на выполнение работ. Подготовка договоров. Анализ планов и расчетов, утверждение отчетов	Заказчик	Договор Отчеты о затратах
Заказчики	Подготовка заказов. Заключение договоров Оплата работ	Администрация ПТО	Заказ, утвержденный договор Данные для смет

Иерархическая информационно-функциональная модель (взаимосвязанные диаграммы потоков данных) АИС

Схема верхнего уровня (контекстная диаграмма) и диаграмма потоков данных (ДПД) АИС на уровне подсистем приведена на рис. I (создана с помощью программной системы Case_Analyst).

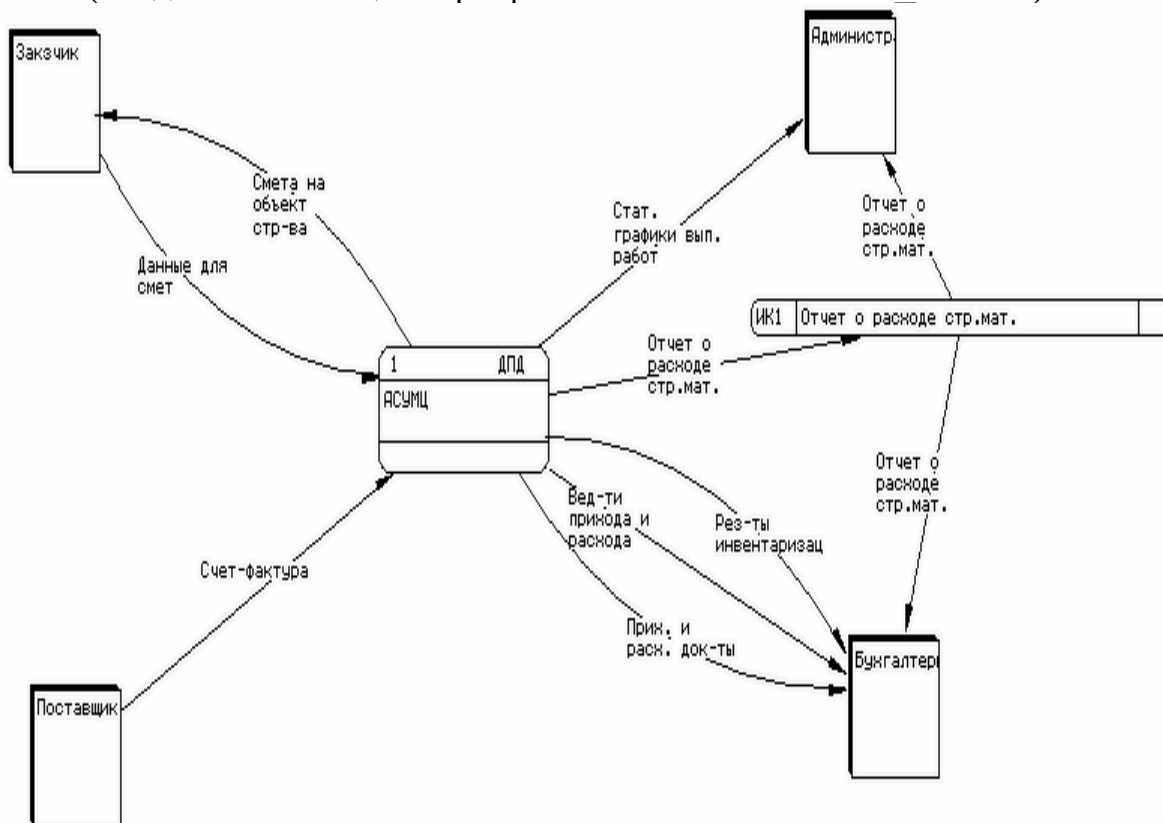


Рис. I

Продолжение приложения

Схема следующего уровня (соответствующая подсистеме) и ДПД на уровне АРМ (склад) приведена на рис. II (создана с помощью программной системы Case_Analyst).

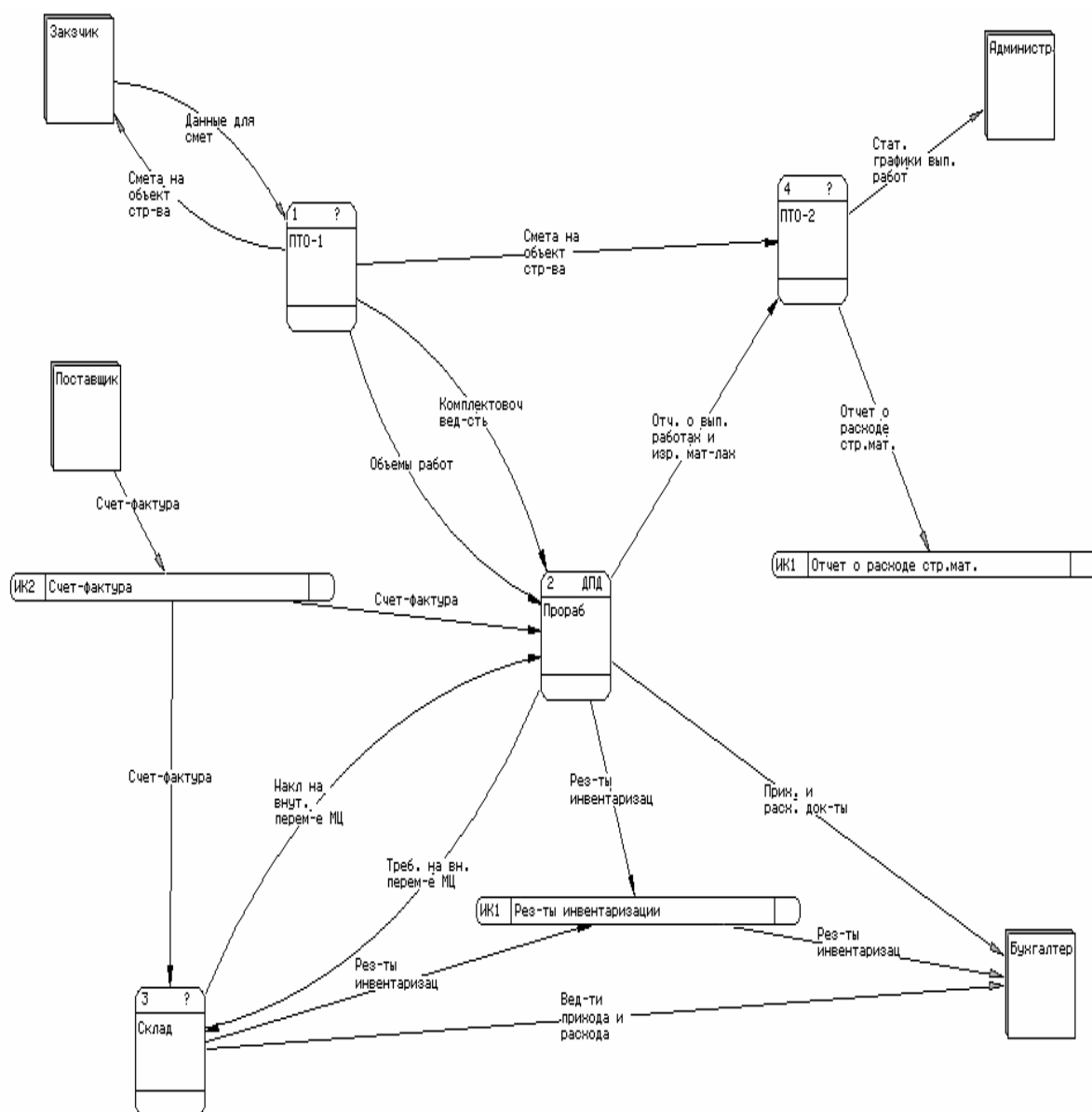


Рис. II

На рис. III приведена ER-модель базы данных АИС, созданная в среде ERWin, которая затем передана в среду СУБД Access 2000 для создания БД и дальнейшего ее использования в системе.

Продолжение приложения

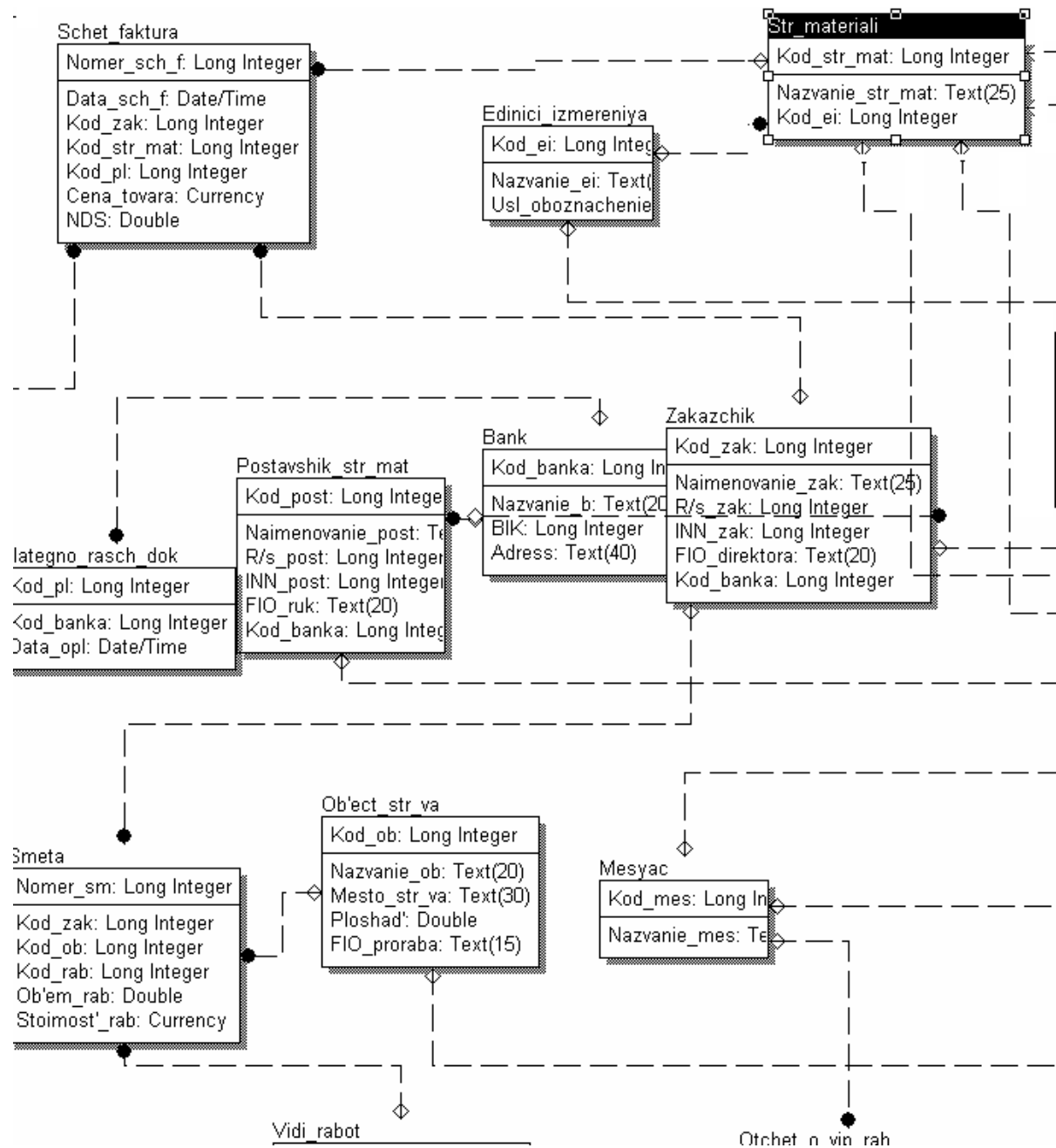


Рис. III (начало)

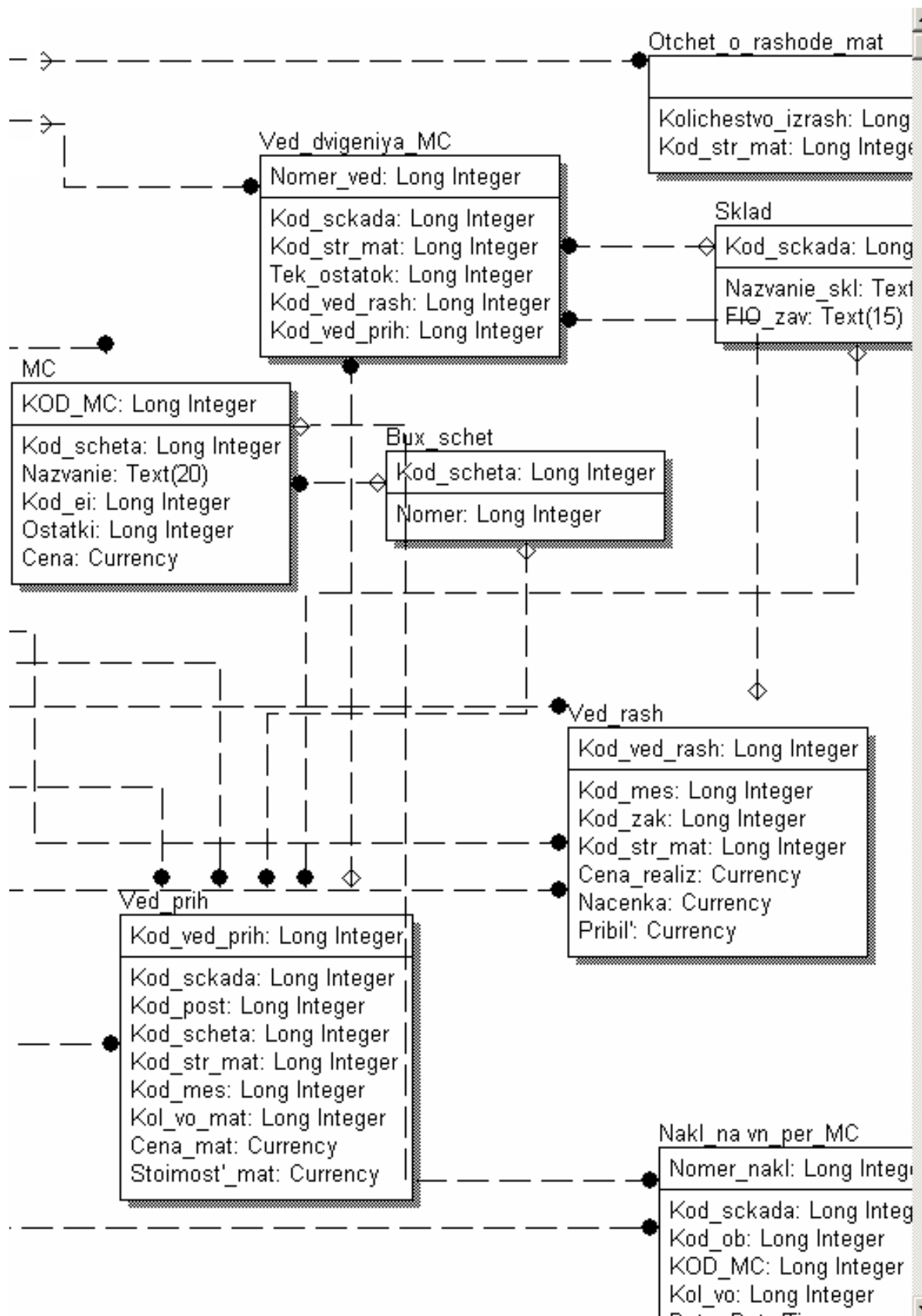


Рис. III (окончание)

Окончание приложения

На основе ER-модели в среде Access 2000 создана БД АИС, фрагменты компонент которой приведены на рис. IV.

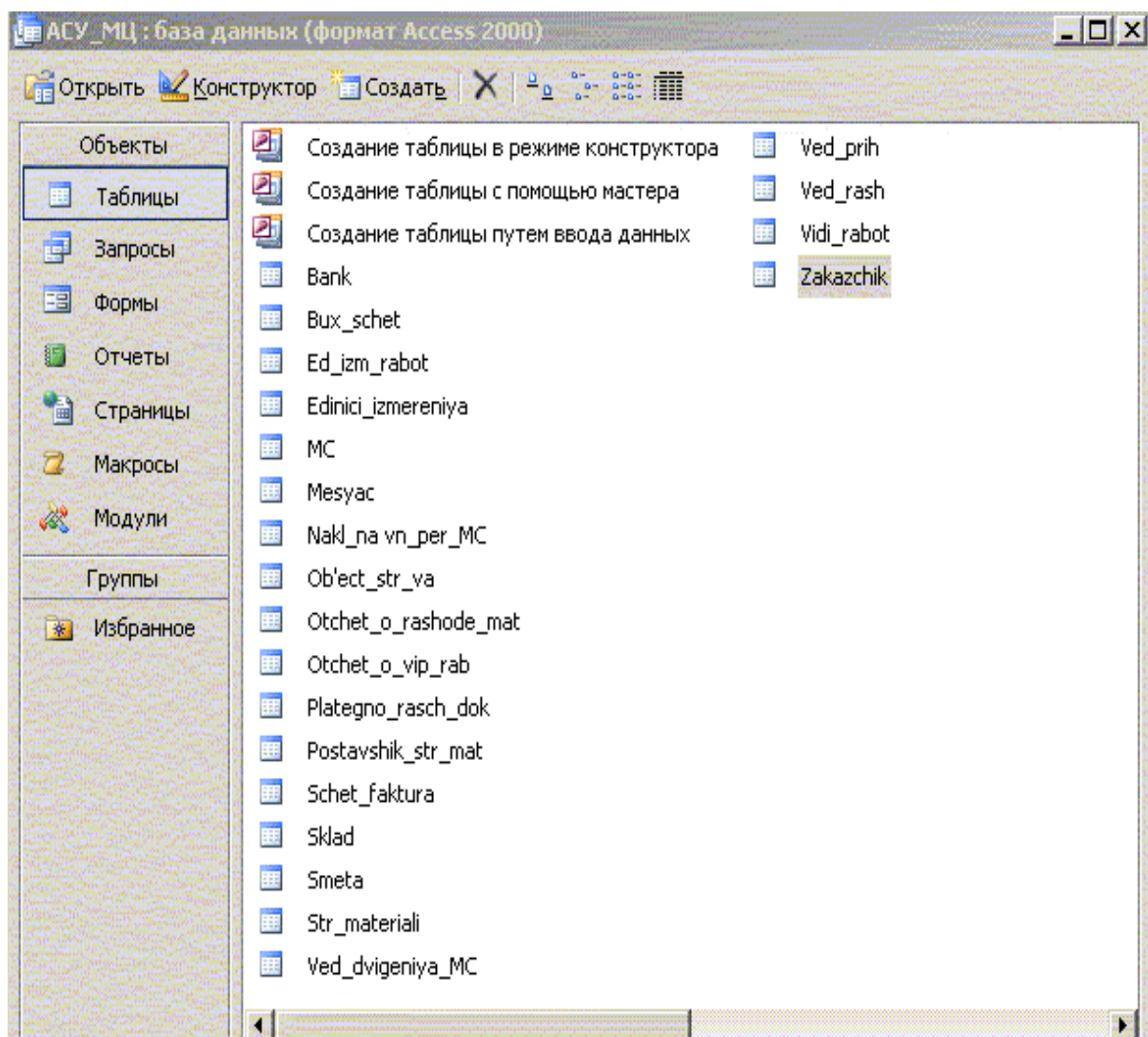


Рис. IV

В среде Access 2000 созданы все инструментальные средства, необходимые для полноценной эксплуатации АИС (формы, запросы, отчеты, макросы и т.п.). Особенности их создания и использования рассматриваются в дисциплине «Управление данными».

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Антонов, А.В. Системный анализ [Текст]: учеб. для вузов / А.В. Антонов. – М.: Высш. шк., 2006.
2. Анфилатов, В.С. Системный анализ в управлении [Текст]: учеб. пособие / В.С. Анфилатов, А.А. Емельянов, А.А. Кукушкин. – М.: Финансы и статистика, 2002.
3. Бородакий, Ю.В. Информационные технологии. Методы, процессы, системы [Текст] / Ю.В. Бородакий, Ю.Г. Лободянский. – М.: Радио и связь, 2004.
4. Бочаров, Е.П. Интегрированные корпоративные информационные системы: принципы построения [Текст]: лаб. практикум на базе системы «Галактика» / Е.П. Бочаров, А.И. Колдина. – М.: Финансы и статистика, 2005.
5. Избачков, Ю. Информационные системы [Текст]: учеб. пособие / Ю. Избачков, В. Петров. – СПб.: Питер, 2006.
6. Куликовский, Л.Ф. Теоретические основы информационных процессов [Текст]: учеб. пособие для вузов / Л.Ф. Куликовский, В.В. Мотов. – М.: Высш. шк., 1987.
7. Колесник, В.Д. Курс теории информации [Текст]: учеб. пособие / В.Д. Колесник, Г.Ш. Полтырев. – М.: Наука, 1982.
8. Криницкий, Н.А. Автоматизированные информационные системы [Текст]: справ. / Н.А. Криницкий, Г.А. Миронов, Г.Д. Фролов. – М.: Наука, 1982.
9. Игнатов, В.А. Теория информации и передачи сигналов [Текст]: учеб. для вузов / В.А. Игнатов. – М.: Радио и связь, 1991.
10. Николаев, В.И. Системотехника: методы и приложения [Текст] / В.И. Николаев, В.М. Брук. – Л.: Машиностроение, 1985.
11. Астафьева, Н.М. Вейвлет-анализ: Основы теории и примеры применения [Текст] / Н.М. Астафьева // Успехи физических наук. – 1996. – Т.166. – № 11. – С. 1145–1170.
12. Дремин, И.Л. Вейвлеты и их использование [Текст] / И.Л. Дремин [и др.] // Успехи физических наук. – 2001. – Т.171. – № 5. – С. 465–501.
13. Математическое моделирование систем связи : учебное пособие [Текст] / К.К. Васильев, М. Н. Служивый. – Ульяновск : УлГТУ, 2008. – 170 с. – Режим доступа: http://www.sernam.ru/book_mm.php
14. Кроновер Р.М. Фракталы и хаос в динамических системах [Текст] / Р.М. Кроновер. – М., Техносфера, 2006
15. Шелухин, О.И. Фрактальные процессы в телекоммуникациях [Текст]: монография / О.И. Шелухин, А.М. Тенякшев, А.В. Осин; под ред. О.И. Шелухина. – М.: Радиотехника, 2003.
16. Шелухин, О.И. Моделирование информационных систем [Текст]: учеб. пособие / О.И. Шелухин, А.М. Тенякшев, А.В. Осин. – М.: Радиотехника, 2005.

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ	3
ВВЕДЕНИЕ.....	4
1. ОСНОВЫ ТЕОРИИ СИСТЕМ.....	6
1.1. Системы и их основные свойства.....	7
1.2. Классификация систем	18
1.3. Особенности функционирования систем.....	20
1.4. Критерии эффективности сложных систем	26
1.5. Основы разработки и исследования сложных систем	28
Контрольные вопросы.....	32
2. ИНФОРМАЦИОННЫЕ СИСТЕМЫ.....	33
2.1. Автоматизированные информационные системы.....	33
2.2. Интегрированные корпоративные ИС	41
2.3. Основные виды обеспечения АИС	48
Контрольные вопросы.....	74
3. ИНФОРМАЦИЯ В СИСТЕМАХ.....	75
3.1. Кодирование информации и алфавиты	76
3.2. Сигналы в системах.....	78
3.3. Математическая модель сигналов.....	80
3.4. Математические модели реализаций случайных процессов	83
3.5. О некоторых свойствах непрерывных сигналов	87
3.6. Цифровое представление непрерывных сигналов	89
3.7. Вейвлеты и вейвлетный анализ сигналов	99
3.8. Фрактальные стохастические процессы	102
3.9. Энтропия.....	107
3.10. Количество информации	113
3.11. Основные результаты теории информации	117
Контрольные вопросы.....	123
4. ДЕКОМПОЗИЦИЯ И АГРЕГИРОВАНИЕ СИСТЕМ	125
4.1. Модели систем как основание декомпозиции	125
4.2. Алгоритмизация процесса декомпозиции.....	128
4.3. Алгоритм декомпозиции	132
4.4. Агрегирование, эмерджентность и внутренняя целостность систем.....	134
4.5. Виды агрегирования.....	136
4.6. Обобщенная модель агрегата.....	141
4.7. Некоторые особенности моделирования процесса функционирования агрегата.....	144
4.8. Агрегативные системы	145

Контрольные вопросы	147
5. ЭКСПЕРИМЕНТ В АНАЛИЗЕ СИСТЕМ	148
5.1. Измерительные шкалы	150
5.2. Расплывчатое описание ситуаций	163
5.3. Вероятностное описание ситуаций. Статистические измерения	164
5.4. Классификационные модели	169
5.5. Числовые модели	170
5.6. Особенности протоколов наблюдений	171
Контрольные вопросы	172
ПРИЛОЖЕНИЕ	174
Пример структуры АИС	174
Иерархическая информационно-функциональная модель (взаимосвязанные диаграммы потоков данных) АИС	176
БИБЛИОГРАФИЧЕСКИЙ СПИСОК	181

Учебное издание

Блинков Юрий Вадимович

**ОСНОВЫ ТЕОРИИ ИНФОРМАЦИОННЫХ
ПРОЦЕССОВ И СИСТЕМ**
Учебное пособие

Редактор М.А. Сухова
Верстка М.А. Сазонова



Подписано в печать 23.11.11. Формат 60×84/16.

Бумага офсетная. Печать на ризографе.

Усл. печ. л. 10,7. Уч.-изд. л. 11,5. Тираж 200 экз. 1-й завод 100 экз.

Заказ № 183.



Издательство ПГУАС.

Отпечатано в полиграфическом центре ПГУАС.

440028, г. Пенза, ул. Г. Титова, 28.