ЛЕКЦИИ

Тема 1. Методологический базис информационной технологии

1.1. Понятие информационной технологии

1.1.1. Технология, цель, система

В реальном мире можно выделить три вида процессов: процессы преобразования вещества, процессы преобразования энергии и информационные процессы. Использование этих процессов в современной целенаправленной деятельности человеческого общества связано с понятием «технология».

Подчеркнем, именно целенаправленной деятельности: цель — это ключ к раскрытию сущности понятия технология. Цель же неразрывно связана с понятиями «система» и «структура». При этом уяснить сущность любого из понятий четверки: **система, структура, цель, технология** по отдельности, без взаимосвязи с остальными, не возможно. Основу составляет тройка: **система, структура, цель;** технология же обусловлена целью системы. Заметим, что степень адекватности представления об этой тройке (система, структура, цель) бездонных по сложности и неподдающихся математическим наскокам понятий определяет не только уровень инженерной зрелости, но и уровень культуры в целом (практика показывает, что человек с низкой культурой не способен добиться высоких результатов).

1.1.2. Понятие "технология"

Термин **технология** происходит от греч. **techno** – искусство, мастерство, умение и **logos** – понятие, учение. Понятие «технология» относится к числу сложных, системных понятий.

Любое понятие имеет три следующих аспекта: **содержание понятия** (совокупностью отличительных признаков), **объем понятия** (совокупность объектов, отображенных в понятии) и **дефиниция** (определение понятия). У любого понятия между его содержанием и объемом существует обратная зависимость, т.е. чем больше признаков включено в содержание понятия, тем меньше объектов оно охватывает. Например, стол вообще и *письменный* стол, или еще уже — *письменный дубовый* стол. Детализация понятия предполагает классификацию его объектов (род, вид и т.д.), при этом в общих рассуждениях часто используют словосочетания: «понятие в широком смысле слова», «понятие в узком смысле слова». Дефиниция же дает лишь самое общее представление о понятии, она не позволяет раскрыть понятие всесторонне и с требуемой полнотой.

Вот некоторые дефиниции понятия технология.

Технология – совокупность методов обработки, изготовления, измерения состояния, свойств, формы сырья, материала или полуфабриката, осуществляемых в процессе производства продукции. *Задача технологии как науки* – выявление химических, физических, механических и др. закономерностей с целью определения и использования на практике наиболее эффективных производственных процессов.

Технология — это комплекс научных и инженерных знаний, реализованных в приемах труда, наборах материальных, технических, энергетических, трудовых факторов производства, способах их соединения для создания продукта или услуги, отвечающих определенным требованиям.

Технология есть любое средство, с помощью которого входящие в производство элементы преобразуются в выходящие; она охватывает машины, механизмы и инструменты, навыки и знания.

Здесь следует хорошо уяснить следующее. Во-первых, технология включает все то, что с ней связано; а, во-вторых, технология это не просто совокупность составляющих ее компонентов, а образованная на их основе система причем, как правило, сложная система.

Пример для сомневающихся. Обувная фабрика понесла убытки и потеряла репутацию из-за клейщицы обуви. Причина: некачественное выполнение технологической операции, обусловленное сдельной оплатой труда (нарушен принцип оплаты труда: зависимость от эффективности работы фабрики в целом). Казалось бы, какое отношение имеет оплата труда к технологии? Оказывается, имеет, поскольку технология это система и охватывает все то, что с ней связано. Именно непонимание системности приводит к подобным ситуациям, о которых лучше, чем крылатой фразой: «хотели как лучше, а получилось как всегда» не скажешь**.**

Наконец, понятие «технология» включает два аспекта: практический, связанный с реально существующими технологиями, и теоретический, составляющий методологию создания высокоэффективных технологий.

Теоретический аспект обязан охватывать понятие технологии некоторой предметной области в целом (например, технология самолетостроения, технология деревообработки и т.д.). В этом случае мы говорим технология, но не технологии, равно как не говорим медицины, математики, философии, т.е. во множественном числе. Забегая вперед, отменим, что в заголовках многочисленных стандартов по информационным технологиям используется термин «информационная технология».

1.1.3. Понятие "информационная технология"

Понятие «информационная технология», являясь одним из видов общего понятия технологии, появилось сравнительно недавно (в конце 70-х годов прошлого столетия) в процессе стремительного развития практики и теории информационных систем. Это понятие имеет самую тесную связь с информатикой, системотехникой и кибернетикой. Более того, именно становлению информатики (в той интерпретации, которую дал международный конгресс, состоявшегося Японии, см. п. 1.2) обязано регулярное использование термина информационной технологии.

Приведем дефиниции понятия «информационная технология».

Информационная технология (Information Technology) — совокупность методов, производственных и программно-технологических средств, объединенных в технологическую цепочку, обеспечивающую сбор, хранение, обработку, вывод и распространение информации.

Информационная технология: приемы, способы и методы применения средств вычислительной техники при выполнении функций сбора, хранения, обработки, передачи и использования информации (ГОСТ 34.003-90).

Информационная технология — это комплекс взаимосвязанных, научных, технологических, инженерных дисциплин, изучающих методы эффективной организации труда людей, занятых обработкой и хранением информации; вычислительную технику и методы организации и взаимодействия с людьми и производственным оборудованием, их практические приложения, а также связанные со всем этим социальные, экономические и культурные проблемы.

Информационная технология – средство преобразования входящей информации (входного сырья) в выходной информационный «продукт». Информационная технология включает аппаратно-программные средства, методы, навыки, знания.

Примечание. В настоящее время (отдельно или в комбинации с другими терминами) широко используется аббревиатура IT (ИТ). Например, **ИТ-компания, ИТ-система, ИТ-специалист, ИТ-архитектура, ИТ-менеджмент, ИТ-ресурс ИТ-процесс, ИТ-услуга,** и т.д. Встречается даже и такое: **ИТ-технология.** При этом сокращение IT (ИТ) используется как для обозначения *информационной технологии*, так и для обозначения *информационных технологий*, так что выяснить что конкретно имеется ввиду можно только из контекста. Отметим, что иногда (в основном в ученых целях) используют термин **КИТ** (**Компьютерные информационные технологии**) признавая тем самым и статус докомпьютерных ИТ.

Практический аспект понятия «информационная технология» отличает небывалая динамика развития и огромное разнообразие современных информационных технологий, которые пронизывают практически все сферы и виды человеческой деятельности.

Теоретический аспект понятия «информационная технология» (информационная технология как наука) охватывает вопросы классификации, описания, анализа и создания информационных технологий.

Между обсуждаемыми аспектами понятия «информационная технология» существует глубокий разрыв, обусловленный сложностью информационных технологий, которые несопоставимо сложнее вещественно-энергетических технологий.

1.1.4 Информационная технология как научное направление

Информационная технология как научное направление, находится в начальной стадии своего развития, но занимает особое положение в современной научной парадигме. Она (технология) способна обеспечить конвергенцию на самом высоком уровне: расширить понятие "технология" на все объекты материального мира и тем самым приблизить понимание и разрешение ряда вековых проблем.

Поясним сказанное. Любая технология содержит информационную составляющую. Например, для приготовления блюда необходим рецепт (это — информация). Даже в таком, простейшем случае проявляется триединство (взаимосвязь) вещественных, энергетических и информационных процессов; но в

названном триединстве реально существуют только вещественно-энергетические процессы. Информационные же процессы (технологии) представляют собой специфический вид вещественно-энергетических процессов, которые носят обеспечивающий (инфраструктурный) характер и управляет этими процессами информационная система.

Эту особенность информационных технологий подтверждает весь опыт компьютерной индустрии, которая на современном этапе начинает интегрироваться с биологическими (генетическими) "технологиями". Забегая вперед, заметим, что в современных компьютерных технологиях роль генетического механизма играют многочисленные стандарты, в которых на основе консенсуса аккумулированы научные исследования и практический опыт. Технология же "изделий" живой природы такова, что их генетический материал содержит "рабочие чертежи" не только жизненного цикла, но и такие признаки, как цвет глаз, рост, зубы и т.д.

Если в примере с блюдом все ясно, то когда дело доходит до сложных и тем более сверхсложных информационных систем, например, таких как нервная система человека, то мы сталкиваемся с уникальным проблемным полем (полем "чудес", мистики и спекуляций). В этом плане, расширение понятия технологии на объекты живой природы позволяет, как минимум, уяснить сущность проблемы "интеллектуального флогистона" (души, сознания); проводя определенную параллель с теорией флогистона (проблема разрешена М. Ломоносовым и А. Лавуазье, до которых считалось, что огонь обусловлен особой горючей жидкостью — флогистоном).

1.1.5 Информационная технология и информационная системна

Информационную технологию реализует информационная система (ИТ-система), которая характеризуется тем, что у нее входным и выходным «продуктом» любой ее технологической операции является информация. При этом энергетические и вещественные процессы, на основе которых реализуются технологические операции ИТ-системы, играют обеспечивающую роль.

В Законе Республики Беларусь «об информатизации» от 6 сентября 1995г. N 3850-XII обсуждаемым терминам даны следующие определения:

информационная технология — совокупность методов, способов, приемов и средств обработки документированной информации, включая прикладные программные средства, и регламентированного порядка их применения;

автоматизированная или автоматическая информационная система — совокупность информационных ресурсов, информационных технологий и комплекса программно-технических средств, осуществляющих информационные процессы в человеко-машинном или автоматическом режиме;

информационные процессы — процессы сбора, обработки, накопления, хранения, актуализации и предоставления документированной информации пользователю.

Для уточнения соотношения понятий информационная технология и информационная система обратимся к Государственному стандарту бывшего СССР ГОСТ 34.003-90 "Информационная технология. Комплекс стандартов на автоматизированные системы. Автоматизированные системы. Термины и

определения", который дает такое определение автоматизированной системы (АС).

«АС: Система, состоящая из персонала и комплекса средств автоматизации его деятельности, реализующая ***информационную технологию*** выполнения установленных функции.

Примечания:

1. В зависимости от вида деятельности выделяют, например следующие виды АС: автоматизированные системы управления (АСУ), системы автоматизированного проектирования (САПР), автоматизированные системы научных исследований (АСНИ) и др.

2. В зависимости от вида управляемого объекта (процесса) АСУ делят, например, на АСУ технологическими процессами (АСУТП), АСУ предприятиями (АСУП) и т. д.».

Отметим, что в Республике Беларусь государственные стандарты бывшего СССР имеют статус государственных стандартов Республики Беларусь (Постановление Госстандарта РБ №3 от 17.12.1992г.)

В заключение заметим, что, несмотря на то, что ГОСТ 34.003-90 является действующим, термин АСУ в настоящее время практически вытеснен термином ERP (Enterprise Resource Planning — планирование ресурсов предприятия); а вместо термина АС специалисты обычно употребляют термин ИТ-система. ERP по сути то же, что и АСУП, но АСУП, в отличие ERP, которые ориентированы на рыночную экономику, разрабатывались применительно к плановой социалистической экономике. Финансирование развития методологии АСУ прекратилось с распадом СССР.

1.2. Взаимосвязь информационной технологии с другими научными направлениями

Информационная технология как научное направление является неотъемлемой частью системы научного знания, поэтому уяснить ее роль и место в этой системе можно только во взаимосвязи с другими родственными направлениями, такими как теория автоматического управления, кибернетика, системотехника и информатика.

В качестве отправной точки выберем системотехнику. Объектом исследования системотехники являются сложные технические комплексы (СТК), а предметом исследования — процессы создания, использования, совершенствования и ликвидации СТК. Системотехника имеет ряд направлений: военная системотехника, радиолокационная системотехника и др. Основным методом системотехники является системный подход, который нашел воплощение в методологии системного анализа.

До появления информатики информационные системы, образно говоря, находились под крылом системотехники. Ситуация изменилась после Международного конгресса (Япония, 1978г.), который дал такое определение информатики:

Понятие информатики охватывает области, связанные с разработкой, созданием, использованием и материально-техническим обслуживанием систем обработки информации, включая машины, оборудование, математическое

обеспечение, организационные аспекты, а также комплекс промышленного, коммерческого, административного, социального и политического воздействия.

В этом широком определении можно выделить три основных аспекта:

1) информатика как особая инфраструктурная (обслуживающая) отрасль народного хозяйства;

2) информатика как совокупность средств информационной технологии (аппаратных, программных, алгоритмических и др.);

3) информатика как область научного знания, охватывающая методологические и практические вопросы, связанные с разработкой, созданием и использованием средств информационной технологии и встраиванием информационных технологий в социальную среду.

Следует отметить, что термин информатика имеет более раннюю историю. В бывшем СССР он использовался трижды. Ф.Е. Темниковым в 1963 году, как дисциплина рассматривающая информационные категории (идеология, морфология, физиология, метрология), информационные процессы (восприятие, передача, обработка, представление) и информационные системы (элементы, структуры, поведение, организация) [Темников Ф.Е. Информатика. Известия ВУЗ, Электромеханика, 1963, № 11]. Затем этот термин использовался для обозначения дисциплины, связанной с научно-технической информацией и, наконец, как калька с французского informatique (наука об ЭВМ и их применении). Синонимом **Informatique** стал термин **Computer Science** (наука о вычислительной технике), который в США использовался значительно раньше французского.

На вопрос о месте информатики среди других научно-практических направлений можно с определенным упрощением утверждать, что информатика выделилась из системотехники в отдельное направление с более узкой предметной областью, но с более широкой проблемной средой, сохранив при этом ее (системотехники) понятийный аппарат, общие принципы и методологию. Заметим, что еще сравнительно недавно выпускникам специальности АСОИ присваивалась квалификация инженер-системотехник, а не инженер по информационным технологиям, как в настоящее время.

Информационная технология имеет непосредственное отношение ко всем приведенным выше аспектам информатики и использует тот же что и информатика методологический аппарат, но имеет более специфичную проблемную среду, связанную с вопросами классификации (таксономии), анализа и создания информационных технологий.

1.3 Введение в понятие "система"

1.3.1. Дефиниции понятия "система"

Неоднократно использованное выше в тех или иных вариациях понятие «система» является фундаментальным понятием всех вышеназванных научных теорий и направлений. Это понятие используется практически во всех сферах человеческой деятельности, от квантовой механики (объект исследования: микрочастицы и их системы), до космологии.

Начиная от Адама понятие «система» занимало лучшие умы человечества, с ним связано необозримое количество литературных источников, но до сих пор так и не существует общепринятого толкования понятия «система». Вместе с этим формирование адекватного представления о данном понятии играет фундаментальную роль в формировании не только инженерной, но и мировоззренческой культуры. Это и ключ к пониманию сущности системного подхода.

Приведем некоторые дефиниции понятия система. Под системой понимается:

комплекс элементов, находящихся во взаимодействии (Л. Берталанфи);

множество объектов вместе с отношениями между объектами и между их атрибутами (А. Хол и Р. Фейджин);

отображение входов и состояний объекта в выходах объекта (М. Месарович);

множество взаимосвязанных элементов, каждый из которых связан прямо или косвенно с каждым другим элементом, а два любых подмножества этого множества не могут быть независимыми (Р. Акофф, Ф. Эмери).

Эти и другие определения, в общем, сводятся к тому, что система представляет собой совокупность взаимосвязанных элементов (объектов) любой природы, образующих единое целое. Однако никакая дефиниция не позволяет уяснить сущность системы, поэтому обычно прибегают к рассмотрению основных свойств системы.

1.3.2. Основные свойства системы

Любая система обладает многочисленными свойствами, однако системообразующих свойств немного. Рассмотрим основные (главные) свойства системы.

1. **Система обладает интегративными (эмерджентными) свойствами.**

Интегративные или эмерджентные свойства — это такие свойства системы, которые присущи только системе в целом, но не свойственны ни одному из ее элементов (или совокупности элементов) в отдельности.

Рассматривая некоторый объект, мы обнаруживаем, что его свойства отличаются от свойств окружающей среды. Исследуя его (объекта) элементы по отдельности мы обнаруживаем, что они такими свойствами не обладают. Кто в детстве не разбирал часы?

В качестве примера рассмотрим таджикский способ перехода через горную реку. То, что не под силу одному человеку может сделать группа людей, соединившихся руками на уровне плеч в кольцо (так танцуют гуцулы). Наконец, вспомните описание охоты первобытных людей на мамонта. Эмерджентным свойством обладает и волчья стая, охотящаяся на крупное животное.

2. **Система обладает целостностью.**

Система хотя и состоит из элементов, но это нечто целое. Элементы, составляющие систему, могут в свою очередь быть системами, но для системы это элементы. Она использует только определенные их свойства.

Например, элементом в таджикском способе перехода через реку является человек. Причем это может любой человек (сегодня один, завтра другой). В испортившихся часах мастер заменит деталь, и они снова приобретут свойство показывать время. Запомните данную особенность системы, это поможет уяснению очень непростого понятия структуры системы.

В свою очередь система может входить в качестве элемента в более сложную (вышестоящую) систему, так что, как правило, имеет место иерархическая взаимосвязь объектов, в том числе и в самой системе. В этом плане часто используют термины: надсистема, система, подсистема.

3. **В системе имеются существенные устойчивые связи.**

Целостность системы обеспечивают устойчивые связи, которые, можно интерпретировать как некоторое множество физических каналов (трактов), по которым между элементами системы (или ее подсистемами) и системы с окружающей средой осуществляется обмен (метаболизм) энергией, веществом и информацией. При этом связи могут различаться мощностью (интенсивность потока вещества или энергии, скоростью передачи информации), направленностью (прямые, обратные) и ролью в системе (усиление, преобразование, координация и др.). Такой метаболизм по существу представляет собой «технологию» существования системы.

Система может быть и абстрактной: моделью или теорией. В этом случае говорят не о связях, а об отношениях между элементами.

Наличие существенных устойчивых связей в системе накладывает определенные ограничения на степень свободы ее элементов. Например, человек в кольце таджикского способа может перемещаться только месте с кольцом. Но именно эта несвобода и обеспечивают такой системе ее эмерджентное свойство: возможность прейти реку. Отметим, что у людей, входящих в кольцо одна цель — прейти реку, она-то их и объединяет; а действия (функции), которые они совершают для достижения этой цели и составляют технологию («технологию кольца»). То же, что остается неизменным в процессе существования данного кольца является его структурой (это традиционное понимание структуры, но если Вы не согласны, то мы это здесь постулируем). Попробуйте выделить в «чистом» виде структуру кольца таджикского способа. А где находится свойство кольца прейти реку? Где расположена его цель? Поставим те же вопросы по отношению к часам. Где свойство показывать время? Где цель?

Ответив на эти, казалось бы, простые вопросы мы сможем ответить и на аналогичные вопросы по отношению к сущности сознания и души человека. В качестве подсказки рассмотрим историю с теорией флогистона. Во времена М. В. Ломоносова считали, что огонь обусловлен особой горючей материей — флогистоном. А. Лавуазье экспериментально опроверг теорию флогистона и показал, что огонь — особое (эмерджентное) свойство материи. Вот и ответ на поставленные вопросы. Душа уходит туда же, куда уходит и огонь. Кстати ученым так и не удается прямо связать ни одну из мыслительных «операций», приписываемых человеческому разуму, с какой-то частью мозга (т.е. не удается найти «интеллектуальный флогистон»).

1.3.2. Пространство свободы системы

Приобретая (посредством связей) эмерджентные свойства, система "платит" несвободой своих элементов, и тем самым ограничивает пространство своих возможных состояний (пространство свободы системы). При этом поведение системы (без потерь качества) возможно только в рамках этого ограниченного пространства.

Примером объекта, обладающего уникальной свободой, основанной на уникальной же несвободе, является человек. Именно несвобода физическая и духовная дает человеку (и еще в большей степени человечеству) эмерджентные свойства, отличающие его от остального живого мира. Заметим, что Спиноза определял свободу как осознанную необходимость, т.е. необходимость находиться в пространстве свободы системы, на которое в конструктивном плане указывает религия. По поводу сложности названных аспектов несвободы применительно к человеку Нодар Думбадзе сказал: "душа человека во сто крат тяжелее его тела…" (не соотношение ли это сложности технологий вещественно-энергетических и информационных?). Во сколько же раз духовная наркота опаснее той, которая преследуется по закону? Затронутые вопросы могут показаться далекими от информационных технологий, но это только на текущий момент; сама же такая постановка свидетельствует о конвергентном характере расширенной трактовки понятия "технология".

Элементов в системе может быть много, при этом количество возможных состояний системы становится невероятно большим, таким, что исследование свойств системы приобретает характер трансвычислительной задачи. Важно представить себе, на сколько связи между элементами системы ограничивают степень ее свободы.

Предположим, что S — число возможных состояний элементов системы, а N — число состояний этих же элементов, но без связей, т.е. предоставленных самим себе (как молекулы газа в некотором сосуде).

В качестве примера рассмотрим множество всех возможных различающихся комбинаций, состоящих из ста букв русского языка. Если не различать буквы: ***е*** и ***ё***, ***ь*** и ***ъ*** и дополнительно ввести пробел, то будем иметь

$$N = 32^{100} \approx 3{,}27 \times 10^{150} \ . \qquad (1.1)$$

Однако русский язык является сложной системой, в которой буквы и слова имеют определенные взаимосвязи, поэтому количество комбинаций из множества (1.1), принадлежащих русскому языку (т.е. используемых на практике), значительно меньше. В этом случае говорят, что система (в данном случае русский язык) обладает избыточностью. Избыточность обычно обозначают буквой R. Наличие избыточности являемся неотъемлемым свойством сложной системы (заметим, что множество (1.1) имеет R=0).

Известно, что избыточность русского языка R > 0,5. Если предположить, что R=0,5, то число типичных, т.е. существующих в русском языке и благозвучных сто буквенных комбинаций составит только $S = 32^{50} \approx 1{,}8 \times 10^{75}$ из общего количества $32^{100} \approx 3{,}27 \times 10^{150}$ (проверить этот результат Вы сможете, изучив основ классической теории информации, приведенные разделе 4).

Оба числа огромны, но первое ничтожно мало по сравнению со вторым. Сравните.

Масса Земли: 6×10^{27} *г*, количество атомов 10^{51};

масса Солнца: 2×10^{33} *г*, количество атомов 10^{57};

масса нашей Галактики: 10^{44} *г*, количество атомов 10^{67};

масса Вселенной: 10^{54} *г*, количество атомов 10^{77}.

Так что, количество слов, которые мы используем или можем использовать ничтожно в сравнении с общим числом комбинаций типа: ъхъъььчюф……ыыыъъ; но и количества типичных комбинаций (при R=0,5) более чем достаточно для присвоения собственного имени каждому атому нашей Галактики. В качестве упражнения подсчитайте, на сколько порядков масса бумаги, необходимая для распечатки всех возможных комбинаций $N \approx 3{,}27\times10^{150}$, превысит массу нашей Вселенной.

Это пример того, как связи между элементами могут уменьшить число возможных состояний системы. Избыточность русского языка составляет примерно 0,7. Точный расчет R именно из-за связей представляет собой трансвычислительную задачу.

1.3.4. Энергетический барьер информационных технологий [14]

Ханс Бремерманн утверждал, что система (процессор) искусственная или естественная с практически недостижимыми предельными возможностями, имеющая массу, равную массе Земли за 10 млрд. лет может обработать не более чем 10^{93} бит информации. Под обработкой Бремерманн понимал пересылку N бит по одному или нескольким каналам. Задача, требующая обработки более чем 10^{93} бит информации, называется трансвычислительной, а число 10^{93} — пределом Бремерманна.

Вот пример трансвычислительной задачи, связанный с рассмотренными выше сто буквенными комбинациями русского языка, полное множество которых составляет 32^{100}. На базе этого полного множества можно составить $2^{32^{100}}$ различающихся подмножеств, включая и исходное полное множество. Теперь предположим, что нам необходимо из этих $2^{32^{100}}$ подмножеств отобрать некоторую систему множеств, удовлетворяющую заданному критерию. Тогда, даже при самом эффективном методе поиска, задача становится трансвычислительной, так как необходимо обработать $\log_2 2^{32^{100}} \approx 3{,}27\times10^{150}$ *бит* информации. И таких задач очень много. Они встречаются при распознавании образов, тестировании современных интегральных схем (чипов) и т. д., так что на практике приходится применять различные методы упрощения их решения.

Однако при огромных объемах вычислений узким местом все же является не быстродействие процессора, а потребляемая энергия. Рассмотрим данный вопрос более детально.

Для записи данных в ячейки памяти необходима энергия. Минимальное значение энергии для записи одного бита определяется принципом неопределенности Гейзенберга, согласно которому энергия может быть измерена с точность ΔE, при условии выполнения неравенства

$$\Delta E \Delta t \geq h, \tag{1.2}$$

где Δt — длительность времени измерения,

$h \approx 6{,}625 \times 10^{-27}$ *эрг·с* — постоянная Планка.

Полагая в неравенстве (1.2) $\Delta t = 1$с имеем

$$E^1_{\min} \approx 6{,}625 \times 10^{-27}. \tag{1.3}$$

Используя формулу Эйнштейна $E = mc^2$, $c \approx 3 \times 10^{10}$ *см/с*, найдем массу соответствующую этой энергии

$$M^1_{\min} = E^1_{\min} / c^2 \approx 7{,}36 \times 10^{-48} г. \tag{1.4}$$

Числовое значение в правой части (1.4) можно интерпретировать как минимальную массу, посредством которой энергетический источник еще способен обеспечить запись одного бита информации.

Используя выражение (1.4) определим количество бит, запись которых способен обеспечить 1 грамм массы

$$N_{1г} = 1.36 \times 10^{47} \text{ бит}. \tag{1.5}$$

Формально это совпадает с главным выводом Бремерманна: “Не существует системы обработки данных, искусственной или естественной, которая могла бы обработать более чем 2×10^{47} бит в секунду информации на грамм своей массы” (Бремерманн округлил 1,36 до 2).

Заметим, что выражение (1.5) предполагает использование ядерной энергии 1 грамма массы и сопряжено с немыслимым давлением и температурой.

Используя (1.3) подсчитаем массу ядерного топлива необходимого для достижения предела Бремерманна.

$$E \approx E^1_{\min} \times 10^{93} = 7{,}36 \times 10^{-48} \times 10^{93} = 7{,}36 \times 10^{45} г. \tag{1.6}$$

Масса в (1.6) более чем в 70 раз превышает массу нашей Галактики, которая составляет примерно $10^{44} г$.

Предел 10^{93} был получен Бремерманном простым перемножением числа 1.36×10^{47} на массу Земли в граммах и на 10^{10} лет в секундах, т.е. без учета потребляемой энергии. В то время как для записи 10^{93} бит информации необходимо $7{,}36 \times 10^{45}$г ядерного топлива. Если предположить, как это сделал Бремерманн, что масса процессорной системы равна массе Земли ($6 \times 10^{27} г$), то при записи 10^{93} бит отношение массы требуемого топлива к массе процессорной системы составит примерно 10^{18}.

Используя предел Бремерманна, У. Росс Эшби подсчитал, что вся информация, используемая мировой наукой, какой бы она не стала в будущем, не сможет превзойти объем 10^{80} *бит*. Нетрудно подсчитать, подставляя 10^{80} *бит* в выражение (1.4), что для достижения этого предела нашей информационной Вселенной, необходимо $7{,}36 \times 10^{32}$г массы ядерного топлива, что составляет примерно треть массы Солнца.

Так что при огромных объемах вычислений узким местом становится не масса и быстродействие процессора, а энергия, необходимая для выполнения вычислений. Это является нормой для вещественно-энергетических технологий. Отнесите, например, массу топлива, израсходованного самолетом за все время его эксплуатации, к массе самого самолета. В области же информационных технологий роль энергетического фактора еще не в полной мере осознана. По-видимому, это говорит о том, что информационные технологии находятся еще в младенческом «возрасте». Для размышления заметим, что масса мозга человека составляет 2% массы его тела, но потребляет 20% получаемой энергии.

Следует иметь в виду, что из-за теплового шума (1.5) не достижимо. Минимально необходимая энергия, которую надо затратить на получение одного бита информации, как показал Л. Бриллюэн, равна

$$E = kT\ln 2,$$

где $k = 1{,}38 \cdot 10^{-23}$ дж/град — постоянная Больцмана,

T — температура по шкале Кельвина.

При комнатной температуре $E = 3 \cdot 10^{-21}$, так что различие составляет 27 порядков.

Наконец возвращаясь к трансвычислительной задаче, со сто буквенными комбинациями русского языка обнаруживаем, что даже записать исходное множество комбинаций $32^{100} \approx 3{,}27 \times 10^{150}$ в некоторый гипотетический регистр мы не сможем по энергетическим соображениям, так как потребуется более чем 10^{100} г массы ядерного топлива, в то время как масса нашей Вселенной составляет всего 10^{54} г.

Что следует из этого, удручающего на первый взгляд, факта. Во-первых, Природа, по-видимому, "идет" другим путем; у нее иная "технология проектирования", тем более, что упомянутая задача не идет ни в какое сравнение с конструированием человеческого мозга (вероятность возникновения жизни случайным перебором составляет 10^{-260}). Во-вторых, традиционные компьютерные технологии должны пополниться, квантовыми компьютерными технологиями (они находятся на стадии лабораторных исследований). Например, задачу, связанную с отбором системы множеств из $32^{100} \approx 3{,}27 \times 10^{150}$ комбинаций, удовлетворяющую заданному критерию, мог бы решить квантовый компьютер, содержащий 500 кубитов (кубит— квантовый бит, который может одновременно находиться в состоянии и "0" и "1").

Поясним сказанное, проводя аналогию с простейшей системой, состоящей из 500 монет. Эта система позволяет сгенерировать (путем бросания монет) любую сто буквенную комбинацию с вероятностью 32^{-100} (5 монет на одну букву). А далее мы имеем то, что описано выше.

Иначе работает гипотетический 500 кубитовый компьютер В процессе эксперимента он имеет состояние в 2^{500}-мерном гильбертовом пространстве (это допланковский уровень). В результате измерения параллельно формируется 500-мерный бинарный вектор (сто буквенная комбинация), но с высокой вероятностью, определяемой перепутыванием (сцеплением) кубитов. Важны два момента: во-первых, большая экономия энергии; а во-вторых, возможность формирования определенных свойств. Однако здесь возникает масса вопросов.

1.4. Структура, архитектура и цель системы

Структура системы — это ее инвариант, это то, что остается неизменным в процессе функционирования системы. Структура сохраняет систему как сущность в процессе вещественного, энергетического и информационного обмена между ее элементами, но до определенного предела, за которым в системе происходят структурные изменения. Описание структуры связано с теми же трудностями, что и описание системы.

Вещественный, энергетический и информационный обмен в системе приводит на определенном этапе к качественному изменению ее элементов, а последнее — к определенному изменению структуры системы. В этом плане различают следующие типы структур:

экстенсивные (рост числа элементов во времени),

редуцирующие (уменьшение числа элементов),

интенсивные (рост числа связей и их мощностей при постоянном числе элементов),

деградирующие (уменьшение числа связей и их мощностей).

Важным классификационным признаком системы является ее сложность. Сложность понятие многогранное. Общепринятого критерия определения сложности системы не существует, так как для того, чтобы его иметь необходимо уметь адекватно описывать системы, а это-то и является проблемой. Обычно используют такие качественные характеристики как **простая, сложная, очень сложная система** (большая система). Отличительным признаком сложной системы является невозможность ее адекватного формализованного описания.

Структура сложной системы характеризуется многочисленными и разноплановыми аспектами, каждому из которых может быть сопоставлена, в свою очередь, определенная структура (структура аспекта системы). Например, структура системы на макро- и микро-уровне; структура морфологического, функционального и информационного описания системы; структура компонентов АСУ (технического, информационного, программного и др. обеспечения; см. ГОСТ 34.003-90).

При разработке сложной системы ее строение описывают при помощи различных структурных схем, так, например, внутренне строение АСУ описывают при помощи следующих схем:

1) функциональных (элементы: функции, задачи, процедуры; связи: информационные);

2) технических (элементы: устройства, компоненты и комплексы; связи: линии и каналы связи);

3) организационных (элементы: коллективы людей и отдельные исполнители; связи — информационные, соподчинения и взаимодействия);

4) документальных (элементы: неделимые составные части и документы АС; связи: взаимодействия, входимости и соподчинения);

5) алгоритмических (элементы: алгоритмы; связи: информационные);

6) программных (элементы: программные модули и изделия; связи: управляющие);

7) информационных (**элементы**: формы существования и представления информации в системе; **связи**: операции преобразования информации в системе).

Порожденная связями элементов, структура системы имеет **цель** (целевую структуру). Для искусственных систем эта цель определяется требованиями пользователя (Заказчика системы), на основе которых формируется архитектура системы. При этом архитектура, характеризуя свойства и принципы построения системы, объединяет («склеивает») разноплановые структурные аспекты системы. Разработку архитектуры системы осуществляет главный конструктор (архитектор) системы или менеджер проекта, который обеспечивает, с одной стороны, интерфейс с Заказчиком системы, с другой стороны, с разработчиками компонентов (подсистем) системы.

При этом качество, стоимость и доверие Заказчика к разрабатываемой системе (информационной технологии) напрямую зависят от «зрелости» процесса разработки и сопровождения программного обеспечения в проектной организации, ее культуры, опыта. Названная зрелость определяется качеством управления процессом проектирования, используемыми методами и степенью автоматизации процесса проектирования, уровнем квалификации специалистов и т.д.

Другими словами, качество изделия во многом обусловлено качеством системы, которая его разрабатывает. Это следует хорошо усвоить и применительно к личной жизни: человек с низкой культурой не способен добиться высоких результатов.

Возникает вопрос: как же оценить качество проектной организации? Для этого существуют стандарты ISO серии 9000 и интегрированная модель уровней зрелости CMMI.

Стандарты ISO серии 9000

К основополагающим стандартам данной серии относятся:

ISO 9000:2005 «Системы менеджмента качества. Основные положения и словарь» (основные положения систем менеджмента качества и терминология);

ISO 9001:2008 «Системы менеджмента качества. Требования» (требования к системам менеджмента качества);

ISO 9004:2009 «Менеджмент для достижения устойчивого успеха организации. Подход с позиции менеджмента качества» (рекомендации по достижению устойчивого успеха любой организацией в сложной, требовательной и постоянно изменяющейся деловой среде);

ISO 19011:2002 «Руководящие указания по аудиту систем менеджмента качества и/или систем экологического менеджмента» (указания по принципам и правилам проведения аудита систем менеджмента качества и систем экологического менеджмента)

Названные стандарты применимы к любой предметной области. Построение адекватной систем качества по **ISO** к конкретной предметной области предполагает использования дополнительных стандартов. Для организаций,

занимающихся разработкой программного обеспечения, к таким стандартам относятся: **ISO 9003, ISO 10007, ISO 10013, ISO 12207.**

Стандарты **ISO серии 9000** признаны во многих странах. Существуют переведенные на национальные языки и адаптированные версии стандартов (например, ГОСТ Р ИСО 9000). Республика Беларусь в лице Госстандарта РБ является 0-членом ISO/TC176. 0-члены имеют право: участвовать в работе комитета в качестве наблюдателя, получать документы комитета ТС 176 и предлагать на рассмотрение замечания.

Стандарт CMMI

Стандарт CMMI (Capability Maturity Model Integration; март 2002) вобрал в себя лучшее из частных моделей **CMM** (Capability Maturity Model), предложенных в 1987 году американским институтом программной инженерии (**SEI**). Эта модель содержит 5 уровней зрелости и позволяет организации не только оценить уровень ее зрелости, но и помогает в совершенствовании процесса управления разработкой программного обеспечения.

Основные модели CMM:

SW-CMM (Capability Maturity Model for Software: модель зрелости процессов разработки ПО),

EIA/IS 731 (Electronic Industries Alliance Interim Standard: модель зрелости процессов для системного реинжиниринга),

PD-CMM (Integrated Product Development Capability Maturity Model: модель зрелости процессов интегрированной разработки продуктов).

Основные характеристики уровней CMM:

1. **Начальный.** Процесс разработки носит хаотический характер. Определены лишь немногие из процессов и успех проектов зависит от конкретных исполнителей.

2. **Управляемый.** Установлены основные процессы управления проектами: отслеживание затрат, графика работ и функциональности. Упорядочены некоторые процессы, необходимые для того, чтобы повторить предыдущие достижения на аналогичных проектах либо проектах с аналогичными приложениями.

3. **Определенный.** Процессы разработки ПО и управления проектами описаны и внедрены в единую систему процессов компании. Во всех проектах используется стандартный для организации процесс разработки и поддержки ПО, адаптированный под конкретный проект.

4. **Количественно-управляемый**. Собираются детальные количественные данные по функционированию процессов разработки и качеству конечного продукта. Анализируется значение и динамика этих данных.

5. **Оптимизированный**. Постоянное улучшение процессов основывается на количественных данных по процессам и на пробном внедрении новых идей и технологий.

Стоимость сертификации по CMMI составляет 100 000 $. Первыми в Европе сертифицированными по 4 и 5 уровням CMMI. (2002–2004гг.) стали такие компаний как «Motorola СПб», «Люксофт», «ЭПАМ Системс», IBA

1.5. Классификация информационных технологий по укрупненным видам и сферам информационной деятельности человека

Современные информационные технологии находят применение практически во всех сферах человеческой деятельности. Отличаясь небывалой динамикой развития и огромным разнообразием используемых аппаратно-программных средств информационные технологии, тем не менее, стабильны по отношению к сферам применения. Введем классификацию информационных технологий по данному аспекту. В качестве классификационных признаков выберем укрупненные виды информационной деятельности (первая строка табл.1. 1) и сферы деятельности человека (первый столбец табл. 1.1).

Таблица 1.1. Виды и сферы информационной деятельности человека

Укрупненные виды информационной деятельности / Сферы информационной деятельности	**Управление**	Информационное обслуживание	Проектирование и конструирование	Научные исследования	Обучение
Промышленность					
Строительство					
Сельское хозяйство					
Транспорт					
Связь					
Торговля					
Общественное питание					
Коммунально-бытовое обслуживание					
Наука					
Образование					
Культура					
Физическая культура и спорт					
Здравоохранение					
Охрана окружающей среды					
Социальное обеспечение					
Государственное управление					
Военное дело					

Выберем, например, второй столбец таблицы 1.1 и получим: «Информационная технология управления». Сайты с таким названием можно найти в Интернет (как правило, в них затрагиваются лишь отдельные частные

вопросы). Это очень сложная область, которая должна охватывать все сферы второго столбца.

Теперь обратим внимание на то, что получилось нечто масло масленое. Дело в том, что управление, по сути, является видом информационной технологии, поэтому слово «информационная» как избыточное можно опустить, записав так: «Технология управления» или «Компьютерная технология управления». Так и поступают в области вещественно-энергетических технологий. Вот, например, названия конкретных литературных источников: Технология сборки самолетов, Технология пылеулавливания, Технология склеивания деталей в самолетостроении. Заметим, что такое «масло масленое» характерно практически для всех видов первой строки таблицы 1.1, поэтому ниже данное обстоятельство будем игнорировать.

Сузим предметную область (сферу) столбца «Управление». Выберем строку «Транспорт», которая в свою очередь включает следующие виды транспорта: железнодорожный, морской, речной, автомобильный, воздушный, нефтепроводный, газопроводный, городской автомобильный, городской электрический. Остановимся, например, на воздушном транспорте, получим: «Информационная технология управления воздушным транспортом».

А теперь развернем столбец «Управление» по горизонтали. Управление подразделяется на организационное и технологическое управление. Выберем одну из функций организационного управления, например, принятие решений. Получаем — «Информационная технология принятия решений в управлении» или по умолчанию (опуская «управление») имеем «Информационная технология принятия решений» (источники с одноименным названием нетрудно найти в Интернет).

Выберем вторую строку таблицы 1.1, получаем: «Информационная технология в промышленности». Здесь слово «информационная» не только уместно, но и необходимо. Книга с таким названием появилась в бывшем СССР в 1988г. и была одной из первых в области информационной технологии. В данном случае имеет место полный охват таблицы по горизонтали, а по вертикали охватываются, не приведенные в таблице, все виды тяжелой и легкой промышленности.

На этом остановимся, предлагая выполнить в качестве упражнения подобную процедуру с поиском в Интернете для других видов и сфер информационной деятельности человека; такая работа окажется особенно полезной при выборе места будущей профессиональной деятельности.

Приведенные в таблице 1.1 виды информационной деятельности человека (первая строка табл. 1.1) известны с незапамятных времен. Обладая определенным методологическим консерватизмом, они во временном аспекте отличаются в основном достигнутым уровнем знаний и состоянием технологических средств и по существу являются укрупненными видами информационных технологий. В подтверждение сказанному, заметим, что даже при отсутствии автоматизации в этих видах деятельности в явном или неявном виде применялись технологические карты.

Каждая из сфер деятельности (первый столбец табл. 1.1) ориентированна на определенную предметную область и на некотором временном отрезке

представляет собой систему с интенсивной структурой. На определенных этапах в этих сферах происходят структурные изменения, объективно обусловленные технологическим прорывом (как в области вещественно-энергетических, так и информационных технологий). Так что развитие сфер деятельности в целом носит экстенсивный характер, приводящий к появлению новых сфер деятельности, например, таких как авиастроение, атомная, ракетно-космическая промышленность, индустрия нанотехнологий и т. д.

1.6. Методология современных информационных технологий

Рассмотрим первый столбец таблицы 1.1. Что можно «вынести за скобки» в столь различных областях человеческой деятельности? Общим во все времена существования этих областей являются ***процессы сбора, накопления, хранения, передачи, поиска и использования информации.*** Взаимосвязанная совокупность таких процессов применительно к конкретной предметной области (подразделения, организации, отрасли, государства) составляет определенную информационную технологию. Во временном аспекте такие технологии, отличаясь используемыми средствами и методами, прошли путь от докомпьютерной эры (ручная, механическая, электрическая технологии) до сложнейших современных компьютерных технологий, основанных на архитектурном подходе. Такие архитектурные информационные технологии ориентированы на корпорации, государственные учреждения и государство в целом.

Сущность архитектурного подхода к построению информационной технологии состоит в следующем. Архитектура информационной технологии представляет собой пирамиду, основание которой составляет архитектура инфраструктуры объекта автоматизации (аппаратное, программное обеспечение, сетевая инфраструктура). На этом основании формируются архитектура данных и архитектура общих сервисов; выше располагаются архитектура интеграции (третий уровень) и архитектура приложений четвертый уровень). И на самом верху пирамиды находится бизнес-архитектура, которая имеет интерфейс с внешними пользователями. Возможны и другие IT-архитектуры. При этом разработка архитектуры информационной технологии осуществляется сверху в низ, т.е. начиная с бизнес-процессов (функций по ГОСТ 34.003-90). Последнее обусловлено тем, что именно бизнес-процессы представляют собой базовые объекты автоматизации, они же являются и основными источниками изменений. Примеры бизнес-процессов: создание продуктов, продажа продуктов и услуг, управление заказами и т.д. Разработка таких архитектурных технологий сложна и требует специального рассмотрения, выходящего за рамки данного курса.

Используемые же в настоящее время на практике информационные технологии отличается большим разнообразием: от простейших офисных и других «коробочных» пакетов до технологий, основанных на методологии ERP/ACY и концепции CALS.

Системы класса ERP (Enterprise Resource Planning — планирование ресурсов предприятия) базируются на принципе единого хранилища корпоративной бизнес-информации и обеспечивают следующие (основные) функции:

- Ведение конструкторских и технологических спецификаций изготовляемых изделий.
- Формирование планов продаж и производства.
- Планирование потребностей в материалах и комплектующих, сроков и объемов поставок.
- Управление запасами и закупками: ведение договоров, реализация централизованных закупок, обеспечение учета и оптимизации складских и цеховых запасов.
- Планирование производственных мощностей.
- Оперативное управление финансами, включая составление финансового плана и осуществление контроля его исполнения; финансовый и управленческий учет.
- Управления проектами, включая планирование этапов и ресурсов, необходимых для их реализации.

Архитектура ERP-систем, как правило, строится по модульному принципу (совокупность интегрированных пакетов, каждый из которых состоит из определенного набора модулей). Основоположник ERP-систем — немецкая компания SAP AG (система SAP R/3). К современным ERP-системам также относятся: People Soft (People Soft Inc), BAAN IV (BAAN), Oracle Applications (Oracle Corporation), One World (J.D.Edwards) и другие. Близки к ERP-системам и такие интегрированные системы, как Галактика ERP (Корпорация Галактика), Парус (Корпорация Парус) и др.

ERP-системы «выросли» из систем класса **MRP II** (Manufacturing Resource Planning — планирование ресурсов предприятия: материальных, мощностных, финансовых; не путать со стандартом **MRP II**), а последние «выросли» из MRP (Material Requirements Planning — планирование потребностей материалов).

Становление Интернет (Web-технологий) привело к созданию ERP II (Enterprise Resource and Relationship Processing: управление ресурсами и внешними отношениями предприятия). ERP II имеет два контура: традиционный внутренний (back-office), управляющий внутренними бизнес процессами предприятия; и внешний (front-office) – обеспечивающий взаимодействие с контрагентами и покупателями продукции.

Повышение эффективности бизнес-процессов современного предприятия предполагает интеграцию системы электронной коммерции **B2B** (Business-to-Business) с ERP/ACY предприятия. B2B — это портал, который предназначен для взаимодействия с фирмами-партнерами, поставщиками, потребителями и инвесторами. B2B-портал, как и любой другой корпоративный Web-портал, представляют собой единую Web-точку доступа к информации, сервисам и приложениям доступную как для внешних, так и для внутренних пользователей. При этом портал, являясь интегратором данных и приложений, не заменяет другие ИТ-системы и приложения. Кроме B2B находят применение и другие виды порталов: **B2C** (Business-to-Consumer: для потребителя), **B2E** (Business-to-Employee: для служащих), **B2B2C** (интеграция на одной площадке **B2B** и **B2C**). Лидер в области ERP система SAP R/3 обеспечивает бесшовную интеграцию с B2B.

Концепция CALS (Computer Aided Logistics Support: компьютерная поддержка процесса поставок) возникла в 80-х годах в оборонном комплексе США. Она совершенствовалась, дополнялась и, сохранив существующую аббревиатуру (CALS), получила более широкую трактовку: Continuous Acquisition and Life cycle Support — непрерывные поставки и информационная поддержка жизненного цикла продукции

Под непрерывными поставками (Continuous Acquisition) понимается постоянное взаимодействие с заказчиком в процессе формализации его требований, формирования заказа и процесса поставки. Поддержка жизненного цикла изделия (Life Cycle Support) отражает системность подхода к информационной поддержке всех процессов жизненного цикла изделия, включая процессы эксплуатации, обслуживания, ремонта и утилизации.

В руководстве по применению CALS в НАТО CALS определяется как «…совместная стратегия государства и промышленности, направленная на совершенствование существующих процессов в промышленности, путем их преобразования в информационно-интегрированную систему управления жизненным циклом изделий». Русскоязычное наименование этой концепции и стратегии — ИПИ (Информационная Поддержка жизненного цикла Изделий).

Поскольку термин CALS имеет военную окраску, то в гражданской сфере используются также термины PLCS (Product Life Support: поддержка жизненного цикла изделия) или PLM (Product Life Management: управление жизненным циклом изделия).

CALS-технологии сложны и имеют высокую стоимость, но их внедрение позволяет резко сократить сроки разработки и изготовления сложных изделий, например, срок разработки аэробуса А-360 был сокращен до 3-х лет против обычных 10-12 лет. Представление о стоимости CALS-технологий можно составить на примере разработки виртуального предприятия (1990-1995гг.), на разработку программного обеспечения которого General Motors затратила 3 млрд. долларов. Поэтому внедрение CALS-технологии является делом государственной политики.

Виртуальное предприятие (ВП) представляет собой группу производств объединенных единым информационным пространством, организующим весь жизненный цикл продукта или услуги. Термин «виртуальное предприятие» был предложен по аналогии с понятием виртуальной машины. Существуют различные интерпретации термина ВП.

1.7. Общие и частные принципы разработки информационной технологии

Создание ИТ-системы, основанной на информационной технологии современного уровня, является процессом сложным и дорогостоящим. Процесс разработки ИТ-системы, который сам по себе является информационной технологией, состоит из стадий и этапов, регламентированных стандартами. При разработке ИТ-системы, на всех ее этапах (начиная от анализа предметной области вплоть до сопровождения ИТ-системы) необходимо соблюдать как общие принципы системного подхода, так и частные принципы разработки информационной технологии. Пренебрежение названными принципами (особенно

на ранних стадиях проектирования) дорого обходится, поскольку исправление ошибок, допущенных на предыдущей стадии, стоит на порядок дороже, чем на текущей стадии. Несоблюдение же общих и частных принципов разработки информационной технологии приводит либо к неработоспособности ИТ-системы, либо к ее низкой эффективности. Именно этим можно объяснить то, что успешными являются менее 50 % ERP/АСУ.

Общие принципы системного подхода применимы к любой сфере человеческой деятельности. **К таким принципам относятся:**

1. Принцип системности (учитываются все существенные аспекты системы; ее элементы, связи, процессы).
2. Принцип иерархичности (рассматривается вышестоящая система, сама система и ее подсистемы).
3. Принцип интегративности (рассматриваются эмерджентные свойства системы).
4. Принцип формализации (используются конструктивные методы описания, анализа и синтеза).

Разработка любой проблемы (системы) предполагает поэтапно-циклическое разрешение следующих подпроблем:

Постановка проблемы (выделение проблемы).

Описание (построение модели).

Формирование критериев (для сравнения альтернативных вариантов разрешения проблемы).

Идеализация (рациональное упрощение проблемы).

Декомпозиция (разделение целого на части).

Композиция (объединение частей в целое).

Решение **(**разрешение проблемы).

Процедура разрешения проблемы носит, как правило, последовательно-циклический характер. Подпроблемы решается поэтапно и вместе с тем одновременно (совместно). Совместное решение подпроблем взаимноограничивает области возможных решений, отсекая большинство нерациональных альтернатив, что обусловлено структурной взаимосвязанностью подпроблем, которая и ограничивает пространство возможных решений.

Яркими примерами системного подхода является разработка Эталонной модели взаимодействия открытых (RM OSI), см. п. 2.2., и концептуальной модели компьютерной поддержки процесса поставок (CALS), см. п. 1.6.

К основополагающим частным принципам построения информационной технологии относятся:

Принцип открытости.

Принцип распределенности (децентрализация процесса обработки информации).

Принцип интеграции (функциональной, информационной, программной, технической, организационной).

Принцип унификации (приведение ИТ к единообразию, позволяющему расширить масштабы внедрения и снизить затраты на создание ИТ).

Важнейшим из названных принципов является принцип открытости. Открытость систем достигается на основе стандартизации их поведения на границах уровней систем и их интерфейсах.

Основные свойства системной открытости:

переносимость (portability*)* программного обеспечения, данных и опыта людей;

интероперабельность (interoperability**),** интерпретируемая как возможность взаимодействия компонентов распределенной системы (обмен информацией, совместное использование информации);

масштабируемость (scalability), т.е. свойство расширяемости и сохранения работоспособности ИТ-системы при модульном ее наращивании.

Методологические основы концепции открытых систем рассмотрены в п. 2.1.

Тема 2. Концептуальные основы информационной технологии

2.1. **Концепция открытых систем**

Методологическую основу концепции открытых, систем составляют:

• концептуальный базис и принципы построения открытых систем;

• эталонная модель взаимосвязи открытых систем RM OSI (Reference Model Open Systems Interconnection);

• эталонная модель окружения (среды) открытых систем RM OSE (Reference Model Open System Environment), эту модель можно представить как расширение RM OS**I** с детализацией верхнего (прикладного) уровня;

• концепция профилирования ИТ (для конструирования открытых систем в пространстве стандартизованных решений);

• концепция тестирования конформности (соответствия) систем ИТ исходным стандартам и профилям;

• таксономия профилей (Taxonomy: классификация, систематизация, для однозначной идентификации).

Согласно ISO/IEC TR 10000–1 “профиль — это один или сочетание нескольких базовых стандартов с идентификацией выбранных классов, подмножеств, факультативных возможностей и параметров этих базовых стандартов, необходимых для выполнения конкретной функции”.

Таксономию ISO/IEC TR 10000-1 трактует как“ систематизация профилей и определение структуры, в которых они должны размещаться”. Профили подразделяются на классы, каждый из которых представляет собой набор функций, независимых (в разумных пределах) от функций других классов.

В ISO/IEC TR 10000-1 определены следующие классы профилей:

F — профили формата обмена и представления данных;

T — транспортные профили с установлением соединения;

U — транспортные профили без установления соединения;

R — ретрансляционные профили;

A — прикладные профили, использующие профили T;

B — прикладные профили, использующие профили U.

Термин “таксономия” предложен в 1813г. швейцарским ботаником О. Декандолем в качестве синонима к слову “систематика” (растений, надо полагать). Однако современные словари русского языка трактуют это понятие уже как “теория классификации и систематизации сложно организованных областей действительности, имеющих обычно иерархическое строение”.

Основополагающие документы, определяющие концепцию открытых систем:

1. ISO/IEC TR 10000 (Framework and taxonomy of International Standardized Profile) — Основы и таксономия международных стандартизованных профилей. Включает три части:

ISO/IEC TR 10000-1 (1995, General Principles and Documentation Framework) — Общие принципы и основы документирования.

ISO/IEC TR 10000-2 (1995, Principles and Taxonomy for OSI Profiles) — Принципы и таксономия профилей взаимосвязи открытых систем (ГОСТ Р ISO/IEC/МЭК ТО 10000-2-99).

ISO/IEC TR 10000-3 (1995, Principles and Taxonomy for Open System Environment Profiles) — Принципы и таксономия профилей окружений открытых систем.

2. ISO 7498 (Open Systems Interconnection - Basic Reference Model. RM OSI, 1983г.) — эталонная модель взаимодействия открытых систем. Документ опубликован в 1984г. Практически одновременно была принята одноименная Рекомендация CCITT X.200 .

3. ISO/IEC DTR 14252 (Portable Operating System Interface for Computer Environments: интерфейс переносимой операционной системы). Этот стандарт фиксирует на международном уровне стандарт POSIX (IEEE 1003.0-1995: Open System Environment). Первая редакция стандарта POSIX была выпущена в 1988 году. Вторая редакция утверждена 12 июля 1996 года. IEEE – организация, в которой разработан стандарт POSIX. POSIX (Portable Operating System Interface) — стандарт на интерфейс (сопряжение) между операционной системой и прикладной программой, обеспечивающий переносимость прикладных программ

Примечание. Используемые сокращения:

PAS — публично доступные,
TS — технические спецификации,
TR — технические отчеты,
PDTR — предложенный черновой технический отчет,
DTR — предварительный технический отчет,
ITA — промышленные технические соглашения.

2.2. Эталонная модель OSI

RM OSI (Reference Model Open Systems Interconnection) — Эталонная модель взаимодействия открытых систем (ЭМ ВОС) разработана международной организацией по стандартизации ISO.

Модель OSI определяет общий подход к построению архитектуры систем распределенной обработки информации. Она описывает и регламентирует структуру взаимодействия открытых систем. Термин «открытая система» подчеркивает тот факт, что если какая-либо система отвечает стандартам, принятым в данной концепции, то эта система будет открыта для взаимодействия с любой другой системой, отвечающей этим стандартам.

Работы по созданию OSI начались в ISO и в CCITT в середине 70-х годов прошлого столетия. В мае 1983 г. подкомитет ПК16 «Взаимосвязь открытых систем ISO» принял стандарт ISO 7498 «Базовая эталонная модель взаимодействия открытых систем» и практически одновременно CCITT принимает одноименную рекомендацию X.200. Оба документа опубликованы в конце 1984 г.

Из существующих сетевых концепций и архитектур комитет ПК16 выбрал сетевую концепцию OSI, близкую к концепции SNA фирмы IBA.

В основу эталонной модели положена идея декомпозиции процесса функционирования открытых систем на отдельные компоненты (подсистемы), называемые уровнями. Взаимодействие между уровнями по горизонтали осуществляется согласно стандартным протоколам. При этом по вертикали каждый нижестоящий уровень обеспечивает вышестоящему определенный набор услуг (межуровневый стандартный интерфейс).

Протоколы регламентируют правила взаимодействия одинаковых уровней у различных пользователей, а межуровневый интерфейс – правила взаимодействия смежных уровней одного пользователя. Набор услуг N-го уровня модели OSI (совокупность функциональных возможностей данного и всех нижележащих уровней, включая средства, реализующие эти возможности) составляет N-ую службу модели OSI (сервис N-го уровня или N-сервис). N-ая служба (N-сервис) предоставляет услуги N+1 уровню, так что в модели OSI можно выделить шесть уровней службы: физическая служба, канальная служба, сетевая служба, транспортная служба, сеансовая служба и представительная служба.

Основные понятия модели OSI

1. Протокол – совокупность семантических и синтаксических правил, определяющих работу функциональных устройств в процессе связи.
2. Интерфейс – граница между двумя функциональными устройствами, определенная своими функциональными характеристиками, общими механическими характеристиками соединения, характеристиками сигналов обмена и другими полезными характеристиками.
3. Пользователь службы – объект в некоторой открытой системе, который использует службу через точку доступа к службе (ТДС).
4. Точка доступа к службе – точка, в которой логический объект уровня предоставляет сервис логическому объекту смежного верхнего уровня.
5. Услуга уровня – функциональная возможность, которую данный уровень вместе с нижерасположенными уровнями обеспечивает смежному верхнему уровню.
6. Служба уровня – совокупность услуг уровня и правил их использования.
7. Поставщик службы – некоторое множество объектов, обеспечивающих службу для ее пользователей.

8. Примитив службы – абстрактное, не зависящее от конкретной реализации, представление взаимодействия между пользователем и поставщиком службы.

9. Примитив запроса – примитив, инициируемый пользователем службы для вызова некоторой процедуры.

10. Примитив индикации – примитив, инициируемый поставщиком службы для вызова некоторой процедуры либо для указания о ее вызове одним из взаимодействующих пользователей сервиса в ТДС данного уровня.

11. Примитив ответа – примитив, инициируемый пользователем службы для завершения в определенной ТДС некоторой процедуры, ранее вызванной посредством примитива индикации в этой ТДС.

12. Примитив подтверждения – примитив, инициируемый поставщиком службы для завершения в определенной ТДС некоторой процедуры, ранее вызванной посредством примитива запроса в этой ТДС.

Уровни OSI. Основные задачи и выполняемые функции

В модели OSI взаимодействие делится на семь уровней или слоев, см. рис.2.1. Каждый уровень имеет дело с одним определенным аспектом взаимодействия, так что проблема взаимодействия декомпозирована на 7 частных проблем, каждая из которых может быть решена независимо от других. Каждый уровень поддерживает интерфейсы с выше- и нижележащими уровнями.

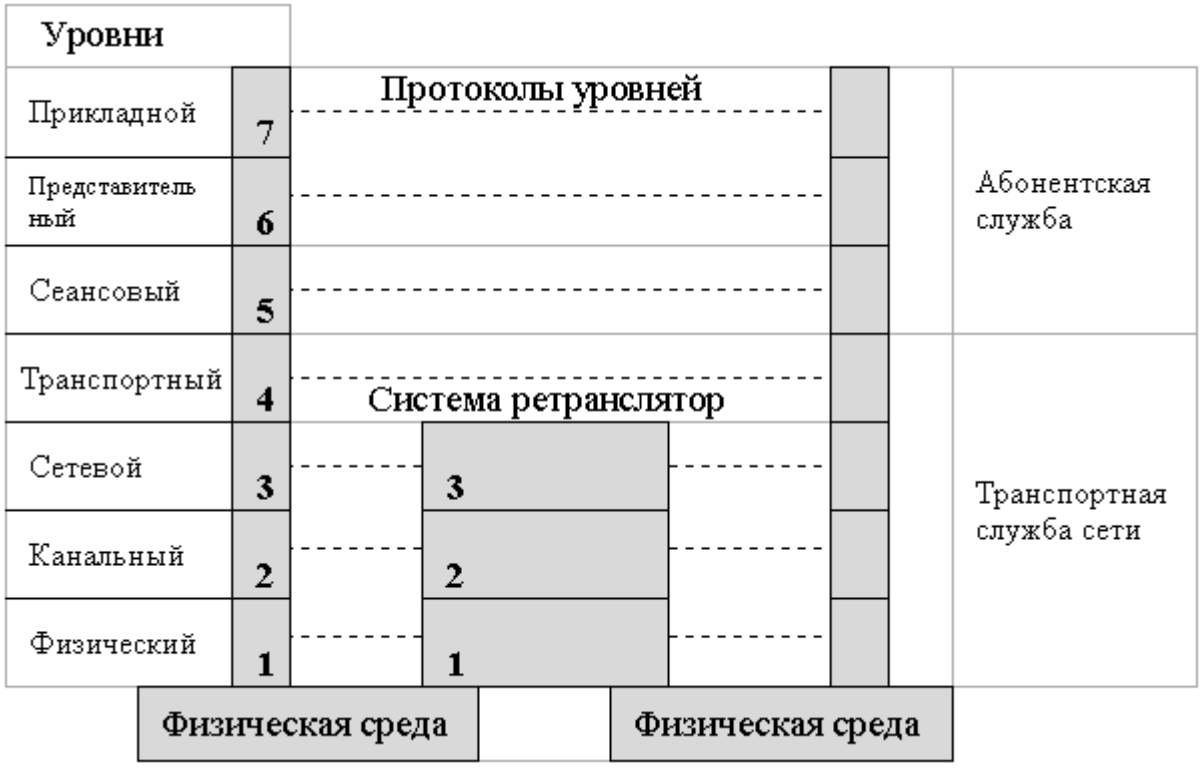

Рис. 2.1. Уровни RM OSI.

Физический уровень выполняет передачу битов по физическим каналам, таким, как коаксиальный кабель, витая пара или оптоволоконный кабель. На этом уровне определяются характеристики физических сред передачи данных и параметров электрических сигналов (такие параметры, как: напряжение в сети, сила тока, число контактов на разъемах и т.п.).

Канальный уровень обеспечивает передачу кадра данных между любыми узлами в сетях с типовой топологией, либо между двумя соседними узлами в сетях с произвольной топологией. В протоколах канального уровня заложена определенная структура связей между компьютерами и способы их адресации.

Сетевой уровень обеспечивает доставку данных между любыми двумя узлами в сети с произвольной топологией, при этом он не отвечает за надежность передачи данных.

Транспортный уровень обеспечивает передачу данных между любыми узлами сети с требуемым уровнем надежности. Для этого на транспортном уровне имеются средства установления соединения, нумерации, буферизации и упорядочивания пакетов.

Сеансовый уровень предоставляет средства управления диалогом, позволяющие фиксировать, какая из взаимодействующих сторон является активной в настоящий момент, а также предоставляет средства синхронизации в рамках процедуры обмена сообщениями.

Уровень представления. Уровень представления имеет дело с внешним представлением данных. На этом уровне могут выполняться различные виды преобразования данных, такие как компрессия и декомпрессия, шифровка и дешифровка данных (этот уровень необходим для преобразования данных из промежуточного формата сессии в формат данных приложения). В Internet это преобразование возложено на прикладные программы.

Прикладной уровень представляет собой набор разнообразных сетевых сервисов, предоставляемых конечным пользователям и приложениям. Примерами таких сервисов являются, например, электронная почта, передача файлов, сетевое подключение удаленных терминалов. Этот уровень определяет протоколы обмена данными прикладных программ.

При построении транспортной подсистемы (транспортная служба сети) наибольший интерес представляют функции физического, канального и сетевого уровней, тесно связанные с используемым в данной сети оборудованием (сетевыми адаптерами, концентраторами, мостами, коммутаторами, маршрутизаторами). Функции прикладного, сеансового уровней и уровня представления (составляющие абонентскую службу) реализуются операционными системами и системными приложениями конечных узлов; при этом транспортный уровень выступает посредником между этими двумя группами протоколов.

В компьютерных сетях широко используется понятие стека протоколов, под которым понимается иерархически организованная совокупность протоколов, решающих задачу взаимодействия узлов сети.

Стек протоколов эталонной модели взаимодействия открытых систем (стек OSI), в отличие от других стеков протоколов, полностью соответствует модели OSI. Он включает спецификации протоколов для всех семи уровней взаимодействия, определенных в этой модели.

Однако основным стеком протоколов сетевых технологий в настоящее время является стек TCP/IP, который является промышленным стандартом протоколов и широко применяется как в Интернет, так и в интранет. Этот стек имеет многолетнюю историю и поддерживается всеми современными операционными системами.

2.3. Общие сведения о стандартах в области информационных технологий

2.3.1. Роль стандартов в области информационных технологий

По определению ISO стандартом является документ, доступный и опубликованный, коллективно разработанный или согласованный и общепринятый в интересах тех, кто им пользуется, основанный на интеграции результатов науки, технологии, опыта, способствующий повышению общественного блага и принятый организациями, полномочными на национальном, региональном и международном уровнях.

Стандартам в области информационных технологий принадлежит особая роль. Они, образно говоря, играют роль генетического механизма, в котором на основе консенсуса аккумулированы практический опыт и результаты научных исследований в области информационных технологий.

Стандарты в области информационных технологий содержат определения основных понятия, описания информационных технологий, моделей, сценариев, функций, правил поведения и представления информации и по существу являются научно-методической основой и фундаментом информационной индустрии. Интеграция мирового научно-технического потенциала в данной области, осуществляется международной системой стандартизации и характеризуется небывалыми для науки и техники масштабами.

2.3.2. Уровни и виды стандартов

Различают 5 уровней стандартов**:**

1) стандарты международных организаций, например, ISO, IEC, ITU;

2) стандарты региональных организаций (международно-групповых объединений), например, CEN, CENELEC, ETSI;

3) национальные стандарты, например, ANSI, BSI, DIN, ГОСТ, ГОСТ Р, СТБ;

4) стандарты профессиональных организаций, например, IEEE, ISA, ISOC, IAB, IETF, IRTF, IESG, ECMA;

5) стандарты отдельных фирм, например, Intel, Xerox, IBM.

Различают также юридические, фактические и промышленные стандарты**.**

Стандарт де-юре (de jure; юридически принятый) – это стандарт, который создан официально признанной организацией (ISO, IEC, ITU). Их иногда называют базовыми или формальными стандартами. Такие стандарты являются открытыми. Они свободны для копирования, а продукция, изготовленная на их основе, не требует лицензии, что обеспечивают независимость от поставщиков изделий.

Стандарт де-факто (de facto; фактический) – стандарт на продукцию поставщика, который захватил большую часть рынка, и который другие поставщики стремятся эмулировать, копировать или использовать.

Промышленный стандарт – это стандарт, который широко применяется в промышленности. Это может быть как стандарт де-факто, так и де-юре. Неудачные стандарты де-юре часто не становятся промышленными.

Промышленные стандарты, как правило, связаны с изделиями, доминирующими на рынке, и в значительной степени зависят от изготовителей продукции.

2.3.3. Международные организации по стандартизации

ISO (International Organization for Standardization; ИСО) — Международная организация по стандартам. Основана в 1946 году. Осуществляет разработку международных стандартов в различных областях человеческой деятельности путем координации деятельности национальных организаций. ISO работает под эгидой ООН и включает представителей более 100 стран. Общее число созданных и сопровождаемых ISO стандартов к 2001г., составляло порядка 13000, из них более 2000 стандартов относятся к области ИТ.

В ISO работает около 3 000 технических комитетов, подкомитетов и рабочих групп, в совещаниях которых ежегодно принимает участие более 30 000 экспертов. ISO сотрудничает с более чем 500 международными организациями.

Отдельные документы ISO мы уже рассматривали (ISO 9000, ISO 7498 и др.). Следует отметить, что наиболее значимые стандарты организаций, не относящихся к международному уровню, становятся международными посредством ISO

IEC (International Electrotechnical Commission, http://www.iec.ch,: Международная Электротехническая Комиссия; МЭК). Организация IEC, образованная в 1906г. Также как и ISO является добровольной неправительственной организацией. Ее деятельность в основном связана со стандартизацией физических характеристик электротехнического и электронного оборудования. Организационное устройство IEC во многом аналогично ISO (ISO создавалась по образу и подобию IEC), аналогичен и процесс создания стандартов в IEC. Как и в ISO основную работу по разработке стандартов в IEC осуществляют технические комитеты (TCs) и подкомитеты (SCs), общая численность которых превышает 200.

JTC1 (Joint Technical Committee 1 — Объединенный технический комитет 1). Обеспечивает формирование системы базовых стандартов в области информационных технологий (ИТ) и их расширений для конкретных сфер деятельности. Образован в 1987г. на основе **I**SO и IEC (ISO /IEC).

Работа в JTC1 над стандартами ИТ, относящимися к окружению открытых систем (Open Systems Environment — OSE), распределена по следующим подкомитетам (Subcommittees – SC; на 2001г.):

SC1 Vocabulary (словарь понятий).

SC2 Corded character sets (Символьные наборы и кодирование информации).

SC6 Telecommunication and information exchange between systems (Телекоммуникация и информационный обмен между системами).

SC7 Software engineering (Программная инженерия).

SC11 Flexible magnetic media for digital data interchange (Гибкая магнитная среда для обмена электронными данными).

SC17 Identification cards and related devices (Идентификационные карты и связанные с ними устройства).

SC22 Programming languages, their environments and system software interfaces (Языки программирования, их окружения и интерфейсы системного программного обеспечения).

SC24 Computer graphics and image processing (Компьютерная графика и обработка изображений).

SC25 Interconnection of information technology equipment (Взаимосвязь оборудования информационных технологий).

SC27 IT Securities techniques (Методы безопасности ИТ).

SC29 Coding of audio, picture, multimedia and hypermedia information (Кодирование аудио, графической, мультимедийной и гипермедиа информации).

SC31 Automatic identification and data capture techniques (Автоматическая идентификация и методы считывания данных).

SC32 Data management and interchange (Обмен и управление данными).

SC34 Document description and processing languages (Языки описания и обработки документов).

SC35 Use interfaces (Пользовательские интерфейсы).

SC36 Learning Technology (Технологии обучения).

Дополнительно к названным подкомитетам была создана группа по функциональным стандартам (Special Group on Functional Standards — SGFS**)** для обработки предложений по Международным стандартизованным профилям (International Standardized Profiles — ISPs), представляющим определения профилей ИТ.

ITU (International Telecommunication Union) – международный телекоммуникационный союз, структурное подразделение ООН. Эту организацию называют также Международным союзом электросвязи (МСЭ). Объединяет более 500 правительственных и неправительственных организаций. Центральный офис ITU расположен в Женеве (Швейцария).

ITU – старейшая международная профессиональная организация. Основана в 1865г. (после подписания 20-ю европейскими государствами первой международной конвенции по телеграфии) под названием Международный союз по телеграфии (International Telegraph Union). Нынешнее название присвоено в 1932г. В 1947г. ITU получила статус специализированного агентства ООН.

В 1956г. в ITU был сформирован Международный консультативный комитет по телеграфии и телефонии (International Telephone and Telegraph Consultative Committee, – CCITT). Документы CCITT носят название "Recommendations" (Рекомендации, с большой буквы). После структурной реформы ITU (декабрь 1992г., внеочередная женевская конференция) функции CCITT начиная с 1993г. были возложены сектор ITU-T так что, например, Рекомендации V.42 CCITT и V.42 ITU-T означают одно и то же.

ITU-T (International Telecommunications Union-Telecommunications Standardization Sector; или иногда ITU-TSS) – один из трех секторов ITU.

Другие сектора: Сектор радиосвязи **(ITU-R),** и Сектор стандартизации телекоммуникаций и развития телекоммуникаций **(ITU-D),** занимающийся вопросами стратегии и политики в области связи.

Для организаций, входящих в состав ITU-T определены следующие пять классов членства:

класс A – национальные министерства и ведомства связи;

класс B – крупные частные корпорации, работающие в области электросвязи;

класс C – научные организации и предприятия, производящие связное оборудование

класс D – международные организации, в том числе, организация ISO;

класс E –организации из других областей деятельности, но заинтересованные в работе в данном секторе.

Право голоса при принятии решений дается только представителям организаций классов A, B

Основная работа по разработке стандартов выполняется тематическими исследовательскими группами (Study Groups — SGs), которые сформированы таким образом, чтобы обеспечить полноту покрытия всех актуальных направлений технологий электросвязи. В 2000г. насчитывалось 14 таких групп. В плане стандартизации ИТ наибольший интерес представляет результаты деятельность следующих групп:

SG7 – Data and open communications systems (Данные и открытые коммуникационные системы).

SG8 – Multimedia Services (Мультимедийные сервисы).

SG10 – Software languages (Языки для программного обеспечения: стандарты языков программирования и языков формальной спецификации, используемых при разработке телекоммуникационных систем).

SG13 – GII principles and structure (Структура и принципы Глобальной информационной инфраструктуры).

Для разрабатываемых ITU-T Рекомендаций введена (со времен деятельности CCITT) серийная классификация документов (24 серии). Вот список этих серий:

Series A: Organization of the work of the ITU-T (Организация работы ITU-T);

Series B: Means of expression: definitions, symbols, classification (Средства выражения: символы, классификация);

Series C: General telecommunication statistics. (Общие статистические данные в телекоммуникации);

Series D: General tariff principles (Общие принципы тарификации);

Series E: Overall network operation, telephone service and human factors (Общая работа сетей, телефонные услуги и человеческие факторы);

Series F: Non-telephone telecommunication services (Нетелефонные службы электросвязи);

Series G: Transmission systems and media, digital systems and networks (Системы передачи и среды, цифровые системы и сети);

Series H: Audiovisual and multimedia systems (Аудиовизуальные и мультимедийные системы);

Series I: Integrated services digital network - ISDN (Цифровая сеть с интеграцией служб);

Series J: Transmission of television, sound programme and other multimedia signals (Передача звукового вещания, телевизионных и мультимедийных сигналов);

Series K: Protection against nterference (Защита от помех);

Series L: Construction, installation and other elements of outside plant (Конструкция, прокладка, защита кабелей и элементов линейных сооружений);

Series M: TMN and network maintenance: international transmission systems, telephone circuits, telegraphy, facsimile and leased circuits (Техническая эксплуатация: международные системы передачи, телефонные каналы, телеграфные, факсимильные и арендуемые каналы);

Series N: Maintenance: international sound programme and television transmission circuits (Техническая эксплуатация: международные каналы звукового и телевизионного вещания);

Series O: Specifications of measuring equipment (Требования к измерительной аппаратуре);

Series P: Telephone transmission quality, telephone installations, local line networks (Качество телефонной передачи, прокладка линий, сети локальных линий);

Series Q: Switching and signalling (Коммутация и сигнализация);

Series R: Telegraph transmission (Телеграфная передача);

Series S: Telegraph services terminal equipment (Оконечное оборудование телеграфных служб);

Series T: Terminals for telematic cervices (Оконечное оборудование и телематические службы);

Series U: Telegraph Switching (Телеграфная коммутация);

Series V: Data communication over the telephone network (Передача данных по телефонной сети);

Series X: Data networks and open system communications (Сети передачи данных и связь открытых систем);

Series Y: Global information infrastructure (Глобальная информационная инфраструктура);

Series Z: Programming languages (Языки программирования).

2.3.4. Региональные организации по стандартизации

К региональным относятся организации, представляющие в глобальном процессе стандартизации ИТ интересы крупных регионов или континентов.

CEN (the European Committee for Standardization — www.cenorm.be) — европейский комитет стандартизации широкого спектра товаров, услуг и технологий, в том числе, и связанных с областью ИТ.

CENELEC (the European Committee for Electrotechnical Standardization — www.cenelec.be) — европейский комитет стандартизации решений в электротехнике, включая коммуникационные кабели, волоконную оптику и электронные приборы.

ETSI (European Telecommunications Standards Institute — www.etsi.org) — европейский институт стандартизации в области сетевой инфраструктуры.

2.3.5. Национальные организации по стандартизации

ANSI (American National Standards Institute) — Американский институт стандартов организация, ответственная за стандарты в США. ANSI является членом ISO.

BSI, DIN – национальные организации по стандартизации Англии и Германии соответственно.

ГОСТ — стандарты бывшего СССР.

ГОСТ Р – стандарты России,

СТБ — стандарты РБ.

Стандарты действующие на территории РБ можно найти на официальных сайтах Госстандарта и БелГИСС, а также используя ИПС «СТАНДАРТ» (информационное обеспечение > тематический или контекстный поиск). В Республике Беларусь государственные стандарты бывшего СССР (ГОСТ) имеют статус государственных стандартов Республики Беларусь (Постановление Госстандарта РБ №3 от 17.12.1992г.).

В РБ международные стандарты ISO и др. обычно применяются через национальные стандарты. Это объясняется не только их англоязычностью, но и дороговизной.

Различаются также отраслевые стандарты (ОСТ), стандарты предприятий (СТП) и руководящие документы отрасли (РД).

2.3.6. Организации по стандартизации профессиональных объединений и промышленных консорциумов

IEEE (Institute of Electrical and Electronic Engineers) — институт инженеров по электротехнике и радиоэлектронике. Профессиональная организация (США), основанная в 1963 году для координации разработки компьютерных и коммуникационных стандартов. IEEE подготовил группу стандартов 802 для локальных сетей (подготовил подкомитет IEEE 802, созданный в 1980г).

ISA – Приборостроительное общество Америки, разрабатывает стандарты локальных сетей реального времени.

Взаимосвязанные организации **ISOC, IAB, IETF, IRTF, IESG** отвечают за стандартизацию в области Интернет-технологий**.**

ISOC (Internet Society — Общество Интернета, www.isoc.org/index.html,) — ассоциация экспертов, отвечающая за разработку стандартов технологий сети Интернет, организовано в январе 1992 год. ISOC — некоммерческая неправительственная международная профессиональная организация (ее членами являются 175 организаций и около 9000 физических лиц из более чем 170 стран мира). **ISOC** относится к верхнему уровню иерархии и его называют также организационным домом (organizational home) для IAB, IETF, IRTF, IESG.

IAB (Internet Architecture Board - Совет по архитектуре сети Интернет) — группа технических советников в составе ISOC, непосредственно отвечающая за развитие архитектуры Интернет, управление разработкой и сопровождением стандартов для протоколов и сервисов сети Интернет. Кроме этого, IAB несет ответственность за управление редактированием и публикацией спецификаций **RFC** (Request for Comments), осуществляемое издательским органом RFC Editor

(http://www.rfc-editor.org), а также за управление присваиванием номеров для RFC (посредством механизма **IANA** — Internet Assigned Numbers Authorities).

Деятельность IAB поддерживается напрямую и косвенно, как правительством США, так и промышленностью. Прямая поддержка осуществлялась, например, через Корпорацию национальных исследовательских инициатив CNRI (Corporation for National Research Initiatives), через которую IAB спонсировался от ряда агентств федерального правительства, включая DARPA, NASA (National Aeronautics and Space Administration**)**, министерство энергетики (Department of Energy), Национальный научный фонд (National Science Foundation).

Подразделение **IETF** (Internet Engineering Task Force) — рабочая группа по проектированию Интернет-технологий (www.ietf.org). **IETF** включает более 40 рабочих подгрупп. IETF выпускает стандарты TCP/IP в виде серии документов, названных **RFC** (Request for Comment). Стандарты TCP/IP всегда публикуются в виде документов RFC, но не все RFC являются стандартами.

IETF по существу является большим открытым международным сообществом разработчиков, операторов, изготовителей и исследователей в области сетевых технологий, занимающихся вопросами развития архитектуры сети Интернет и способов ее использования. Она открыта для всех, кто интересуется Интернет-технологиями.

IRTF **(**Internet Research Task Force) отвечает за исследования и разработку набора протоколов Internet.

IESG (Internet Engineering Steering Group — группа технического управления сети Интернет, www.ietf.org) отвечает за техническое управление процессом стандартизации Интернет-технологий, осуществляет экспертизу проектов спецификаций, разрабатываемых IETF, несет ответственность за принятие Интернет-стандартов и их дальнейшее продвижение.

ECMA (European Computer Manufacturers Association) – европейская ассоциация изготовителей вычислительных машин или позже: европейская ассоциация производителей компьютеров, ЕАПК. ECMA организована в 1961г. по инициативе ведущих западноевропейский компаний в области средств обработки данных.

2.4. Стратегическое направление информационных технологий

Качественно новый подход в области стандартизации информационных технологий сформировался в 1995 году в рамках Концепции Глобальной информационной инфраструктуры **(GII —** Global Information Infrastructure**).** Согласно этой концепции GII будет представлять собой интегрированную общемировую информационную сеть массового обслуживания населения планеты на основе интеграции глобальных и региональных информационно-коммуникационных систем, а также систем цифрового телевидения и радиовещания, спутниковых систем и подвижной связи. При этом реализации GII предполагает эволюционный путь развития, основанный на последовательной модернизации и интеграции существующих систем и технологий на базе новых принципов и стандартов.

Разработка GII осуществляется на основе концепции открытых систем и относится к числу наиболее крупномасштабных разработок, охватывающих несколько десятков взаимосвязанных проектов. Международные стандарты **GII** представляются в виде Рекомендаций ITU-T серии Y (Рекомендации Y.100 — Y.799, разработаны исследовательской группой SG13).

Рекомендации серии Y:

Y.100 -Y.199 General;

Y.200 -Y.299 Services, applications and middleware;

Y.300 -Y.399 Networks aspects;

Y.400 -Y.499 Interfaces and protocols;

Y.500 -Y.599 Numbering, addressing and naming;

Y.600 -Y.699 Operation, administration and maintenance;

Y.700 -Y.799 Security;

К важнейшим системным стандартам относятся Рекомендации:

Y.100:1998 General overview of the Global Information Infrastructure standards development.

Y.110:1998 Global Information Infrastructure principles and framework architecture.

Y.120:1998 Global Information Infrastructure scenario methodology.

Современная система стандартов в области информационных технологий имеет сложную структуру, развивается очень быстрыми темпами и содержит огромный объем стандартов (например, только число RFC приближается к 5000).

Тема 3. Основные подходы и методы описания информационных процессов и явлений

3.1. Информация и данные

Понятие «информация» является одним из наиболее сложных и многогранных понятий. Уяснить сущность этого понятия, которому посвящено огромное количество литературных источников, в отрыве от таких понятий как система, цель и технология не представляется возможным. Информационная технология реальной системы представляет собой вещественно-энергетические информационные процессы, реализующие сбор, хранение, преобразование, обработку, передачу и использование информации для достижения определенной цели. Поэтому ценность информации и другие ее свойства преломляются (оцениваются) именно через цель системы, в этом то и сложность, поскольку надо иметь описание (модель) системы.

Ввиду разнообразия трактовок понятия «информация», равно как и других родственных понятий, далее, по возможности, будем придерживаться дефиниций, приведенных в стандартах.

Информация (для процесса обработки данных) — содержание, присваиваемое данным, посредством соглашений, распространяющихся на эти данные (ИСО 2389/1-84).

Данные — событие, понятие или команда, представленные в формализованном виде, позволяющем передачу, интерпретацию или обработку, как вручную, так и с помощью средств автоматизации **(**ИСО 2382/1-84).

Информацию классифицируют по многочисленным признакам. По физической природе восприятия информацию подразделяют на зрительную, слуховую, тактильную; по метрическим свойствам — на параметрическую, топологическую и абстрактную. Параметрическая информация может быть нульмерной (событие), одномерной (величина), двумерной (функция), трехмерной (комплекс), n-мерной (n-мерное пространство). Топологическую информацию составляют геометрические образы, карты местности, различные изображения и объемные объекты. К абстрактной информации относятся обобщенные образы и понятия, математические соотношения.

Наиболее сложными и разнообразными видами информации являются биологическая (обеспечивает жизнедеятельность отдельно взятого живого организма) и социальная информация (связана с практической деятельностью человека).

Социальная информация делится на массовую (для всех членов общества) и специальную **(**для определенных социальных групп). Специальная социальная информация подразделяется на политическую, экономическую, технологическую, научно-техническую и т.д.

Социальная информация характеризуется тремя основными аспектами: синтаксическим, семантическим и прагматическим.

Синтаксический аспект характеризуют количественные и структурные свойства информации. Эти свойства связаны с формированием носителей информации и организацией информационных потоков в ИТ-системах.

Семантический аспект характеризуют смысловое содержание информации и отношением между ее элементами. Семантические свойства информации связаны с организацией процесса обработки информации.

Прагматический аспект связан с ценностью информации для ее получателя. Прагматические свойства информации проявляются на стадии принятия решений (при разработке и эксплуатации ИТ-систем).

3.2. Сигналы и знаки

Под сигналом в широком смысле слова будем понимать материальный носитель информации. В этом смысле сигнал – это некоторый физический объект или процесс, содержащий информацию. Сигналы в широком смысле можно подразделить на сигналы в узком смысле и знаки.

Вот определение сигнала в узком смысле. Сигнал – изменение физической величины, используемой для передачи данных (ИСО 2382/I-84). Здесь по умолчанию под передачей данных понимается перенос информации в пространстве. Но под передачей данных можно понимать и перенос данных во времени, т.е. хранение информации. Такое расширение обсуждаемого понятия вполне оправдано, поскольку и при передаче и при хранении информации по существу используются одни и те же алгоритмы сжатия данных и помехоустойчивого кодирования. Забегая вперед, заметим, что разработка этих алгоритмов была инициирована основными теоремами Шеннона для дискретных каналов связи.

Сигналы классифицируют по многочисленным признакам. По происхождению сигналы делятся на естественные, которые не связанны с

деятельностью человека и искусственные, являющиеся результатом прямой или косвенной деятельности человека.

При математическом моделировании сигнал, независимо от его физической природы, заменяется математической моделью в виде функции одного или нескольких аргументов.

По пространственно-временным свойствам математические модели сигналов подразделяются на:

непрерывные (непрерывная функция непрерывного аргумента);

непрерывно-дискретные (непрерывная функция дискретного аргумента);

дискретно-непрерывные (дискретная функция непрерывного аргумента);

дискретные (дискретная функция дискретного аргумента).

По степени знания исходных (априорных) данных о сигнале сигналы делятся на:

детерминированные (регулярные), которые подразделяются на периодические и непериодические;

случайные (стохастические), подразделяющиеся на стационарные, нестационарные, с априорной неопределенностью (параметрической и непараметрической неопределенностью).

Изучением сигналов в явном виде занимается теория сигналов, которая рассматривает методы построения математических моделей сигналов, а также изучает свойства и характеристики сигналов.

Знаки делятся на символы и диакритики. Форма символа отражает его значение. Например, стрелка как указатель направления. Диакритики не имеют прямой связи между формой и значением или такая связь утеряна в результате переноса его значения на другие сущности или в результате изменения формы знаков. Диакритиками являются знаки математических операции, буквы русского языка и т.д.

Различают знаковые системы (естественные и искусственные языки) и внесистемные знаки (остатки некогда существовавших систем; знаки, созданные временно в небольших коллективах людей; междометия, восклицания, жесты). Диакритики обычно входят в состав знаковых систем, а знаки-одиночки, как правило, являются символами. Имеются и исключения. Например, система знаков дорожного движения.

Изучением знаков и знаковых систем во всех их проявлениях занимается **семиотика** (основатель семиотики Ч. Пирс), которая имеет следующие направления:

Биосемиотика, посвящена изучению знаковой коммуникации в животном мире (танцы пчел, поведенческие акты животных).

Этносемиотика изучает этнознаки в человеческом обществе (жесты, позы, ритуалы, обряды).

Лингвосемиотика изучает человеческие языки, рассматриваемые как знаковые системы.

Семиотика делится на три области: синтактику семантику и прагматику (полезность и ценность знаков).

Синтактика знаковых систем занимается изучением их структуры и правил соединений отдельных знаков.

Семантика рассматривает отношения между знаками и тем, что они обозначают (замещают). Для уяснения этих отношений обратимся к семантическому треугольнику (треугольнику Фреге), см. рис. 3.1

Рис. 3.1. Треугольник Фреге.

Денотатом принято называть то, заместителем чего является знак, а представление, которое вызвано этим знаком, называют концептом или смыслом. Существуют знаки, которые имеют концепт, но не имеют денотата, например флогистон, баба Яга.

3.3. Информационные подходы, меры и теории

Рассмотрим основные информационные подходы к описанию информационных явлений и процессов, развитие которых было инициировано возникновением теории информации, связанной с именем Клода Шеннона. Эту теорию иногда называют шенноновской, классической или статистической теорией информации. Она по существу сыграло роль катализатора, стимулировавшего исследования различных аспектов феномена информации, которые образовали три, возрастающих по сложности формализации, направления: синтаксическое, семантическое и прагматическое. Сущность этих направлений (аспектов) рассмотрена выше. Основные направления и меры информации приведены в табл. 3.1.

Таблица 3.1. Основные направления и меры информации

Направление	Мера информации
Синтаксическое	Хартли, Шеннона, Колмогорова, Кульбака, Реньи
Семантическое	Карнапа и Бар-Хиллела, Шрейдера
Прагматическое	Харкевича, Бонгарда, Стратоновича

Р.Хартли в 1928 году заложил первый кирпич в «здание» теории информации, введя логарифмическую меру количества информации для равновероятных событий. Основу здания теории информации построил в 1948г. Клод Шеннон. В создании теории информации, которая сыграла выдающуюся роль в развитии теории и практики информационных систем значительный вклад сделали и русские ученые А.Н. Колмогоров, А.Я. Хинчин и др. Основные положения теории информации будут рассмотрены в следующем 4 разделе.

Информационные меры С. Кульбака и А. Реньи, связаны с мерой Шеннона и в основном используются в математической статистике. Мера А.Н. Колмогорова (колмогоровская сложность), величайшего русского математика (создателя современной теории вероятностей), предназначена для оценки сложности объекта. Сложность некоторого объекта оценивается как минимальная длина программы, которая необходима для повторного воспроизведения этого объекта.

Семантическая мера информации Р. Карнапа и У. Бар-Хиллела является одной из первых попыток оценки смыслового содержания информации. Сущность этой меры сводится к вычислению логарифма вероятности истинности оцениваемого предложения ($\inf(s) = -\log p(s)$, где $p(s)$ — вероятность истинности предложения s).

Более широкое признание получил подход А. Шрейдера, в котором мера количества семантической информации оценивается как степень изменения тезауруса получателя при приеме некоторого сообщения. Тезаурус — это словарь, в котором указаны семантические отношения между лексическими единицами.

Одним из первых меру ценности информации, как логарифм отношения вероятности достижения цели после получения информации к вероятности достижения цели до получения информации, ввел А. Харкевич. Связь между полезностью информации и мерой шенновского количества установил М. Бонгард. Р. Стратонович установил взаимосвязь между шенноновской теорией информации и теорией статистических решений. Ценность информации по Стратоновичу определяется как максимальная польза, которую данное количество информации может принести в деле уменьшения среднего риска.

Тема 4. Основы классической теории информации

4.1. Количество информации при конечном числе равновозможных исходов. Мера Хартли

Рассмотрим три примера, каждый из которых имеет доопытную (*априорную*) неопределенность относительно возможного исхода опыта. В результате проведения эксперимента (опыта) послеопытная (*апостериорная*) неопределенность либо полностью устраняется (примеры 1 и 2), либо устраняется частично (пример 3). Интуитивно ясно, что снятие неопределенности связано с получением некоторого количества информации. Возникает вопрос, какое количество информации мы получаем в результате того или иного опыта? Рассматривая приведенные ниже примеры в такой постановке вопроса, нам предстоит оказаться в позиции первооткрывателей и попутно уяснить различие между проблемой и задачей.

Пример 1. Монета. Число равновозможных априорных исходов $n=2$, а апостериорных (послеопытных) — $n_c=1$. Для вероятностей событий соответственно имеем $p=\frac{1}{2}$, $p_c=1$.

Пример 2. Игральная кость. Число равновозможных априорных исходов $n=6$, а апостериорных — $n_c=1$. Для вероятностей событий соответственно имеем $p=\frac{1}{6}$, $p_c=1$.

Пример 3. Деталь. Нас интересует диаметр детали с точностью до 1 мм. До опыта о диаметре детали нам известно, что он составляет 125 ± 25 мм, а после опыта — 125 ± 5 мм. Будем считать размеры диаметра детали равновозможными и аналогично предыдущим примерам запишем, $n=51$, $n_c=11$; $p=\frac{1}{51}$, $p_c=\frac{1}{11}$.

Необходимо определить количество информации, получаемое в результате проведения опыта в примерах 1-3. Перед нами проблема, имеющая множество альтернативных решений. В подобной ситуации находились и создатели теории информации.

Попробуем сконструировать количественную меру информации. Первое что приходит в голову это взять отношение числа априорных событий к числу апостериорным событиям. Вроде бы не плохо, но тут же, обнаруживаем принципиальный недостаток: достоверному событию ($p=1$) соответствует единица информации. Значит, этот вариант не подходит. Попробуйте, не читая дальше, рассмотреть другие варианты это поможет более глубоко усвоить основные положения теории информации.

Классическая теория информации построена на основе следующей количественной меры информации

$$I(x_i)=k\cdot\log_a\frac{p_C(x_i)}{p(x_i)}, \qquad (4.1)$$

где $p_c(x_i)$, $p(x_i)$ — соответственно, апостериорная и априорная вероятность i - го события. Коэффициент *k* обычно полагают равным единице.

Основание логарифма a может принимать различные значения. На практике в основном используют a=2. В этом случае единицей измерения информации будет двоичная единица (бит). В теоретических исследованиях (для непрерывных распределений) используют также $a=e$. Ниже будем использовать a=2, при этом выражение (4.1) примет вид

$$I(x_i)=\log_2\frac{p_C(x_i)}{p(x_i)}\ \text{бит.} \qquad (4.2)$$

Теперь наша проблема превратилась в тривиальную задачу. Вычислим, используя (4.2), количество информации для наших примеров.

Пример 1. $I=\log_2\frac{p_C}{p}=\log\frac{1}{0.5}=\log_2 2=1$ бит. Этот пример позволяет уяснить смысл единицы измерения информации. Получение одной двоичной единицы

информации соответствует тому, что мы узнаем, какое из двух равновероятных событий произошло, или какая из двух равновероятных гипотез справедлива.

Пример 2. $I = \log_2 \frac{1}{1/6} = \log_2 6 \approx 2{,}6$ бит.

Пример 3. $I = \log_2 \frac{1/11}{1/51} = \log_2 \frac{51}{11} \approx 2{,}2$ бит.

Обобщим примеры 1 и 2 на случай любого n. Имеем: число равновозможных априорных исходов n, а апостериорных — $n_c = 1$. Для вероятностей событий соответственно имеем $p = \frac{1}{n}$, $p_c = 1$. Подставим эти данные в формулу (4.2) $I = \log_2 \frac{1}{1/n} = \log_2 n$.

Мы получили знаменитую формулу Хартли (1928 г.):

$$I = \log_2 n . \qquad (4.3)$$

Эта формула позволяет определить количество информации в случае, если события равновероятны. Именно поэтому мы опустили аргументы при вычислениях. Возникает вопрос, а как определить количество информации для событий с произвольными вероятностями? Эта задача была решена Клодом Шенноном спустя 20 лет.

4.2. **Количество информации как случайная величина. Энтропия**

Рассмотрим дискретную случайную величину X, закон распределения вероятностей которой задан следующей таблицей (рядом распределения)

$$X = \begin{pmatrix} x_1 & x_2 & \dots & x_i & \dots & x_n \\ p(x_1) & p(x_2) & \dots & p(x_i) & \dots & p(x_n) \end{pmatrix}. \qquad (4.4)$$

Здесь $p(x_i)$ вероятность события x_i, $\sum_{i=1}^{n} p(x_i) = 1$. Запишем выражение для математического ожидание случайной величины X (оно нам понадобится)

$$M(X) = \sum_{i=1}^{n} x_i p(x_i) . \qquad (4.5)$$

Определим, используя формулу (4.2), количество информации, доставляемое произвольным событием x_i (напомним, что в данном случае $p_c(x_i) = 1$)

$$I(x_i) = \log_2 \frac{P_C(x_i)}{P(x_i)} = -\log_2 p(x_i) \text{ бит.} \qquad (4.6)$$

А дальше, поскольку $p(x_i)$ может принимать любые значения на интервале (0,1), мы оказываемся в тупике. Дело в том, что в такой модели $I(x_i)$ принимает случайные, не предсказуемые значения, а наука, как известно, имеет дело только с

закономерностями. Выход из тупика был найден Клодом Шенноном. Ниже мы сделаем то же, но более простым способом.

Учитывая, что значения $I(x_i)$ в рассматриваемой схеме случайны, сформируем новую дискретную случайную величину I, ряд распределения которой имеет следующий вид

$$I = \begin{pmatrix} I(x_1) & I(x_2) & \dots & I(x_i) & \dots & I(x_n) \\ p(x_1) & p(x_2) & \dots & p(x_i) & \dots & p(x_n) \end{pmatrix}. \qquad (4.7)$$

Используя формулу (4.5) запишем выражение для математического ожидания случайной величины I

$$M[I(x_i)] = \sum_{i=1}^{n} [-\log_2 p(x_i)] \cdot p(x_i) = -\sum_{i=1}^{n} p(x_i) \cdot \log_2 p(x_i).$$

Введем для этого математического ожидания специальное обозначение и запишем последнее выражение в виде

$$H(X) = = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i) \text{ бит/опыт.} \qquad (4.8)$$

Мы получили знаменитую формулу шенноновской энтропии. В данном случае $H(X)$ — это энтропия дискретной случайной величины X, которая является математическим ожиданием количества информации (4.6) и характеризует степень неопределенности (непредсказуемости) этой случайной величины.

Обсудим интерпретацию $H(X)$. Предположим, что к обыкновенной батарейке подключен амперметр. Стрелка стоит как вкопанная. А почему? Да потому, что амперметр, как инерционный прибор, усредняет количество электронов пробегающих по проводнику, т.е., упрощенно говоря, он показывает математическое ожидание (среднее значение) количества пробегающих электронов. А теперь представьте, что по «абстрактному проводнику» бегут биты и стоит такой же инерционный, усредняющий их, «прибор». Так вот показанием этого прибора и будет энтропия. В каждом конкретном опыте количество получаемой информации может быть больше или меньше $H(X)$, но статистическое среднее при большом числе опытов будет близко к значению $H(X)$.

И еще один важный момент. Обратите внимание на первые строки рядов распределения случайных величин X и I. Размерность первой строки ряда распределения X, как модели некоторого реального объекта, может быть какой угодно (безразмерной, кг, см. градусы и т.д.), в то время как размерность первой строки I всегда одна и та же — биты. Это важнейшее, можно сказать революционное, свойство, позволяющее абстрагироваться от физической сущности исследуемого объекта и рассматривать только свойства его неопределенности.

В качестве примера определим энтропию $H(X)$ случайной величины X, закон распределения вероятностей которой задан следующим рядом распределения

$$X = \begin{pmatrix} x_1 & x_2 \\ p & 1-p \end{pmatrix}.$$

Используя (4.8) запишем $H(X) = -p\log_2 p - (1-p)\log_2(1-p)$ и построим график зависимости $H(X)$ от вероятности p

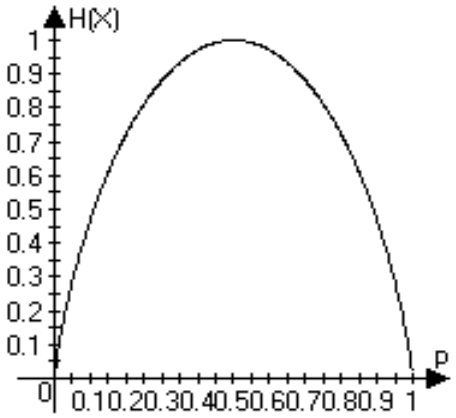

Рис. 4.1. График зависимости $H(X)$ от вероятности p.

Заметим, что характер зависимости $H(X)$ от p хорошо согласуется с нашими интуитивными представлениями о неопределенности возможных исходов альтернативных событий. Максимум $H(X)$=1 при p=0,5, т.е. в случае, если события равновероятны, и мы не можем отдать предпочтение ни одному из них. И, наоборот, при малых и близких к единице значениях p энтропия $H(X)$ мала, хотя количество информации в отдельном опыте может быть большим. Так, например, если вероятность события x_1 составляет $2^{-20} \approx 10^{-6}$ и это событие произойдет, то полученное количество информации будет составлять 20 бит. Такие события в обыденной жизни мы рассматриваем как сенсацию.

Примечание. Отмеченные свойства $H(X)$ не случайны. Дело в том, что при конструировании меры неопределенности Шеннон предъявил к свойствам $H(X)$ определенные требования и посредством соответствующей теоремы доказал, что (4.8) является единственной функцией (в нашем случае с точностью до множителя К), удовлетворяющей предъявленным требованиям. А то обстоятельство что $H(X)$ совпала с математическим ожиданием от количества информации, существенно упростило знакомство с этой важнейшей функцией.

4.3. **Основные свойства энтропии**

Рассмотрим основные свойства энтропии дискретных случайных величин.

1. Энтропия одномерной дискретной случайной величины X может принимать только неотрицательные значения, т.е. $H(X) \geq 0$. Равенство нулю имеет место тогда и только тогда, когда все вероятности случайной величины X, за исключением одной, равны нулю, а эта единственная — равна единице. В остальных случаях $H(X) > 0$. Последнее следует из того, что для всех $i = 1..n$ имеет место $-p(x_i) \cdot \log_2 p(x_i) > 0$.

2. Энтропия одномерной дискретной случайной величины X максимальна и равна

$$H_{\max}(X) = \log_2 n, \qquad (4.9)$$

(где n – число событий данной случайной величины X) только в том случае, если события этой случайной величины равновероятны. Заметим, что правая часть (4.9) есть ни что иное как формула Хартли. Данное свойство доказывается методом неопределенных множителей Лагранжа (для n=2 см. рис 4.1).

Обобщим свойство (4.9). Для n-мерной дискретной случайной величины $(X_1, X_2,..., X_n)$, у которой компоненты и события независимы, имеет место

$$H_{\max}(X_1, X_2,..., X_n) = \log_2(n_1 * n_2 * ... * n_n) = \log_2 N, \qquad (4.10)$$

где $n_1, n_2,..., n_n$ — число событий случайных величин $X_1, X_2,..., X_n$, соответственно.

Это свойство мы уже использовали при рассмотрении основных аспектов сложной системы. Значимость обсуждаемого свойства для теории и практики трудно переоценить. Выражение (4.9) или (4.10), задает верхнюю границу неопределенности, что позволят без особых усилий оценить на постановочном (рекогносцировочном) этапе сложность и реализуемость задачи (проблемы) и при необходимости редуцировать (упростить) задачу.

3. Энтропия двумерной дискретной случайной величины (X,Y) удовлетворяет следующему неравенству

$$H(X,Y) \le H(X) + H(Y), \qquad (4.11)$$

где $H(X,Y)$ — энтропия двумерной случайной величины (X,Y), определяемая по формуле

$$H(X,Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)\log_2 p(x_i, y_j). \qquad (4.12)$$

Равенство в (4.11) имеет место только если случайные величины X и У независимы, т.е. если $p(x_i, y_j) = p(x_i)p(y_j)$.

Обобщим свойство (4.11). Для n-мерной дискретной случайной величины $(X_1, X_2,..., X_n)$ имеет место

$$H(X_1, X_2,..., X_n) \le H(X_1) + H(X_2) + ... + H(X_n), \qquad (4.13)$$

где $H(X_1, X_2,..., X_n) = -\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}...\sum_{l=1}^{n_n} p(x_i^1, x_j^2,..., x_l^n)\log_2 p(x_i^1, x_j^2,..., x_l^n)$.

Равенство в (4.13) имеет место, только если случайные величины $X_1, X_2,..., X_n$ независимы.

Сущность этого свойства прояснится при рассмотрении следующего свойства энтропии. Здесь же отметим, что если между компонентами X и Y имеется статистическая зависимость, то знание результата опыта для одной компоненты снижает неопределенность по второй компоненте. К примеру, если вес и диаметр арбуза представить в дискретном виде (например, в килограммах и сантиметрах), то, зная вес арбуза (например, 10 кг), можно составить довольно хорошее представление о его диаметре.

4. Для энтропии двумерной дискретной случайной величины (X,Y) справедливы следующие равенства:

$$H(X,Y) = H(X) + H(Y \mid X) \qquad (4.14)$$

$$H(X,Y) = H(Y) + H(X \mid Y). \qquad (4.15)$$

Здесь $H(X) == -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i)$, $H(Y \mid X) = -\sum_{i=1}^{n} p(x_i)\sum_{j=1}^{m} p(y_j \mid x_i)\log_2 p(y_j \mid x_i)$;

$H(Y) == -\sum_{j=1}^{m} p(y_j)\log_2 p(y_y)$, $H(X \mid Y) = -\sum_{j=1}^{m} p(y_j)\sum_{i=1}^{n} p(x_i \mid y_j)\log_2 p(x_i \mid y_j)$.

Прокомментируем это свойство на примере равенства (4.15). Первое слагаемое правой части (4.15) представляет собой энтропию случайной величины Y, второе слагаемое: $H(X \mid Y)$ называется условной энтропией случайной величины X и характеризует среднюю неопределенность X при известной величине У . Остановимся на этом более подробно. Представим условную энтропию $H(X \mid Y)$ в виде

$$H(X \mid Y) = -\sum_{j=1}^{m} p(y_j)\sum_{i=1}^{n} p(x_i \mid y_j)\log p(x_i \mid y_j) = \sum_{j=1}^{m} p(y_j)H(X \mid y_j), \qquad (4.16)$$

где $H(X \mid y_j) = -\sum_{j=1}^{m} p(x_i \mid y_j)\log p(x_i \mid y_j)$ — частная условная энтропия случайной величины X. Эта величина характеризует неопределенность X при конкретном значении y_j. Вспомним пример с арбузом и предположим, что Y характеризует вес арбуза. Тогда, например, при y_j=10 кг $H(X \mid y_j)$ будет характеризовать неопределенность диаметра 10 килограммового арбуза, но поскольку и вес арбуза заранее неизвестен, то необходимо выполнить усреднение по всем возможным значениям веса арбуза, (см. правую часть (4.16)).

В заключение заметим, что формулы (4.14), (4.15) легко получить путем элементарных преобразований выражения (4.12) после подстановки в него соотношений $p(x_i, y_j) = p(x_i)p(y_J \mid x_i)$, $p(x_i, y_j) = p(y_j)p(x_i \mid y_j)$ соответственно.

4.4. Взаимная информация (дискретный случай)

В реальной жизни об интересующих нас событиях X нам сплошь и рядом приходится судить не непосредственно, а косвенно по связанным с X событиям У, что обусловлено либо невозможностью прямого измерения физической величины (например, температура, напряжение и т.д.), либо пространственной или физической недоступностью объекта. Очевидно, что часть информации об интересующих нас событиях при таком способе получения информации теряется. Возникает вопрос: как оценить в данном случае количество получаемой информации? Решение этого вопроса позволит не только оценивать количество получаемой информации, но и решать задачи, связанные с оптимизацией средств доставляющих информацию.

Рассмотрим эту задачу в следующей формальной постановке. Пусть дискретная случайная величина X, заданная рядом распределения

$$X = \begin{pmatrix} x_1 & x_2 & \dots & x_i & \dots & x_n \\ p(x_1) & p(x_2) & \dots & p(x_i) & \dots & p(x_n) \end{pmatrix},$$

описывает интересующие нас события (X ненаблюдаемая величина) и пусть связанная с X случайная величина Y (наблюдаемая величина)

$$Y = \begin{pmatrix} y_1 & y_2 & \dots & y_j & \dots & y_m \\ p(y_1) & p(y_2) & \dots & p(y_j) & \dots & p(y_m) \end{pmatrix}$$

характеризует события, по которым мы судим о том, какое из событий X имело место. Предположим также, что нам известны вероятностные характеристики взаимосвязи X и У.

В такой постановке необходимо определить количество информации, которое содержится в событиях Y относительно интересующих нас событий X.

Предположим, что имело место событие y_j, и определим, используя формулу (4.2), количество информации, которое содержится в событии y_j относительно некоторого события x_i

$$I(y_j \to x_i) = \log_2 \frac{p_C(x_i)}{p(x_i)} = \log_2 \frac{p(x_i \mid y_j)}{p(x_i)},$$

Здесь $p(x_i \mid y_j)$ — условная вероятность события x_i при имевшем место событии y_j, она то и характеризует связь между результатом наблюдения и истинным, непосредственно не наблюдаемым событием.

Очевидно, что величина $I(y_j \to x_i)$ для различных событий y_j и x_i будет принимать случайные значения, поэтому найдем для нее математическое ожидание

$$M\{I(y_j \to x_i)\} = \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j) \log_2 \frac{p(x_i \mid y_j)}{p(x_i)}.$$

Введем для этого выражения специальное обозначение

$$I(Y \to X) = \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j) \log_2 \frac{p(x_i \mid y_j)}{p(x_i)}. \qquad (4.17)$$

Мы получили фундаментальное соотношение теории информации, значимость которого трудно переоценить. Заметим, что на основе этого выражения построена не одна прикладная теория.

В чем притягательность (4.16)? Если предположить, что соотношение (4.17) связанно с параметрами средств доставляющих информацию, то весьма привлекательной является вариационная задача на максимум $I(Y \to X)$.Отметим также, что методы и аппарат шенноновской теории информации оказались настолько плодотворными, что были востребованы не только в прикладных областях, но и усилиями А.Н. Колмогорова и А.Я. Хинчина стали одним из разделов теории вероятностей, а также оказали существенное влияние на развитие методов математической статистики.

Воспользуемся свойством логарифмов и представим (4.17) в виде

$$I(Y \to X) = H(X) - H(X \mid Y), \qquad (4.18)$$

где $H(X) == -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i)$, $H(X \mid Y) = -\sum_{j=1}^{m} p(y_j)\sum_{i=1}^{n} p(x_i \mid y_j)\log_2 p(x_i \mid y_j)$.

Здесь энтропия $H(X)$ характеризует неопределенность интересующей нас, но непосредственно не наблюдаемой, случайной величины X; в то время как условная энтропия $H(X \mid Y)$ характеризует остаточную неопределенность, которая обусловлена действием помех или несовершенством средства наблюдения.

Умножим числитель и знаменатель подлогарифмического выражения в (4.17) на $p(y_j)$ и, учитывая что $p(x_i, y_j) = p(y_j)p(x_i \mid y_j)$, запишем (4.17) в виде

$$I(Y \to X) = \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)\log_2 \frac{p(x_i, y_j)}{p(x_i)p(y_j)} \qquad (4.19)$$

Мы получили каноническую форму основного соотношения теории информации. Это простое и красивое выражение позволяет путем элементарных преобразований получить любую из приведенных выше формул. Воспользуемся свойством логарифмов и представим (4.19) в виде

$$I(Y \to X) = H(X) + H(Y) - H(X,Y).$$

Если в (4.19) подставить $p(x_i, y_j) = p(x_i)p(y_J \mid x_i)$, то после элементарных преобразований будем иметь

$$I(Y \to X) = \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)\log_2 \frac{p(y_j \mid x_i)}{p(y_j)},$$

$$I(Y \to X) = H(X) - H(X \mid Y). \qquad (4.20)$$

Сравнивая (4.18) и (4.20) замечаем, что $I(Y \to X) = I(X \to Y)$. Именно поэтому $I(Y \to X)$ называют взаимной информацией и иногда обозначают так: $I(Y \leftrightarrow X)$ или так $I(Y; X)$.

Основные свойства $I(Y \to X)$ можно представить следующим неравенством

$$0 \le I(Y \to X) \le H(X). \qquad (4.21)$$

В этом неравенстве $I(Y \to X) = 0$ только в том случае, если X и Y независимы, а $I(Y \to X) = H(X)$ в случае, если события Y достоверно определяют события X.

4.5. Взаимная информация (непрерывный случай)

Рассмотрим ту же, что и в п. 4.4 задачу, но в предположении, что случайные величины X и Y являются непрерывными и описываются совместной плотностью распределения вероятностей $f(x, y)$.

Эта задача решается аппроксимацией плотности $f(x, y)$ путем дискретизации областей определения случайных величин X и **Y** с последующим использованием соотношения (4.17) и выполнением хорошо известного в математике предельного перехода. Опуская названные операции, запишем аналог выражения (4.16) для данного случая

$$I(Y \to X) = \int_{L_X}\int_{L_Y} f(x,y)\log_2 \frac{f(x \mid y)}{f(x)}\,dxdy\,. \qquad (4.17^*)$$

Запишем аналоги формул, приведенных в п. 4.4.

$$I(Y \to X) = H(X) - H(X \mid Y); \qquad (4.18^*)$$

$$H(X) == -\int_{L_X} f(x)\log_2 f(x)dx\,, \quad H(X \mid Y) = -\int_{L_X} f(x)\int_{L_Y} f(x \mid y)\log_2 f(x \mid y)dxdy\,.$$

Здесь $H(X)$— дифференциальная энтропия, а $H(X \mid Y)$ — условная дифференциальная энтропия случайной величины X.

Каноническая форма основного соотношения теории информации для непрерывного случая получается аналогично (4.19), но вместо $p(y_j)$ необходимо использовать $f(y)$:

$$I(Y \to X) = \int_{L_X}\int_{L_Y} f(x,y)\ \log_2 \frac{f(x,y)}{f(x)f(y_j)}\,dxdy. \qquad (4.19^*)$$

Используя свойства логарифмов для (4.19*) запишем

$$I(Y \to X) = H(X) + H(Y) - H(X,Y),$$

где $H(X,Y) = -\int_{L_X}\int_{L_Y} f(x,y)\ \log_2 f(x,y)dxdy$— дифференциальная энтропия двумерной случайной величины (X,У).

Свойства же $I(Y \to X)$ в данном случае обличаются от свойств $I(Y \to X)$ для дискретных распределений. Справедливой оказывается только левая часть неравенства (4.21), т.е. $I(Y \to X) \ge 0$, в то время как, входящая в правую часть (4.21), дифференциальная энтропия $H(X)$ может принимать даже отрицательные значения. Приведем пример. Пусть непрерывная случайная величина X распределена по равномерному закону на интервале длиной $L_X = 10$ см (a=10, b=20). Определим дифференциальную энтропию $H(X)$ этой случайной величины

$$H(X) == -\int_{L_X} f(x)\log_2 f(x)dx\ = -\int_{10}^{20}\frac{1}{b-a}\log_2\frac{1}{b-a}dx = -\int_{10}^{20}\frac{1}{L_X}\log_2\frac{1}{L_X}dx = \log_2 L_X = \log_2 10\,.$$

Как видно в данном случае $H(X) = \log_2 10 > 0$, но если L_X измерять в метрах ($L_X = 0{,}1$ м), то будем иметь $H(X) = \log_2 0{,}1 < 0$, т.е. $H(X)$ будет отрицательной. Отличаются и другие свойства дифференциальной энтропии, в частности, аналогом формулы Хартли для всех законов заданных на одном и том же отрезке является именно дифференциальная энтропия равномерного закона, но если ввести ограничение по дисперсии, то максимальной дифференциальной энтропией среди всех непрерывных законов распределения будет обладать нормальный закон.

4.6. Информационные характеристики источников дискретных сообщений

4.6.1. Энтропия источника дискретных сообщений

Под источником дискретных сообщений будем понимать некоторый объект, который случайным образом вырабатывает сообщения x_i из конечного набора сообщений n. Далее сообщения x_i будем называть буквами, а набор различающихся сообщений n — объемом алфавита этого источника. Необходимо оценить энтропию источника дискретных сообщений.

Простейшей моделью такого источника является дискретная случайная величина X, заданная рядом распределения вероятностей

$$X = \begin{pmatrix} x_1 & x_2 & \dots & x_i & \dots & x_n \\ p(x_1) & p(x_2) & \dots & p(x_i) & \dots & p(x_n) \end{pmatrix},$$

энтропия, которой вычисляется по формуле

$$H(X) == -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i).$$

Однако оценка энтропии по этой формуле может оказаться существенно завышенной, так что практическое ее использование может привести к большим и неоправданным материальным затратам. Дело в том, что буквы источника могут иметь определенную взаимосвязь, так, например, в русском языке мы практически не встречаем два подряд идущих твердых знака и т.д. Оценка влияния не величину энтропии таких взаимосвязей и составляет предмет дальнейшего рассмотрения**.**

Предположим, что вероятностные характеристики источника дискретных сообщений не зависят от времени (т.е. источник является стационарным) и рассмотрим общий случай. Очевидно, что в таком общем случае вероятность очередной буквы источника x_i зависит от ряда предшествующих букв $\dots x_h x_l x_j$ и поэтому она является условной вероятностью $p(x_i \mid x_j, x_l, x_h, \dots)$. О таком источнике говорят, что он обладает памятью (память источника характеризуют так называемой связностью, обозначаемой буквой М).

Рассмотрим частные случаи. Если буквы источника не зависимы, то $p(x \mid x_j, x_l, x_h, \dots) = p(x_i)$ и мы имеем рассмотренный выше случай. Это источник без памяти, связность которого М=0. Если предположить, что очередная буква зависит только от одной предшествующей, то условная вероятность примет вид $p(x_i \mid x_j, x_l, x_h, \dots) = p(x_i \mid x_j)$, М=1; если же очередная буква зависит от двух предшествующих, то будем иметь $p(x_i \mid x_j, x_l, x_h, \dots) = p(x_i \mid x_j, x_l)$, М=2 и т.д.

Теперь определим, какое количество предшествующих букв при М=1 могут оказать влияние на вероятность появления буквы x_i. Очевидно, что в общем случае такое влияние могут оказать все буквы алфавита. Так что в данном случае

мы имеем n различающихся состояний (характерных состояний), только от которых зависит вероятность появления буквы x_i. Определим число характерных состояний для M=2: $n \times n = n^2$. Что это означает в плане построения вероятностной модели для такого случая? Это значит, что нам необходимо иметь не только n вероятностей букв алфавита, но и n^2 условных вероятностей для каждой буквы источника. Заметим, что при M= k число характерных состояний будет составлять n^k. Чтобы представить сложность вычисления энтропии источника при больших значениях k отметим, что если в качестве источника дискретных сообщений рассматривать русский язык , то при M=70 только на запись условных вероятностей в память компьютера потребуется энергия эквивалентная массе ядерного топлива, превышающей массу нашей Галактики (см. п. 1.2).

Энтропия источника дискретных сообщений в установившемся состоянии определяется по формуле

$$H(X) = -\sum_{K} p(S_K) \sum_{l/K} p(S_l \mid S_K) \log_2 p(S_l \mid S_K), \qquad (4.22)$$

где $p(S_K)$ — вероятность пребывания источника в состоянии K,

$p(S_L \mid S_K)$ — условная вероятность перехода источника из состояния K в состояние L.

Формула (4.22) справедлива для любого M. Рассмотрим частные случаи.

Для случая n =5 и M=1 формула (4.20) принимает вид

$$H(X) = -\sum_{k=1}^{5} p(x_k) \sum_{l=1}^{5} p(x_l \mid x_k) \log_2 p(x_l \mid x_k),$$

где $p(S_K) = p(x_k)$ — вероятность пребывания источника в состоянии k,

$p(S_L \mid S_K) = p(x_l \mid x_k)$ — условная вероятность перехода источника из состояния k в состояние l.

Для случая n =5 и M=2 имеем

$$H(X) = -\sum_{k=1}^{5} \sum_{j=1}^{5} p(x_k, x_j) \sum_{l=1}^{5} p(x_l \mid x_k, x_j) \log_2 p(x_l \mid x_k, x_j),$$

где $p(S_K) = p(x_k, x_j)$ — вероятность пребывания источника в состоянии K, которое в данном случае определяется двумя ранее полученными буквами x_k, x_j,

$p(S_L \mid S_K) = p(x_l \mid x_k, x_j)$ — условная вероятность перехода источника из состояния К в состояние l.

4.6.2. Понятие избыточности источника сообщений

Для оценки эффективности функционирования источника (в смысле среднего количества информации, вырабатываемого источником на одну букву) используют понятие избыточности. Избыточность источника дискретных сообщений определяется по формуле

$$R = \frac{H_{\max}(X) - H(X)}{H_{\max}(X)}, \qquad (4.23)$$

где $H_{\max}(X) = \log_2 n$ — максимальная энтропия источника дискретных сообщений, использующего n различающих сообщений,

$H(X)$ — энтропия этого источника дискретных сообщений.

Нетрудно заметить, что R принимает значения на отрезке [0,1]. Минимальное значение $R = 0$ имеет место при $H(X) = H_{\max}(X)$. В данном случае, как нам известно, все сообщения источника должны быть равновероятными. Если $R = 1$, то $H(X) = 0$ и источник вообще никакой информации не вырабатывает.

Возникает вопрос. Так что, избыточность — это плохо? Плохо при передаче информации, но, как правило, она (избыточность) неизбежна. Остановимся на этом более подробно.

Мы уже сталкивались с понятием «избыточности» при обсуждении свойств сложной системы, отмечая, что возникновение существенных устойчивых связей приводит к ограничению степени свободы элементов системы, и это, естественно, приводит к снижению энтропии системы и, следовательно, к повышению избыточности? Нет, в данном случае к повышению структурированности и возникновению эмерджентных свойств. Устранить такую избыточность (структурированность) системы не разрушив ее нельзя. Но избыточность накладна при передаче и хранении информации, так как в первом случае возрастает время передачи, а во втором — объем требуемой памяти. Как с этим бороться, по-видимому, впервые додумался Морзе, но и Вы постоянно пользуетесь этим способом. Скажите, как Вы размещаете наиболее часто используемые при работе инструменты, или наиболее часто используемые вещи? Подумайте, чем обусловлена частота использования инструментов или вещей и можно ли эту частоту изменить?

В современных информационных технологиях борьба с избыточностью информации осуществляется посредством сжатия данных и при этом для обеспечения требуемой надежности передачи (хранения) данных вводится специальным образом организованная избыточность, направленная на борьбу с помехами. Такая избыточность является платой за новое эмерджентное свойство: высокая достоверность передачи (хранения) данных.

4.6.3. Понятие скорости создания информации источником дискретных сообщений

Запишем формулу для энтропии источника дискретных сообщений (см. п. 4.6.1):

$$H(X) = -\sum_K p(S_K) \sum_{l/K} p(S_l \mid S_K) \log_2 p(S_l \mid S_K) \ \textit{бит/буква}$$

и обратим внимание на то обстоятельство, что два источника с одинаковой энтропией могут различаться объемом передаваемой информации за один и тот же промежуток времени. Это приводит к необходимости использования еще одной характеристики источника сообщений: скорости создания информации, называемой также потоком информации источника сообщений.

Скорость создания информации источником дискретных сообщений определяется по формуле

$$\overline{H}(X) = \frac{H(X)}{\tau_{и}} \quad \textit{бит/сек,} \qquad (4.24)$$

где $\tau_{и} = \sum_{K} p(S_K) \sum_{l/K} \tau(S_L \mid S_K) p(S_L \mid S_K)$ — математическое ожидание длительности передачи одной буквы источника.

Здесь $p(S_K)$ — вероятность пребывания источника в состоянии *K*,

$\tau\ (S_L \mid S_K)$ — длительность передачи буквы, переводящей источник из состояния *K* в состояние *L*.

$p(S_L \mid S_K)$ — условная вероятность перехода источника из состояния *K* в состояние *L*.

Если длительность передачи всех букв источника одинакова, то $\tau_{и} = \tau$ и (4.24) принимает вид

$$\overline{H}(X) = \frac{1}{\tau} H(X).$$

В заключение перечислим основные характеристики источников дискретных сообщений, рассмотренные в п. 4.6. К таким характеристикам относятся: энтропия, избыточность и скорость создания информации.

4.7. Понятия скорости передачи информации и пропускной способности канала связи

Представьте себе водный канал, который по тем или иным соображениям запрятали в трубу соответствующего диаметра. Возникает вопрос, какой максимальный поток воды (кубометров в секунду) способна пропустить эта труба? Т.е. какова пропускная способность трубы? Очевидно, что пропускная способность трубы будет определяться диаметром трубы, ее прочностью (допустим, будут использоваться мощные насосы), а так же качеством ее внутренней поверхности. Названные параметры трубы являются ограничениями, без которых задача на поиск максимума лишена смысла, так что мы не зря запрятали водный канал в трубу.

Аналогичные вопросы возникают и при передаче информации по каналам связи, которые организуются с использованием различных физических сред (телефонной пары, витой пары, коаксиального кабеля, оптоволокна и т.д.). Введем понятия скорости передачи информации и пропускной способности канала связи для обобщенного канала (рис. 4.2).

Рис. 4.2. Обобщенная схема канала связи.

Обозначим через $I(Y_T \to X_T)$ среднее количество информации (математическое ожидание), содержащееся в сообщении Y_T на выходе канала связи относительно сообщения X_T на входе канала за время T (X_T и Y_T могут быть как дискретными, так и непрерывными). Очевидно, что $I(Y_T \to X_T)$ будет зависеть от вероятностных характеристик источника информации, воздействия помех, а также времени передачи T.

Рассмотрим предел

$$\bar{I}(Y \to X) = \lim_{T\to\infty}\left[\frac{I(Y_T \to X_T)}{T}\right] \text{ бит/сек}, \qquad (4.25)$$

Этот предел не зависит от времени передачи и называется **скоростью передачи информации** (в известном смысле это аналог потока воды). Максимальное же значение (точнее верхняя грань) скорости передачи информации $\bar{I}(Y \to X)$ при заданных ограничениях называется **пропускной способностью** канала связи, формула для которой в общем случае записывается так

$$C = \sup[\bar{I}(Y \to X)] \qquad (4.26)$$

Ограничениями в данном случае являются выделенная для передачи полоса частот, ограничения по мощности передаваемого сигнала и т.д. Очевидно, что чем больше ограничений, тем меньше при прочих равных условиях будет пропускная способность канала связи C.

Для определения пропускной способности реального канала связи необходимо задать вероятностное описание сигнала и помех, а также фиксированные ограничения. В седьмой теме нашего курса мы используем формулу (4.26) для определения пропускной способности канала связи в наиболее простых случаях.

Тема 5. Восприятие информации

Процесс восприятия информации представляет собой целенаправленную последовательность операций, связанных с извлечением и анализом информации о некотором объекте, событии, процессе.

Под восприятием в широком смысле слова понимают процесс формирования целостного образа предметной области. Такая трактовка восприятия близка к пониманию восприятия в психологии и приобретает все большую актуальность в связи с интеллектуализацией роботов как технических, так и роботов реализующих функции интеллектуального интерфейса (почтовые, поисковые и др.).

Процесс восприятия в зависимости от цели, сложности воспринимающего и воспринимаемого объектов и др. особенностей включает определенную совокупность этапов восприятия информации.

Различают следующие этапы восприятия информации:

1. Первичное восприятие информации. На этом этапе осуществляются поиск, измерение, преобразование и необходимая обработка информации.

2. Обнаружение, распознавание образов, анализ сцен (восприятие совокупности трехмерных образов).

3. Морфологический, синтаксический и семантический анализ.

Первичное восприятие связано с использованием различных чувствительных элементов и устройств (рецепторов, сенсоров, приемников): тактильных, локационных, силометрических датчиков; приемников визуальной и звуковой информации и др. Существует отдельный класс сложных систем (так называемых систем извлечения информации), которые связаны с восприятием информации. Такие системы находят применение в радио- и гидролокации, навигации, метеорологии, астрономии, разведке и т.д.

К системам извлечения информации можно отнести и поисковые машины Интернет: Google, AltaVista, Rambler, Aport и др. Здесь мы не будем останавливаться на вопросах технологии поиска информации в Интернет, так как они освящены в лабораторном практикуме курса, там же приведены и дополнительные источники информации, позволяющие более глубоко изучить вопросы поиска и анализа информации.

Рассмотрим обобщенную схему процесса восприятия информации (рис. 5,1)

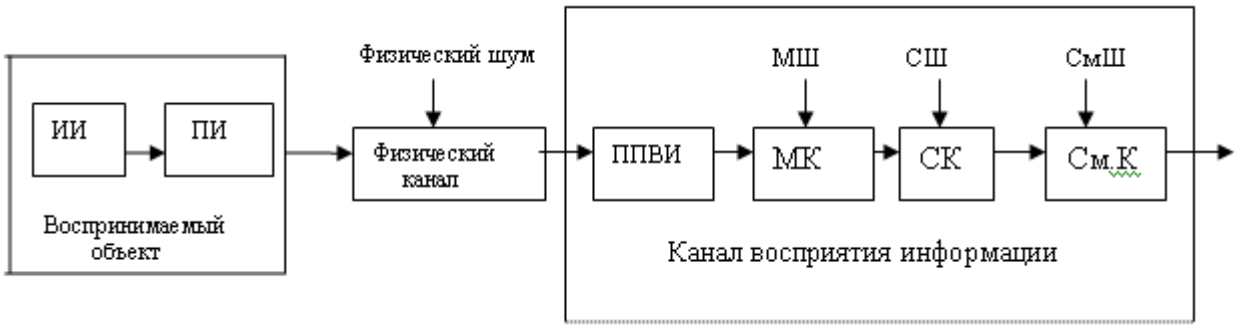

Рис. 5.1. Обобщенная схема процесса восприятия информации: ИИ – источник информации; ПИ – преобразователь информации; ППВИ – преобразователь первичного восприятия информации; МК – морфологический канал; МШ – морфологический шум; СК – синтаксический канал; СШ – синтаксический шум; См.К – семантический канал; См.Ш – семантический шум.

Преобразователь первичного восприятия информации (ППВИ) выполняет физическое преобразование сигнала, а также необходимую обработку информации, включая обнаружение, распознавание (лексический анализ для лингвистического процессора). В процессе лексического анализа информация разбивается на абзацы, предложения, слова. При этом выявляются тип лексических выражений (сленг, бранные слова) и т.д.

В процессе морфологического анализа осуществляется сопоставление каждого слова воспринимаемого текста определенной синтаксического группе (лексико-грамматическому классу).

Сущность синтаксического анализа состоит в построении синтаксической структуры входного предложения на основе морфологической информации и синтаксических правил объединения слов и словосочетаний.

Цель семантического анализа состоит в установлении смысла предложения, т.е. в выявлении и оценке смыслового содержания текста. Семантический анализ является наиболее сложным этапом процесса восприятия, поскольку является трудно формализуемым и поэтому предполагает использование экспертных систем.

Тема 6. Преобразование информации

6.1. Цели и виды преобразования информации

Преобразование информации, различаясь по целям, видам и способам реализации, находит очень широкое применение в современных информационных технологиях. Являясь разновидностью информационных процессов, преобразование информации обычно используется в других информационных процессах (см., например, предыдущую тему) в качестве некоторых промежуточных операции (этапов). В свою очередь преобразование информации предполагает определенную ее обработку, так что провести четкую границу между этими двумя видами процессов не представляется возможным.

Такие промежуточные операции обычно связаны с преобразованием некоторых синтаксических свойств информации и параметров ее носителя (параметров и свойств сигнала, формата данных, формы представления информации и т.д.). По существу эти операции реализуют аналого-цифровое или цифровое преобразование сигналов, но так уж сложилось, что устройства, которые ранее назывались устройствами преобразования сигналов в настоящее время все чаше и чаще называют устройствами преобразования информации (например, устройство преобразования информации СУПИ 54, которое является аналого-цифровым преобразователем)

Заметим, что преобразование информации в узком смысле слова, т.е. преобразование, связанное с изменением семантических и прагматических свойств информации, могут осуществлять только средства обладающие интеллектом.

Основными целями преобразования информации являются:

1. Обеспечение интерфейса, необходимого для реализации информационного взаимодействия различных функциональных элементов информационных систем (см., например, определение интерфейса эталонной модели OSI).
2. Формирование специальных (эмерджентных) свойств сигнала, связанных со спецификой технологического процесса. Наиболее известным примером является модуляция, посредством которой, например, в радиовещании осуществляется разнесение несущих частот радиостанций.
3. Извлечение полезной информации, содержащейся в сигнале. Например, демодуляция, декодирование.
4. Устранение, исходя из требований по качеству, излишних данных, содержащихся в сигнале (редукция данных).
5. Обеспечение специальных свойств, позволяющих повысить скорость передачи, помехоустойчивость, а также обеспечить конфиденциальность.

Основные виды преобразования информации:

1. Функциональное преобразование. При функциональном преобразовании осуществляется изменение характеристик сигнала без потерь полезной информации (усиление, дифференцирование, интегрирование и т.д.). К этому виду преобразования можно отнести также различные виды модуляции гармонических колебаний, импульсной и импульсно-кодовой модуляции. Вопросы модуляции, касающиеся сетевых технологий, рассматриваются в курсе «Аппаратное и программное обеспечение ЭВМ и сетей».

2. Квантование сигнала по уровню.

3. Дискретизация сигнала во времени.

4. Статистическое и помехоустойчивое кодирование. Вопросы этих видов кодирования будут рассмотрены в разделе, посвященном передаче информации.

5. Криптографическое шифрование.

6.2. Модуляция

6.2.1. Основные виды модуляции

Под модуляцией понимают процесс изменения одного или нескольких параметров физического процесса по закону передаваемого сообщения. Так, например, если моделью физического процесса является функция $f(a,b,c,d;t)$, то параметры a, b, c, d можно использовать для осуществления модуляции.

Как правило, физический процесс (несущее колебание) является высокочастотным, а передаваемое сообщение (модулирующее колебание) — низкочастотным. Такое соотношение частот позволяет реализовать одно из важнейших свойств модуляции: управляемый перенос спектра низкочастотного колебания в область высоких частот. На этом принципе основано радиовещание, разделение диапазонов частот в коаксиальных и волоконно-оптических средах. Принцип переноса спектра низкочастотного модулирующего сигнала в область высоких частот на примере амплитудной модуляции рассмотрен в п. 6.2.2.

На практике получили распространение следующие основные виды модуляции: **модуляция гармонических колебаний**, **импульсная модуляция** (ИМ) и **импульсно-кодовая модуляция** (ИКМ). Заметим, что используемых на практике типов модуляции в настоящее время более ста.

6.2.2. Модуляция гармонических колебаний

При модуляции гармонических колебаний по закону передаваемого сообщения $c(t)$ изменяется один из параметров гармонического колебания

$$a(t) = A_0 \cos(\omega_0 t + \theta_0), \qquad (6.1)$$

так что в данном случае возможны три вида модуляции:

амплитудная модуляция (АМ): изменяется пропорционально $c(t)$ амплитуда A гармонического колебания (6.1);

частотная модуляция (ЧМ): изменяется частота ω;

фазовая модуляция (ФМ): изменяется начальная фаза θ.

Ввиду неразрывной взаимосвязи ЧМ и ФМ их объединяют понятием угловая модуляция (УМ), понимая под УМ изменение по закону передаваемого сообщения $c(t)$ полной фазы $\varphi(t)=(\omega_0 t+\theta_0)$ гармонического колебания (6.1).

На рис. 6.1 приведены виртуальные осциллограммы модулирующего колебания $c(t)=\cos 10t$, амплитудной модуляции (АМ) и угловой модуляции (УМ). Виртуальные осциллограммы получены на S-модели "Модуляция гармонических колебаний" лабораторного практикума по курсу "Основы информационных технологий". Для более глубокого изучения свойств модуляции гармонических колебаний следует выполнить лабораторную работу №7 названного практикума.

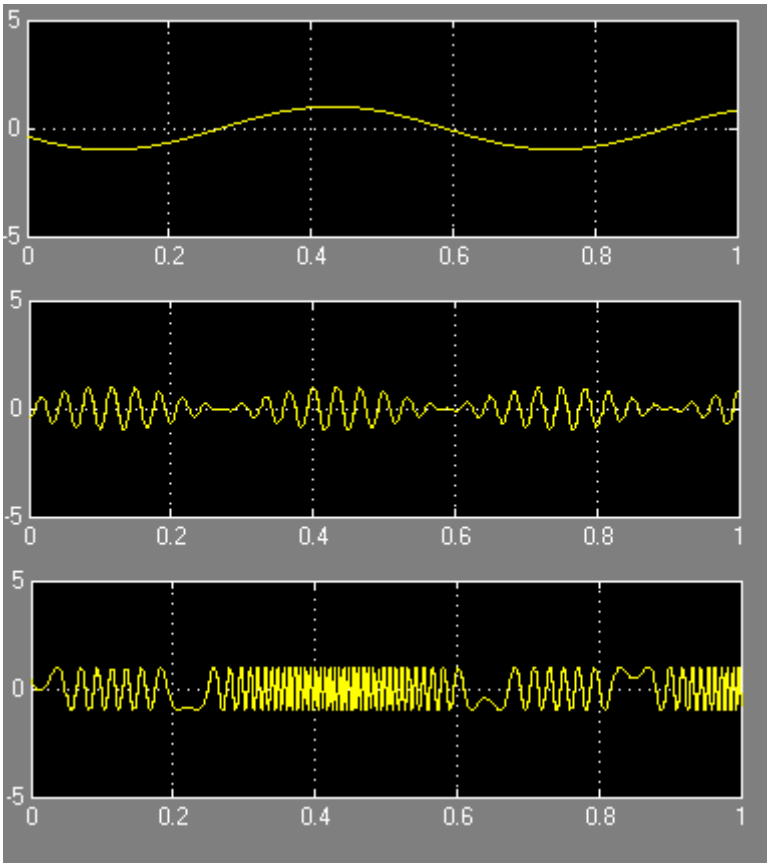

Рис. 6.1. Виртуальные осциллограммы (сверху вниз) модулирующего колебания $c(t)$, амплитудной модуляции (АМ) и угловой модуляции (УМ),

Широко применяется модуляция гармонических колебаний и при передаче данных ($c(t)$ в данном случае представляет собой последовательность сигналов "0", "1"), называемая в данном случае манипуляцией, например, частотная манипуляция, фазовая манипуляция, квадратурно-амплитудная манипуляция. Дальнейшее развитие технология манипуляции нашла в модемной связи (протоколы модуляции модемной связи).

Рассмотрим принцип переноса спектра низкочастотного модулирующего сигнала в область высоких частот на примере амплитудной модуляции. Пусть модулирующий сигнал задан в виде

$$c(t)=C_m\cos(\Omega t+\gamma), \qquad (6.2)$$

тогда, см. (6.1), для АМ колебания имеем

$$a(t) = [A_0 + K_{AM} c(t)]\cos(\omega_0 t + \theta_0) = [A_0 + K_{AM} C_m \cos(\Omega t + \gamma)]\cos(\omega_0 t + \theta_0), \quad (6.3)$$

где K_{AM} – коэффициент модуляции.

Представим (6.3) в виде

$$a(t) = A_0[1 + M\cos(\Omega t + \gamma)]\cos(\omega_0 t + \theta_0) \ , \quad (6.4)$$

где $M = \frac{K_{AM} C_m}{A_0}$ – коэффициент глубины модуляции ($K_{AM} \le 1$). Выполним в (6.4) элементарные тригонометрические преобразования

$$a(t) = A_0 \cos(\omega_0 t + \theta_0) + \frac{A_0 M}{2}\cos[(\omega_0 - \Omega)t + \theta_0 - \gamma)] + \frac{A_0 M}{2}\cos[(\omega_0 + \Omega)t + \theta_0 + \gamma)]. \quad (6.5)$$

Выражение (6.5) не содержит низкочастотных составляющих ($\omega_0 >> \Omega$), в то время как исходное (модулирующее) колебание $c(t)$ является низкочастотным. Для более глубокого усвоения рассмотренного постройте и сравните графики спектра амплитуд колебаний (6.2) и (6.5). В заключение заметим, что несущей частотой ω_0 можно варьировать, это и позволяет реализовывать совместную работу радиостанций путем разноса их несущих частот.

6.2.3. Импульсная модуляция

При импульсной модуляции по закону передаваемого сообщения $c(t)$ изменяется один из параметров периодической последовательности прямоугольных импульсов (см. рис. 6.2).

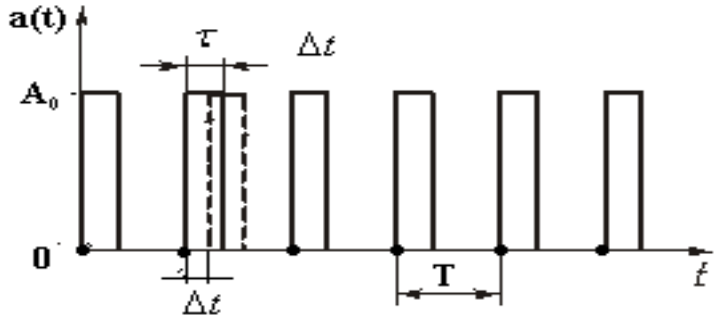

Рис. 6.2. Периодическая последовательность прямоугольных импульсов.

В данном случае (см. рис.6.2.) возможны четыре вида импульсной модуляции (см. рис. 6.3):

Амплитудно-импульсная модуляция (АИМ): изменяется амплитуда периодической последовательности прямоугольных импульсов **A.**

Широтно-импульсная модуляция (ШИМ): по закону предаваемого сигнала изменяется длительность прямоугольных импульсов τ .

Частотно-импульсная модуляция (ЧИМ): изменяется период следования прямоугольных импульсов **T.**

Фазово-импульсная модуляция (ФИМ): изменяется смещение импульса Δt относительно тактовой точки ("жирные" точки на рис. 6.2).

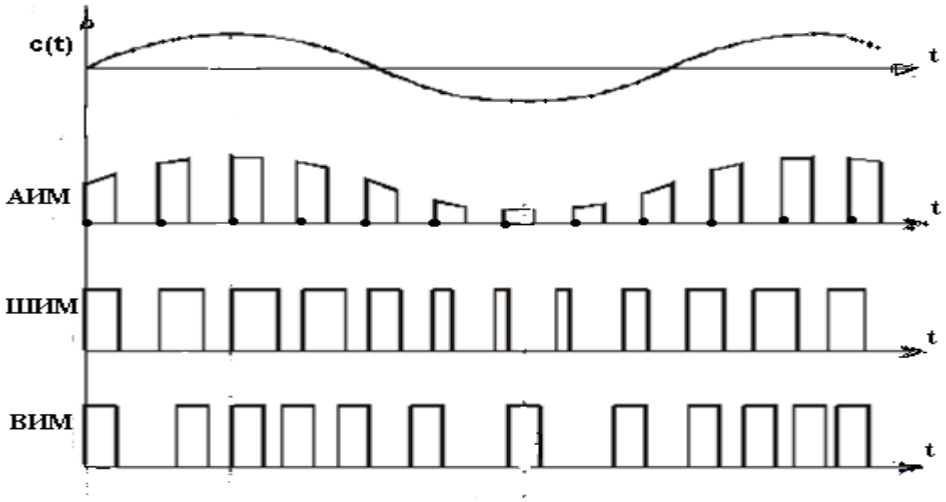

Рис. 6.3. Основные виды импульсной модуляции

Импульсные модуляции ЧИМ и ФИМ взаимосвязаны (подобно ЧМ и ФМ), поэтому их объединяют понятием время-импульсная модуляция (ВИМ).

6.2.3. Импульсно-кодовая модуляция

При импульсно-кодовой модуляции (ИКМ) сообщение $c(t)$ подвергается дискретизации во времени (см. п. 6.4) и квантованию по уровню (см. п. 6.3) , а затем преобразуется в цифровой код (см. рис 6.4). Как правило, все эти операции осуществляются аналого-цифровым преобразователем (АЦП).

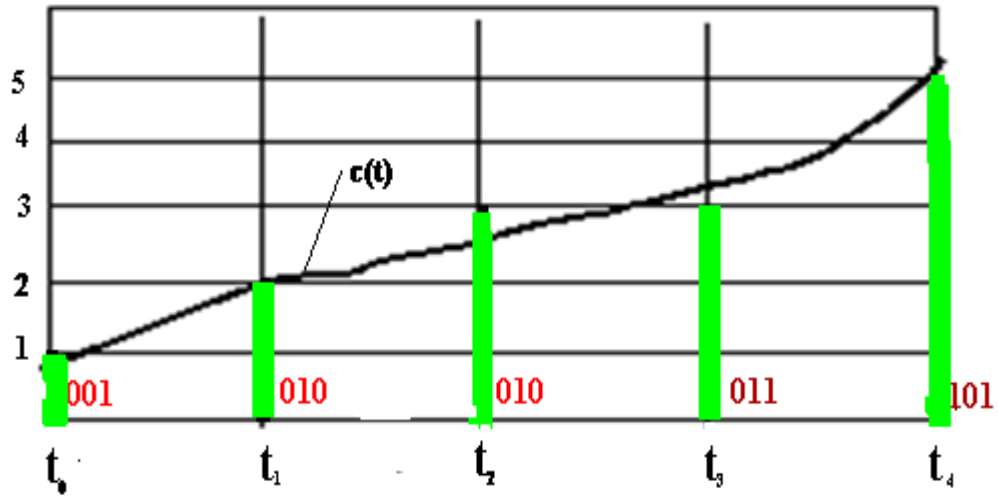

Рис. 6.4. Импульсно-кодовая модуляция.

На рис. 6.4 зеленым цветом показано первичное преобразование сигнала $c(t)$ (сигнал представляется разрешенными уровнями в тактовые моменты времени t_0, t_1, t_2, t_3, t_4); красный цвет — двоичный код, соответствующий первичному преобразованию сигнала $c(t)$. Такое представление сигнала $c(t)$ сопряжено с определенной погрешностью, как по уровню, так и во времени, но эти погрешности являются контролируемыми. Представление же сигнала $c(t)$ в цифровом виде позволяет обеспечить ему при необходимости любую требуемую помехозащищенность.

6.3. Квантование сигнала по уровню

С квантованием по уровню мы постоянно встречаемся в обыденной жизни. Прогноз погоды дается в градусах, вес тела мы, как правило, измеряем в килограммах, рост в сантиметрах и т.д. Попутно заметим, что это нас интересует лишь в отдельные, довольно редкие моменты времени.

При вводе информации в компьютер квантование по уровню и дискретизацию сигнала во времени должны осуществлять технические средства, преобразующие, в конечном счете, аналоговый сигнал в цифровую форму. При этом возникают вопросы, связанные с выбором методов преобразования и оценкой погрешностей таких преобразований. Названные погрешности довольно легко оцениваются теоретически, но выбор значений параметров дискретизации во многом определяется здравым смыслом и практическими соображениями.

Обратим внимание на одно важное свойство цифрового представления сигнала. Правильно выбрав параметров дискретизации, мы приобретаем высокое качество сигнала (нам хорошо известна разница между цифровым и аналоговым звуком), обусловленное высокой помехозащищенностью цифрового представления. Но не думайте, что все так просто. Проблем с чтением компакт-дисков нет потому, что для коррекции ошибок, связанных с мелкими дефектами диска используется корректирующий код Рида-Соломона. В сетевых же технологиях из-за специфики помех дело обстоит еще сложнее: для обнаружения ошибок используется циклический код, а для их исправления применяется метод обратной связи между получателем и источником сообщений. Названные способы повышения помехоустойчивости имеют самое прямое отношение к преобразованию информации, и они будут рассмотрены частично в следующем разделе нашего курса, а также в курсе «Аппаратное и программное обеспечение ЭВМ и сетей».

Вернемся к квантованию сигнала по уровню, которое, как мы это уже выяснили, является одним из промежуточных этапов преобразования аналогового сигнала в цифровую форму. Это хорошо разработанный вид преобразования, сущность которого сводится к представлению диапазона изменения аналогового сигнала определенным количеством уровней. Каждому уровню назначается так называемый шаг квантования. Квантуемый аналоговый сигнал отождествляется с тем уровнем, в шаг квантования которого он попадает ((см. рис 6.5), уровень квантования всегда располагается по середине шага квантования); при этом, поскольку квантуемый сигнал является случайным, то и ошибка квантования также является случайной и поэтому называется шумом квантования.

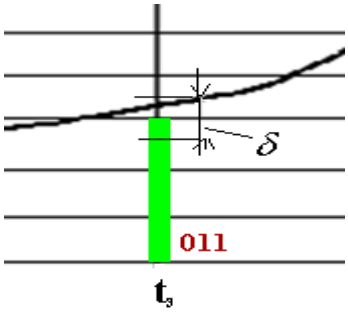

Рис. 6.5. Шаг квантования δ.

Ошибка квантования обычно оценивается дисперсией шума квантования, которая при равномерных шагах квантования вычисляется по очень простой формуле:

$$D_\delta \approx \frac{\delta^2}{12}\ , \qquad (6.6)$$

где δ— где шаг квантования. Минимизация D_δ путем использования неравномерных шагов квантования обычно не приводит к снижению D_δ более чем на 30 % и зависит от плотности распределения вероятностей квантуемой величины; так что неправильный выбор названной плотности может привести к обратному эффекту.

Так обстоит дело в идеальной схеме, когда мы пренебрегаем действием помех. Если же квантуемый сигнал является смесью полезного сигнала и помехи, то квантовать сигнал точнее, чем это позволяет помеха бессмысленно. Но и в таком случае задача легко решается, если известны статистические свойства помехи.

А теперь применительно к квантованию сигнала по уровню воспользуемся методами теории информации. Предположим, что аналоговый сигнал с диапазоном изменения $[X_{\min}, X_{\max}]$квантуется на n уровней. Тогда, используя формулу (4.8), для энтропии такого сигнала имеем

$$H_\delta(X) == -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i), \qquad (6.7)$$

где $p(x_i)$ вероятность попадания квантуемого сигнала в $i-$ый шаг квантования. Напомним, что $H_\delta(X)$ в данном случае это энтропия квантованого источника информации поэтому, чем выше значение $H_\delta(X)$, тем лучше. Используя второе свойство энтропии (4,9) запишем

$$H_\delta(X) \le \log_2 n\,. \qquad (6.8)$$

Равенство в (6.8) имеет место только в случае если $p(x_1) = \ldots = p(x_i) = \ldots = p(x_n) = \frac{1}{n}$. Но применительно к квантованному сигналу такое возможно только в двух случаях.

В первом случае квантуемый сигнал имеет равномерную плотность распределения вероятностей на интервале $[X_{\min}, X_{\max}]$ и равномерные шаги квантования; при этом (6.8) является точным равенством. Заметим (см. пример в п. 4.5) что в данном случае энтропия сигнала (среди всех возможных законов распределения вероятностей на интервале $[X_{\min}, X_{\max}]$) будет максимальной. Так что при заданном количестве уровней квантования n мы имеем верхнюю границу $H_\delta(X)$. Причем дифференциальная энтропия исходного сигнала равна $H(X) = \log_2(X_{\max} - X_{\min}) = \log_2 L$, а дифференциальная энтропия остаточной неопределенности составит $\log_2 \delta$ (не забывайте о негативном свойстве дифференциальной энтропии). Обратите внимание на простоту применения информационного подхода и его конструктивность. До квантования имели

неопределенность на интервале L, а после квантования — на интервале δ, естественно, при одном и том же равномерном законе. Сказанное станет еще более ясным, если напомнить, что дисперсия равномерного закона равна $\frac{L^2}{12}$.

Во втором случае квантуемый сигнал имеет отличную от равномерного закона плотность распределения вероятностей, но шаги квантования выбираются таким образом, чтобы обеспечивалось равенство вероятностей попадания сигнала в любой шаг квантования. В данном случае шкала квантования будет неравномерной. Это упомянутый выше случай снижения D_δ до 30%. При обычном аналого-цифровом преобразовании такая экономия несущественна, но в отдельных случаях она может быть весьма значимой. Представьте себе, например, случай дорогостоящего поиска утерянного объекта с дискретизацией площади поиска. Заметим, что уменьшение D_δ достигается за счет того, что шаги квантования выбираются таким образом, чтобы там, где плотность вероятности больше они были меньше и наоборот. Очевидно, что если квантуемый сигнал окажется в области малой плотности вероятности, то ошибка квантования будет выше средней.

В заключение обсудим случай квантования сигнала на том же интервале $[X_{\min}, X_{\max}]$, но при наличии помех. Воспользуемся выражением (4.18*): $I(Y \to X) = H(X) - H(X \mid Y)$. Применительно к этому случаю $H(X)$, как и в случае без помех, представляет собой дифференциальную энтропию исходного сигнала, в то время как $H(X \mid Y)$ является дифференциальной энтропией остаточной неопределенности, обусловленной воздействием помех. При этом случайная величина Y (смесь полезного сигнала и помехи) интерпретируется как величина, в которой содержится полезная информация о случайной величине X.

Теперь предположим, что мы вычислили значение $I(X \to Y)$, тогда количество уровней квантования (при оптимальной шкале квантования) определится формулой $n = 2^{I(X \to Y)}$. Мы положили $H_\delta(X) = \log_2 n = I(X \to Y)$.

Наконец, если предположить, что и полезный сигнал и помеха распределены по равномерному закону, а помеха является шумом (случай сигнал плюс помеха), то шаг квантования δ будет интерпретироваться также как и в случае квантования без помех, но с одним принципиальным различием. Если в случае без помех увеличение n приведет к возрастанию количества информации на выходе квантующего устройства, то при наличии шума такое увеличение n бессмысленно, ибо шум не позволит это сделать.

В заключение заметим, что тема дискретизации выходит далеко за рамки рассмотренных выше вопросов. Объектами дискретизации могут быть поверхность, объем пространства, пространство параметров произвольной природы и т.д. При этом используются такие понятия как разрешающая способность, порог различимости, $\varepsilon-$сеть. Во всех этих случаях информационный подход незаменим, поскольку позволяет не только описать исходную неопределенность, но и оценить качество последующей обработки информации.

6.4. Дискретизация сигнала во времени

Как уже отмечалось для представления сигнала в цифровой форме аналоговый сигнал необходимо не только проквантовать по уровню, но и представить его в виде отдельных временных отсчетов. При этом возникают вопросы выбора частоты отсчетов, воспроизведения сигнала, а также оценки погрешности дискретизации. Эти вопросы имеют вековую историю и связаны с именами многих ученых, но наиболее значимым результатом в этой области считается так называемая **теорема отсчетов**, авторство которой приписывают Найквисту (1928 г.), Котельникову (1933 г.), Уиттакеру (1935 г.), Шеннону (1949). Котельников использовал эту теорему при разработке теории потенциальной помехоустойчивости (1946 г.), а Шеннон для решения вопроса о пропускной способности непрерывного канала с шумом (1949).

Войдем в проблему. Начнем с хорошо известного из курса высшей математики ряда Фурье

$$f(t) = \frac{A_0}{2} + \sum_{i=1}^{\infty} A_n \cos(n\Omega_1 t - \Psi_n), \qquad (6.9)$$

который позволяет представить любую периодическую функцию $f(t)$, $(-\infty < t < \infty)$, с периодом T, удовлетворяющую условиям Дирихле, в виде суммы гармонических составляющих (гармоник) плюс $\frac{A_0}{2}$.

Что здесь представляет интерес? Обратите внимание на то, что вся информация о периодической функции содержится в значениях ее множества амплитуд, частот и фаз (это множество называется спектром функции). Так что, зная спектр $f(t)$, функцию $f(t)$ можно, используя (6.9), полностью восстановить. В этом случает (6.9) играет роль воспроизводящей функции. Заметим, что кроме (6.9) существуют и другие ряды (Уолша, Хаара, Лаггера и т.д.), позволяющие осуществлять подобное представление функций.

Практическое применение ряда Фурье в качестве модели реального сигнала связано с определенной погрешностью, поскольку реальные сигналы финитны (конечны) и поэтому по определению не могут быть периодическими. К тому же такая модель сигнала является детерминированной и, следовательно, сама по себе не содержит информации. С точки зрения получателя информации эту модель можно рассматривать лишь как знак (букву или иероглиф), начертание которого заранее известны, так как заданы спектром модели.

А как быть в реальной ситуации, в случае передачи по каналу связи заранее неизвестного сигнала (именно такие сигналы и содержат информацию)? Существует ли более компактное представление такого сигнала, спектр которого (в отлитие от моделей Фурье для непериодического сигнала) всегда ограничен? Ограничена и полоса частот любого физического канала, поэтому все частоты выше максимальной частоты этой полосы канал попросту обрежет.

По-видимому, первым кто ответил на этот вопрос, был Найквист, который отмечал, что для такого представления достаточно приблизительно $2T^*F$ чисел, где T^* — время передачи, а F — максимальная частота в спектре сигнала

(утверждение основывалось на разложении функции в ряд Фурье на интервале T^*).

А далее последовала теорема отсчетов (она базируется на моделях Фурье) которая гласит:

Функция $x(t)$, не содержащая частот выше $F_m гц$ полностью определяется своими значениями отстоящими друг от друга на

$$\Delta t = \frac{1}{2F_m}. \qquad (6.10)$$

Эта теорема справедлива также и в случае если F_m является шириной спектра функции $x(t)$. Восстановление (воспроизведение) функции $x(t)$ осуществляется с помощью ряда

$$x(t) = \sum_{k=-\infty}^{\infty} x(k\Delta t)\frac{\sin \omega_m (t - k\Delta t)}{\omega_m (t - k\Delta t)}, \quad \omega_m = 2\pi F_m. \qquad (6.11)$$

Практическое использование (6.11) связано с погрешностью, которая может оцениваться как отношение энергии ошибки к полной энергии сигнала.

Существуют и другие методы дискретизации сигнала во времени, например, адаптивная дискретизация (называемая также дельта-модуляцией), при которой временной отсчет сигнала производится в момент пересечения им уровня шкалы квантования. Так что в этом случае квантование по уровню и дискретизация во времени, как теперь говорят, осуществляется в одном «флаконе». Достоинство адаптивной дискретизации состоит не только в контроле качества цифрового представления сигнала, но и в минимизации затрат на передачу сигнала. Поскольку в данном случае можно передавать не само значение сигнала в точке отсчета, а только его отклонение от предыдущего состояния (+1, -1), В этом случае количество передаваемой информации в одном отсчете не будет превышать 1 бит.

6.5. Криптографическое шифрование

6.5.1. Основные понятия и методы криптографического шифрования

Криптографическое шифрование (преобразование) информации предназначено для защиты информации от ее незаконного или нежелательного использования (прочтения или изменения) и является одним из важнейших средств защиты информации в современных информационных технологиях. Более того, сегодня криптографическое шифрование это и технология и индустрия.

В развитии методов и средств шифрования можно выделить три основных периода (шифрование возникло практически одновременно с письменностью):

1. Начальный период, который характеризуется ручным шифрованием и использованием простых обычных (симметричных) шифров. Примером такого древнего шифра является шифр Юлия Цезаря. Симметричный шифр при шифровании и дешифровании используют один и тот же ключ.

2. Второй период начинается с 30-х гг. XX века и отличается от предыдущего периода использованием более сложных симметричных шифров с применением механических, электромеханических и электронных средств шифрования, а также засекреченных сетей связи. Примером шифровального средства этого периода является известная по кинофильмам немецкая шифровальная машина «Энигма».

3. Начало третьего (современного) периода относят к 1976 году и связывают с технологией ассиметричного шифрования как принципиально нового метода шифрования, основанного на паре ключей, не требующего для организации конфиденциальной связи предварительного обмена абонентами секретным ключом.

Разработкой методов и алгоритмов шифрования занимается **криптография.** Различают также **криптоанализ** – научную область, посвященную криптографическому анализу и взлому. Иногда используют понятие **криптология** *как* объединение криптографии и криптоанализа. Еще одной областью, связанной с шифрованием является **стеганография**, которая посвящена скрытному встраиванию информации в изображения, текст и т.д. Обнаружить такой встроенный объект, например в графическом файле, довольно сложно, к тому же он может быть еще и надежно зашифрован. Стеганография очень древний способ скрытия сообщения. Известен, например, такой способ: сообщение писали на бритой голове раба, и отправляли раба по адресу после отрастания волос.

Становление криптографии как науки связывают с появлением шифровальных машин (начало второго периода). Понимание же математической сущности криптографических задач пришло после работы Клода Шеннона «Теория связи в секретных системах» (рассекреченный материал доклада «Математическая теория криптографии, 1945г.), в которой Шеннон обобщил накопленный до него опыт разработки симметричных шифров.

Последующее развитие криптографии связывают с работой американских математиков У. Диффи и М. Хеллмана «Новые направления в криптографии», которая привела не только к перестройке здания криптографии, но и вызвала к жизни новые направлений в математике. Так что современная криптография использует самые последние достижения науки и, в первую очередь в области математики, при этом практика шифрования во многом определяются состоянием средств информационной технологии.

Сущность шифрования состоит в функциональном преобразовании синтаксических свойств открытого сообщения в целях сокрытия его семантического содержания или конкретных конфиденциальных данных, составляющих тайну (государственную, военную, коммерческую, юридическую, врачебную и т.д.).

Существующие методы криптографического шифрования можно классифицировать по многочисленным признакам:

по виду используемых ключей (симметричные, асимметричные),

по методам преобразования информации (подстановки, перестановки, гаммирование и др.),

по технологии режима шифрования (блоковое, поточное) и т.д.

Рассмотрим основные понятия шифрования.

Шифрование – процесс преобразования открытого текста в шифрованное сообщение (**криптограмму, шифровку**), осуществляемый по определенному криптографическому алгоритму.

Дешифрование – процесс, обратный шифрованию по криптографическому алгоритму.

Криптографический алгоритм (шифр) — математическая модель, содержащая функцию для шифрования и функцию для дешифрования.

Алгоритм шифрования называется **ограниченным**, если его надежность обеспечивается только за счет сохранения в тайне самого алгоритма шифрования. Такие алгоритмы шифрования представляют лишь исторический интерес и совершенно не пригодны в современных условиях.

В современной криптографии алгоритм шифрования может быть доступен любому. Секретным является только ключ (специальная величина), который выбирается из так называемого **ключевого пространства**. Использование алгоритмов с ключевым пространством (в отличие от ограниченных алгоритмов) позволяют наладить массовое производство криптографических средств, поскольку знание криптографического алгоритма не позволяет злоумышленнику расшифровать сообщение. Понятие криптосистема включает алгоритм шифрования, множество всевозможных ключей, а также открытые и шифрованные тексты.

Существуют две разновидности алгоритмов шифрования: **симметричные алгоритмы шифрования** и **алгоритмы шифрования с открытым ключом** (ассиметричные). В современных криптосистемах используются обе разновидности.

6.5.2. Симметричные алгоритмы шифрования

К симметричным относят криптографические алгоритмы, в которых для шифрования и дешифрования используется один ключ или такие, в которых ключ для дешифрования получается из ключа для шифрования и наоборот. В симметричных шифрах в основном используются алгоритмы подстановки (символы шифруемого текста заменяются символами того же или другого алфавита), алгоритмы перестановки (символы шифруемого текста переставляются по определенному правилу в пределах некоторого блока этого текста) и гаммирование (символы шифруемого текста складываются с символами некоторой случайной последовательности).

Рассмотрим пример простого симметричного алгоритма шифрования, основанного на подстановке символов. Допустим, что для шифрования сообщений на русском языке используются шифр простой подстановки, состоящий в том, что в открытом тексте A осуществляется взаимнооднозначная замена букв подстановкой других букв русского алфавита (33 буквы), т.е. взаимно однозначное отображение множества букв алфавита на себя. Количество таких подстановок, как известно из комбинаторики, составляет 33! Это множество является ключевым пространством K данного алгоритма (данное множество является симметрической группой по операции подстановка). Ключом же будет

являться конкретный экземпляр k_i ключевого пространства K, используемый при шифровании, в результате которого открытый текст A преобразуется в криптограмму B.

Если предположить, что k_i выбирается из множества K по равномерному закону, то вероятность угадать ключ составит $\frac{1}{33!} \approx 5*10^{35}$. Однако энтропия криптограммы B будет такой же как, и открытого текста A (см. тему энтропия источника дискретных сообщений); так что при достаточно большой длине текста, поскольку русский язык (как и другие европейские языки) обладает большой избыточностью (примерно 0,8), то, используя статистических свойства языка, можно на много порядков сократить необходимое число переборов.

Алгоритмы шифрования такого типа являются простейшими и легко описываются математически. Пусть X — алфавит открытого, а Y — алфавит шифрованного текста (объемы алфавитов совпадают) и пусть $g: X \to Y$ — взаимно-однозначное отображение X в Y. Тогда преобразование открытого текста $x_1 x_2 \ldots x_n$ в криптограмму осуществляется так $g(x_1)g(x_2)\ldots g(x_n)$.

Существует большое количество симметричных алгоритмов шифрования, многие из них являются комбинацией нескольких простых шифров. Наибольшее практическое применение получили стандарт DES (государственный стандарт США) и ГОСТ 28147-89.

Компьютерные алгоритмы симметричного шифрования подразделяются на два вида: блоковые алгоритмы и потоковые алгоритмы. Потоковые алгоритмы обрабатывают открытый текст побитно, в блоковых же алгоритмах обработка осуществляется поблочно, длина блока обычно составляет 64 бита.

DES (Data Encryption Standard) разработан в IBM в 1977г. и принят в качестве стандарта в ANSI в 1980г. Используется для защиты коммерческой информации. DES относится к симметричным блочным алгоритмом шифрования и предусматривает несколько режимов алгоритма шифрования. Наиболее простым режимом является ECB (электронная кодовая книга), в котором каждый блок исходного текста (64 бита) кодируется с помощью одного и того же 56-битового ключа. В режиме CBC каждый блок исходного текста перед обработкой суммируется с предыдущим, уже зашифрованным; режим FSB обеспечивает шифрование с обратной связью блоками менее 64 бит; режим OFB обеспечивает потоковый режим шифрования. Стандарт DES имеет несколько разновидностей.

В российском стандарте ГОСТ 28147-89 (принят в 1989г. в СССР) используются несколько связанных процедур шифрования с защитой данных от внесения в них несанкционированных изменений; шифруются 64-битовые блоки данных, используется ключ длиной 256 бит. ГОСТ 28147-89 по большинству параметров превосходит DES.

6.5.3. Алгоритмы шифрования с открытым ключом

При шифровании с открытым ключом для шифрования и дешифрования используются разные ключи. Один из них является открытым (общедоступным) и

используется для шифрования, другой, посредством которого осуществляется дешифрование, сохраняется в секрете.

Получили признание и закреплены в стандартах два метода шифрования с открытым ключом: метод, основанный алгоритме RSA (ISO/IEC/DIS 9594-8 и ITU-T X.509) и метод, основанный на алгоритме ЭльГамаля (MD 20899; стандарт NIST: бывший ANSI).

Алгоритм RSA используется во многих коммерческих продуктах, поддерживается современными операционными системами, реализуется на аппаратном уровне (смарт-карты, сетевые карты Ethernet), находит применение в стандартах Интернет. Системы RSA используются правительственными учреждениями, большинством корпораций, коммерческими сетями. Например, международная сеть электронного перечисления платежей SWIFT требует от банковских учреждений, применения именно этой криптографической системы.

Алгоритм RSA

Алгоритм RSA разработан в 1977г. Ривестом (R. **R**ivest), Шамиром (A. **S**hamir) и Адлеманом (L. **A**dleman). Надежность RSA основана на трудности разложения очень больших целых чисел на простые сомножители. Алгоритм RSA работает следующим образом:

1. Отправитель выбирает два очень больших простых числа P и Q и вычисляет произведения: $N = PQ$ и $M = (P-1)(Q-1)$.

2. Затем Отправитель выбирает случайное целое число D, взаимно простое с M (D не должно иметь делителей >1, общих с M), и вычисляет E, удовлетворяющее условию $DE = 1(\mathrm{mod}\, M)$.

3. После этого Отправитель публикует D и N как свой открытый ключ шифрования, сохраняя E как закрытый ключ.

4. Если S - сообщение в интервале $(1, N)$, то оно превращается в шифровку возведением в степень D по модулю N и отправляется получателю $S^* = S^D (\mathrm{mod}\, N)$.

5. Получатель сообщения расшифровывает его, возводя S^* в степень E по модулю N, так как $S = S^{*E} (\mathrm{mod}\, N) = S^D (\mathrm{mod}\, N)$.

Рассмотрим пример, используя на первом шаге алгоритма (для наглядности) малые простые числа.

1. Выбираем два простых числа $P = 7$; $Q = 17$ и вычисляем произведения: $N = PQ = 119$, $M = (P-1)(Q-1) = 96$.

2. Выбираем целое число D=5 (взаимно простое с M) и вычисляем E, удовлетворяющее условию $DE = 1(\mathrm{mod}\, M)$. Используя расширенный алгоритм Эвклида получаем E=77.

3. Публикуем D=5 и N =119 как свой открытый ключ шифрования (допустим, посылаем другу), сохраняя E как закрытый (секретный) ключ.

4. Пусть Ваш друг желает передать сообщение S=25, тогда он шифрует это сообщение по формуле $S^* = S^D (\mathrm{mod}\, N)$=9. и посылает Вам $S^* = 9$

5. Вы получаете S^*= 9 и дешифруете полученное сообщение, используя формулу $S = S^{*E} (\mathrm{mod}\, N)$. В результате имеете открытое сообщение друга S=25

Высокая надежность в системе RSA обеспечивается только при очень больших N (порядка 200 десятичных разрядов), что требует огромных вычислительных ресурсов для шифрования каждого блока информации. Поэтому при передаче больших объемов информации обычно используют симметричное шифрование (например, стандарт DES), а секретный ключ DES шифруют открытым ключом RSA.

Для практической реализации шифрования RSA разрабатываются специальные процессоры, лучшие из которых позволяют выполнять возведение в степень целого числа из 300 десятичных знаков за доли секунды.

Алгоритм Эль-Гамаля

Алгоритм Эль-Гамаля может использоваться как для шифрования данных, так и для формирования электронной подписи. На этом алгоритме построен российский стандарт на цифровую подпись ГОСТ Р 34.10-94.

В алгоритме Эль-Гамаля генерация пары ключей осуществляется следующим образом. Выбираются: простое число p и два случайных числа $g < p$ и $x < p$, на основе которых вычисляется $y = g^x (\mathrm{mod}\, p)$. Открытыми ключами являются y, g и p, а секретным ключом — x.

Для электронной подписи сообщения M выбирается случайное число k, взаимно простое с $p-1$ и вычисляется $a = g^k (\mathrm{mod}\, p)$. Затем определяется b, удовлетворяющее уравнению: $M = (xa + kb)(\mathrm{mod}(p-1))$.

Электронную подпись сообщения M составляют числа a и b, при этом число k является секретным. Верификация подписи осуществляется проверкой равенства $y^a a^b (\mathrm{mod}\, p) = g^M (\mathrm{mod}\, p)$. Дешифровка сообщения M выполняется по формуле $M = \dfrac{b}{a^x (\mathrm{mod}\, p)}$.

Криптосистема, основанная на алгоритме Эль-Гамаля, обеспечивает большую степень защиты, чем на алгоритме RSA. Оба алгоритма основаны на использовании односторонней функции, т.е. такой функции, которая легко выполняется в одном направлении (прямая операция), но для которой очень трудно выполнить обратную операцию. Однако найти показатель степени, в которую нужно возвести заданное число для получения искомого задача значительно более трудная (задача дискретного логарифмирования), чем разложение больших чисел на простые сомножители в алгоритме RSA. Алгоритм Эль-Гамаля позволяет практически на порядок увеличить скорость шифрования и дешифрования данных.

Основным недостатком алгоритмов шифрования с открытым ключом является их низкое быстродействие (на 3-4 порядка ниже, чем у симметричных алгоритмов), поэтому их обычно используют совместно с симметричными алгоритмами шифрования. Предаваемое сообщение шифруется посредством

симметричного алгоритма, а сам ключ (передаваемый совместно с этим зашифрованным сообщением) шифруется ассиметричным алгоритмом.

Тема 7. Передача информации

7.1. Передача информации и коммуникационные технологии

Начиная с глубокой древности трудно найти область человеческой деятельности, которая бы не была связана с передачей информации. Папирусы, глиняные таблички, берестяные грамоты, наскальные надписи — все это примеры *передачи информации во времени* **(хранение информации).** Сигнальные костры; то, что связано с марафонским бегом; стеганография на голове раба и т. п. — это уже примеры *передачи информации в пространстве* ***(*связь или коммуникация*).***

Оба метода передачи информации имеют много общего (сжатие данных, помехоустойчивое кодирование, шифрование и др.), но индустриальное развитие вначале получила передача информации в пространстве, т.е. то, что изначально называется связью (электросвязь и радиосвязь), а в настоящее время, по большей части применительно к передаче данных именуется **коммуникационной технологией.** Методы же хранения информации на машинных носителях получили распространение лишь с появлением компьютеров.

Роль коммуникационных технологий в современном мире такова, что в последнее время все чаще и чаще, особенно когда речь идет о дистанционном обучении, употребляют термин «информационно-коммуникационные технологии» (ИКТ). Происхождением этот термин, по-видимому, обязан II Международному конгрессу ЮНЕСКО «Образование и информатика» (Москва, 1996 г.), на котором речь шла о роли информационных и коммуникационных технологий в обучении.

Уясним "местоположение" современных коммуникационных технологий по отношению к эталонной модели взаимодействия открытых систем RM OSI (см. п. 2.2). Шесть нижних уровней RM OSI по существу обеспечивают транспорт и интерфейс для многочисленных прикладных программ, относящихся к 7-му уровню, и ничего более (для конечного пользователя технологии этих нижних уровней по существу скрыты). Транспортная же служба сети (четыре нижних уровня модели RM OSI) организуется путем использования (аренды) каналов связи коммуникационной сети. В Республике Беларусь услуги доступа по таким каналам предоставляет национальный оператор электросвязи Республики Беларусь РУП «Белтелеком».

Ниже, в подразделе 7.2, мы ознакомимся с принципами построения коммуникационных сетей, а далее рассмотрим фундаментальные положения шенноновской теории информации, связанные с передачей информации. Основные теоремы Шеннона для пропускной способности каналов связи сыграли выдающуюся роль в развитии теории и практики систем передачи информации.

7.2. Коммуникационные сети

По особенностям архитектуры информационные сети делятся на коммуникационные (транспортные) и компьютерные (компьютерные сети изучаются в курсе «Аппаратное и программное обеспечение ЭВМ и сетей»).

Транспортные сети предназначены для обеспечения связи и информационного обмена между территориально разнесенными пользователями. Транспортные сети подразделяются на первичные и вторичные, которые могут быть аналоговыми или цифровыми**.**

Первичная сеть (опорная сеть) представляет собой совокупность типовых каналов, групповых трактов, сетевых узлов, сетевых станций.

Вторичные сети организуются на базе первичной сети и подразделяются на телефонные, телеграфные, телевизионные, звукового вещания, передачи данных.

Основным типовым каналом аналоговой транспортной сети является канал тональной частоты (ТЧ). Это типовой канал с полосой частот от 300 до 3400 Гц. Для увеличения пропускной способности каналы тональной частоты согласно Рекомендациям CCITT объединяют в группы: первичная группа – 12 каналов ТЧ, вторичная группа – 60 каналов ТЧ, третичная группа – 300 каналов ТЧ. Допускается создание других групп, например, четвертичных – три 300-канальные группы.

Цифровые каналы первичной сети строятся на принципах:

Плезиохронной цифровой иерархии (PDH — Plesiochronous Digital Hierarchy; 1960 гг. старая, обладающая недостатками технология).

Синхронной цифровой иерархии (SDH — Synchronous Digital Hierarchy, 1988 г., определена комитетом ITU или SONET — Synchronous Optical NETwork, стандарт ANSI, совместима с SDH).

В цифровых сетях цифровой канал представляет собой битовый тракт с цифровым (импульсным) сигналом на входе и выходе канала. Оконечное оборудование таких каналов работает только с цифровыми сигналами. Базовым цифровым каналом является канал DS0 (Digital Signal, Level 0: цифрой сигнал нулевой уровень) со скоростью передачи информации 64 Кбит/с. (стандарт для канала голосовой телефонии).

На его основе DS0 строятся следующие каналы PDH с более высокими скоростями передачи:

DS1 (Digital Signal, Level 1). Объединение (уплотнение) 24-х каналов DS0, скорость -- 1.544 Мбит/с.

На такой скорости работает североамериканская линия (технология) T1. Технологическим аналогом T1 в Европе являются E1 (ИКМ30 в СНГ). Объединение 30-ти DS0, 2.048 Мбит/с;

Линии T1 были разработаны в 60-х годах американской компанией Bell Telephone System и использовались для голосовой связи и передачи факсов (речевой сигнал дискретизируется с частотой 8000 1/с). При передаче используется временное разделение (уплотнение) каналов. Дискретные отсчеты (длительностью 5,2 миллисекунды) передаются последовательно с использование всех 24 каналов. В T1 возможно дробление каналов: каждый из 24 каналов может выть передан в индивидуальное распоряжение отдельного пользователя.

DS2. Объединение 4-х DS1 (96 DS0), скорость — 6.312 Мбит/с. Линия Т2. В Европе Е2. Объединение 4-х Е1, 8.448 Мбит/с;

DS3. Объединение 7-ми DS2 (672 DS0), скорость — 44.736 Мбит/с. Линия Т3. В Европе Е3 Объединение 4-х Е2, 34.368 Мбит/с;

DS4. Объединение 6 DS3 (4032 DS0), скорость – 274,176 Мбит/с. Линия Т4. В Европе Е4. Объединение 4-х Е3, 139.264 Мбит/с.

В Европе имеется так же Е5. Объединение 4-х Е4, 564.992 Мбит/с.

Цифровые каналы первичной сети, построенные на SDH/SONET, характеризуются следующими скоростями.

SDH	SONET	Скорость
STM-0	OC-1	51,840 Мбит/с,
STM-1	OC-3	155,52 Мбит/с,
STM-3	OC-9	466,560 Мбит/с,
STM-4	OC-12	662 Мбит/с,
STM-6	OC-18	933,120 Мбит/с,
STM-8	OC-24	1,244 Гбит/с,
STM-12	OC-36	1,866 Гбит/с,
STM-16	OC-48	2,488 Гбит/с,
STM-64	OC-192	9,953 Гбит/с,
STM-256	OC-768	39,81 Гбит/с.

Здесь STM-x: Synchronous Transport Module Level x; OC-x: Optical Carrier Level x.

По масштабу транспортные сети подразделяются:

1. Сетевое ядро, соединяющее города, страны и континенты. Требование по пропускной способности: от единиц до сотен Гбит/с. Верхняя граница для США, Западной Европы и Японии; нижняя – для развивающихся стран.

2. Городская транспортная сеть (от сотен Мбит/с до десятков Гбит/с).

3. Сеть доступа (от единиц до сотен Мбит/с). Конечные пользователи могут подключаться к сети доступа в диапазоне скоростей от десятка Кбит/с до единиц Мбит/с.

Иерархия первичных транспортных сетей образуется путем использования оконечных мультиплексоров, промежуточных и узловых мультиплексоров, мульдексов (мультиплесоров/демультиплексоров) и другой аппаратуры. При этом используются как каналы PDH, так и SDH (каналы PDH объединяются посредством SDH). Доступ к сети осуществляется с использованием технологий ISDN или xDSL.

ISDN (Integrated Services Digital Network) — цифровая интегрированная сеть передачи информации (видеоинформации, речи, данных и т.д.). Стандартное подключение линии ISDN осуществляется по интерфейсам BRI (Basic Rate Interface) или PRI (Primary Rate Interface).

Интерфейс BRI (базовый доступ, служба 2B+D, ISDN2) обеспечивает два дуплексных B-канала по 64 Кбит/с каждый и один служебный канал (16 Кбит/с). BRI использует телефонную сеть общего пользования (ТСОП). Каждому B-каналу присваивается номер, аналогичный телефонному. B-каналы используются

индивидуально и коммутируются по вызову. Абонентская линия ISDN заканчивается блоком сетевого окончания NTBA, к которому можно подключить до 8-ми цифровых оконечных устройств, но одновременная работа возможна только для двух устройств.

Интерфейс PRI (первичный доступ ISDN, служба 30B+D, ISDN30) обеспечивает 30 дуплексных B-канала по 64 Кбит/с и специализированный D-канал с пропускной способностью 64 Кбит/с (пропускная способность PRI составляет 2,048 Мбит/с). PRI может применяться для соединения удаленных филиалов с центральным офисом, а также для подключения учрежденческих АТС к цифровой телефонной сети. В РБ услуги ISDN предоставляют Белтеком, БелПак, Атланттелеком и др.

Технологии xDSL отличаются от ISDN большим разнообразием принципов физической транспортировки данных, более высокими скоростями передачи данных и постоянно включенной (некоммутируемой) линией. В РБ используются следующие xDSL технологии:

ADSL (Asymmetric Digital Subscriber Line) — ассиметричная цифровая абонентская линия (скорость на расстоянии 1,5 км: вх. 8,44 Мбит/с, исх. 1 Мбит/с).

G. HDSL (High Bit-Rate Digital Subscriber Line) — последняя разработка HDSL (скорость на одной паре (1,5 км): вх. 2,324 Мбит/с, исх. 2,324 Мбит/с).

SDSL (Single Digital Subscriber Line) — однолинейная цифровая абонентская линия (скорость на расстоянии 1,5 км: вх. 2,0 Мбит/с, исх. 2.0 Мбит/с).

7.3. Информационные характеристики каналов связи

7.3.1. Информационная модель канала связи

Под каналом связи понимают совокупность средств, предназначенных для передачи информации из одного пункта пространства в другой. Источник информации, канал связи и получатель информации составляют систему связи. Различают также линию связи (ЛС), под которой понимают тракт, соединяющий оконечные точки передающей и приемной стороны.

Обсудим названные понятия на примере обобщенной схемы одноканальной системы связи (рис. 7.1). Вначале следует хорошо уяснить сущность понятия канала связи. Согласно приведенному выше определению каналом связи является и линия связи (канал связи в узком смысле слова), и часть системы связи, включающая передающую и приемную сторону (канал связи в широком смысле слова). Применительно к рисунку 7.1 можно указать еще два частных случая канала связи. Укажите их.

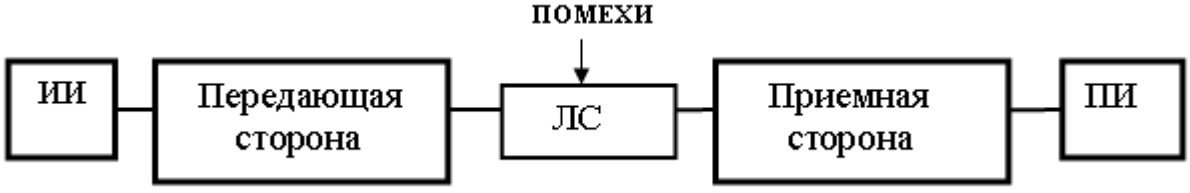

Рис. 7.1. Обобщенная схема одноканальной системы связи: ИИ — источник информации; ЛС — линия связи; ПИ — получатель информации.

Если раскрыть блоки этой простой схемы, то возможных вариантов будет еще больше. Так, например, в системе передачи данных передающая и приемная стороны содержат кодек (кодер/декодер) и модем (модулятор/демодулятор), что приводит к дополнительным вариантам. Смысл такого гибкого подхода к понятию канала связи состоит в возможности единообразного исследования и оптимизации любой, представляющей интерес, части системы передачи информации.

Каналы связи можно классифицировать по многочисленным признакам. Для нас представляет интерес классификация каналов по числу возможных состояний на входе и выходе канала. С этой точки зрения различают:

Дискретные каналы. Это каналы, у которых число состояний (объем алфавита) на входе и выходе канала является конечным. Важным частным случаем являются бинарные (двоичные) каналы, которые имеют два состояния: «0» и «1».

Дискретно-непрерывные каналы. У таких каналов число состояний на входе — конечно, а на выходе — непрерывно.

Непрерывно-дискретные каналы (число состояний на входе — непрерывно, а на выходе — конечно).

Непрерывные каналы. Каналы, у которых состояния на входе и выходе непрерывны.

Забегая вперед, заметим, что у канала передачи данных можно выделить все четыре, названных выше вида каналов, каждый из которых можно представить в виде схемы, приведенной на рис. 7.2. Различаться такие каналы будут лишь статистическим описанием и ограничениями, накладываемыми на канал (см. п. 4.7).

Рис. 7.2. Обобщенная схема канала связи.

Уточним понятие помеха. Под помехой понимают стороннее возмущение, искажающее полезный сигнал. Понятия помеха и полезный сигнал относительны. Например, для преподавателя разговаривающие на лекции студенты являются помехой, и наоборот. В этой связи, очевидно, что для описания и полезного сигнала и помехи применяются одни и те же математические модели. Различают следующие виды помех:

1. Аддитивная помеха, ее часто называют шумом, $y(t) = x(t) + \eta(t)$. Здесь $x(t)$ — полезный сигнал, $\eta(t)$ — помеха.

2. Мультипликативная помеха: $y(t) = \mu(t)x(t)$, где $\mu(t)$ — помеха.

3. Смешанная помеха: $y(t) = \mu(t)x(t) + \eta(t)$.

Классическая теория информации (как и общая теория связи) традиционно имеет дело с первым случаем (мультипликативную помеху всегда можно свести к эквивалентной аддитивной), так что обобщенная информационная модель канала связи будет иметь следующий вид

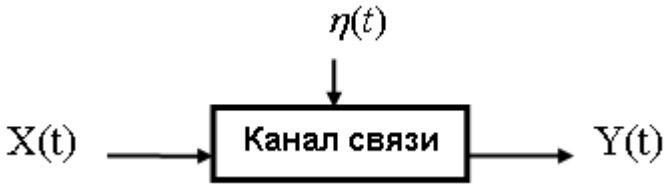

Рис. 7.3. Обобщенная информационная схема канала связи.

При этом по определению (см. п. 4.7) пропускная способность канала связи

$$C = \sup[\bar{I}(Y \to X)], \tag{7.1}$$

где скорость передачи информации

$$\bar{I}(Y \to X) = \lim_{T\to\infty}\left[\frac{I(Y_T \to X_T)}{T}\right]. \tag{7.2}$$

Здесь $I(Y_T \to X_T)$ среднее количество информации, содержащееся в сообщении Y_T на выходе канала связи относительно сообщения X_T на входе канала за время T.

Напомним, см. п. 4.7, что для определения пропускной способности канала связи необходимо знать статистическое описание сигнала и шума, а также ограничения, накладываемые на канал.

Заметим, что (7.1) и (7.2) получили прописку в качестве математических моделей в современной теории вероятностей, а по своей практической значимости они далеко выходят за пределы области передачи информации.

7.3.2. Пропускная способность дискретного канала без шума

В идеализированном случае, при отсутствии шума, схема канала связи имеет вид, приведенный на рис. 7.4.

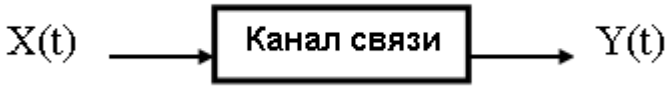

Рис. 7.4. Информационная схема канала связи без шума.

Учитывая, что шум отсутствует запишем

$$I(Y_T \to X_T) = I(X_T \to X_T) = H(X_T), \tag{7.3}$$

где $H(X_T)$— энтропия источника информации за время T. Подставим (7.3) в (7.2) и полученный результат в (7.1)

$$\mathrm{C} = sup \lim_{\mathrm{T}\to\infty} \frac{H(X_T)}{T}. \tag{7.4}$$

Обозначим через N_T количество различающихся последовательностей, которое источник информации способен выработать за время T. Тогда учитывая, что $H_{\max}(X_T) = \log_2 N_T$, для (7.4) имеем

$$C = \lim_{T\to\infty} \frac{\log_2 N_T}{T} . \qquad (7.5)$$

Это выражение является исходным при определении пропускной способности дискретного канала без шума. Рассмотрим примеры.

Пример 1. Определить пропускную способность дискретного канала без шума, если известно, что число состояний на входе и выходе канала связи равно a (a — объем алфавита входа); длительность передачи любой буквы канала одинакова и равна τ.

Определим, используя исходные данные, N_T. Учитывая, что количество переданных букв в канале за время T определяется отношением $M = \frac{T}{\tau}$, запишем: $N_T = a^M = a^{\frac{T}{\tau}}$. Наконец, поставляя полученный результат в (7.5) получаем

$$C = \lim_{T\to\infty} \frac{\log_2 a^{\frac{T}{\tau}}}{T} = \lim_{T\to\infty} \frac{T\log_2 a}{\tau T} = \frac{1}{\tau}\log_2 a \ . \qquad (7.6)$$

Таким образом, пропускная способность дискретного канала без шума, при заданных ограничениях на канал связи, определяется по следующей формуле

$$C = \frac{1}{\tau}\log_2 a . \qquad (7.7)$$

Пример 2. Определить пропускную способность дискретного канала без шума и скорость передачи информации, если известно, что $a = 2$, длительность передачи любой буквы канала $\tau = 1{,}25\times 10^{-3} c$, а вероятностные характеристики источника информации описываются следующим рядом распределения вероятностей

$$X = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ 0{,}5 & 0{,}25 & 0{,}125 & 0{,}125 \end{pmatrix}. \qquad (7.8)$$

Определим, используя формулу (7.7) пропускную способность канала

$$C = \frac{1}{\tau} = 800 \quad \text{бит/с (бод).} \qquad (7.9)$$

Для определения скорости передачи информации запишем: $I(Y_T \to X_T) = I(X_T \to X_T) = M_{\text{и}} H(X)$, где $M_{\text{и}}$ — количество букв, вырабатываемое источником за время T. Подставим полученный результат в (7.2)

$$\bar{I}(Y \to X) = \lim_{T\to\infty}\left[\frac{I(Y_T \to X_T)}{T}\right] = \lim_{T\to\infty} \frac{M_{\text{и}} H(X)}{T} = \frac{H(X)}{\tau_{\text{и}}} . \qquad (7.10)$$

$$H(X) == -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i) = -0{,}5\log_2 0{,}5 - 0{,}25\log_2 0.25 -$$

$$-0{,}125\log_2 0{,}125 - 0{,}25\log_2 0{,}25 = 1{,}75,$$

$\tau_и = 2\tau = 2\times 1{,}25\times 10^{-3}$ (Передача одной буквы источника информации требует двух букв канала связи).

Подставим значения $H(X)$ и $\tau_и$ в (7.10)

$$\bar{I}(Y \to X) = \frac{H(X)}{\tau_и} = \frac{1{,}75}{2{,}5\times 10^{-3}} = 700 \text{ бод}. \qquad (7.11)$$

Таким образом $C > \bar{I}$, так что канал связи используется неэффективно.

7.3.3. Основная теорема Шеннона для дискретного канала без шума

В примере 2 скорость передачи информации оказалась меньше пропускной способности канала. В этой связи возникает вопросы:

1. Можно ли приблизить скорость передачи информации к пропускной способности канала связи?
2. И, если да, то, как это осуществить практически?

Основная теорема Шеннона для дискретного канала без шума является типичной теоремой существования. Она отвечает на первый вопрос, при этом одно из доказательств этой теоремы основано на построении кода (эффективного кода), который дает ответ и на второй вопрос. Теорема имеет прямое и обратное утверждение.

Прямое утверждение. Если скорость создания информации источником дискретных сообщений равна:

$$\overline{H}(X) = C - \alpha,$$

где C – пропускная способность дискретного канала без шума;

α - сколь угодно малая положительная величина,

то существует такой способ кодирования, при котором все вырабатываемые источником сообщения будут переданы; при этом скорость передачи информации будет равна

$$\bar{I} = C - \alpha.$$

Обратное утверждение. Если $\overline{H}(x) > C$, то такого способа кодирования не существует.

7.3.4. Эффективное кодирование

Под эффективным кодированием понимают такое кодирование, которое позволяет приблизить скорость передачи информации к пропускной способности канала. Исторически первыми такими кодами являются код Шеннона-Фано и код Хаффмена.

Методика построения кода Шеннона-Фано

Буквы алфавита источника выписываются в столбец в порядке убывания их вероятностей. Столбец разбивается на две подгруппы с равными (по возможности) суммарными вероятностями. Каждая подгруппа, которая содержит более одной буквы, в свою очередь разбивается таким же образом на две подгруппы и т.д. Описанный процесс продолжается до тех пор, пока во всех подгруппах очередного шага разбиения не останется по одной букве.

Код формируется следующим образом. Всем верхним подгруппам каждого шага разбиения приписывается символ «0», а нижним — символ «1» (можно и наоборот). Длина кодовой комбинации буквы определяется числом шагов разбиения, в которых эта буква участвовала. Кодовая комбинация буквы формируется слева неправо путем записи символов 0 или 1 в зависимости от того, в какую подгруппу (верхнюю или нижнюю) попала данная буква в соответствующем шаге.

Проиллюстрируем методику построения кода Шеннона-Фано (см. табл.7.1) на примере источника сообщений, который описывается рядом распределения (5.12)

$$X = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ 0{,}5 & 0{,}25 & 0{,}125 & 0{,}125 \end{pmatrix}$$

Таблица 7.1. Пример построения кода Шеннона–Фано

X	P(xi)	Номера деления на группы			Символы кода			Длина
		1	2	3	1	2	3	
x_1	0,5	0			0			1
x_2	0,25	1	0		1	0		2
x_3	0,125		1	0	1	1	0	3
x_4	0,125			1	1	1	1	3

Определим скорость передачи информации в примере 2 п. 7.3.3 при использовании кода Шеннона–Фано по формуле

$$\bar{I}(Y \to X) = \frac{H(X)}{\tau_{\text{ср}}} . \qquad (7.12)$$

Вычислим $\tau_{\text{ср}}$ (см. п. 4.6.3)

$$\tau_{\text{ср}} = \sum_{i=1}^{4} \tau(x_i)\, p(x_i) = \tau \times 0{,}5 + 2\tau \times 0{,}25 + 2 \times 3\tau \times 0{,}125 = 1{,}75\tau .$$

Подставим $\tau_{\text{ср}}$ в (7.12)

$$\bar{I}(Y \to X) = \frac{H(X)}{\tau_{\text{ср}}} = \frac{1{,}75}{1{,}75\tau} = \frac{1}{\tau} = \frac{1}{1{,}25 \times 10^{-3}} = 800 \text{ бод.}$$

Мы получили $C = \bar{I}$. Этот результат обусловлен специально подобранными исходными данными. В реальной ситуации вероятности подгрупп при делении не равны, поэтому, как следствие, $C > \bar{I}$.

Методика построения кода Хаффмена

Код Хаффмена можно строить таблично, подобно коду Шениона-Фано, или графически. Графическое построение более наглядно. Рассмотрим методику графического построения кода Хаффмена для источника сообщений, который описывается следующим рядом распределения вероятностей

$$X = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ 0{,}22 & 0{,}20 & 0{,}16 & 0{,}16 & 0{,}10 & 0{,}10 & 0{,}06 \end{pmatrix}$$

Выпишем буквы алфавита источника в столбец в порядке убывания их вероятностей. Две буквы с наименьшими вероятностями объединяем так, как это показано на рис.7.5 в одну вспомогательную букву, которой припишем суммарную вероятность объединяемых букв. Среди оставшихся букв, включая вспомогательную, вновь находим две буквы с наименьшими вероятностями и повторяем описанную выше процедуру до тех пор, пока не будет получена единственная вспомогательная буква с вероятностью, равной единице.

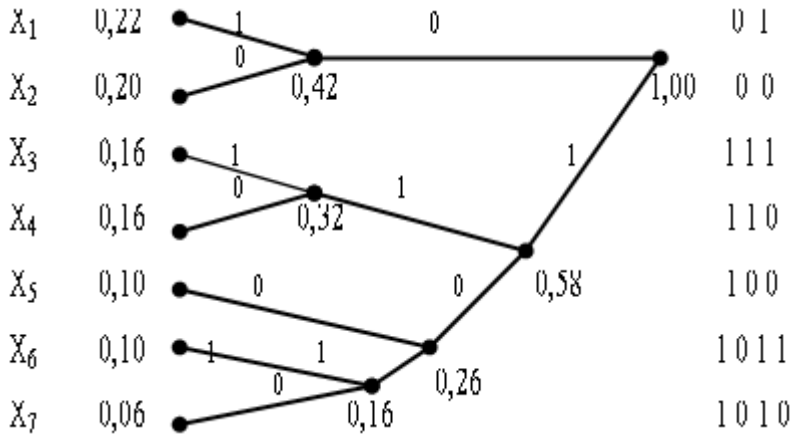

Рис. 7.5. Кодовое дерево Хаффмена

Далее, всем ребрам полученного таким образом кодового дерева приписываются символы «0» или «1» по следующему правилу: из двух ребер, выходящих влево из одной вершины, ребру, соединенному с вершиной, имеющей большую (или равную) вероятность, приписывается 1, а ребру, соединенному с вершиной, имеющей меньшую вероятность, — 0. Кодовая комбинация буквы составляется, начиная с вершины с вероятностью "единица", путем последовательной записи символов, находящихся на ребрах цепи, соединяющей эту вершину с соответствующей буквой алфавита.

В архиваторах, использующих метод Хаффмена, могут применяться алгоритмы, основанные либо на применении готовых частотных таблиц, либо такие таблицы строятся в процессе статистического анализа содержимого сжимаемого файла.

7.3.5. Современные методы сжатия

Рассмотренные выше коды Шеннона-Фано и Хаффмена, являются примерами классических методов сжатия данных, основанных на статистических свойствах источника информации; в этой связи их часто называют статистическими кодами. Заметим, что проблематика сжатия информации имеет такую же древнюю историю, как кодирование и шифрование. Вопросы сжатия информации мы уже частично обсуждали при изучении понятия избыточности источника сообщений (см. п. 4.6.2). Ярким примером дошенноновского эвристического метода сжатия информации является код Морзе.

Современные методы сжатия находят широкое применение при хранении и передаче текстовых, графических, аудио и видео данных. Объектами сжатия могут быть файлы, папки и диски. Методы сжатия информации делятся на методы сжатия без потерь (обратимое сжатие) информации и с потерями информации (необратимое сжатие).

Методы сжатия без потерь информации делятся на методы сжатия источников информации без памяти (метод Хаффмена, арифметическое сжатие, и др.) и методы сжатия источников информации с памятью (см. п. 4.6.1). К последним относятся алгоритмы Лемпеля-Зива (LZ), Лемпеля-Зива-Велча (LZW) и др. На основе названных и других методов сжатия работают различные программы сжатия данных (архиваторы). К форматам сжатия без потери информации относятся: .ZIP, .ARJ, .RAR, .LZH, .LH, .CAB и др. (сжатие любых типов данных); .GIF, .TIF, .PCX и др. (сжатие графических данных

Методы сжатия с потерей информации используются для архивации графических, аудио и видео данных, которые в несжатом виде требуют огромных объемов дискового пространства. С этими методами связано понятие качества сжатия, понимаемое как степень соответствия исходного и воспроизведенного изображения. Оценки качества сжатия обычно субъективны (см. темы квантование сигнала по уровню и дискретизация сигнала во времени). Форматы сжатия с потерей информации: .JPG для графических данных; .MPG для видеоданных; .MP3 для звуковых данных.

7.3.6. Пропускная способность дискретного канала с шумом

Для дискретного канала с шумом (см. рис.7.3) справедливы общие соотношения (7.1) и (7.2). Однако для определения пропускной способности канала связи в данном случае необходимо задать статистические свойства шума.

Если предположить, что шум в канале носит стационарный характер, то выход канала связи можно рассматривать как стационарный источник дискретных сообщений (см. п. 4.6.1), а соотношение для пропускной способности канала (7.1) после достаточно очевидных преобразований можно представить в виде двух равносильных равенств:

$$C = \sup\{\overline{H}(Y) - \overline{H}(Y \mid X)\}, \qquad (7.13)$$

$$C = \sup\{\overline{H}(X) - \overline{H}(Y \mid X)\}. \qquad (7.14)$$

Рассмотрим, например, (7.13). В этом выражении $\overline{H}(Y)$ и $\overline{H}(Y \mid X)$ соответственно скорость и условная скорость создания информации на входе канала связи, определяемые по формулам:

$$\overline{H}(Y) = \frac{H(Y)}{\tau_{CP}}; \qquad \overline{H}(Y \mid X) = \frac{H(Y \mid X)}{\tau_{CP}},$$

где τ_{CP} – средняя длительность получения одной буквы на выходе канала;

$H(Y) = -\sum_{K} p(Q_K) \sum_{l/K} p(Q_l \mid Q_K) \log p(Q_l \mid Q_K)$ – энтропия выхода канала связи;

$H(Y \mid X) = -\sum_{i=1}^{n} p(x_i) \sum_{K} p(Q_K) \sum_{l/K} p(Q_l \mid Q_K, x_i) \log_2 p(Q_l \mid Q_K, x_i)$ – условная энтропия выхода канала связи.

Здесь $p(Q_K)$ — вероятность пребывания канала в состоянии К;

$p(Q_L \mid Q_K)$ — условная вероятность перехода канала из состояния *K* в состояние *L*.

Задача на вычисление пропускной способности дискретного канала с шумом далеко не такая тривиальная, как для дискретного канала без шума. Наиболее простым случаем является *бинарный симметричный канал без памяти*. Использование для этого случая выражения (7.13) дает следующий результат

$$C = \frac{1}{\tau} \left[+ p_o \log_2 p_o + (1 - p_o) \log_2 (1 - p_o) \right], \qquad (7.15)$$

где p_o – вероятность ошибочного приема буквы канала.

Проиллюстрируем возможные ошибки при приеме сообщений в бинарном канале без памяти (рис. 7.6)

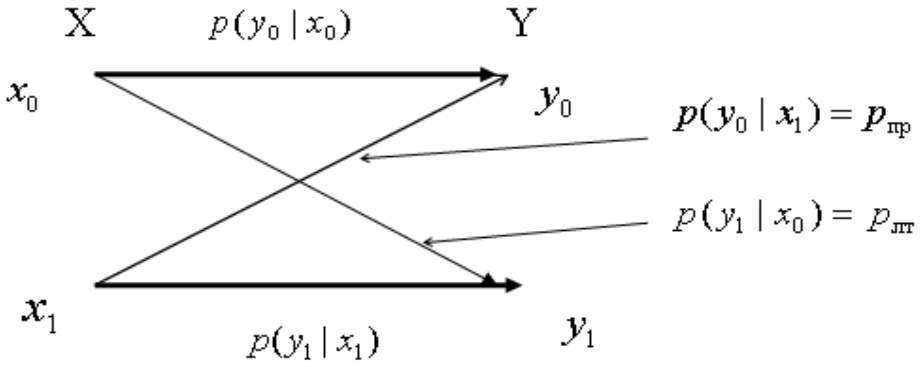

Рис. 7.6. Граф бинарного канала без памяти.

Приемник бинарного канала по результатам обработки на некотором интервале времени (0,Т) сигнала

$$y(t) = x(t) + \eta(t), \qquad (7.16)$$

где $\eta(t)$ — помеха, $x(t)$ — полезный сигнал (полезный сигнал может иметь значения: $x(t) = x_0$ или $x(t) = x_1$) должен принять решение либо в пользу x_0, либо в пользу x_1. При этом возможны два вида ошибок (у канала без памяти эти ошибки не зависят от состояния канала):

Ошибочное решение в пользу x_1 с условной вероятностью $p(y_1 \mid x_0) = p_{\text{лт}}$, которая называется *вероятностью ложной тревогой.*

Ошибочное решение в пользу x_0 с условной вероятностью $p(y_0 \mid x_1) = p_{\text{пр}}$, называемой *вероятностью пропуска сигнала.*

При этом вероятность ошибки в бинарном канале без памяти (если известны априорные вероятности сообщений на входе канала) определяется по формуле

$$p_{\text{ош}} = p(x_0)p_{\text{лт}} + p(x_1)p_{\text{пр}}. \qquad (7.17)$$

Заметим, что задача синтеза оптимального приемника (в данном случае бинарного обнаружителя), которая состоит в минимизации (7.17), т.е.

$$\min\, p_{\text{ош}} = p(x_0)p_{\text{лт}} + p(x_1)p_{\text{пр}}, \qquad (7.18)$$

была решена в 1946 В.А. Котельниковом в его теории потенциальной помехоустойчивости. Критерий (7.18) носит имя Котельникова, его также называют **критерием идеального наблюдателя.**

Кроме названного на практике применяются также следующие критерии оптимальности бинарного обнаружения:

Критерий минимального среднего риска

$$\min R = ap(x_0)p_{\text{лт}} + bp(x_1)p_{\text{пр}}, \qquad (7.19)$$

где a и b коэффициенты нежелательности ошибок $p_{\text{лт}}$ и $p_{\text{пр}}$.

Критерий минимальной взвешенной вероятности ошибки

$$\min Q = cp_{\text{лт}} + dp_{\text{пр}}, \qquad (7.20)$$

где c и d аналогичные предыдущему критерию, выбираемые из практических соображений коэффициенты.

Критерий Неймана-Пирсона

$$p_{\text{лт}} = const, \quad \min\, p_{\text{пр}}. \qquad (7.21)$$

Существуют и другие критерии оптимальности, в частности критерии, основанные на информационном подходе. Достоинство информационных критериев оптимальности, которые взаимосвязаны с традиционными критериями, состоит в том, что они позволяют оптимизировать процесс обработки информации в целом.

У бинарного симметричного канала без памяти помехи одинаково действуют на передаваемые буквы, поэтому имеет место

$$p_{\text{лт}} = p_{\text{пр}} = p_{\text{ош}} = p_{\text{о}}.$$

На рис 7.7 приведен график зависимости $C\tau$ от вероятности ошибки $p_{\text{о}}$ для бинарного симметричного канала без памяти, построенный по формуле (7.15)

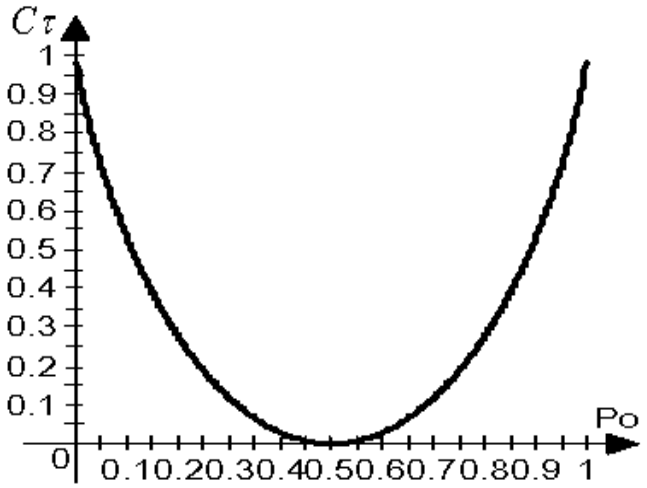

Рис. 7.7. График зависимости $C\tau$ от вероятности ошибки $p_{\text{о}}$.

Сравните этот график с графиком энтропии $H(X)$ в примере п. 4.2

7.3.7. Основная теорема Шеннона для дискретного канала с шумом

Основная теорема Шеннона для дискретного канала с шумом также как и аналогичная теорема для канала без шума является типичной теоремой существования и имеет прямое и обратное утверждение.

Прямое утверждение.

Если скорость создания информации источником дискретных сообщений равна:

$$\overline{H}(X) = C - \alpha,$$

где C – пропускная способность дискретного канала с шумом;

α – сколь угодно малая положительная величина,

то существует такой способ кодирования, при котором все вырабатываемые источником сообщения будут переданы, а вероятность ошибочной передачи любого сообщения может быть сколь угодно малой величиной, т.е.

$$p_{ОШ} < \eta,$$

где η - сколь угодно малая положительная величина.

Обратное утверждение.

Если $\overline{H}(x) > C$, то такого способа кодирования не существует.

7.3.8. Пропускная способность непрерывного канала с шумом

Для непрерывного канала с шумом справедливы те же, что и для дискретных каналов общие соотношения (7.1) и (7.2). Однако для определения пропускной способности канала связи в этом случае необходимо использовать статистические описания сигнала и шума в виде непрерывных случайных процессов. При традиционных предположениях о характере этих процессов выражения для пропускной способности непрерывного канала сводятся к внешне схожим с (7.13) и (7.14) выражениям.

Для важного частного случая, когда шум в канале является «белым» и гауссовым, пропускная способность непрерывного канала, как это показал К. Шеннон, определяется по следующей формуле:

$$C = F \log_2 (1 + \frac{P_C}{P_{Ш}}),$$

где F – полоса частот, в которой работает канал; P_C – средняя мощность сигнала; $P_{Ш}$ – средняя мощность шума.

К этой теме еще предстоит вернуться в курсе «Аппаратное и программное обеспечение ЭВМ и сетей» при обсуждении свойств аналоговых модемов.

7.3.9. Основные методы повышения помехоустойчивости передачи данных

Существует следующие основные методы повышения надежности (помехоустойчивости) передачи данных:

Методы оптимального приема сигналов.

Методы помехоустойчивого кодирования.

Методы, основанные на использовании обратной связи между выходом и входом канала.

Надежная передача данных в пространстве предполагает использование всех трех видов названных методов. Методы оптимального приема сигналов применяются для минимизации ошибок при приеме элементарных кодовых посылок («0» и «1»). Для этого приемник сигнала осуществляет оптимальную, в смысле одного из критериев (7.18) — (7.21), обработку смеси $y(t) = x(t) + \eta(t)$, которая образуется на выходе непрерывного канала связи (НКС).

Центральную роль в борьбе с помехами играют методы помехоустойчивого кодирования, развитие которых по существу было инициировано основанной теоремой Шеннона для дискретного канала с шумом, утверждающей, что существует способ кодирования (система кодирования, по Шеннону), позволяющий практически безошибочно передавать информацию по дискретному каналу со скоростью, не превышающей пропускную способность этого канала. Но тогда возникает вопрос: зачем нужна обратная связь? Остановимся на этом вопросе более подробно.

Как уже отмечалось в п. 7.1. передачу информации можно рассматривать в двух аспектах: как передачу во времени или хранение информации и как передачу информации в пространстве или связь. В обоих этих случаях применимы теоремы Шеннона.

Для дискретного канала без шума методологические детали, связанные со сжатием данных в обоих случаях незначительны. При борьбе же с шумом различие в технологиях носит принципиальный характер. В случае хранения информации используются только корректирующие коды, исправляющие ошибки, что обусловлено простой и стабильной статистикой ошибок в средах систем хранения информации. При передаче же данных в пространстве используются коды обнаруживающие ошибки, а для их исправления применяется обратная связь. Такая схема вызвана сложным нестационарным характером «поведения» помех в линиях связи, что не позволяет эффективно использовать свойства корректирующих кодов.

Поясним сказанное. Дело в том, что корректирующий код может исправлять лишь ограниченное множество возможных ошибок и оптимальным он будет лишь в том случае, если сумма вероятностей ошибок этого множества будет максимальной (по теореме Шеннона она должна быть сколь угодно близка к единице). Однако сложный нестационарный характер помех в линии связи как раз и не позволяет реализовать выполнение этого условия. Если же исправлению будут подвергаться не самые вероятные ошибки, то эффект может оказаться даже обратным. С этим обстоятельством исследователи столкнулись еще на заре передачи данных.

На рис 7.8 приведена схема передачи данных с обратным каналом, отражающая только сам принцип использования обратной связи. На практике возможны различные подходы к организации взаимодействия между передающей и приемной стороной, но сущность такой схемы состоит в том, что приемник прямого канала должен информировать передающую сторону (отправить квитанцию) по обратному каналу о правильном получении порции данных. В противном случае передатчик прямого канала вынужден будет при определенных условиях повторить передачу этой порции данных (так «работает» в скользящем окне протокол TCP). Решение о безошибочной передаче принимается приемником ПК посредством корректирующего кода, при этом как прямой, так и обратный канал, как правило, являются виртуальными.

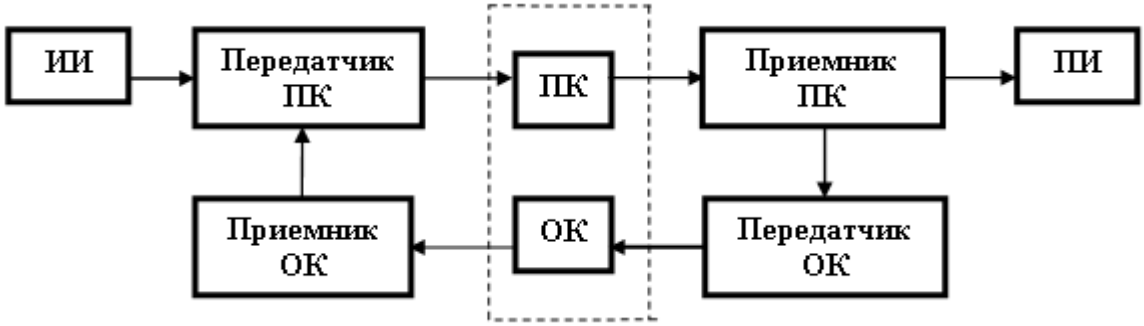

Рис. 7.8. Схема передачи данных с обратным каналом: ИИ — источник информации; ПИ — получатель информации; ПК – прямой канал; ОК – обратный канал.

Организация обратной связи при передаче информации в современных сетевых технологиях сложна и многогранна. Она реализуется как аппаратно, так и программно, как на канальном, так и на транспортном уровне модели OSI (см. п. 2.2), а также может осуществляться самими прикладными программами, т.е. на прикладном уровне. Технология обратной связи в локальных, городских и глобальных сетях рассматривается в курсе «Аппаратное и программное обеспечение ЭВМ и сетей».

Борьба с помехами посредством корректирующих кодов связана с использованием специально организованной избыточности (см. п. 4.6.2). Рассмотрим общие принципы использования избыточности при построении корректирующих (помехоустойчивых) кодов.

Пусть вектор $\vec{u} = (u_1,..u_i,..,u_L)$, поступающий на вход кодека (см. рис. 7.9), представляет собой последовательность, состоящую из L двоичных символов, а

вектор $\vec{x}=(x_1,..x_j,..,x_N)$ – последовательность N двоичных символов на выходе кодека ($N>L$), осуществляющего помехоустойчивое кодирование.

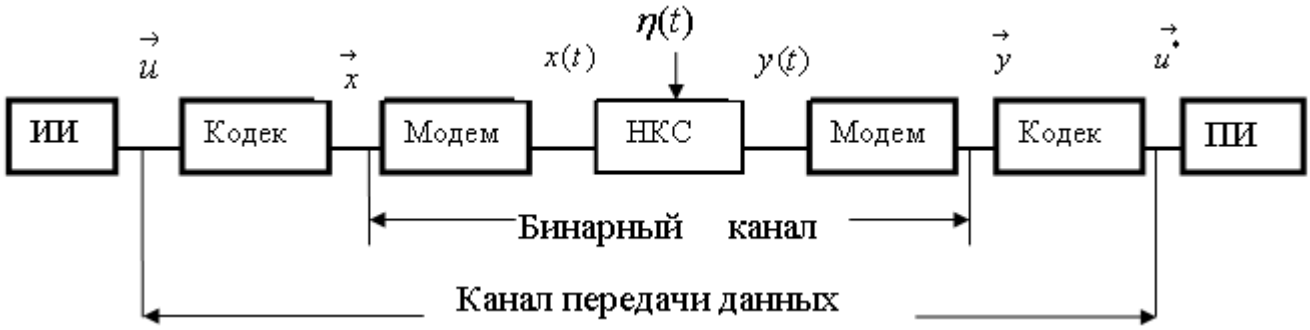

Рис. 7.9. Обобщенная схема канала передачи данных: кодек (КПД) — кодер/декодер; модем — модулятор/демодулятор; НКС – непрерывный канал связи.

На вход кодека приемной стороны канала передачи данных поступает вектор $\vec{y}=(y_1,..y_j,..,y_N)$, который преобразуется (декодируется) кодеком в вектор $\vec{u}^*=(u_1^*,..u_i^*,..,u_L^*)$. Вектора $\vec{u}^*=(u_1^*,..u_i^*,..,u_L^*)$ и $\vec{u}=(u_1,..u_i,..,u_L)$ из-за воздействия помех могут не совпадать. Вектор $\vec{u}^*=(u_1^*,..u_i^*,..,u_L^*)$ поступает получателю информации.

Свойства помехоустойчивости у кода возникает только при условии $N>L$, поскольку в этом случае для передачи по бинарному каналу из полного множества 2^N кодовых комбинаций (слов) вектора $\vec{x}=(x_1,..x_j,..,x_N)$ используются только 2^L слов (как правило, имеет место $2^N >> 2^L$), остальные же кодовые комбинации являются запрещенными (ошибочными). При этом можно так выбрать 2^L слов, что будет обеспечена максимальная помехозащищенность.

Поясним сказанное предельно простым примером. Предположим, что L=1 (т.е. источник информации передает только два сообщения: $u_1=0$ и $u_2=1$). Выберем N=10, а для передачи по бинарному каналу будем использовать: $x_1=0000000000$ и $x_2=1111111111$. Мы выбрали кодовые слова, которые отстоят друг от друга на максимальном расстоянии по Хэммингу d=10 (расстояние по Хэммингу определяется как число единиц при сложении двух векторов по модулю два). Очевидно, что на приемной стороне все комбинации вектора $\vec{y}=(y_1,..y_j,..,y_N)$, за исключением слов $y_1=0000000000$ и $y_2=1111111111$ будут восприняты как ошибки (количество таких ошибок составит $2^{10}-2=1022$). В данном случае единственными не обнаруживаемыми ошибками будут десятикратные ошибки, переводящие $x_1=0000000000$ в $x_2=1111111111$, и наоборот. Подумайте, как организовать исправление ошибок.

Построение корректирующих кодов для больших L и N требует применение математического аппарата, основанного на теории групп, полей и многочленов.

7.3.10. Помехоустойчивое кодирование

7.3.10.1 Основные разновидности корректирующих кодов

Роль помехоустойчивого кодирования в деле повышения надежности передачи данных рассмотрена в п. 7.3.9 (см. также п.4.6.2, п. 6.3, п. 7.1), там же обсуждены и особенности применения корректирующих кодов. Заметим, что понятие помехоустойчивый код шире понятия корректирующий код; но корректирующий код может как обнаруживать и исправлять ошибки (т.е. корректировать; что характерно для хранения данных), так и только обнаруживать ошибки (как при передаче данных, см. п. 7.3.9). В этой связи именно корректирующие коды получили широкое распространение, отдельные их разновидности нашли применение в стандартах на локальные и глобальные сети.

Корректирующие коды многочисленны и разнообразны. Классификация корректирующих кодов, отражающая состояние этой области в 60-е годы прошлого столетия, приведена в книге А. Харкевича [10].

Корректирующие коды подразделяются на два класса: **блочные** и **непрерывные**. В блочных кодах каждому сообщению ставится с соответствие блок, состоящий из N символов (см. п. 7.3.9); непрерывные коды (рекуррентные или цепные коды) образуют непрерывную последовательность символов, не подразделяемую на блоки.

Блочные коды подразделяются на **линейные** и **нелинейные**. Наибольшее распространение получили двоичные линейные коды или **групповые коды.** Важнейшей разновидностью последних в практическом отношении являются циклические коды. Именно эти коды благодаря простоте технической реализации нашли применение в стандартах на локальные и глобальные сети.

Циклический код характеризуется тем, что если $(x_1, x_2, x_3, ..., x_j, .., x_{N-1}, x_N)$ является разрешенным кодовым словом, то и образованное посредством циклического сдвига кодовое слово $(x_2, x_3, ..., x_j, .., x_{N-1}, x_N, x_1)$ также будет являться разрешенным словом. По части практической значимости циклических кодов отметим, например, следующее. Кадр самого распространенного стандарта локальных сетей IEEE 803 содержит 4-х байтную контрольную последовательность кадра (CRC — Cyclic redundancy code), которая предназначена для обнаружения ошибок. Формирование кодового вектора и проверка на наличие ошибок осуществляется с использованием образующего полинома: $G(x) = x^{32} + x^{26} + x^{23} + x^{22} + x^{16} + x^{12} + x^{11} + x^{10} + x^8 + x^7 + x^5 + x^4 + x^2 + x + 1$. Вероятность ошибки в принятом кадре составляет примерно 2^{-32}.

При хранении информации, как уже отмечалось, используются **Коды Рида — Соломона,** которые также являются разновидностью циклических кодов**,** но это недвоичные циклические коды. Они позволяют исправлять ошибки в блоках (байтах) данных. Коды Рида — Соломона является частным случаем кодов **Боуза — Чоудхури — Хоквингема** (БЧХ), отличительная особенность которых состоит в возможности построения кода с минимальным расстоянием не меньше заданного.

Изучение названных кодов предполагает ознакомление с понятиями группы и поля.

7.3.10.2. Понятия группы и поля

Понятие группы.

Группой называется множество элементов ***a, b, c, …*** с операцией обозначаемой символом "точка" ($\bullet$) удовлетворяющее следующим свойствам:

1. Для любых элементов a, b принадлежащих данному множеству $a \bullet b$ также принадлежит этому множеству.

Примечание. Множество элементов, обладающее таким свойством, называется замкнутым по данной операции.

2. Для любых элементов a, b, c принадлежащих множеству выполняется ассоциативный закон:

$$a \bullet (b \bullet c) = (a \bullet b) \bullet c.$$

3. Множеству принадлежит нейтральный элемент ***e,*** такой, что имеет место $a \bullet e = e \bullet a = a$ для всех a принадлежащих множеству.

4. Для любого элемента a множества существует принадлежащим множеству противоположный (обратный) элемент удовлетворяющий соотношениям:

$$a \bullet a^{-1} = a^{-1} \bullet a = a.$$

Если в дополнение к этим свойства в группе для любых элементов a, b принадлежащих данному множеству выполняется и коммутативный закон:

$$a \bullet b = b \bullet c,$$

то такая группа называется коммутативной или Абелевой.

Примеры групп. Множество целых чисел с операцией сложения (нейтральный элемент ─ нуль; противоположными элементами для положительных чисел являются соответствующие им отрицательные числа, и наоборот). Множество действительных чисел с операцией сложения.

Понятие поля.

Полем называется множество, по крайней мере, двух элементов, замкнутое по двум операциям называем сложением (+) и умножением **($*$)** и обладающее следующими свойствами:

1. Множество элементов образует абелеву группу по операции сложения.

2. Множество элементов, исключая нулевой (нейтральный) элемент по операции сложения, образует абелеву группу по операции умножения.

3. Для всех элементов множества выполняется дистрибутивный закон:

$$(a+b) * c = a * c + b * c$$

Поле, содержащее конечное число элементов, называется полем Галуа. В дальнейшем поле с q элементами будем обозначать ***GF(q).***

Примеры полей. Множество действительных чисел с традиционными операциями сложения и умножения (заметим, что в данном случае нуль не является элементом группы по операции умножения). Множество двоичных чисел 0 и 1 с операцией сложения по модулю два и обычным умножением (в этом поле противоположным элементом является сам этот же элемент).

7.3.10.3. Групповые коды

Представим обобщенную схему канала передачи данных (см. рис. 7.9) эквивалентной схемой, приведенной на рисунке 7.10. На этой схеме бинарный канал заменен эквивалентной схемой в виде сумматора по модулю два, так что вектор $\vec{y} = \vec{x} \oplus \vec{e}$. В данной схеме помеха $\eta(t)$ (см. рис. 7.9) заменена эквивалентным бинарным вектором ошибки $\vec{e} = (e_1, ...e_j, .., e_N)$, который образует сумму по модулю два с вектором $\vec{x} = (x_1, ...x_j, .., x_N)$.

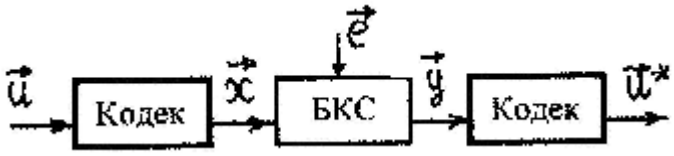

Рис. 7.10. Эквивалентная схема канала передачи данных. БКС— бинарный канал связи.

Итак, имеем. На вход кодека поступает вектор $\vec{u} = (u_1^*, ...u_i^*, .., u_L^*)$. На выходе кодека формируется кодовый вектор $\vec{x} = (x_1, ...x_j, .., x_N)$, поступающий на вход БКС. На выходе БКС образуется вектор $\vec{y} = (y_1, ...y_j, .., y_N) = \vec{x} \oplus \vec{e}$, который поступает на вход кодека приемной стороны. Этот кодек декодирует вектор $\vec{y} = (y_1, ...y_j, .., y_N)$ и формирует информационный вектор $\vec{u}^* = (u_1^*, ..u_i^*, .., u_L^*)$. Вектор $\vec{u}^* = (u_1^*, ..u_i^*, .., u_L^*)$ из-за влияния помех может не соответствовать информационному вектору $\vec{u} = (u_1^*, ...u_i^*, .., u_L^*)$, поступающему на вход канала передачи данных.

Заметим, что записанные в общем виде вектора применительно к схеме рис. 7.9 являются бинарными, но последующее ниже определение группового кода применимо и к недвоичному алфавиту.

Групповым (N,L) кодом называется такой код, в котором кодовое слово $\vec{x} = (x_1, ...x_j, .., x_N)$ соответствующее $\vec{u} = (u_1^*, ...u_i^*, .., u_L^*)$ образуется по правилу

$$x_i = \sum_{l=1}^{L} u_l g_{l,i}; \qquad 1 \le i \le N, \tag{7.22}$$

где u_l, $g_{l,i}$ являются элементами поля GF(q), а сложение и умножение является операциями в *GF(q).*

Частным случаем группового кода является систематический групповой код, у которого первые L компонент кодового слова $\vec{x} = (x_1, ...x_j, .., x_N)$ соответствуют кодовому слову $\vec{u} = (u_1^*, ...u_i^*, .., u_L^*)$. Приведем определение этого кода.

Систематическим групповым (N,L) кодом называется такой код, в котором кодовое слово $\vec{x} = (x_1, ...x_j, .., x_N)$ соответствующее $\vec{u} = (u_1^*, ...u_i^*, .., u_L^*)$ удовлетворяет соотношениям

$$x_i = u; \qquad 1 \le i \le L, \tag{7.23}$$

$$x_i = \sum_{l=1}^{L} u_l g_{l,i}; \qquad L+1 \le i \le N, \tag{7.24}$$

где u_l, $g_{l,i}$ являются элементами поля GF(q), а сложение и умножение является операциями в *GF(q).*

Нетрудно заметить, что (7.22) представляет собой умножение вектора строки на матрицу, поэтому (7.22) можно представить так

$$\vec{x} = \vec{u} * G,$$

где G — порождающая матрица группового кода, которая имеет следующий вид

$$G = \begin{pmatrix} g_{11} & g_{12} & \cdots & g_{1j} & \cdots & g_{1N} \\ g_{21} & g_{22} & \cdots & g_{2j} & \cdots & g_{2N} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ g_{L1} & g_{L2} & \cdots & g_{Lj} & \cdots & g_{LN} \end{pmatrix}. \tag{7.25}$$

Для систематического группового кода, учитывая (7.23), (7.23) порождающая матрица G принимает вид

$$G = \begin{pmatrix} 1 & 0 & \ldots & 0 & g_{1,L+1} & \cdots & g_{1,N} \\ 0 & 1 & \ldots & 0 & g_{2,L+1} & \cdots & g_{2,N} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 1 & g_{L,L+1} & \cdots & g_{L,N} \end{pmatrix}. \tag{7.26}$$

Проверка кодового слова на наличие ошибок осуществляется с помощью проверочной матрицы ***H,*** которая для систематического группового кода имеет вид

$$H = \begin{pmatrix} g_{1,L+1} & g_{1,L+2} & \cdots & g_{1,N} \\ \ldots & \ldots & \ldots & \ldots \\ g_{L,L+1} & g_{L,L+2} & \cdots & g_{L,N} \\ 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 1 \end{pmatrix}. \tag{7.27}$$

Сравнивая матрицы (7.26) и (7.27) замечаем, что имеет место $\vec{x} * H = \vec{0}$, т.е. результатом умножения кодового слова $\vec{x}_{=(x_1,\ldots x_j,\ldots,x_N)}$ на проверочную матрицу H является вектор нуль, размерность которого $N - L$.

Рассмотрим произведение $\vec{y} * H$. В силу дистрибутивности элементов поля *GF(q)*, имеем

$$\vec{y} * H = (\vec{x} \oplus \vec{e}) * H = \vec{x} * H \oplus \vec{e} * H = \vec{0} \oplus \vec{e} * H = \vec{e} * H = \vec{s} \ . \qquad (7.28)$$

Здесь вектор $\vec{s}$ – синдромом или опознавателем ошибки. Синдром $\vec{s}$ может принимать 2^{N-L} значений, $2^{N-L}-1$ из которых свидетельствуют об ошибке при передаче данных. Именно такое количество ошибок, т.е. $2^{N-L}-1$, можно исправить, сопоставив конкретным ошибкам определенные синдромы; но ошибок намного больше, поэтому возникает проблема выбора множества исправляемых ошибок (см. п.7.3.9; там эта проблема обсуждена).

Дальнейшее изложение продолжим на примере систематического группового кода (6,3), порождающая матрица которого имеет вид

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix} \qquad (7.29)$$

Используя (7.29), выполним матричную операцию $\vec{x} = \vec{u} * G$ для множества информационных кодовых слов $\vec{u}$. Результаты сведены в таблицу 7.2 (здесь и далее вычисления выполнены с использование лабораторной работы №8).

Таблица 7.2. Соответствие кодовых слов $\vec{u}$ и $\vec{x}$.

$\vec{u}$	000	001	010	011	100	101	110	111
$\vec{x}$	000000	001110	010101	011011	100011	101101	110110	111000

Проверочная матрица, соответствующая порождающей матрице (7.29), имеет вид

$$H = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} . \qquad (7.30)$$

Множество векторов $\vec{y}$ составляет: $2^6 = 64$, 8 из них соответствуют множеству векторов $\vec{x}$ (см. табл. 7.2) и являются разрешенными, остальные 56 кодовых векторов являются ошибочными. Множество синдромов $\vec{s}$, см. (7.28), составляет 2^3=8, один из восьми синдромов соответствует правильному приему. Таким образом, из 56 ошибок можно исправить только 7. Если имеется вероятностное описание ошибок, то, очевидно, что следует отобрать такие ошибки, которые обеспечивают максимальную суммарную вероятность. Если вероятностное описание отсутствует, то произвольный выбор ошибок может ухудшить качество передачи.

Предположим, что бинарный канал связи (рис. 7.10) является симметричным каналом без памяти (в этом случает достаточно знать лишь вероятность

ошибочного приема одного бита p_o; см. п. 7.3.6). В данном случае исправлению подлежат, прежде всего, однократные ошибки. Рассмотрим технологию исправления ошибок.

Однократным ошибкам ставятся в соответствие их синдромы (см. табл. 7.3; однократных ошибок всего шесть, поэтому добавлена двукратная ошибка 100100). Кодек приемной стороны, вычисляет синдром и если этот синдром отличается от нулевого вектора, то исправляет ошибку, прибавляя к вектору $\vec{y}$ вектор $\vec{e}$, соответствующий найденному синдрому.

Таблица 7.3. Соответствие векторов ошибок и синдромов.

$\vec{e}$	000000	100000	010000	001000	000100	000010	000001	100100
$\vec{s}$	000	011	101	110	100	010	001	111

Пусть, например, при передаче кодового слова $\vec{x}$=101101 на вход кодека приемной стороны поступает $\vec{y}$=101111. Кодек находит синдром $\vec{s} = \vec{y} * H =$ 010. Этому синдрому в табл. 7.3 соответствует кодовое слово ошибки $\vec{e}$ = 000010. Кодек исправляет ошибку: $\vec{y} \oplus \vec{e} = (101111) \oplus (000010) = 101101$.

Но синдрому 010 могут соответствовать ошибки и другой кратности, поэтому выбор модели ошибок в канале является ответственным делом. Рассмотренная же в данном примере модель является предельно упрощенной и не соответствующей реальным каналам.

Разновидностью групповых (линейных) кодов являются циклические коды, практическая значимость которых обсуждена в п. 7.3.10.1. У таких кодов строки порождающих матриц связаны дополнительным условием цикличности, поэтому количество циклических (N,L) кодов значительно меньше общего числа групповых (N,L) кодов. Циклическим (N,L) кодом над полем *GF(q)* называется такой групповой код, у которого при любом циклическом сдвиге какого-либо кодового слова получается другое кодовое слово. Так что если $(x_1, x_2, x_3, ..., x_N)$ — кодовое слово, то и $(x_2, x_3, ..., x_N, x_1)$ кодовое слово. Детальное рассмотрение циклических кодов выходит за рамки нашего курса; вопросы же практического использования этих кодов рассматриваются в курсе "Аппаратное и программное обеспечение ЭВМ и сетей".

Тема 8. Обработка информации

8.1. Понятие обработки информации

Понятие обработка информации является очень широким и многоаспектным; с ним связаны два родственных понятия: обработка данных и переработка информации.

Если обработка данных (Data processing), понимаемая как процесс выполнения последовательности операций над данными, практически

отождествляется с понятием обработки информации, то переработка информации имеет более широкую трактовку.

Под переработкой информации обычно понимают аналитико-синтетическое или эвристическое преобразование информации, связанное с систематизацией, анализом, обобщением, прогнозированием и другой интеллектуальной деятельностью.

По виду обработки, обработку информации подразделяют на аналоговую, цифровую и символьную.

Аналоговая обработка применяется, например, при восприятии и преобразовании информации. Аналоговую обработку часто называют обработкой сигнала.

Цифровая обработка связана с выполнением вычислительных операций (логических и арифметических) над переменными, векторами, матрицами, многомерными массивами и т.д.

Символьная обработка информации связана с выполнением нечисловых операций над такими объектами как файлы, записи, поля, иерархии, сети, отношения и т.д.

Объектами обработки могут быть текстовые материалы, изображения, речь и т.д. Обработка данных осуществляется прикладными процессами в интерактивном или фоновом режиме и в настоящее время может осуществляться сложными ассоциациями информационных сетей, включающих тысячи компьютеров.

Конкретная обработка информации предполагает, прежде всего, наличие определенной цели, затем необходим алгоритм обработки («конечный набор предписаний, определяющий решение задачи посредством конечного количества операций», ISO 2382/1-84), основанный на определенном методе, и, наконец, технологические средства обработки (ручные, автоматизированные или автоматические), реализующие заданную цель.

Очевидно, что классификация обработки информации предполагает рассмотрение всех названных аспектов. Однако классификация по алгоритмам и средствам обработки выходит за рамки нашего курса. К тому же на современном этапе это непосильная для одной дисциплины задача.

Заметим, что обработка данных по существу сама по себе является ни чем иным как преобразованием информации; но в отличие от ранее рассмотренных видов преобразования информации она (обработка данных) не имеет собственной цели.

8.2. Технологический процесс обработки данных

Технологический процесс обработки данных обычно отображается технологической сетью обработки данных, которая состоит из взаимосвязанных по входам и выходам технологических операций (рис. 8.1), к таким операциям могут относиться сбор и регистрация, ввод и накопление информации и т.д.

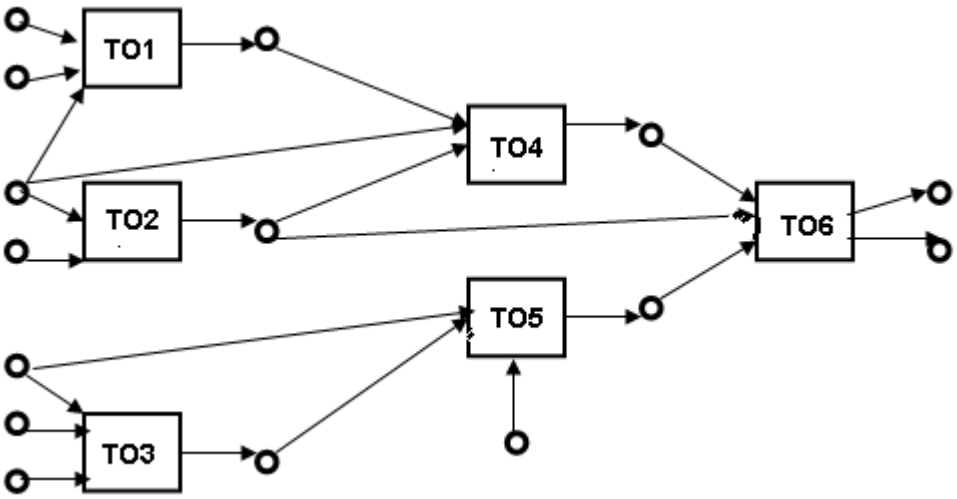

Рис. 8.1. Фрагмент технологической сети обработки данных: ТО — технологическая операция; — входная и выходная информация (данные)

Рассмотрим интерпретацию понятия технологическая операция обработки данных (рис. 8.2).

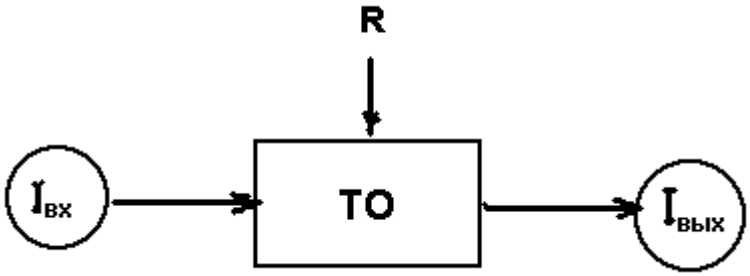

Рис. 8.2. Технологическая операция обработки данных".

На рисунке 8.2 введены следующие обозначения:

$I_{вх}$ – структурированная совокупность входных данных на машинном (или не машинном) носителе ("входной продукт").

$I_{вых}$ – совокупность выходных данных (выходной "продукт").

R – вещественно–энергетические и информационные ресурсы, которые включают: материальные и трудовые ресурсы, аппаратно-программные средства, базы данных и знаний.

ТО – технологическая операция обработки данных, представляющая собой информационный процесс преобразования входных данных в выходные, реализуемый по заданной (сформированной) программе.

8.3. Типовые операции обработки информации

Обработка данных в широком смысле включает следующие виды типовых операций обработки информации:

1. Сбор и регистрация.
2. Перенос на машинные носители.
3, Ввод и контроль.
4. Накопление.
5. Сортировка.

6. Обработка.
7. Корректировка.
8. Вывод.

1. Операции сбора и регистрации данных в АСОИ отличаются большой трудоемкостью и требуют участия специалистов различных подразделений. Эти операции могут выполняться ручным, автоматизированным и автоматическим способами.

При ручном способе специалисты соответствующих служб обеспечивают регистрацию качественных и количественных характеристик, отражающих результаты работы исполнителей или оборудования на бумажном носителе без применения технических средств. Этот способ характеризуется наибольшей трудоемкостью и предполагает последующий перенос данных на машинные носители. Такой способ сильно снижает эффективность технологического процесса обработки информации в целом.

Автоматизированный способ сбора и регистрации информации предполагает использование технических средств, позволяющих регистрировать данные на машинном носителе, либо непосредственно вводить в их компьютер.

Автоматический способ сбора и регистрации информации позволяет формировать исходные данные без участия человека. Устройства регистрации устанавливаются непосредственно в местах возникновения информации. Специальные датчики измеряют время, давление, температуру, подсчитывают количество деталей и т.д. Полученная информация может выводиться на машинные носители, либо передаваться по каналам связи для последующей обработки.

2. Операция переноса данных на машинные носители применяется в случае, если сбор и регистрация данных выполняется вручную.

3. Ввод и контроль данных связан с формированием автономных или предбазовых файлов.

4. Накопление данных представляет собой процесс периодического добавления данных в существующие файлы с целью получения сведений за определённый промежуток времени. Эту операцию можно считать частным случаем корректировки данных.

5. Сортировка обеспечивает расположение данных в определённом порядке. В АСОИ 25% машинного времени тратится на сортировку. Различают следующие виды сортировки:

5.1. Упорядочение – процесс, в результате которого записи сортируемого файла располагаются в порядке возрастания или убывания ключевых признаков.

5.2. Распределение – процесс разнесения записей сортируемого файла по группам с одинаковым значением признаков.

5.3. Слияние (объединение) – процесс, в результате которого несколько упорядоченных файлов сливаются в один с записями, расположенными в определённой логической последовательности.

6. Собственно обработка данных (обработка данных в узком смысле слова) включает в себя выполнение вычислительных операций. Обработка данных в

АСОИ характеризуется тем, что только примерно 20% общего машинного времени затрачивается на алгебраическую обработку, а остальные 80% – на управление данными.

7. Корректировка – процесс модификации сформированных файлов данных, позволяющий поддерживать их соответствие реально существующим условиям обработки. При корректировке могут выполняться следующие операции: добавление, исключение, изменение. Корректировка связана с поиском местоположения данных.

8. Вывод данных. По способу отображения результатов информации различают вывод данных на бумагу, на машинные носители и на видеотерминальные устройства.

Тема 9. Хранение информации

До боли знакомое Вам со школьной скамьи хранение информации (папирус, пергамент… винчестеры, CD- и DVD-диски, флэш-память и т.д.) как процесс и тем более как современная технология является едва ли не самым сложным видом среди рассмотренных выше информационных процессов. Это обусловлено тем, что любой из ранее рассмотренных процессов, так или иначе, связан с хранением информации; но наиболее сложны процессы хранения информации в современных корпоративных системах.

Напомним, что в пирамиде архитектуры ИТ (см. п. 1.6) архитектуре данных отведен второй ярус, который "покоится" на инфраструктуре объекта автоматизации, состоящей из аппаратного, программного обеспечения и сетевой инфраструктуры (это первый уровень). Выше находятся: архитектуры интеграции, приложений и бизнеса. Существуют и другие подходы к разработке ИТ-архитектуры. Например, на втором уровне — архитектура приложений, на третьем — архитектура данных, на четвертом — бизнес-архитектура. Но, так или иначе, все уровни имеют отношение к хранимой информации. При этом наиболее "прозрачным" является первый уровень, который охватывает средства хранения данных на внутренних (оперативная память, кэш-память и память адаптеров) и внешних носителях (жесткие магнитные диски, CD- и DVD-диски, флэш-память, Zip-диски, магнитооптические диски и т.д.) , а также определенные программные средства (это физический и логический или частично логический уровень данных). К современным системам хранения данных, которые организуются на вышележащих уровнях ИТ-архитектуры, предъявляются высокие требования, обусловленные ценностью хранимой информации.

Процесс хранения информации, как отмечалось в п. 7.1, имеет много общего с процессом передачи информации (шифрование, сжатие, помехоустойчивое кодирование). Вместе с этим между обсуждаемыми видами процессов имеются принципиальные различия. Рассмотрим эти различия.

Передача информации, как правило, связана с синтаксической стороной информации, поэтому она как технология сравнительно легко формализуется и, хотя передача информации является все же информационной технологией, ее можно рассматривать и как отдельный вид технологии. Именно так и обстоит дело, когда употребляется термин «информационно-коммуникационные

технологии». Для уяснения правомерности такого подхода представьте себе технологию пересылки обычных писем, бандеролей, да и вообще любого груза. Ну, чем для оператора связи отличается пересылка вышеназванных объектов от передачи байтов? Только сложностью технологии, которая лишь частично «ложится» на оператора электросвязи. Эта технология регламентирована многочисленными стандартами, специфицирующими форматы пакетов и кадров, в полях данных которых (как в контейнерах) размещаются фрагменты передаваемых файлов (сборка полученных фрагментов осуществляется на приемной стороне). Современная технология передачи данных сложна, требует отдельного рассмотрения, и будет изучаться в курсе «Аппаратное и программное обеспечение сетей».

Специфика хранения информации обусловлена тем, что, как отмечается в многочисленных источниках, стоимость самой информации современного предприятия дороже средств ее обработки, передачи и хранения, а расходы на поддержку постоянно увеличивающихся объемов данных превышают половину общих расходов на информационную технологию. Но это внешняя сторона дела, сущность же в том, что информация является важнейшим ресурсом современного предприятия (по данным Gartner, 43% компаний потерпевших крупную необратимую потерю корпоративных данных прекратили свою деятельность). Состояние данного ресурса и доступность к нему аналитиков и руководителей предприятия в современных рыночных условиях напрямую связаны с конкурентоспособностью продукции предприятия. Так что в отличие от передачи информации (где превалирует синтаксический аспект) в данном случае важную роль играют семантический и прагматический аспекты информации, а это требует создания адекватной структуры хранения данных, позволяющей осуществлять необходимый поиск и анализ информации. В этой связи получили распространение так называемые хранилища данных.

Чем же отличается ***хранилище данных*** (Data Warehouse) от обычной базы данных? Реляционная база данных состоит из совокупности связанных таблиц с короткими записями и предназначена для выполнения коротких транзакций. Такие базы данных непригодны для выполнения аналитических запросов, анализа накапливаемой информации и формирования отчетов, поэтому на основе традиционных баз данных и других форм создается хранилище данных ориентированное на технологию **OLAP** (On-Line Analytical Processing: оперативный анализ информации). В основе этой технологии лежит идея многомерной модели данных. OLAP представляет собой удобное и быстродействующее средство анализа и просмотра информации, основанное на естественной и интуитивно понятной модели данных в виде многомерного куба, осями координат которого служат параметры анализируемого процесса. Технология OLAP применяется для анализа тенденций и закономерностей, а также принятия управленческих решений.

OLAP может применяться везде, где имеет место анализ многофакторных данных. Вот некоторые сферы применения этой технологии:

продажи, закупки, цены, движение денежных средств, бухгалтерские счета, финансовая отчетность;

потребление электроэнергии, использование помещений, потребление расходных материалов;

текучесть кадров в регионе, текучесть кадров на предприятии, уровень жизни населения, заболеваемость населения;

результаты социологических опросов и выборов;

пассажирские перевозки, грузовые перевозки, простои транспорта;

и т.д.

В заключение отметим, что с понятием хранилище данных связана не только технология OLAP, но и другие системы и средства анализа и представления данных. Например, такие как *информационная система руководителя* (ИСР), *системы поддержки принятия решений* (СППР), которые появились еще до концепции хранилища данных. Наконец, понятие *технология хранилища данных* включает такие объекты как *хранилища данных*, *витрины данных* и соответствующее программное обеспечение. Витрины данных используются для увеличения скорости работы системы. В витринах содержится выгружаемая из хранилища данных информация, ориентированная на определенную группу пользователей. В небольших хранилищах само хранилище одновременно является и витриной.