МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«ОРЛОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»
ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«ОРЛОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

ТЕОРИЯ ИНФОРМАЦИИ

*Рекомендовано Учебно-методическим объединением вузов
Российской Федерации по образованию в области историко-архивоведения
в качестве учебника для студентов высших учебных заведений,
обучающихся по специальностям
090103 «Организация и технология защиты информации»
и 090104 «Комплексная защита объектов информатизации»*

Орел 2010

УДК 621.391(075)
ББК 32.811я7
Т33

Рецензенты:

доктор технических наук, профессор учебно-научного комплекса «Автоматизированные системы и информационные технологии» Академии Государственной противопожарной службы МЧС России
Н.Г. Топольский,

доктор физико-математических наук, декан факультета «Компьютерные системы и информационные технологии» Российского нового университета
А.С. Крюковский

Т33 **Теория информации:** учебник для вузов / В.Т. Еременко, В.А. Минаев, А.П. Фисун, И.С. Константинов, А.В. Коськин, В.А. Зернов, ЮА. Белевская, С.В. Дворянкин; под общей научной редакцией В.Т. Еременко, В.А. Минаева, А.П. Фисуна, В.А.Зернова, А.В. Коськина. – Орел: ОрелГТУ, ОГУ, 2010. – 443 с.

ISBN 978-5-93932-311-6

В учебнике представлены основные положения классической теории информации. Системно изложены фундаментальные понятия информации, раскрыто содержание ее свойств, количественных и качественных характеристик, знания по современным процедурам кодирования информации и математической теории передачи знаков, лежащей в основе теории связи. Определены границы применимости классической теории информации. Рассмотрены вопросы формирования квантовой теории информации.

Материал рассчитан на студентов, аспирантов и специалистов в области разработки и эксплуатации информационных телекоммуникационных систем и обеспечения их информационной безопасности.

УДК 621.391(075)
ББК 32.811я7

ISBN 978-5-93932-311-6

СОДЕРЖАНИЕ

ПРЕДИСЛОВИЕ

Учебник разработан в рамках программы естественнонаучной учебной дисциплины «Теория информации» государственного образовательного стандарта высшего профессионального образования (ГОС ВПО) по специальности 090104 – «Комплексная защита объектов информатизации». Его содержание отражает также ряд разделов программ общих профессиональных дисциплин ГОС ВПО по специальности 090103 – «Организация и технология защиты информации», специальной дисциплины для подготовки научно-педагогических кадров по специальностям 05.13.17 «Теоретические основы информатики», 05.13.19 «Методы и системы защиты информации, информационная безопасность».

Учебник входит в серию задуманных авторами учебных изданий для технических и информационных направлений подготовки специалистов информационной сферы, которая включает, соответственно естественнонаучные и специальные дисциплины государственного образовательного стандарта: «Информатика», «Аппаратные средства вычислительной техники», «Правовые основы информационной безопасности», «Основы информационной безопасности», «Организационные основы информационной безопасности» и другие, объединенных системным замыслом достаточно полного представления знаний о содержании информации, ее значимых для деятельности человека и общества аспектах, и, прежде всего, одного из ее основных свойств – безопасности.

Основной целью учебника является представление системных знаний по классическим основам теории информации, положенной в основу построения современных средств информационных телекоммуникационных систем обработки информации, составляющей материальную основу объектов информатизации развивающегося информационного общества.

Содержание учебника разработано на основе научных, учебных, методических материалов, публикаций известных отечественных и зарубежных ученых и специалистов, приведенных в библиографии, а также научных результатов, практического опыта по подготовке специалистов, прочитанных лекций и разработанных авторами учебных, учебно-методических материалов в области в информатики, вычислительной техники, информационных систем и обеспечения их информационной безопасности.

Новизна издаваемого учебника характеризуется:

- системным изложением фундаментальных вопросов теории информации;
- достаточной методической систематизацией результатов научных направлений теории информации;
- учетом методологических аспектов информатики, определяющих характер научного мышления и мировоззрения обучающихся, специалистов информационной сферы;
- отражением логических причинно-следственных связей изучаемой дисциплины с другими научными направлениями информационной сферы, а также учебными дисциплинами рассматриваемых специальностей;
- структурированностью изложенного учебно-методического материала на уровне принципов, методов, моделей, алгоритмов.

Представленная структура книги, включает 12 глав, раскрыты основные положения классической теории информации, фундаментальные понятия информации, содержание свойств, количественных и качественных характеристик информации, а также знания по современным процедурам кодирования информации, математической теории передачи знаков, лежащей в основе теории связи. Показаны границы применимости классической теории информации и рассмотрены источники и проблемы формирования квантовой теории информации.

Первая глава знакомит читателя со взглядами на информацию современной информационной науки и техники. Рассматриваются базовые понятия, объект, предмет и задачи теории информации. Раскрываются особенности формирования содержания понятий информации как объектов теории информации и взаимосвязанных с ней информатики, информатизации и информационной безопасности социотехнических систем.

Вторая глава посвящена изучению качества информации, оцениваемого принятыми в теории информации рядом основных понятий и показателей качества и свойств информации. Подробно рассмотрены варианты классификации мер и единиц изменения информации. Изложены методы количественной оценки информации. Достаточно подробно раскрыто содержание статистической, семантической и других мер оценки количества информации.

В третей главе представлены основные понятия системной классификации информации и рассмотрен вариант классификации информации по различным основаниям.

Четвертая глава раскрывает общие характеристики и описывает алгоритмы процессов сбора, подготовки, обработки, передачи, и хранения информации.

В пятой главе излагаются основы обработки данных в информационных системах. На основе рассмотрения содержания базовых понятий сообщения, сигналов, носителей информации и сигналов, системно изложены их виды, характеристики, особенности кодирования, квантования сигналов, основные виды и способы обработки аналоговой и цифровой информации, способы и виды модуляции сигналов, а также принципы построения, работы и характеристики устройств обработки данных.

Шестая глава посвящена системному изложению известного классического подхода оценки количества информации, ее источников и передаваемой по телекоммуникационным каналам, обладающих различными характеристиками. Рассмотрены энтропии вероятностных схем, аксиомы Хинчина и Фаддеева, свойства энтропии, энтропия при непрерывном сообщении, условная энтропия и взаимная информация.

В седьмой главе представлены известные основы кодирования информации, предназначенной для эффективной передачи по телекоммуникационным каналам. Рассмотрена классическая постановка задачи кодирования и теоремы Шеннона об источниках информации. Раскрыто содержание Марковских и эргодических источников информации, эффективного кодирования, алфавитного неравномерного двоичного кодирования сигналами равной длительности. Описаны префиксные, линейные коды, коды Хемминга, сверточные коды, циклические и БЧХ – коды, а также параметры кодов и их границ, корректирующие свойства кодов.

Восьмая глава дает представление о математической модели канала связи, пропускной способности непрерывного канала связи и представляет математических аппарат теоремы Шеннона для его описания.

В девятой главе раскрыты подходы к формализации понятия ценности информации, в том числе достаточно глубоко изложены апостериорный подход и формализация понятий неопределенности и количества информации, отличные от подхода Шеннона.

Десятая глава «Определение границ применимости классической теории информации» описывает случайный выбор, как источник генерации информации, современные взгляды на характеристики информации, синергетический подход к классификации информации и дает представление о методологическом анализе рецепции информации.

В одиннадцатой главе раскрыты взгляды на информацию в контексте постнеклассической науки, в которых раскрывается многозначность определения понятия информации и дается представление об информационных процессах в нелинейном мире.

Двенадцатая глава, завершающая курс «Теория информации» представлена как его заключение. Она содержит сведения об источниках формирования квантовой информатики, ее основных понятиях, достижениях, проблемах. Раскрывается взгляд на квантовую информацию как основной объект квантовой информатики. Дано представление о классической информации в квантовых каналах, квантовой различимости, создании и преобразовании запутанности, как важного динамического процесса квантовой теории информации.

Каждая из глав учебника отвечает отдельной теме лекционного курса. Большинство параграфов и глав приближается по своему объему к отдельной лекции, а пункты ряда параграфов содержат определенный завершающий вопрос. Однако строгой зависимости здесь нет. Главы заканчиваются контрольными и проблемными вопросами, которые обеспечат активизацию самоконтроля полученных знаний, а также помогут обучаемым систематизировать свои знания и подготовиться к экзаменам и зачетам.

По мере изучения вопросов курса читателю предлагаются для постоянного обращения примеры решения задач, что повышает эффективность усвоения материала в ходе самостоятельной работы по тематике.

В качестве заключения в учебник включен раздел более сложного содержания по квантовой теории информации, который не только отражает направление дальнейшего развития искомой дисциплины, углубляет, расширяет знания обучаемых по теории информации, но и развивает методический аппарат книги в части рекомендаций по использованию материала студентами, которые интересуются вопросами теории информации и имеют склонность к научным исследованиям.

Содержание учебника основано на систематизации материалов различных литературных источников, результатах исследований авторов в области информатики и обеспечения информационной безопасности, авторских разработках по проблемам информатики и информационной безопасности, а также на базе курсов лекций, прочитанных авторами в вузах России.

Авторы учебника выражают благодарность за ценные материалы авторам, указанным в библиографии, а также за предложения, замечания и представленные отдельные материалы в следующие главы учебника: В.А. Лобановой (главы 6, 7, 8, 9), О.В. Третьякову (главы 10, 11, 12), А.Е. Георгиевскому (глава 6), И.Ю. Баранову (главы 4,5,8), К.А. Джеваге (главы 4,5,8), Д.С. Мишиной (главы 6-8, 10-12), В.Е. Фисенко (глава 5), А.В. Тютякину (глава 7), В.В. Митяеву (глава 11).

Авторский коллектив*: В.А. Минаев (предисловие, введение, главы 7,10,12), В.Т. Еременко (предисловие, введение, главы 6, 10-12), А.П. Фисун (предисловие, введение, главы 1-5,12), В.А. Зернов (главы 7, 10), И.С. Константинов (главы 7, 8, 10, 12), А.В. Коськин (глава 7, 9, 12), С.В. Дворянкин (глава 6, 7), Ю.А. Белевская (главы 1-4).*

ВВЕДЕНИЕ

В последние десятилетия мир переживает информационный взрыв, характеризующий один из признаков перехода индустриального общества к информационному, основой которого являются новые информационные технологии, информационные ресурсы, информатизация, информационная сфера, информационное пространство. Интенсивно разрабатываются концепции и программы информатизации государства, регионов, ведомств. Происходит кардинальная смена способа производства, мировоззрения людей, межгосударственных отношений, сфер и видов деятельности личности, общества и государства. Основным объектом этой деятельности является информация.

В XX в. бурное развитие получили всевозможные средства связи (телефон, телеграф, радио), назначение которых заключалось в передаче сообщений. Однако эксплуатация их выдвинула ряд проблем: как обеспечить надежность связи при наличии помех, какой способ кодирования сообщения применять в том, или ином случае, как закодировать сообщение, чтобы при минимальной его длине обеспечить передачу смысла с определенной степенью надежности. Эти проблемы требовали разработки теории передачи сообщений, иными словами, теории информации. Одним из основных вопросов этой теории был вопрос о возможности измерения количества информации.

Попытки количественного измерения информации предпринимались неоднократно. Первые отчетливые предложения об общих способах измерения количества информации были сделаны Р. Фишером (1921 г.) в процессе решения вопросов математической статистики. Проблемами хранения информации, передачи ее по каналам связи и задачами определения количества информации занимались Р. Хартли (1928 г.) и Х. Найквист (1924 г.). Р. Хартли заложил основы теории информации, определив меру количества информации для некоторых задач. Наиболее убедительно эти вопросы были разработаны и обобщены американским инженером Клодом Шенноном в 1948 г.

Для этих исследователей общим был подход, связанный с отвлечением от смысла содержания информации, так как чистая математика оперирует с количественными соотношениями, не вдаваясь в физическую природу тех объектов, за которыми стоят соотношения. При рассмотрении различных ситуаций имеет место неопределенность различных событий. С другой стороны неопределенность неотъемлема от понятия вероятности. Уменьшение неопределенности всегда связано с выбором (отбором) одного или нескольких элементов (альтернатив) из некоторой их совокупности. Такая взаимная обратимость понятий вероятности и неопределенности послужила основой для использования понятия вероятности при измерении степени неопределенности в теории информации.

На количество информации, получаемой из сообщения, влияет фактор неожиданности его для получателя, который зависит от вероятности получения того или иного сообщения. Чем меньше эта вероятность, тем сообщение более неожиданно и, следовательно, более информативно. Сообщение, вероятность которого высока и, соответственно, низка степень неожиданности, несет немного информации. Наиболее широкое распространение при определении среднего количества информации, которое содержится в сообщениях от источников самой разной природы, получил подход К. Шеннона. Оказалось, что формула, предложенная Хартли, представляет собой частный случай более общей формулы Шеннона. Кроме этой формулы, Шенноном была предложена абстрактная схема связи, состоящая из пяти элементов (источника информации, передатчика, линии связи, приемника и адресата), и сформулированы теоремы о пропускной способности, помехоустойчивости, кодировании.

В результате развития теории информации и ее приложений идеи Шеннона быстро распространяли свое влияние на самые различные области знаний. Было замечено, что от текста к тексту увеличиваются упорядоченность и информация, которой мы располагаем о тексте, а энтропия (мера неупорядоченности) уменьшается. Используя различие формул количества информации Шеннона и энтропии Больцмана (разные знаки), Л. Бриллюэн охарактеризовал информацию, как отрицательную энтропию, или негэнтропию. Так, как энтропия является мерой неупорядоченности, то информация может быть определена, как мера упорядоченности материальных систем.

Постепенно теория информации «переросла» рамки поставленных первоначально перед ней задач. Ее начали применять к более широкому кругу явлений. Увеличение количества информации стали связывать с повышением сложности системы, с ее прогрессивным развитием. Процесс развития в определенном аспекте можно моделировать, используя процесс передачи информации. Применение информационной модели развития дает возможность прояснить механизм прогресса с учетом усложнения, упорядочения и повышения степени организации материальных систем. Так, по данным некоторых исследований, при переходе от атомного уровня к молекулярному количество информации увеличивается в 10^3 раза. Известно, что количество информации, относящейся к организму человека, примерно в 10^{11} раз больше информации, содержащейся в одноклеточном организме.

Шенноновская теория информации основана на вероятностных, статистических закономерностях явлений. Она дает полезный, но не универсальный аппарат. Поэтому множество ситуаций не укладываются в информационную модель Шеннона. Не всегда представляется возможным заранее установить перечень всех состояний системы и вычислить их вероятности. Кроме того, в теории информации рассматривается только формальная сторона сообщения, в то время как смысл его остается в стороне. Этот подход и основанная на нем мера количества информации выражают, прежде всего, «структурно-синтаксическую» сторону ее передачи, т.е. выражают отношения сигналов. Однако понятия «вероятность», «неопределенность», с которыми связано понятие информации, предполагают процесс выбора. Этот процесс может быть осуществлен только при наличии множества возможностей. Без этого условия, как можно предположить, передача информации невозможна.

Р. Эшби осуществил переход от толкования информации, как «снятой» неопределенности к «снятой» неразличимости. Он считал, что информация есть там, где имеется (дано или выявляется) разнообразие, неоднородность. В данном случае единицей измерения информации может быть элементарное различие, т.е. различие между двумя объектами в каком-либо одном фиксированном свойстве. Чем больше в некотором объекте отличных (в строго определенном смысле) друг от друга элементов, тем больше этот объект содержит информации. Информация есть там, где имеется различие хотя бы между двумя элементами. Информации нет, если элементы неразличимы. В середине 50-х годов, используя материал статистической теории информации, Р. Эшби изложил концепцию разнообразия, согласно которой под разнообразием следует подразумевать характеристику элементов множества, заключающуюся в их несовпадении. Суть концепции разнообразия, по Эшби, заключается в утверждении, что теория информации изучает процессы «передачи разнообразия» по каналам связи, причем «информация не может передаваться в большем количестве, чем это позволяет количество разнообразия».

Исходя из идей основоположника кибернетики Н. Винера и результатов, полученных К. Шенноном, Эшби открыл закон, названный законом необходимого разнообразия, который так же, как закон Шеннона для процессов связи, может быть общим для процессов управления. Суть этого закона состоит в том, что для управления состоянием кибернетической системы нужен регулятор, ограничивающий разнообразие возмущений, которые могут разрушить систему. При этом регулятор допускает такое их разнообразие, которое необходимо и полезно для системы. Регулирование, возмущения – это термины, связанные с процессом управления. Поэтому закон необходимого разнообразия является одним из основных в кибернетике – науке об управлении.

Если понятие информации первоначально рассматривалось применительно только к процессам связи, а затем использовалось для характеристики сложности и упорядоченности материальных систем, то теперь уже речь идет об управлении ими. Впитывая всевозможные взгляды и концепции, понятие информации становится более емким и «дорастает» до уровня философских категорий – самых общих понятий, которыми можно оперировать.

Очень близка к трактовке разнообразия информации идея *алгоритмического измерения ее количества*, выдвинутая в 1965 г. А.Н. Колмогоровым. Суть ее заключается в том, что количество информации определяется, как минимальная длина программы, позволяющей преобразовать один объект (множество) в другой (множество). Чем больше различаются два объекта между собой, тем сложнее (длиннее) программа перехода от одного объекта к другому. Этот подход, в отличие от подхода Шеннона, не базирующийся на понятии вероятности, позволяет, например, определить прирост количества информации, содержащейся в результатах расчета, по сравнению с исходными данными. Вероятностная теория информации на этот вопрос не может дать удовлетворительного ответа.

Рассматриваемые подходы, связанные с количественным аспектом понятия информации без учета ее смысловой стороны, позволили привлечь к изучению информации точные математические методы. В результате были созданы всевозможные кибернетические устройства, вычислительные машины. С другой стороны, теория информации Шеннона, значительно дополненная и обогащенная новыми подходами, все же не может охватить всего многообразия понятия информации и, в первую очередь, ее содержательного аспекта. Теория информации К. Шеннона также не занимается определением ценности информации. Количество информации ее интересует лишь с точки зрения возможности передачи сообщений оптимальным образом.

Попытки оценить не только количественную, но и содержательную сторону информации дали толчок к развитию *семантической (смысловой) теории информации*. Исследования в этой области теснее всего связаны с семиотикой – теорией знаковых систем. Семиотика исследует знаки, как особый вид носителей информации. При этом знаком является условное изображение элемента сообщения, словом – совокупность знаков, имеющих смысловое значение, языком – словарь и правила пользования им. Таким образом, рассуждая о количестве, содержании и ценности информации, содержащейся в сообщении, можно исходить из возможностей соответствующего анализа знаковых структур.

В качестве знаковых систем используются естественные и искусственные языки, в том числе информационные и языки программирования, различные системы сигнализации, логические, математические и химические символы. Они служат средством обмена информацией между высокоорганизованными системами (способными к обучению и самоорганизации). Рассматривая знаковые системы, выделяют три основных аспекта их изучения: синтактику, семантику и прагматику.

Синтактика изучает синтаксис знаковых структур, т.е. способы сочетаний знаков, правила образования этих сочетаний и их преобразований безотносительно к их значениям.

Семантика изучает знаковые системы, как средства выражения смысла, определенного содержания, т.е. правила интерпретации знаков и их сочетаний, смысловую сторону языка.

Прагматика рассматривает соотношение между знаковыми системами и их пользователями, или приемниками-интерпретаторами сообщений. Иными словами, к прагматике относится изучение практической полезности знаков, слов и, следовательно, сообщений, т.е. потребительской стороны языка.

Основная идея семантической концепции информации заключается в возможности измерения содержания (предметного значения) суждений. Но содержание всегда связано с формой, поэтому синтаксические и семантические свойства информации взаимосвязаны, хотя и различны. Получается, что содержание все-таки можно измерить через форму, т.е. семантические свойства информации выразить через синтаксические. Поэтому и исследования семантики базировались на понятии информации, уменьшении или устранении неопределенности. Необходимо отметить, что методы точного количественного определения смыслового содержания информации еще не разработаны.

Первую попытку построения теории семантической информации предприняли Р. Карнап и И. Бар-Хиллел. Они положили начало применению идей и методов символической логики и логической семантики к анализу информационного содержания языка науки и предложили определять величину семантической информации посредством так называемой логической вероятности, которая представляет собой степень подтверждения той или иной гипотезы. При этом количество семантической информации, содержащейся в сообщении, возрастает по мере уменьшения степени подтверждения априорной гипотезы.

Если вся гипотеза построена на эмпирических данных, полностью подтверждаемых сообщением, то такое сообщение не приносит получателю никаких новых сведений. Логическая вероятность гипотезы при этом равна единице, а семантическая информация оказывается равной нулю. Гипотеза здесь полностью вытекает из данных опыта. И, наоборот, – по мере уменьшения степени подтверждения гипотезы, или запаса знаний, количество семантической информации, доставляемой сообщением, возрастает. Чем больше логическая вероятность высказывания, тем меньше должна быть мера его содержания, т.е. чем больше описаний состояния «разрешает» то или иное высказывание, тем меньше должна быть его семантическая информативность и, наоборот, чем больше описаний состояния им исключается, тем больше должна быть его информативность. Таким образом, семантико-информационное содержание высказывания определяется не тем, что содержит данное высказывание, а тем, что оно исключает.

Концепция Карнапа – Бар-Хиллела, получившая впоследствии развитие в трудах Кемени, является только началом исследований в области измерения содержания передаваемой информации. Эта концепция позволяет, например, выявить связь гипотезы с начальным достоверным значением, в частности, сделать заключение о степени подтверждения гипотезы.

Финский ученый Я. Хинтикка распространил основные идеи семантической теории информации Карнапа и Бар-Хиллела на логику высказываний. Для многих ситуаций (наблюдения, измерения, подтверждения гипотезы, научного предсказания, объяснения) он предложил метод определения уменьшения неопределенности. Однако, несмотря на определенные достижения, концепция Карнапа – Бар-Хиллела оказалась малопригодной для анализа содержания естественного языка. Эта теория, основанная на вероятностной логике, неприменима к анализу основного массива научного знания – достоверного знания.

Изучение отношений между знаками и их потребителями с точки зрения использования получаемой информации и влияния знаков на поведение систем составляет основу *прагматической теории информации*. Для всех подходов здесь характерно стремление связать понятие прагматической информации с целью, целенаправленным поведением и выдвинуть те или иные количественные меры ценности информации.

Исходя из этих соображений, А.А. Харкевич предложил связать меру ценности информации с изменением вероятности достижения цели при получении этой информации. А.А. Харкевич первым подчеркнул фундаментальный характер связи прагматических свойств информации с категорией цели, понимаемой, как опережающее отражение, модель будущего результата деятельности. Другой подход к проблеме ценности информации осуществлен М.М. Бонгардом. Он вводит понятие «полезная информация», связывая сообщение с тем, какую задачу решает получатель, что он знает до прихода сообщения и как его истолковывает. Этот подход имеет вероятностно-алгебраическую сущность и носит более общий характер, чем подход, предложенный А.А. Харкевичем.

Значительную роль в развитии прагматической теории информации сыграли работы американского логика Д. Харраха, поставившего перед собой цель показать, как символическая логика и теория семантической информации могут быть использованы для анализа некоторых аспектов человеческой коммуникации. Эту цель он пытается достигнуть путем создания «модели того, как разумный получатель оценивает последовательность сообщений на основе определенных семантических и прагматических свойств». Ученым предлагается обеспечить получателя «программой обработки сообщений», с помощью которой извлекается из получаемых сообщений «годная к употреблению сумма сообщений». Именно к этому результату переработки сообщений, а не к сообщениям в их первоначальной форме могут быть применены количественные меры информации. Созданная Харрахом логическая модель коммуникации служит тем языковым каркасом, в рамках которого программа может быть образована и применена.

Прагматические и семантические оценки зачастую трудно разделить и в некоторых случаях они сливаются. Так, семантические оценки характеризуют смысл, содержательность сообщений, а прагматические – их ценность, полезность. Как семантические, так и прагматические теории информации могут быть практически применены пока только к очень небольшому числу явлений реальной действительности. Но не следует забывать, что они имеют еще и теоретическое значение. В борьбе идей, мнений, гипотез и выводов, в их соперничестве и сотрудничестве рождается истина.

По каждому из перечисленных направлений исследований в теории информации написано много трудов. Несмотря на это, фронт наступления на понятие информации широк: его пристально изучают философы, биологи, физики, математики. Исследования, проводимые в разных направлениях, способствуют углублению понятия информации, подчеркивая в нем оттенки, специфичные для той или иной области знаний. Огромна практическая ценность полученных результатов. Интенсивными исследованиями представителей самых разнообразных наук – от математики и физики до биологии и философии – шаг за шагом собирается воедино образ, пожалуй, самого исключительного феномена в истории науки – понятия информации.

ГЛАВА 1. ПОНЯТИЕ ИНФОРМАЦИИ. ЗАДАЧИ И ПОСТУЛАТЫ ТЕОРИИ ИНФОРМАЦИИ

1.1. Понятия информации, данных и сообщения в современной науке, технике и теории информации

Информация (от лат. informatio – разъяснение, изложение) является основным понятием информатики. Несмотря на значительные достижения этой науки, ее составных частей, до настоящего времени нет четкого, однозначного и всеми принятого определения информации, отсутствует методология определения значений ее характеристик, что затрудняет решение задач информатики. Это одно из первичных неопределенных понятий науки, что подтверждается большим множеством дефиниций понятия информации, которые рассматриваются в приведенной литературе: от наиболее общего, философского (информация – есть отражение реального мира) до узкого практического (информация есть сведения, являющиеся объектом хранения, передачи и преобразования), в том числе и определенного нормативными правовыми актами.

Понимание информации как передачи сведений сохранялось на протяжении более двух тысячелетий. В связи с прогрессом технических средств массовых коммуникаций (телеграф, телефон, радио, телевидение и т.д.), в особенности с ростом объема передаваемых сведений, появилась необходимость их измерения. В 20-х годах XX века делались попытки измерения информации и высказывались идеи, которые затем были использованы в вероятностно-статистической теории информации (Фишер, 1921 г., Найквист, 1924 г., Хартли, 1928 г., Сциллард, 1929 г.). Однако только в 1948 г., в статье К.Э. Шеннона «Математическая теория связи» было дано вероятностно-статистическое определение понятия количества информации, предложена абстрактная схема связи, сформулированы теоремы о пропускной способности, помехоустойчивости, кодировании и т.д., что позволило сформировать вероятностно-статистическую теорию информации – одну из наиболее развитых среди других математических теорий информации.

Математические теории информации выступают как совокупность количественных (и, в первую очередь, статистических) методов исследования передачи, хранения, восприятия, преобразования и использования информации. Используемые методы преследуют цель измерения информации. Проблема количества информации как первоочередного вопроса теории информации, неразрывно связана и с ее качественно-содержательным аспектом, т.е. выявлением содержания понимания информации, к которому применяются количественные методы исследования.

Существует, в частности, направление теории информации, основанное на использовании положений топологии – раздела математики, изучающего свойства пространства, которые сохраняются при взаимно однозначных непрерывных преобразованиях (растяжении, деформации и т.п.). Одним из топологических объектов является граф. Топологическое пространство информации определяется в зависимости от различия вершин графов, количества ребер, выходящих из них, ориентации этих ребер.

Развитие социального прогресса, науки, техники, объемов информации обусловили необходимость дальнейшего уточнения понятия «информация» и количественной оценки передаваемых сообщений, выявления наиболее характерных свойств информации, что привело к принципиальным изменениям и в трактовке самого понятия «информация».

Как уже говорилось, в соответствии с шенноновской теорией понятие информации определялось на вероятностной основе.

Информация – сведения, сообщения, которые снимают существовавшую до их получения неопределенность полностью или частично. Это представление об информации как снимаемой неопределенности является наиболее распространенной трактовкой понятия информации.

Одними из признаков, характеризующих информацию, являются *различие* и *разнообразие*. Если в статистической теории понятие информации определяется как уничтоженная неопределенность, то в самом общем случае можно сказать, что информация есть уничтожение тождества, однообразия. Переход от трактовки информации как противоположности неопределенности к трактовке ее как противоположности тождеству диктуется всем ходом развития наших знаний, развитием кибернетики (где информация выступает как синоним разнообразия, которое получает и использует кибернетическая система), а также психологии, биологии, химии, экономики и многих других наук. Согласно данной концепции, информация существует там, где имеется разнообразие и различие.

Рассмотрим ряд определений понятий информации, используемых в известных источниках, приведенных в библиографии.

Информация – сведения об объектах и явлениях окружающей среды, их параметрах, свойствах и состоянии, которые уменьшают имеющуюся о них степень неопределенности, неполноты знаний.

Нетрудно увидеть, что по определению информация есть отражение или представление реального мира (чего-нибудь) с помощью сведений (сообщений).

Сообщение – это форма представления информации в виде речи, текста, изображения, цифровых данных, графиков, таблиц и т.п.

Тогда в широком смысле можно привести еще одно определение информации.

Информация – это общенаучное понятие, включающее в себя обмен сведениями между людьми, обмен сигналами между живой и неживой природой, людьми и устройствами.

Обобщенное понятие и классификация сообщений приводятся в словарях по информатике, указанных в библиографии.

Сообщение (*message*) – упорядоченная последовательность символов, предназначенная для передачи информации.

Классификация сообщений может осуществляться по ряду показателей.

По видам информационных коммуникаций (каналов передачи сообщений):

– формальные (созданные для организаций, учреждений и т.п.);

– неформальные (формируемые при личных встречах, беседах, телефонных разговорах и др.).

По форме представления сообщения делятся на виды:

– недокументальные: личные беседы, конференции, совещания, реализованные жестами, звуками, знаками, речью и др.);

– документальные:

1) кодированные: текстовые (опубликованные, неопубликованные), идеографические, аудиальные (запись речи), машиночитаемые;

2) некодированные: иконические, документы трех измерений, аудиальные (кроме записи речи).

– формализованные (formalized message) – сообщения, представленные в формализованном виде, требуемом для передачи по линиям связи или для ввода в ЭВМ.

По адресам рассылки:

– групповое сообщение (groupe message) – сообщение, адресуемое более чем одному абоненту или множество (пакет) сообщений, объединенных в одно сообщение, передаваемое по линиям связи;

– одноадресное (single-address message) – сообщение, передаваемое только в один пункт назначения;

– входное (incoming message) – сообщение, поступающее на вход ЭВМ от терминала или абонентского пункта.

По функциональному предназначению:

– диагностическое (сообщение об ошибке) (diagnostic (error) message) – сообщение, выдаваемое управляющей, обрабатывающей, или обслуживающей программой и содержащее сведения о месте, типе и характере ошибки в программе;

– квитирующее – сообщение системы, предназначенное пользователю и содержащее информацию о результатах выполнения запроса, но не сами результаты;

– контрольное (fox message) – стандартное сообщение, включающее полный набор допустимых знаков и используемое для проверки линий связи;

– наводящее (prompting message) – подсказка пользователю.

Информатика рассматривает информацию как концептуально связанные между собой сведения (сообщения), данные, изменяющие наши представления о явлении или объекте окружающего мира. Наряду с понятием «*информация*» в информатике часто употребляется понятие *«данные»*. В чем же их общность и различия?

В учебнике Макаровой Н.В. приводятся следующие аргументы в пользу их различия. «Данные могут рассматриваться как признаки или записанные наблюдения, которые по каким-то причинам не используются, а только хранятся. В том случае, если появляется возможность использовать эти данные для уменьшения неопределенности о чем-либо, данные превращаются в информацию. Поэтому можно утверждать, что информацией являются используемые данные».

Насколько неоднозначны такие определения данных, сведений и информации, можно судить исходя из следующих известных определений этих понятий.

Данные: 1) – сведения, необходимые для какого-нибудь вывода, решения; 2) – факты, идеи, выраженные в формальном виде, обеспечивающем возможность их хранения, обработки, передачи; 3) – факты, идеи, представленные в формальном виде, позволяющем передавать или обрабатывать их при помощи некоторого процесса и соответствующих технических средств; 4) – информация, представленная в формализованном виде, пригодном для автоматической обработки при возможном участии человека; 5) – обобщенное имя информационных продуктов, являющихся предметов труда в информационном производстве.

Представляют интерес содержание понятий информации и данных, изложенных известными авторами Преснухиным Л.Н. и Нестеровым В.П.

Данные – представление фактов и идей в формализованном виде, пригодном и удобном для фиксации, передачи и переработки в процессе их использования.

Информация – смысл, который приписывается данным посредством соглашений, принятых при их представлении.

Здесь данные рассматриваются как изображение информации, которая несет в себе некоторое сообщение относительно состояния и свойств объектов реального мира. Создаваемые человеком абстрактные и материально-энергетические представления (сообщения) об окружающем мире, среде, отличаются от реальных объектов живой и неживой природы. При этом содержание (смысл) такого сообщения представляет информацию, а знаки (физические формы) представления этого сообщения представляет данные.

Сведения – знания, представление чего-нибудь, известие, сообщение.

Сообщение – то, что сообщается, известие.

Известие – сообщение о чем-нибудь.

Сообщение – набор данных, объединенных смысловым содержанием и пригодных для обработки и передачи.

Информация как совокупность фактов представлена на рис. 1.1.

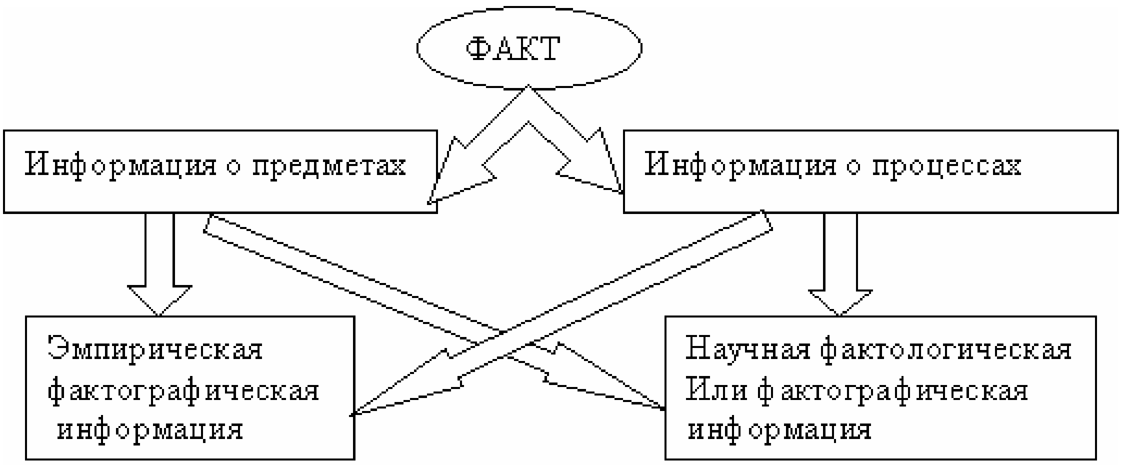


Рис. 1.1. Информация – как совокупность фактов

Анализируя определения данных и информации, можно отметить, что первичным (базисным) определением является понятие информации, с помощью которого последовательно и логично выводятся другие понятия: информация – данные – сообщение. Этот, далеко не полный, анализ понятий информации дает основание утверждать о большом разнообразии и неоднозначности их содержания, а также о разнообразии содержания информатики.

При работе с информацией всегда имеется ее источник, среда передачи и потребитель (получатель). Пути и процессы, обеспечивающие передачу сообщений от источника информации к ее потребителю, называются *информационными коммуникациями*.

С учетом рассмотренных понятий информации можно утверждать, что в наиболее общем виде сама по себе информация может быть отнесена к области абстрактных категорий, которая проявляется в материально-энергетической форме в виде сигналов, методологическая схема образования которых представлена на рис. 1.2.

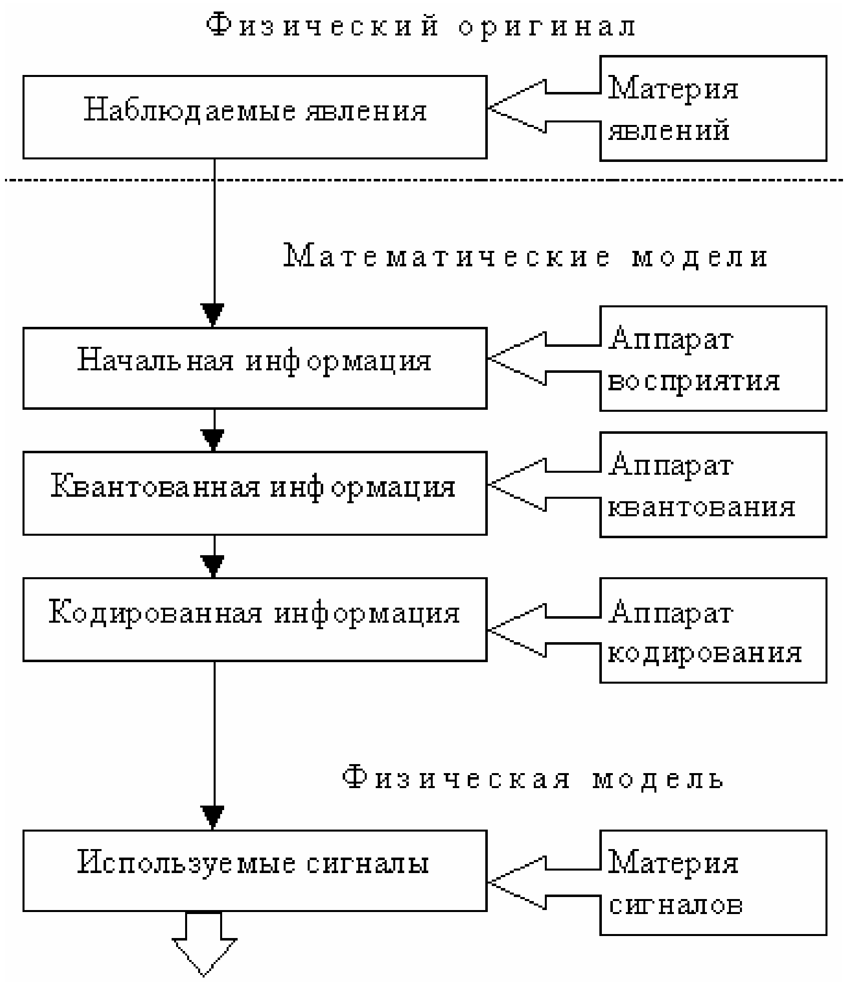


Рис. 1.2. Методологическая модель формирования и материализации информации

1.2. Формирование содержания понятий информации как объектов теории информации, информатики, информатизации и информационной безопасности

Теоретические исследования проблем, связанных с уточнением научного содержания понятия информации, проводятся по трем основным направлениям:

1) разработка математического аппарата, отражающего основные свойства и закономерности информации;

2) разработка теоретических аспектов информации на базе существующих математических средств, обеспечивающих исследование различных свойств информации, в том числе - измерение ее ценности и полезности с применением методов экспертной оценки;

3) разработка и использование информационных методов как в научно-информационной деятельности, так и в лингвистике, биологии, психологии, социологии, педагогике и др.

В некоторых философских работах была выдвинута следующая концепция информации.

Информация – является одним из основных универсальных свойств материи – атрибутом всей материи.

Такой подход связывает понятие информации с понятием отражения. Он получил название атрибутивного. Сторонниками этой концепции являются В.М. Глушков, В.И. Сифоров, А.Д. Урсул и ряд других исследователей. Другой концепции придерживаются Б.С. Украинцев, Н.И. Жуков, П.В. Копнин, В.С. Тюхнин, А.М. Коршунов, Д.И. Дубровский и др. Они исходят из неразрывной связи информации с управлением, с функционированием самоорганизующихся и самоуправляемых схем. Такая концепция получила название функциональной. Ее сторонники не признают существования информации в неживой природе как таковой, самой по себе. Противоборство рассмотренных концепций объективно способствует их совершенствованию.

Представляет интерес методологический подход к рассмотрению содержания информации В.А. Герасименко, в котором:

– принимается утверждение, что понятие информации является чрезвычайно широким, четко и однозначно не определенным;

– определение формируется путем перечисления наиболее характерных признаков соответствующего понятия;

– в качестве базовой признается характеристика информации как сведений, передаваемых от одного объекта или субъекта к другому;

– в зависимости от уровня развития информация представляется в виде зафиксированной структуры, импульсов, сигналов или знаний;

– информационные системы делятся на системы неживой природы, биологические, технические, социальные и др.

В контексте известных взглядов Д.И. Блюменау, содержание информации рассматривается как знания о предметах и явлениях реального мира. Однако, в связи с ограниченностью возможностей человеческой памяти и стремлением к экономии ее ресурсов, знания отражаются в тезаурусе субъекта на уровнях метазнания и знания. Эти уровни свернутых знаний называются *метаинформативным* и *информативным*. С учетом этого Д.И. Блюменау предлагает ряд понятий информации и особенности формирования ее содержания, которые в общем виде можно представить следующими положениями:

– *информация* (конкретная) – факты и концепции, необходимые для принятия решений, достройки психической модели в любой конкретной ситуации;

– *метаинформация* – знания о том, как организована информация, информация об информации, то, что позволяет ориентироваться в мире знаний;

– метаинформация в свою очередь подразделяется: 1) на «внешнюю» (эксплицитную), зарегистрированную на носителях – библиографические источники, программы курсов, литературная критика, правила общения с ЭВМ и многое другое; 2) «внутреннюю» (имплицитную), которой располагает тезаурус каждого субъекта и без которой он практически неспособен воспринять никакой конкретной информации;

– для понимания содержания рассматриваемой информации, необходимо соотнести (включить) сигнал по ряду признаков с определенной областью структуры наших знаний, в которой и произойдет переход с так называемого синтаксического уровня восприятия на уровень поверхностно-семантический. Таким образом, произойдет декодирование сигнала в сообщение. Здесь явно прослеживается необходимость учета проблемы полисемии языка. При восприятии сообщения в форме знака или других форм (например, сообщение – «коса») и отсутствии при этом ситуативной метаинформации, позволяющей отнести данное сообщение к определенной области структуры нашего тезауруса, невозможно однозначно понять содержание переданного сообщения.

Предметом дисциплины «Информатика» являются технические и социальные системы, поэтому в дальнейшем ее содержание будет излагаться применительно к системам этого типа. С учетом этого определение широких понятий сводится к указанию наиболее характерных их признаков, которые структурно могут быть представлены табл. 1.1.

Таблица 1.1

Структуризация понятия «информация»

Уровень развития информации	Виды систем объективного мира			
	Неживая природа	Биологическая	Техническая	Социальная
Знания				
Сигналы				
Импульсы				
Зафиксированная структура				

Для социальных систем значимыми являются все четыре названных выше вида информации, однако наибольшее значение имеет информация в виде знаний. Ее характеристики могут быть определены следующими показателями:

– характером точки зрения на информацию (мировоззренческий, науковедческий, управленческий и технический);

– характером интереса к информации (познавательный, прагматический, семантический и синтаксический);

– масштабом интереса к информации (глобальный, региональный (ведомственный), объектовый и личностный). На основе этих показателей строится общая классификационная структура характеристик информации в виде декартового произведения значений перечисленных показателей. Однако полное множество характеристик значительно шире, многообразнее и требует дальнейшей дифференциации. В плане информационного обеспечения жизнедеятельности общества, личности и государства интересующее нас подмножество характеристик может быть выделено путем наложения ограничений на значения предложенных показателей, например:

1) относительно характера точки зрения на информацию – основной является управленческая и лишь отчасти – остальные три;

2) относительно характера интереса к информации – определяющими являются прагматический и семантический и лишь отчасти – познавательный и синтаксический;

3) относительно масштаба интереса к информации – основными являются региональный (ведомственный) и объектовый и лишь отчасти – глобальный и личностный.

Возможны и другие сочетания ограничений.

Мы ограничили сферу внимания к информации потребностями информационного обеспечения управления в прагматическом (т.е. прикладном), семантическом (т. е. смысловом) плане, в масштабе региона (ведомства) и конкретного объекта. Для каждого конкретного или типового случая можно заблаговременно определить сумму сведений (знаний), необходимых для информационного обеспечения деятельности. Эти сведения могут быть структурированы, например, в виде перечня элементов (объектов), имеющих значение для обеспечения соответствующего вида деятельности и перечня характеристик, полностью описывающих эти элементы (объекты) с точки зрения данного вида деятельности.

Указанная выше сумма сведений может быть представлена множеством всех значений характеристик для всех элементов (объектов). В упорядоченном виде такое множество может быть названо информационным кадастром системы, информационное обеспечение которой является предметом рассмотрения. Поскольку информационный кадастр имеет определенное назначение и структуру, то для оценки находящейся в нем информации может быть определена совокупность конкретных показателей: полнота, оцениваемая по степени заполнения элементов информационного кадастра; достоверность, оцениваемая степенью соответствия элементов информационного кадастра действительным значениям, соответствующим характеристикам соответствующих объектов реального мира; актуальность, определяемая соответствием значений характеристик в информационном кадастре текущему значению соответствующих характеристик реальных объектов.

Нетрудно показать, что для определения значений перечисленных характеристик могут быть разработаны достаточно точные методы, причем в этих методах может быть учтена важность каждого элемента информационного кадастра, которая определяется с учетом двух параметров: важности задачи, для решения которой используется оцениваемый элемент кадастра, и важности оцениваемого элемента для решения задачи. Эти методы достаточно детально изложены В.А. Герасименко. Кроме такой классификации информацию можно различать:

– по областям знаний (биологическая, техническая, социальная, экономическая, правовая и др.);

– по физической природе восприятия (зрительная, слуховая, вкусовая, и др.);

– по структурно-метрическим свойствам (табл. 1.2).

Таблица 1.2

Классификация информации по структурно-логическим свойствам

Виды информации	Обозначение	Формы представления информации		
		Топологическая	Абстрактная	Лингвистическая
Событие Величина Функция Комплекс ………… Поле	$Ф^0$ $Ф^1$ $Ф^2$ $Ф^3$ … $Ф^п$	Точка Линия Поверхность Объем …………… Пространство	Суждение Понятие Образ Система ………… Универсум	Знак Буква Слово Предложение ……………… Фонд

Параметрическая информация – наборы численных оценок значений каких-либо параметров (измеряемые величины), результаты количественных определений при исследовании, анализе, контроле, учете.

Такой информацией пользуются в науке, технике и инженерной практике для выражения результатов измерения. Ее можно свести к таким видам как событие, величина, функция, комплекс.

Элементарное двоичное событие – первичный и неделимый элемент информации, представляющий выбор из утверждения или отрицания, истины или лжи, согласия или несогласия, наличия или отсутствия какого-либо явления.

Величина – упорядоченное в одном измерении (по шкале значений) множество событий, каждое из которых отвечает принятию величиной какого-либо одного значения.

Величина может быть дискретной, тогда множество событий счетное, или непрерывной – множество событий несчетно. Геометрическое представление величины – линия.

Функция – соотношение между величинами.

Интерпретация функции – двумерное поле событий.

Комплекс информации – соответствие между величиной, временем и пространством.

Полный комплекс информации может быть представлен трехмерным полем событий.

Топологическая информация – геометрические образы, карты местности, различные плоские и объемные изображения, объекты.

Топологические виды информации классифицируются по размерности информационных множеств на информацию различного порядка: нулевого порядка (нуль-мерная информация), соответствующая мощности точки; первого порядка (одномерная информация), соответствующая мощности линии; второго порядка (двумерная информация), соответствующая мощности поверхности; третьего порядка (трехмерная информация), соответствующая мощности объема; n-го порядка (n-мерная информация), соответствующая мощности n-мерного пространства. Топологической информацией удобно выражать образы и ситуации, подлежащие распознаванию.

Абстрактная информация применяется в исследованиях на высоком теоретическом уровне, когда нужны отвлечения, обобщения и символизация. С учетом рассмотренных классификаций и критериев структуризации рассматривается следующее понятие информации.

Информация – специфический атрибут объективного мира, создающий условия, необходимые для обеспечения устойчивости и развития систем различной природы.

При этом, чем сложнее система, тем разнообразнее и сложнее виды информации, обеспечивающие достижение целей системы. Конкретизация рассматриваемого понятия информации осуществляется в зависимости от предметных областей философии, управления, техники права и др. (табл. 1.3).

В отечественном законодательстве впервые было определено следующее содержание информации, имеющее достаточно высокую степень общности.

Таблица 1.3

Содержание понятия информации для различных предметных областей

Предметные аспекты содержания информации				
Философский	Управленческий	Технический	Экономический	Информационный
1. Информация как одна из реальностей объективного мира. 2. Происхождение и сущность информации. 3. Информация как мера сущностей объективного мира	1. Информация как непременный атрибут всякого управления. 2. Информационные процессы как основное содержание управления	1. Информация как совокупность символов, зафиксированных на носителях. 2. Проблемы сбора, хранения, передачи, переработки информации	1. Информация как совокупность сведений о социально-экономических процессах. 2. Информационные процессы как основное содержание управления коллективами людей в производственной и непроизводственной сферах. 3. Сопровождает процессы производства, распределения, обмена и потребления материальных благ и услуг	1. Информация как важнейший атрибут жизнедеятельности личности, общества, государства. 2. Проблемы определения информационных потребностей. 3. Проблемы рационализации информационных процессов. 4. Проблемы информационного обеспечения деятельности личности, общества, государства

Информация – сведения о лицах, предметах, фактах, событиях, явлениях и процессах независимо от формы их представления.

С учетом этого определения рассматривается понятие правовой информации.

Правовая информация – сведения о фактах, событиях, предметах лицах, явлениях, протекающих в правовой сфере жизни общества, содержащихся как в нормах права, так и в других источниках, и используемых государством и обществом для решения правотворчества, правоприменительной и правоохранительной деятельности, защиты прав и свобод личности.

Носителями правовой информации являются: правовые нормы и институты (совокупность взаимоувязанных норм, регулирующих качественно однородные общественные отношения), отрасли права и массивы законодательных актов, право и законодательство в целом.

Информацией является конкретное юридическое и социальное содержание правовых норм (предписания, разрешения, запреты, санкции, формы ответственности др.) и правоположений (элементы нормативно-правового текста, в том числе определения, юридически закрепленные цели, декларации и др.).

Источником правовой информации являются нормативные акты (федеральные и иные законы, указы, распоряжения Президента РФ, постановления Правительства РФ, ведомственные акты и др.), а также формы ненормативной информации, в том числе: проявления правовой активности толкования (логическое, систематическое) норм и институтов права, обобщение правовой практики, правоприменительная деятельность (судебное доказательство), процессуальные документы (решение, приговор, определение).

1.3. Задачи и постулаты прикладной теории информации

Возникновение теории информации связывают с появлением фундаментальной работы американского ученого К. Шеннона «Математическая теория связи» (1948). Однако в теорию информации вошли результаты, полученные Р. Хартли, впервые предложившего количественную меру информации (1928), акад. В.А. Котельниковым, сформулировавшим важнейшую теорему о возможности представления непрерывной функции совокупностью ее значений в отдельных точках отсчета (1933) и разработавшим оптимальные методы приема сигналов на фоне помех (1946), акад. Н.А. Колмогоровым, внесшим огромный вклад в статистическую теорию колебаний, являющуюся математической основой теории информации (1941).

К теории информации относятся результаты решения ряда фундаментальных теоретических вопросов, касающихся повышения эффективности функционирования систем связи:

1) анализ сигналов как средства передачи сообщений, включающий вопросы оценки переносимого или «количества информации»;

2) анализ информационных характеристик источников сообщений и каналов связи и обоснование принципиальной возможности кодирования и декодирования сообщений, обеспечивающих предельно допустимую скорость передачи сообщений по каналу связи, как при отсутствии, так и при наличии помех.

В теории информации исследуются системы связи при четко сформулированных условиях (постулатах):

1. Источник сообщения осуществляет выбор сообщения из некоторого множества с определенными вероятностями.

2. Сообщения могут передаваться по каналу связи в закодированном виде. Кодированные сообщения образуют множество, являющееся взаимно однозначным отображением множества сообщений. Правило декодирования известно декодеру (записано в его программе).

3. Сообщения следуют друг за другом, причем число сообщений, от которых зависит кодовый символ (длина кодовых ограничений), может быть сколь угодно большим.

4. Сообщение считается принятым верно, если в результате декодирования оно может быть в точности восстановлено. При этом не учитывается, сколько времени прошло с момента передачи сообщения до момента окончания декодирования, и какова сложность операций кодирования и декодирования.

5. Количество информации не зависит от смыслового содержания сообщения, от его эмоционального воздействия, полезности и даже от его отношения к реальной действительности

Контрольные вопросы

1. Каковы особенности формирования содержания понятия информации в широком и узком смыслах?
2. В чем общности и различие концепций формирования содержания базового понятия информации в теории информации Фишера, Найквиста, Хартли, Сцилларда, Эшби, Шеннона?
3. Какова особенность содержания математических теорий информации?
4. Раскрыть содержание и соотношение понятий информации и сообщения.

5. Какие основания классификации используются в современной информационной науке?

6. Каковы соотношения содержания понятий данных и информации?

7. Какими показателями характеризуется информация?

8. Привести классификацию информации по структурно-логическим свойствам.

9. Дать определение понятиям: элементарное двоичное событие, величина, функция, интерпретация функции, комплекс информации, топологическая информация.

10. Дать определение информации, используемое в Федеральном законе «Об информации, информатизации и защите информации».

11. Дать определение сообщения и привести классификацию сообщений по различным показателям.

12. По каким показателям можно классифицировать сообщения?

13. Дать определения понятиям: сведения, известия, данные, содержание.

14. Привести варианты определений понятий «сведения», «сообщение», «известие».

15. Перечислить и раскрыть сущность современных теоретических исследований проблем, связанных с уточнением научного содержания понятия информации.

16. Раскрыть содержание методологической модели формирования содержания информации.

17. Какие существуют методологические подходы к рассмотрению содержания информации?

18. Дать краткую характеристику методологических подходов к рассмотрению содержания информации.

19. В чем сущность взглядов Д.И. Блюменау на содержание информации?

20. Какова сущность метаинформативных и информативных знаний?

21. В чем различие между содержанием информации и метаинформации?

22. Раскрыть сущность структуризации понятия «информация.

23. Представить вариант классификации информации по структурно-логическим свойствам.

24. Дать определение параметрической информации.

25. Раскрыть содержание понятий «Элементарное двоичное событие», «Величина» «Функция», «Интерпретация функции», «Комплекс информации», «Топологическая информация», «Абстрактная информация».

26. Показать различие содержания понятий информации для различных предметных областей, в том числе физической, управленческой, технической, экономической, информационной.

27. Раскрыть содержание понятия правовой информации, ее носителей, значение для развития информационного общества.

28. Сформулировать задачи и постулаты прикладной теории информации.

ГЛАВА 2. КАЧЕСТВО ИНФОРМАЦИИ

2.1. Основные понятия и показатели качества информации

Решение проблемы выбора методов и подходов определения количества информации неразрывно связано с уточнением, обоснованием и выбором системы показателей, характеризующих разносторонние свойства (качества) информации.

В учебнике Макаровой Н.В. общие свойства информации представлены через формы ее синтаксической, семантической, прагматической адекватности.

Синтактика изучает формальные, внешние свойства знаков и их комбинации (отношение знака к знаку). В синтактике отвлекаются от смысла знаков, рассматривая только их геометрическую конфигурацию и распределение в пространстве и времени.

Семантика изучает знаки как средство обозначения реальных предметов (отношение знака к предмету), то есть вопросы, касающиеся смысла знаков.

Знак – какой-либо материальный предмет или явление, которым обозначают (представляют) другие предметы и явления в процессах коммуникации и познания, т.е. условное обозначение чего-либо.

Классификацию знаков можно представить табл. 2.1.

Таблица 2.1

Классификация знаков

ЗНАКИ	
Языковые	Неязыковые
Буква	Знаки-копии
Морфема	
Слово	Знаки-признаки
Словосочетание	
Предложение	Знаки-символы
Текст	

Особенности знаков заключаются: в их способности выступать в качестве заменителя обозначаемого; нетождественности знака и денотата; многозначности соответствия знак – денотат.

Денотат (лат. denotatys – обозначенный), предметное значение имени (знака), т.е. то, что названо этим именем, представителем чего оно является в языке.

Например, денотат имени «Утренняя звезда» – планета Венера или телевизионная музыкальная передача.

Концепт – выраженные знаком свойства денотата.

Прагматика – изучает знаки с точки зрения их участия в практической деятельности людей.

Информация характеризуется частными потребительскими показателями качеств (свойств): репрезентативностью, содержательностью, достаточностью (полнотой), доступностью, актуальностью, своевременностью, точностью, достоверностью, устойчивостью и др. Однако такая система показателей имеет недостатки, определяемые неполнотой представления связей между общими и частными показателями.

По мнению авторов, этого недостатка лишена система показателей качества информации, приведенная в работах известного ученого Герасименко В.А. Здесь качество информации представляется сложным понятием, которое характеризуется базовой системой показателей трех классов:

1) выдачи (своевременность, актуальность, полнота, релевантность, толерантность);

2) обработки (глубина, достоверность, адекватность);

3) защищенности (целостность физическая, целостность логическая, доверие, безопасность).

Каждый из этих показателей может рассматриваться с синтаксических, семантических, прагматических позиций и ряда других показателей. В.А. Герасименко предложена относительно полная система показателей свойств информации (рис. 2.1), образованная множеством групп показателей свойств информации и методов их определения.

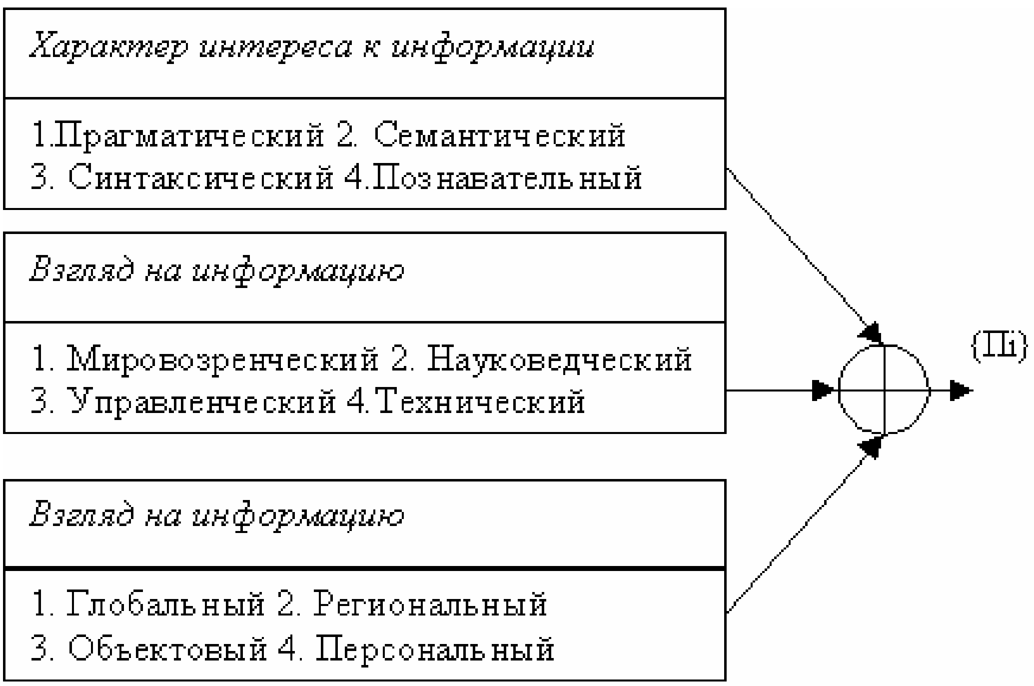


Рис. 2.1. Структура системы показателей свойств информации

2.2. Свойства информации

1) классификация абстрактных свойств информации

Представляет интерес известная система классификации свойств информации, включающая атрибутивные, прагматические, динамические показатели качества информации.

Существует классификация свойств информации на примерах живых систем различной сложности и информационных детерминантов структурогенеза, завершающего определенный этап самоорганизации неживой системы. Здесь наряду с относительно узким утверждением о связи информации только с живыми системами, рассматриваются общие свойства информации для всех ее видов, включая и информацию, генерируемую в процессах самоорганизации неживых систем. Такой подход представляет рассмотрение свойств информации с известных позиций некоторой абстрактной информации, свойства которой присущи всем ее видам и могут быть представлены двумя группами (рис. 2.2), внутри которых определены составляющие свойства, тесно связанные между собой.

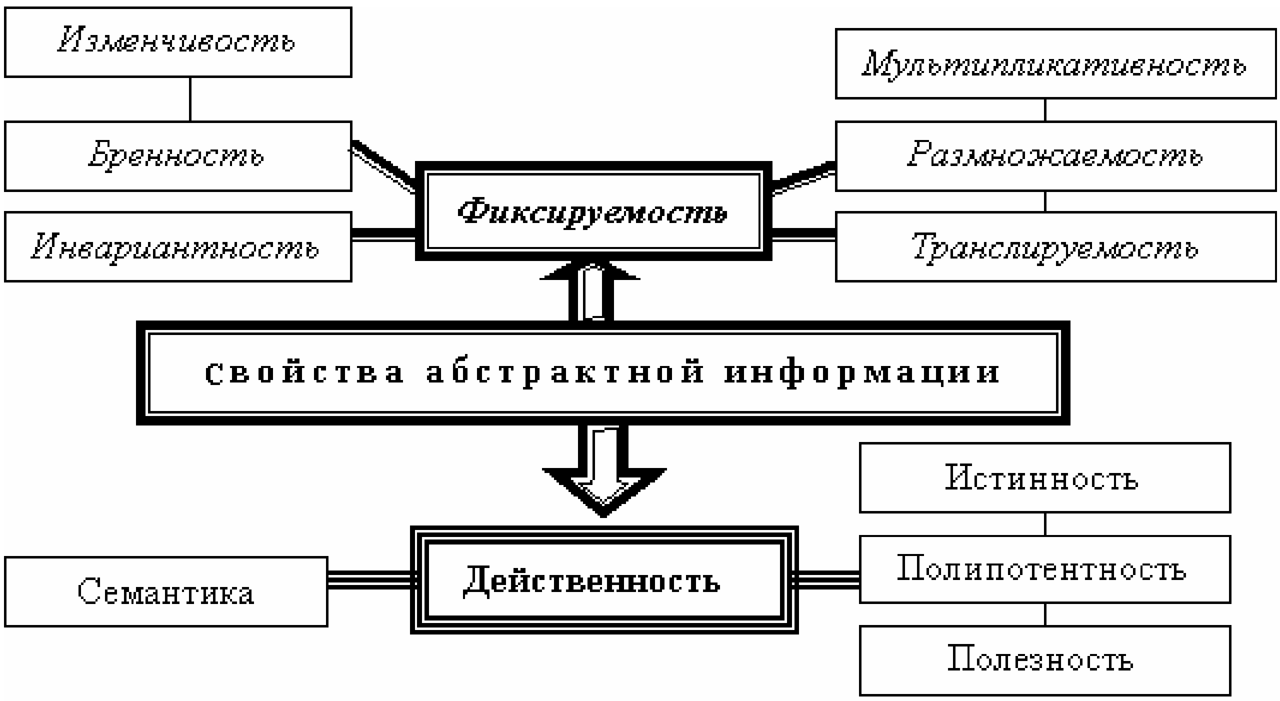


Рис. 2.2. Классификация абстрактных свойств информации

Комплексным свойством первой группы является *фиксируемость* информации, а второй группы – *действенность* информации.

Фиксируемость – способность любой информации, не будучи материей, восприниматься только в зафиксированном состоянии в виде записи на том или ином физическом носителе.

Формой фиксации информации являются предложения, составленные на любом языке в виде последовательности букв (знаков) любого алфавита, нанесенных любым способом на любой носитель.

Заметим, что это представление о фиксации информации хорошо согласуется с представлением о ней как о некоторой программе, но исключает случай рецепции в виде образования некоторых структур, например в автокаталитических процессах.

Инвариантность информации по отношению к носителю – свойство, связанное с ее фиксируемостью, отражающее возможность фиксации (записи) информации на любом языке, любым алфавитом.

Инвариантность отражает внутреннее свойство информации, ее автономность, независимость от материальных носителей и обусловливает возможность осуществления различных элементарных процессов обработки информации.

Инвариантность информации обусловливает возможность осуществлять разные элементарные информационные акты: создания, приема, передачи, хранения и использования информации. Отсюда ясно, что свойство инвариантности лежит в основе возможности понимания генетической информации. Расшифровка генетического кода, потребовавшая многолетних усилий биологов, физиков, химиков и математиков, является крупнейшим научным достижением XX столетия, ярко иллюстрируя значимость свойства инвариантности информации.

Бренность информации – свойство, определяемое связью зафиксированной информации со свойствами материальных носителей, в том числе ограниченностью времени запоминания, относительной ограниченностью времени жизненных циклов и разрушаемостью видов носителей.

Изменчивость – свойство, связанное с бренностью, характеризующее возможность исчезновения существующей и возникновения новой, отличающейся от исходной, информации вследствие ее разрушения, изменения ее физических свойств, повреждения носителей. Исчезновение информации может происходить не только из-за разрушения, но и благодаря ее изменению, например вследствие мутаций генов. При этом прежняя информация гибнет и рождается новая. Изменения могут иметь различные механизмы, определяющиеся спецификой информационных систем.

Информация способна изменяться при ее хранении из-за повреждения носителей, а в состоянии репликации из-за ошибок, которые можно назвать информационными дефектами. В этом случае репликация является фактором повышения изменчивости с последующим продлением существования уже новой информации.

Транслируемость – свойство, противостоящее бренности и характеризующее возможность передачи информации с одного носителя на другой, т.е. возможность размножения информации.

При этом жизнеспособность информации L определяется отношением скоростей рождения (V_P) и гибели ($V_Г$) носителей:

$$L = \frac{V_P}{V_Г}.$$

Если $L > 1$, число копий записи будет возрастать. При $L < 1$ информация обречена на вымирание, при $L = 1$ состояние нестабильно.

Это свойство отражает жизнеспособность информации L, определяемой отношением скорости рождения и гибели носителей. Если $L>1$, то число копий записи будет возрастать, в противном случае информация обречена на вымирание. При $L>1$ проявляется свойство *размножаемости* информации, как следствие ее транслируемости. В свою очередь, следствием размножаемости является *мультипликативность* – свойство, отражающее возможность одновременного существования одной и той же информации на различных носителях.

Действенность – комплексное свойство, связанное с использованием информации, включенной в свою информационную систему для построения различных операторов, способных совершать целенаправленные действия.

Рассматриваемое свойство характеризует способность любой информации материализоваться, воплощаясь в оператор, проявляющий действенность закодировавшей его информации.

Семантика – свойство, проявляющееся в специфике кодируемого информацией оператора, причем каждая данная информация однозначно определяет оператор, для построения которого она использовалась.

Природа целенаправленных действий такова, что она должна повышать вероятность воспроизведения кодирующей его информации. В этом смысле семантика информации всегда представляет собой отражение условий, необходимых и достаточных для ее воспроизведения. Эволюция семантики происходит в направлении улучшения условий воспроизведения информации. Для примитивных самоорганизующихся неорганических систем (по сравнению с живыми) роль семантики информации играет ее прасемантика, основанная на функциональной упорядоченности, которая обеспечивает сохранение относительно устойчивого состояния системы.

Полипотентность – свойство, проявляющееся в том, что оператор, закодированный данной информацией, может быть использован для осуществления различных действий. Это означает возможность использования одной и той же информации для решения различных задач.

Полезность – способность информации быть полезной для целенаправленных действий.

Следствием этого потенциального свойства является вывод о полезности любой информации, что оправдывает ее накопление впрок.

Полезность информации определяется возможностью использовать ее для достижения какой-либо цели. Это «потенциальное» свойство, ибо полезность – свойство содействовать событию, которое еще не произошло.

Истинность – свойство, выявляемое в ходе реализации полезности.

Критерием истинности является практика. Из полипотентности информации следует относительность ее истинности, т.е. зависимость ее от ситуации и цели. Если целью является трансляция информации, то истинность оказывается условием существования информации. Отсюда жизнеспособной может быть только истинная информация, а выявление истинности возможно только в случае, когда информация кому-то полезна. Следовательно, для жизнеспособности информации необходимо сочетание ее истинности и полезности, т.е. гармония объективного и субъективного аспектов информации, отражаемых этими терминами.

С учетом известных вариантов классификации качеств информации, рассматриваются системно и детально атрибутивные, динамические, прагматические комплексные свойства информации, составными частными свойствами которых являются и рассмотренные выше свойства информации.

2) атрибутивные свойства информации

Атрибутивные свойства – необходимые свойства, без которых информация не может существовать и включающие синтаксическую адекватность, неотрывность от физического (материального) носителя, свойства языковой природы, дискретность, непрерывность.

Синтаксическая адекватность – отображает формально-структурные характеристики информации и не затрагивает ее смыслового содержания.

Адекватность информации это:

1) определенный уровень соответствия, создаваемого с помощью полученной информации образа реальному объекту, процессу, явлению и т.п.;

2) степень соответствия информации, полученной в информационном процессе, реальному объективному состоянию дела.

Неадекватная информация может образовываться при создании новой информации на основе неполных или недостоверных данных. Однако и полные, и достоверные данные могут приводить к созданию неадекватной информации в случае применения к ним неадекватных методов.

Синтаксическая адекватность декомпозируется на показатели нижнего уровня:

1) тип носителя;

2) способ представления информации; 3) скорость передачи и обработки;

4) размеры кодов представления информации;

5) надежность и точность преобразования этих кодов и т.п.

Эти показатели отражают внешние структурные характеристики синтаксической стороны информации.

Информативность – характеризуется отношением количества синтаксической информации (по Шеннону) к объему данных $Y = I/V_{\partial}$.

Неотрывность от физического (материального) носителя и языковая природа – свойство информации, заключающееся в том, что содержание одной и той же информации может быть изложено как на разных носителях, так и на разных языках, и от этого ее смысл не должен изменяться.

Относительная независимость содержания и выражения – одно и то же содержание может быть отражено в различных знаковых формах.

Со свойствами неотрывности от физического носителя и относительной независимости содержания и выражения информации связаны рассмотренные выше свойства инвариантности, бренности и изменчивости.

Неаддитивность – свойство информации, состоящее в том, что ее содержание, соответствующее целому объекту, не равно сумме содержания соответствующих его частей, независимо от способов разбиения объекта: $A = \{a_1, a_2, a_3\}$, $A \neq a_1 + a_2 + a_3$.

Некоммутативность – свойство информации, заключающееся в исключении переместительного закона для логически связанных элементов информации: $A \rightarrow B \rightarrow C \neq A \rightarrow C \rightarrow B$.

Неассоциативность – свойство информации, заключающееся в отсутствии соединительного и сочетательного закона (связи) между логически связанными элементами информации: $A \rightarrow B \rightarrow C \neq A \rightarrow D$, $D = B + C$.

Информатика изучает особенности различных носителей и языков информации и разрабатывает новые, более совершенные. Хотя информация неотрывна от физического носителя и имеет языковую природу, она не связана жестко ни с конкретным носителем, ни с конкретным языком. Это свойство носителя не зависит от носителя и языка.

К атрибутивным свойствам информации относится такое ее свойство, как дискретность. Важными атрибутивными свойствами информации, отражающими характеристики физического (материального) носителя, являются достоверность, точность и устойчивость информации.

Достоверность информации определяется ее свойством отражать реально существующие объекты с необходимой точностью.

Измеряется достоверность информации доверительной вероятностью необходимой точности, т.е. вероятностью того, что отображаемое информацией значение параметра отличается от истинного значения этого параметра в пределах необходимой точности.

Несколько специфический смысл имеет термин достоверность правовой информации, который отражает соответствие имеющихся текстов закона первоначальному (аутентичному) тексту, образцу («эталону»).

Точность информации определяется степенью близости получаемой информации к реальному состоянию объекта, процесса, явления и т.п.

Для информации, отображаемой цифровым кодом, известны четыре классификационных понятия точности: формальная точность, измеряемая значением единицы младшего разряда числа; реальная точность, определяемая значением единицы последнего разряда числа, верность которого гарантируется; максимальная точность, которую можно получить в конкретных условиях функционирования системы; необходимая точность, определяемая функциональным назначением показателя.

Устойчивость информации отражает ее способность реагировать на изменения исходных данных без нарушения необходимой точности. Устойчивость информации, как и репрезентативность, обусловлена выбранной методикой ее отбора и формирования.

Дискретность информации – свойство конкретных знаний об окружающем мире, характеризующих отдельные фактические данные, закономерности и свойства изучаемых объектов, распространяемые в виде различных сообщений, состоящих из слов, фраз, параграфов, глав и других фрагментов, объединенных в статьи, журналы, книги и др.

Непрерывность информации – свойство новой информации, зафиксированной в отдельных сообщениях, сливаться с уже накопленной ранее, способствуя поступательному развитию общества.

3) динамические свойства информации

Динамические свойства информации – к данным свойствам следует отнести свойство ее роста, непрерывное создание и накопление новой информации за счет активной и все возрастающей общественно-политической, научно-технической, производственной или другой деятельности людей.

Повторяемость – это способность информации к многократному распространению, обусловленная ее независимостью в сетях коммуникаций от создателя языка и носителя.

Многократная используемость – это свойство одной и той же информации одновременно быть полученной и использованной несколькими потребителями с сохранением ее автором после передачи информационного сообщения всей суммы знаний, составивших это сообщение.

Полезность и многократность использования информации в определенной степени по содержанию адекватны свойствам транслируемости, размножаемости, мультипликативности.

Старение информации – условное свойство, характеризующее старение носителей информации, уменьшающее со временем интенсивность использования конкретного источника информации, так или иначе отвергающей, дополняющей или изменяющей данную, и по другим причинам.

Рассеиваемость по различным источникам – это способность информации в результате проведения научной работы фиксироваться в отчете ученого, затем в статье, монографии и т.д. Она также рассеивается по публикациям в смежных областях знаний.

Рассеянию способствуют многие из отмеченных ранее свойств информации (независимость от создателя, языка и носителя, кумулятивность, старение и др.), а также различные условия ее распространения во времени и пространстве.

Динамические свойства информации заметно влияют на функционирование информационных потоков и на установление коммуникаций между создателем и потребителем информации. Их необходимо учитывать при осуществлении очень важного информационного процесса – сбора информации, от полноты которого во многом зависит и качество функционирования информационных систем.

4) прагматические свойства информации

Прагматические свойства – качества, характеризующиеся потребительской адекватностью информации, отражающей отношение информации и потребителя, соответствие информации цели управления, эффективность ее использования, потребности для решения задач информационного обеспечения деятельности личности, общества, государства.

Прагматические свойства информации проявляются только в ходе использования информации при наличии единства информации (объекта), пользователя и цели управления. Прагматический аспект рассмотрения связан с ценностью, полезностью использования информации при выработке потребителем решения для достижения своей цели. С этой точки зрения анализируются потребительские свойства информации.

Важность информации – обобщенный показатель, характеризующий значимость информации для решаемых задач и организации обработки информации.

Это свойство информации оценивается двумя группами показателей. Первая группа включает: важность задач обеспечения деятельности, для которых используется информация, $K_{ви}$; важность информации для эффективного решения соответствующих задач, $K_{вз}$. Вторая группа состоит из следующих частых показателей: уровень потерь при нарушении качества и нежелательных изменений в процессе обработки под воздействием дестабилизирующих факторов, $K_{пи}$; важность информации с учетом затрат на восстановление ее качества, $K_{св}$.

Для оценки важности информации по этим показателям в настоящее время нет строгих количественных методов. В силу этого используют качественные методы, в том числе методы экспертных оценок. С этой целью значения переменных показателей (шкалы) выражаются лингвистическими переменными.

Показатели смысла и новизны характеризуют способность информации перемещаться в социальных коммуникациях, сопоставлять одну информацию с другой и находить ту ее часть, которой не знает потребитель.

Содержательность (релевантность) информации характеризует соответствие ее содержания потребностям решаемой задачи и отражает семантическую емкость, равную отношению количества семантической информации $I_{сд}$ в сообщении к объему обрабатываемых данных $V_д$, т.е. $C = I_c / V_д$.

Релевантность – объективно существующее смысловое соответствие между содержанием информационного сообщения и содержанием информационного запроса.

Релевантность информации (документа, данных) – объективная характеристика, отражающая степень соответствия искомой информации (документа, данных) потребностям решаемой задачи.

Объективность и субъективность информации. Понятие объективности информации является относительным в силу наличия субъективности методов ее оценки. Более объективной принято считать ту информацию, в которую методы вносят меньший субъективный элемент. В ходе информационного процесса степень объективности информации всегда понижается. Это свойство учитывают, например, в правовых дисциплинах, где по-разному обрабатываются показания лиц, непосредственно наблюдавших события или получивших информацию косвенным путем (посредством умозаключений или со слов третьих лиц).

В такой же степени объективность информации учитывают в исторических дисциплинах. Одни и те же события, зафиксированные в исторических документах разных стран и народов, выглядят совершенно по-разному. У историков имеются свои методы для тестирования объективности исторических данных и создания новых, более достоверных данных путем сопоставления, фильтрации и селекции исходных данных. Необходимо обратить внимание на то, что здесь речь идет не о повышении объективности данных, а о повышении их достоверности (это совсем другое свойство).

Полнота информации во многом характеризует качество информации и определяет достаточность данных для принятия решений или для создания новых данных на основе имеющихся. Чем полнее данные, тем шире диапазон методов, которые можно использовать, тем проще подобрать метод, вносящий минимум погрешностей в ход информационного процесса.

Толерантность информации (документа) – субъективная характеристика, отражающая удобство восприятия и использования пользователем искомой информации (документа) в процессе решения задачи потребителя.

Сложность определения коэффициента релевантности C сопряжена с количественным расчетом объемов информации. В сфере научно-технической информации принято за $V_д$ считать общее количество документированных данных, что требует, в свою очередь решения известных в литературе задач классификации источников информации и документов по видам (опубликованные и неопубликованные первичные и вторичные), определения и оценки свойств (атрибутивных, прагматических и динамических) документов.

Информационный кадастр объекта – полная и хорошо структурированная совокупность данных, необходимых и достаточных для эффективного функционирования объекта в соответствии с его целевым назначением.

Такой кадастр может быть представлен объектно-характеристическими таблицами, которые включают: классификационный словарь понятий; массив понятий; массив значений характеристик; массив сообщений.

В качестве меры полноты информации можно использовать соответствующий коэффициент K_n, равный:

$$K_n = \sum_{\forall\nu}\sum_{\forall\mu} d_{\mu\nu} / \mu\nu ,$$

где $d_{\mu\nu}$ – элемент объектно-характеристической таблицы в μ-й строке и ν-м столбце, равный 1 или 0, при наличии или отсутствии информации об элементе объекте; m и n – число строк и столбцов таблицы.

С увеличением содержательности информации растет семантическая пропускная способность информационной системы, так как для получения одних и тех же сведений требуется преобразовать меньший объем данных.

Репрезентативность (адекватность) информации – степень ее соответствия действительному состоянию объектов материального мира (явлений, процессов, свойств, задач, систем и др.), которые отражает информация, определяемая правильностью, обоснованностью и объективностью ее генерирования (отбора, формирования, съема, определения, установления) и продолжительностью интервала времени до момента оценивания ее адекватного отражения свойств объекта.

Достаточность (полнота) информации означает, что она содержит минимальный, но достаточный для решения задачи или принятия правильного решения состав (набор) показателей.

Пертинентность – субъективное свойство информации, отражающее соответствие содержания информационного сообщения информационным интересам данного потребителя.

Ценность – определяет, в какой степени получение информации способствует достижению целей.

Понятие *полноты* информации связано с ее смысловым содержанием (семантикой) и прагматикой. Как неполная, т.е. недостаточная для принятия правильного решения, так и избыточная информация снижает эффективность принимаемых пользователем решений. Полнота информации оценивается относительно вполне определенной задачи, группы задач. Следовательно, для определения показателя полноты необходимо для каждой задачи определить соответствующий перечень в виде объектно-характеристической таблицы, представляющей матрицу со строками-наименованиями объектов и столбцами-наименованиями характеристик объектов, входящих в решаемую задачу. Совокупность таких объектов будет составлять информационный кадастр.

Для учета важности и значимости элементов таблицы вводится соответствующий коэффициент важности элемента $K_{\mu\nu}$. Тогда взвешенная полнота информации в рассматриваемой объектно-характеристической таблице будет определяться согласно следующему выражению:

$$K_b = \sum_{\forall\nu}\sum_{\forall\mu}\frac{d_{\mu\nu}K_{\mu\nu}}{\mu\nu} * \sum_{\forall\nu}\sum_{\forall\mu}\frac{K_{\mu\nu}}{\mu\nu}$$

Доступность (*толерантность*) информации – характеризует удобство восприятия, и пользования информацией в процессе решения задач, для которых она используется.

Это свойство обеспечивается выполнением соответствующих процедур получения и преобразования информации. Например, в информационной системе информация преобразуется к доступной и удобной для восприятия пользователя форме путем согласования ее семантической формы с тезаурусом пользователя.

Понятие толерантности является широким, весьма субъективным и в значительной степени неопределенным. В силу этого использование строгих математических методов для получения количественных значений этого показателя весьма затруднено. Поэтому наиболее подходящими считаются методы экспертных лингвистических оценок.

Актуальность информации определяется степенью сохранения ценности информации для управления в момент ее использования и зависит от динамики изменения ее характеристик и от интервала времени, прошедшего с момента возникновения данной информации.

Актуальность информации – это степень соответствия информации текущему моменту времени. Нередко с актуальностью, как и с полнотой, связывают коммерческую ценность информации. Поскольку информационные процессы растянуты во времени, то достоверная и адекватная, но устаревшая информация может приводить к ошибочным решениям. Необходимость поиска (или разработки) адекватного метода для работы с данными может приводить к такой задержке в получении информации, что она становится неактуальной и ненужной. На этом, в частности, основаны многие современные системы шифрования данных с открытым ключом. Лица, не владеющие ключом (методом) для чтения данных, могут заняться поиском ключа, поскольку алгоритм его работы доступен, но продолжительность этого поиска столь велика, что за время работы информация теряет актуальность и, соответственно, связанную с ней практическую ценность.

Точность информации определяется степенью близости получаемой информации к реальному состоянию объекта, процесса, явления и т.п. Для информации, отображаемой цифровым кодом, известны четыре классификационных понятия точности:

– формальная точность, измеряемая значением единицы младшего разряда числа;

– реальная точность, определяемая значением единицы последнего разряда числа, верность которого гарантируется;

– максимальная точность, которую можно получить в конкретных условиях функционирования системы;

– необходимая точность, определяемая функциональным назначением показателя.

Достоверность информации определяется ее свойством отражать реально существующие объекты с необходимой точностью. Измеряется достоверность информации доверительной вероятностью необходимой точности, т.е. вероятностью того, что отображаемое информацией значение параметра отличается от истинного значения этого параметра в пределах необходимой точности.

Устойчивость информации отражает ее способность реагировать на изменения исходных данных без нарушения необходимой точности. Устойчивость информации, как и репрезентативность, обусловлена выбранной методикой ее отбора и формирования.

Своевременность информации означает ее поступление не позже заранее назначенного момента времени, согласованного со временем решения поставленной задачи.

Старение информации – уменьшение ее ценности со временем.

Одним из существенных показателей качества информации является ее безопасность.

Безопасность информации – состояние информации, информационных ресурсов и информационных систем, при котором с требуемой вероятностью обеспечивается защита информации (данных) от утечки, хищения, утраты, несанкционированного уничтожения, искажения, модификации (подделки), копирования, блокировки и т.п.

Этот показатель позволяет оценить степень защищенности информации, имеющей статус защищаемой, от случайного или злоумышленного ее получения, лицами или осуществляемыми в системе процессами, не имеющими на это полномочий. Особенность рассматриваемого аспекта качества информации состоит в том, что ее свойство, называемое безопасностью информации, определяется угрозами и дестабилизирующими факторами естественного и искусственного происхождения, среди которых важное место занимают злоумышленные действия людей, которые носят случайный характер и являются трудно предсказуемыми. Это затрудняет получение априорной оценки безопасности информации. Детальное рассмотрение вопросов защиты информации будет представлено в последующих главах.

5) познавательные свойства информации

Высокий и сложный уровень организации производства, науки, техники и культуры современного общества связан с циркуляцией потоков информации самого разного характера. Объемы новой информации непрерывно растут, растет и скорость распространения информации. Увеличилось многообразие средств сбора, хранения, переработки и распространения информации. Через различные коммуникационные каналы человек получает информацию, использует ее и согласует свою деятельность с интересами и потребностями общества.

Важно напомнить, что познавательные свойства информации тесно связаны с первичной и вторичной семантической информацией. Как уже отмечалось ранее, первичная семантическая информация представляет зафиксированное отображение выделенной человеком стороны (сторон) объекта. При этом человек руководствуется выделением относительно устойчивых категорий, образующих логическую структуру первичной семантической информации. Здесь категория это – обобщенное отражение объективной реальности и существенное определение аспекта объекта. Как известно вторичная семантическая информация отражает результаты аналитико-синтетического и логического преобразования первичной семантической информации с помощью знаков заданной формы представления.

Получение вторичной семантической информации можно рассматривать как продолжение процесса познания объекта, заключающееся в выявлении наиболее устойчивых и характерных признаков.

С учетом этого познавательные свойства информации определяются ее научностью. Это понятие определили известные ученые А.И. Михайлов, А.И. Черный, Р.С. Гиляревский.

Научная информация – информация, получаемая в процессе познания и адекватного отображения явлений и законов природы, общества и мышления и используемая в общественно-исторической практике.

Научная информация формируется в результате деятельности отдельных ученых и специалистов или их коллективов и фиксируется в системе точных понятий, суждений, умозаключений, теорий, гипотез, т.е. информация становится научной лишь тогда, когда она подвергнута обработке и обобщению абстрактно-логическим мышлением. Именно этим научная информация отличается от сведений или данных, получаемых человеком в процессе чувственного познания.

Научная информация подразделяется на виды по областям ее получения и использования (техническая, экономическая, медицинская, культурная, социальная и др.).

Техническая информация характеризует физические процессы в различных объектах при создании продукции из исходных составляющих.

Экономическая информация дает сведения о стоимости объекта, его производительности, трудовых затратах на его производство, эффективность работы и т.д.

Медицинская, культурная, социальная информация несут сведения и факты о человеке, коллективе или обществе в целом как объектах исследования и управления.

Научно-техническая информация – это сведения о документах и фактах, получаемых в ходе научной, научно-технической, производственной и общественной деятельности.

По форме восприятия, как и любая семантическая информация, научно-техническая информация может быть визуальной, звуковой, тактильной (осязательной).

Научно-техническую информацию (НТИ) различают по назначению, типу передаваемой информации, способу ее распространения, степени ее аналитико-синтетической переработки, области получения или использования.

Структура научно-технической информации представляет содержательный и формальный аспекты.

По содержанию научно-техническая информация делится на информацию:

1) о научных фактах (класс А);

2) научных гипотезах, концепциях и теориях (когда объединяется некоторая совокупность научных фактов и объясняется взаимосвязь между ними) (класс В);

3) объединяющую некоторую совокупность научных фактов, гипотез, концепций, теорий и законов, образующую основу данной науки или области знания (класс С);

4) отображающую и формирующую общий подход к познанию и измерению окружающего нас мира (класс D).

Содержательная структура научно-технической информации в достаточной степени условна. В одних и тех же фрагментах НТИ может содержаться информация разных классов.

Формальная структура так же иерархична, как и содержательная. Низшие уровни являются общими и для всей семантической информации: отдельные буквы, слова, фразы, смысловые комплексы, произведения. На высших уровнях иерархии в формальной структуре научной информации находится научно-техническая литература.

Следует отметить, что содержательный и формальный аспекты научной информации находятся во взаимосвязи друг с другом: каждый из содержательных классов тяготеет к определенным видам документов.

Основными свойствами научно-технической информации являются:

1) неотрывность от физического носителя;

2) неаддитивность, некоммутативность и неассоциативность (т.е. содержащаяся в каком-либо сообщении информация не является арифметической суммой составляющих это сообщение элементов, эти элементы нельзя без искажения смысла сообщения располагать в сообщении в любой произвольной последовательности и группировать в разные сочетания);

3) ценность (информация тем ценнее, чем больше способствует достижению цели, стоящей перед ее получателем);

4) общественная природа (источником информации является познавательная деятельность людей, общества);

5) семантический характер;

6) языковая природа (информация выражается с помощью языка, т.е. знаковой системы любой природы, служащей средством общения, мышления, выражения мысли. Язык может быть естественным, используемым в повседневной жизни и служащим формой выражения мыслей и средством общения между людьми, и искусственным, созданным людьми для определенных целей, (например, язык математической символики, информационно-поисковый, алгоритмический и др.);

7) независимость от языка и носителя;

8) дискретность (единицами информации как средствами выражения являются слова, предложения, отрывки текста, а в плане содержания – понятия, высказывания, описания фактов, гипотезы, теории, законы и др.);

9) кумулятивность (связана с одной из основных закономерностей развития науки – ее преемственностью и интернациональным характером, с концентрацией информации во времени, т.е. переходом научного знания к все более высоким уровням абстракции);

10) независимость от создателей;

11) старение (основной причиной старения информации является не само время, а появление новой информации, с поступлением которой прежняя информация оказывается неверной, перестает адекватно отображать явления и закономерности материального мира, человеческого общества и мышления);

12) рассеяние (т.е. существование в многочисленных научных произведениях за счет процессов дифференциации и интеграции познания, что является важной закономерностью развития науки).

Ценность (полезность) информации определяется пользой, которую она приносит конкретному потребителю, исходя из наличия неизвестных ему знаний, использование которых повышает эффективность его общественно-политической, научной, технической, производственной или иной деятельности.

Целесообразность обусловливает социальную значимость информационного обслуживания и меру его научно-познавательного и идеологического влияния на членов общества.

Ценность одной и той же информации для разных потребителей, как правило, разная. Ее определяют: объем содержащихся в ней знаний; информационные потребности и объем знаний потребителя; время поступления; возможность использования с минимальным расходом времени, трудовых и материальных затрат.

Информация может иметь нулевую или отрицательную ценность, когда она или бесполезна, или содержит ложные или устаревшие сведения, использование которых не приближает, а иногда и отдаляет потребителя от правильного решения стоящих перед ним задач.

Для информации, особенно научно-технической, характерно в ретроспективном плане стремление к сжатию: все основное получает простое и краткое выражение, а второстепенное и частное – отсеивается. В этом процессе проявляется еще одно важное прагматическое свойство информации – кумулятивность.

Кумулятивность – свойство информации накапливаться, откладываться. Оно постоянно используется учеными и специалистами, которые, наряду с получением новых знаний, проводят также суммирование, систематизацию, оценку и обобщение имеющихся данных.

В заключение следует отметить, что такие параметры качества информации, как репрезентативность, содержательность, достаточность, доступность, устойчивость целиком определяются на методическом уровне разработки информационных систем. Параметры актуальности, своевременности, точности и достоверности обусловливаются в большей степени также на методическом уровне, однако на их величину существенно влияет и характер функционирования системы, в первую очередь ее надежность. При этом параметры актуальности и точности жестко связаны соответственно с параметрами своевременности и достоверности.

2.3. Количество информации и классификация ее мер

Важным вопросом теории информации является установление меры количества и качества информации. Решение этой задачи вызвало появление различных известных направлений теории информации.

Структурная теория рассматривает дискретное построение массивов информации и их измерение простым подсчетом информационных элементов (квантов) или комбинаторным методом, предполагающим простейшее кодирование массивов информации. Эта теория применяется для оценки возможностей аппаратуры ИС, в том числе каналов связи (КС), запоминающих (ЗУ) и регистрирующих устройств вне условий их применения, ИС в конкретных применениях (при передаче по системам связи информации с определенными статистическими характеристиками).

Меры информации применяются для оценки дискретных и непрерывных источников информации и создаваемых ими сообщений.

Дискретный источник информации в конечное время создает конечное множество сообщений, которые имеют счетное множество элементов, создаваемых источником последовательно во времени.

Набор элементов называют алфавитом источника, а элементы – буквами, включающими цифры и знаки.

Объем алфавита – число букв в алфавите.

Непрерывные сообщения отражаются какой-либо физической величиной, изменяющейся в заданном интервале времени. Получение конечного множества сообщений за конечный промежуток времени достигается путем дискретизации (во времени) и квантования (по уровню).

Статистическая теория оперирует понятием энтропии как меры неопределенности, учитывающей вероятность появления, а, следовательно, и информативность тех или иных сообщений.

Семантическая теория учитывает целесообразность, ценность, полезность или существенность информации.

В рамках этих теорий в настоящее время для измерения информации вводятся показатели количества информации (объема данных) $V_д$, и степени информативности Y, между которыми существует следующее отношение $Y = I/V_д$.

В зависимости от рассматриваемых теоретических направлений эти показатели имеют различные выражения, интерпретацию, меру количества информации и объема данных:

– синтаксическая мера оперирует объемом данных и количеством информации, выраженной через энтропию;

– семантическая мера оперирует количеством информации, выраженной через ее объем и степень содержательности;

– прагматическая мера, определяемая ее полезностью, выраженной через соответствующие экономические эффекты.

2.4. Единицы измерения информации

Информационные меры, как известно, рассматриваются в трех аспектах: структурном, статистическом и семантическом.

В структурном аспекте рассматривается строение массивов информации и их измерение простым подсчетом информационных элементов или комбинаторным методом. Структурный подход применяется для оценки возможностей информационных систем вне зависимости от условий их применения.

При статистическом подходе используется понятие энтропии, как меры неопределенности, учитывающей вероятность появления и информативность того или иного сообщения. Статистический подход учитывает конкретные условия применения информационных систем.

Семантический подход позволяет выделить полезность или ценность информационного сообщения.

1) структурная мера информации

Структурная мера предполагает учет только дискретного строения данного информационного комплекса, представляющего количество содержащихся в нем информационных элементов, связей между ними или их комбинации (можно представить декартовым произведением множеств).

Информационный элемент – неделимая часть – квант информации в дискретных моделях реальных информационных комплексов, а также элементы алфавитов в числовых системах.

В структурной теории различают *геометрическую, комбинаторную, аддитивную* меры информации.

Наибольшее распространение получила двоичная аддитивная мера Хартли, измеряющая количество информации в двоичных единицах – битах.

Геометрическая мера определения количества информации представляет метод измерения длины линии, площади, или объема геометрической модели информационного комплекса в количестве дискретных единиц (квантов).

Этим методом определяется потенциальное, т.е. максимальное количество информации в заданных структурных габаритах по всем измерениям, которое называется информационной емкостью, представляемой количеством квантов в полном массиве исследуемой информационной системы.

Так, если для параметрического X пространства N во времени T, представляющего трехмерный информационный комплекс $\Phi^3(X, T, N)$ дискретные отсчеты осуществляются через интервалы Δ_X, Δ_T, Δ_N, то непрерывные координаты распадаются на элементы (кванты), количество которых равно: $m_X = X / \Delta_X$; $m_T = T / \Delta_T$; $m_N = N / \Delta_N$. Общее количество информации комплекса X, T, N, в квантах, будет определяться геометрическим методом и равно: $M = m_X \cdot m_T \cdot m_N$.

Так, геометрическая мера предполагает измерение параметра геометрической модели информационного сообщения (длины, площади, объема и т. п.) в дискретных единицах.

Например, геометрической моделью информации может быть линия единичной длины (рис. 2.3, *а* – одноразрядное слово, принимающее значение 0 или 1), квадрат (рис. 2.3, *б* – двухразрядное слово) или куб (рис. 2.3, *в* – трехразрядное слово). Максимально возможное количество информации в заданных структурах определяет информационную емкость модели (системы), которая определяется как сумма дискретных значений по всем измерениям (координатам).

Комбинаторная мера – количество информации, определяемое количеством комбинаций элементов и характеризующее комбинаторное свойство потенциального структурного разнообразия информационных комплексов.

Комбинирование возможно в комплексах с неодинаковыми элементами, переменными, связями, разнообразными позициями, отличающимися один от другого любым признаком – размером, формой, цветом, местоположением, позицией и т.п.

Примером могут служить позиционная двоичная система счисления, 11110, 01111, система образования геометрических фигур и т.п. Для подсчета числа возможных комбинаций рассматриваемых типов элементов в математике существуют соответствующие понятия. Математическая интерпретация видов соединения элементов может быть представлена известными выражениями комбинаторики (выборкой, перестановкой, размещением, сочетанием). Выборка из элементов данного множества – то же, что совокупность выбранных элементов. Возможны две схемы выбора – без возврата элементов множества и с возвратом. В первом случае выборка не содержит одинаковых элементов, а во втором – может содержать. В соответствии с этим говорят о выборках с повторениями и без повторений. Число элементов выборки называется ее объемом. Если в выборке зафиксирован порядок следования элементов, то выборка называется упорядоченной, в противном случае – неупорядоченной.

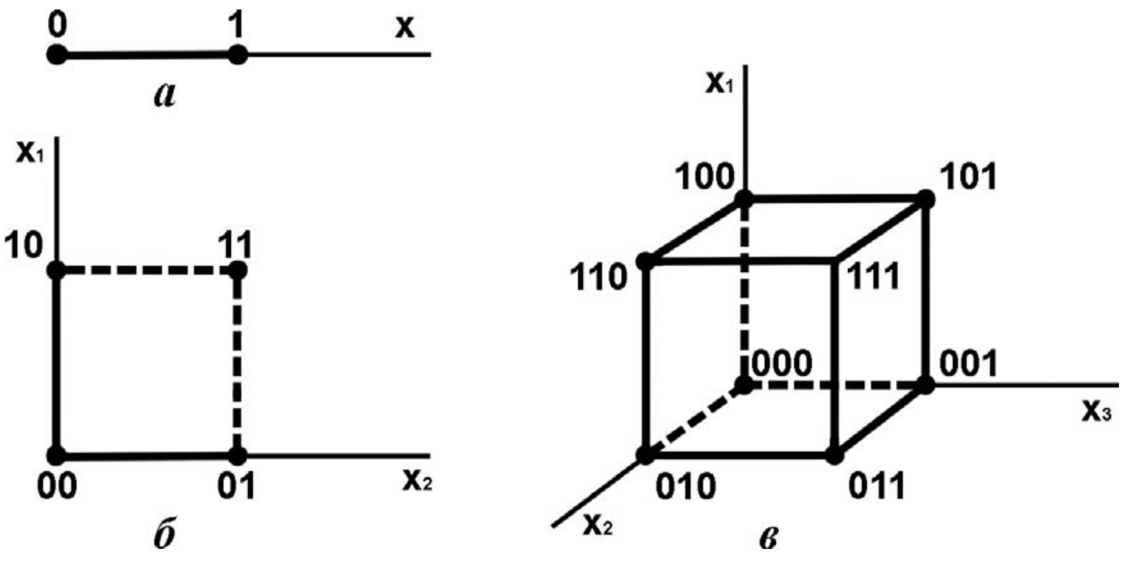


Рис. 2.3. Геометрическая модель информации

Пусть имеем множество S_n из n элементов (например, множество букв алфавита, множество чисел системы счисления и т.п.). *Перестановкой* из n элементов называется упорядоченная выборка без повторений из всех элементов множества S_n. Число всех перестановок из n элементов равно $P_n=n!=1\cdot 2\cdot 3\cdot\ldots\cdot n!$.

Размещение из n элементов по m это упорядоченная выборка объема m из S_n (без повторения элементов). Число размещений из n элементов по m различаются, составом, их порядком и равно: $A_n^m = \frac{n!}{(n-m)!} = n(n-1)...(n-m+1)$. Возможное число размещений с повторениями по n из m элементов равно: $(A_n^m)_{повт} = n^m$.

Сочетания из n элементов по m это неупорядоченная выборка объема m из S_n (без повторения элементов). Число сочетаний из n элементов по m различается составом элементов и равно:

$$C_n^m = \binom{n}{m} \equiv \frac{n!}{m!(n-m)} = \frac{n(n-1)...(n-m+1)}{1*2*3*...*m}.$$

Если при образовании выборок используется схема с возвратом элементов, то получаются перестановки, размещения и сочетания с повторениями. Приведем соответствующие формулы для числа этих выборок.

Сочетания с повторениями различаются составом элементов, но элементы в них могут повторяться. Число таких сочетаний с повторениями из n элементов по m равно: $(C_n^m)_{повт} = \frac{(n+m-1)!}{m!(n-1)!} = \binom{m}{n+m-1}$.

Число перестановок с повторениями элементов (один из элементов повторяется α, другой β, а последний – γ раз) равно: $(P_n)_{повт} = \frac{n!}{\alpha!\beta!...\gamma!} = \frac{(\alpha+\beta+...+\gamma)!}{\alpha!\beta!...\gamma!}$, где α – число повторений 1-го элемента, β – число повторений второго элемента и т.д.

Использование комбинаторной меры позволяет определить количество информации, совпадающее с числом возможных или действительно существующих комбинаций, т.е. оценить структурное разнообразие.

Аддитивная мера Хартли – представляет двоичную логарифмическую меру числового поля глубиной h и длиной l, позволяющую определить количество информации I в двоичных единицах – битах: $I = log_2 Q = log_2 h^l = l*log_2 h$, где h – глубина числа равная количеству различных элементов (знаков), содержащихся в принятом алфавите; l – длина числа, равная количеству числовых гнезд (количество повторений алфавита), необходимых и достаточных для представления чисел нужной величины.

Длина числа соответствует разрядности системы счисления и кодирования.

В качестве вариантов информационных систем для аддитивного расчета информации можно представить:

– роликовый счетчик с l роликами и h цифрами на ободке каждого ролика;

– комбинированный коммутатор с l переключателями, из которых каждый переключает h цепей;

– запоминающее устройство с l ячейками, каждая емкостью h единиц;

– изображение, состоящее из l дискретных элементов, каждый из которых характеризуется h градациями цвета и тона;

– страница печатного текста, состоящая из l_1 строк и l_2 букв в каждой строке емкостью, т.е. всего $l = l_1 \cdot l_2$ числовых или буквенных гнезд, глубина каждого из которых равна h условных единиц.

Для двоичной системы счисления (глубина h=2) и количества разрядов (длины числа) l = 1 при использовании двоичного логарифма потенциальное количество информации равно одному биту: log_2 2 = 1 *бит*, где 1 бит – единица информации в принятой системе, соответствующая одному элементарному событию, которое может произойти или не произойти.

Для того чтобы сообщение можно было передать получателю, необходимо воспользоваться некоторым физическим процессом, способным с той или иной скоростью распространяться от источника к получателю сообщения. Изменяющийся во времени физический процесс, отражающий передаваемое сообщение *называется сигналом* (рис. 2.4).

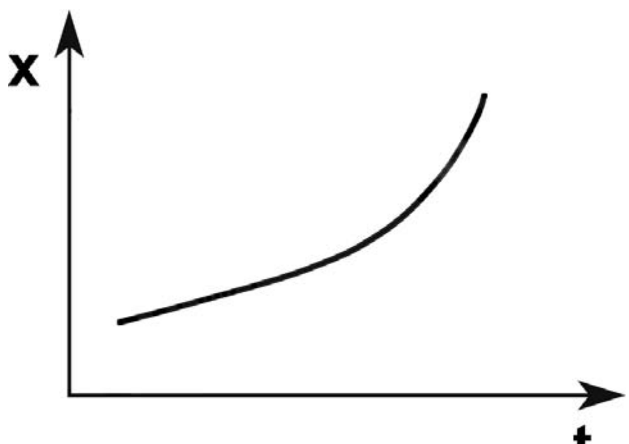


Рис. 2.4. Представление сигнала

Сообщения могут быть функциями времени (когда информация представлена в виде первичных сигналов: речь, музыка) и не является ими (когда информация представлена в виде совокупности знаков).

Сигнал всегда является функцией времени. В зависимости от того, какие значения могут принимать аргумент (время t) и уровни сигналов их делят на 4 типа.

Непрерывный или аналоговый сигналы (случайные сигналы этого типа называются непрерывными случайными процессами). Они определены для всех моментов времени и могут принимать все значения из заданного диапазона. Чаще всего физические процессы, порождающие сигналы, являются непрерывными. Этим и объясняется второе название сигналов данного типа – аналоговый, т.е. аналогичные порождающим процессам (рис. 2.5).

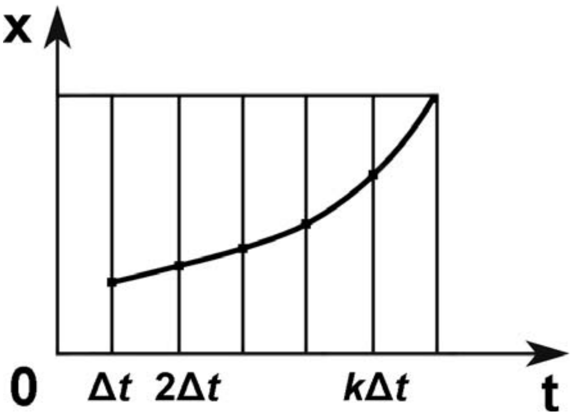


Рис. 2.5. Аналоговый сигнал

Дискретизированный или дискретно непрерывные сигналы (случайные сигналы этого типа называют процессами с дискретным временем или непрерывными случайными последовательностями). Они определены лишь в отдельные моменты времени и могут принимать любые значения уровня. Временной интервал Δt между соседними отсчетами называется шагом дискретизации. Часто такие сигналы называют дискретными по времени, рис. 2.6.

Дискретные по уровню или квантованные сигналы (случайные сигналы этого типа называют дискретными случайными процессами). Они определены для всех моментов времени и принимают лишь разрешенные значения уровней отделенные от друг друга на величину шага квантования $\Delta x = xk + 1 + xk$.

Дискретные по уровню и по времени сигналы (случайные сигналы этого типа называют дискретными случайными последовательностями) (рис. 2.7). Они определены лишь в отдельные разрешенные моменты времени и могут принимать лишь разрешенные значения уровней.

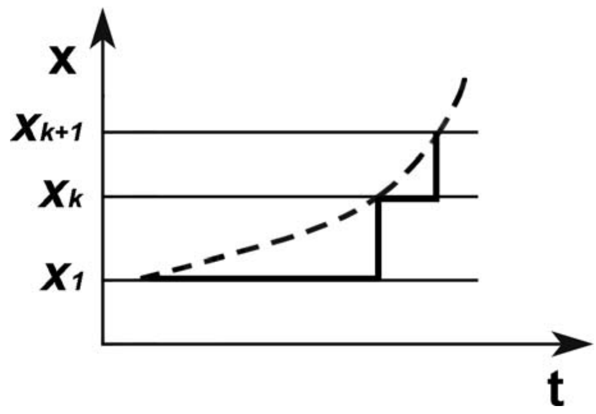


Рис. 2.6. Дискретизированный сигнал

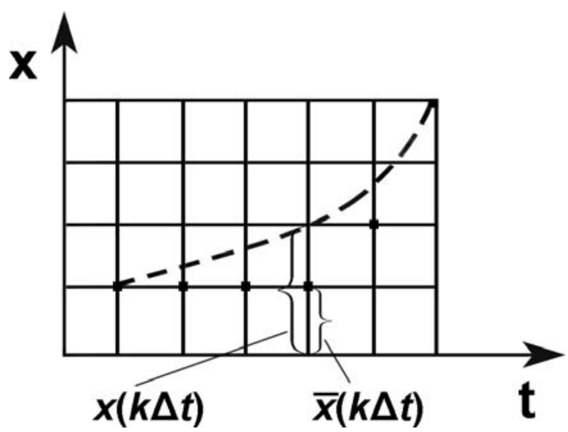


Рис. 2.7. Дискретный сигнал

2) семантическая мера информации

Семантика – наука о смысле, содержании информации.

Место семантики и ее общее содержание в системе понятий семиотики (греч. – знак, признак – науки о знаках, словах, языках) можно представить соответственно рис. 2.8 и табл. 2.2.

Этой структуре семиотики соответствуют и соответствующие аспекты теории информации.

Знак – условное изображение элементов сообщения.

Слово – совокупность знаков, имеющая смысловое (предметное) значение.

Язык – словарь и правила пользования им.

Рассмотренные выше структурная и статистическая оценки информации относятся и к синтаксическому аспекту.

Сигматический аспект представляет теорию сигналов и кодов, рассматривающую условные обозначения элементов информации.

Сигналы – физические носители знаков, слов, языка.

Коды – обозначения знаков, слов, кодов.

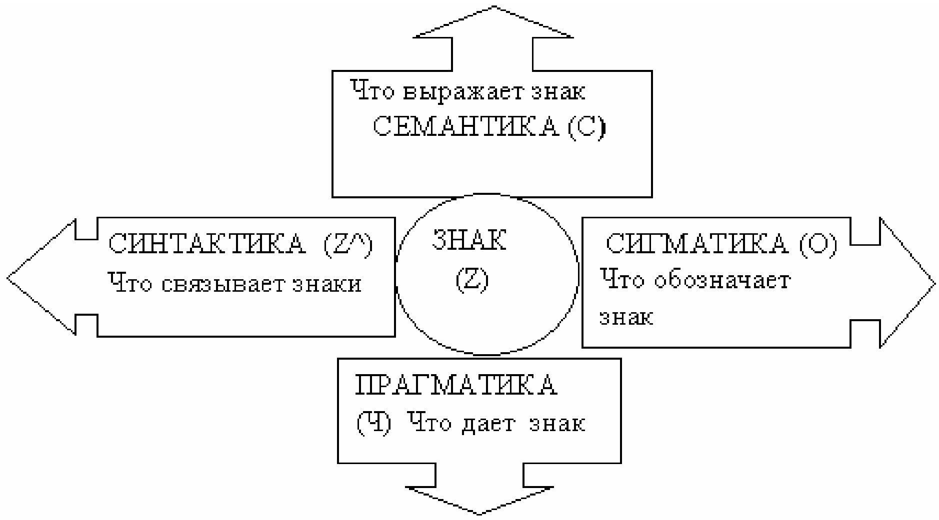


Рис. 2.8. Основные понятия семиотики

Таблица 2.2

Характеристики и содержание семиотики

Характеристики понятий	Содержание понятий семиотики			
	Синтактика	Семантика	Сигматика	Прагматика
Дефиниция	Синтаксис (греч – составление) – отношение между знаками и словами	Семантикос (греч. – обозначающий) - значение знаков и слов	Сигматика (греч. – учение о знаках) – отношение между знаками (словами) и объектами отражения	Прагма (греч. – действие, практика) – практическая полезность знаков и слов
Отношение к языку	Структурная сторона языка Z→ Z^	Смысловая сторона языка Z→C	Словарная сторона языка Z→O	Потребительская сторона языка Z→Ч
Объект исследования	Z^ – соотносящийся знак	C – смысл	O – объект	Ч – человек

Особенностью семантической и прагматической оценки количества информации, обрабатываемой системами организационно-технического типа (человеко-машинными, эргатическими), является слияние этих оценок, так как не имеющие смысла сведения бесполезны, а бесполезные сведения бессмысленны.

Оценка эффективности логического вывода, степени приближения к истине требует некоторой формализации смысла. В настоящее время существует несколько путей такой формализации, которые составляют семантическую теорию информации.

Одним из путей формализации смысла заключается в использовании для целей измерения смысла функции истинности и ложности логических высказываний (предложений), а полученная оценка называется содержательностью информации.

3) содержательность информации

При оценке содержательности информации в качестве основы дискретного описания объекта берется неделимое (аналог атомарному) предложение, подобное элементарному событию теории вероятности и соответствующее неделимому кванту сообщения. Тогда мера содержательности cont события i (от английского content – содержание) выражается через функцию меры m как содержательность его отрицания: $Cont(i) = \bar{m}(i) = 1\text{-}m(i)$. Логические функции истинности $m(i)$ и ложности $\bar{m}(i)$ сходны с функциями вероятностей событий и антисобытий и для них имеют место условия: $m(i) + \bar{m}(i) = 1;\ p(i) + q(i) = 1;\ 0 \le m(i) \le 1.$

Сходны также методы определения статистического (I) и логического (Inf) количества информации:

$$I = \log_2 \frac{1}{p(i)} = -\log_2 p(i)\,;$$

$$Inf = \log_2 \left(\frac{1}{1 - cont(i)} \right) = \log_2 \frac{1}{m(i)} = -\log_2 m(i)\,.$$

Отличие этих оценок состоит в том, что статистическая учитывает вероятности реализации событий, а логическая – меры истинности или ложности событий, что приближает к оценке смысла информации.

4) целесообразность информации

Для управленческих систем полезность информации желательно оценивать с позиций получаемого от нее эффекта. Для этого Харкевичем А.А. была предложена мера целесообразности информации, определяемая изменением вероятности достижения цели управления при получении дополнительной информации. При этом предполагается три условия и результата исходов:

1) полученная информация пустая – не изменяется вероятность достижения цели – мера равна нулю;

2) полученная информация изменяет дела в худшую сторону (дезинформация) – уменьшение вероятности достижения цели – измеряемая в отрицательных значениях количества информации;

3) полученная информация является благоприятной, добротной – увеличивается вероятность достижения цели – измеряется положительной величиной количества информации.

Аналитическое выражение меры целесообразности можно представить в следующем виде: $I_{цел} = \log_2 p_1 - \log_2 p_0 = \log_2 \frac{p_1}{p_2}$,

где p_0, p_1 – начальная (до получения информации) и конечная (после получения информации) вероятности достижения цели.

5) существенность информации

Полученные значения величин, точки пространства событий и моменты времени не равносущественны как сами по себе, так и во взаимных отношениях. С учетом этого различают: существенность самого события; существенность времени совершения события или его наблюдения (рано – поздно – момент); существенность места, адреса, номера, точки пространства, координаты совершения события.

Измерение величины X можно характеризовать несколькими функциями величины x: вероятностью, погрешности измерения, существенностью. Каждой из этих функций можно поставить в соответствие меру информации: меру Хартли, оцениваемую функцией погрешности при фиксированных значениях функций вероятности; меру Шеннона, оцениваемую функциями вероятности при фиксированных значениях функции погрешности и существенности; меру существенности, оцениваемую фиксированными функциями погрешности и вероятности.

6) тезаурусная мера

Для измерения смыслового содержания информации, т.е. ее количества на семантическом уровне, наибольшее признание получила тезаурусная мера, которая связывает семантические свойства информации со способностью пользователя принимать поступившее сообщение.

Количество семантической информации, содержащейся в некотором сообщении (тексте), оценивается степенью изменения индивидуального тезауруса под воздействием этого сообщения.

Тезаурус – список слов (названий объектов и названий их свойств), совокупность сведений, в которых указаны смысловые связи между этими словами и сведениями, которыми располагает пользователь или система.

В зависимости от соотношений между смысловым содержанием информации S и тезаурусом пользователя S_p изменяется количество семантической информации 1_c, воспринимаемой пользователем, включаемой им в свой тезаурус. Характер такой зависимости показан на рис. 2.9.

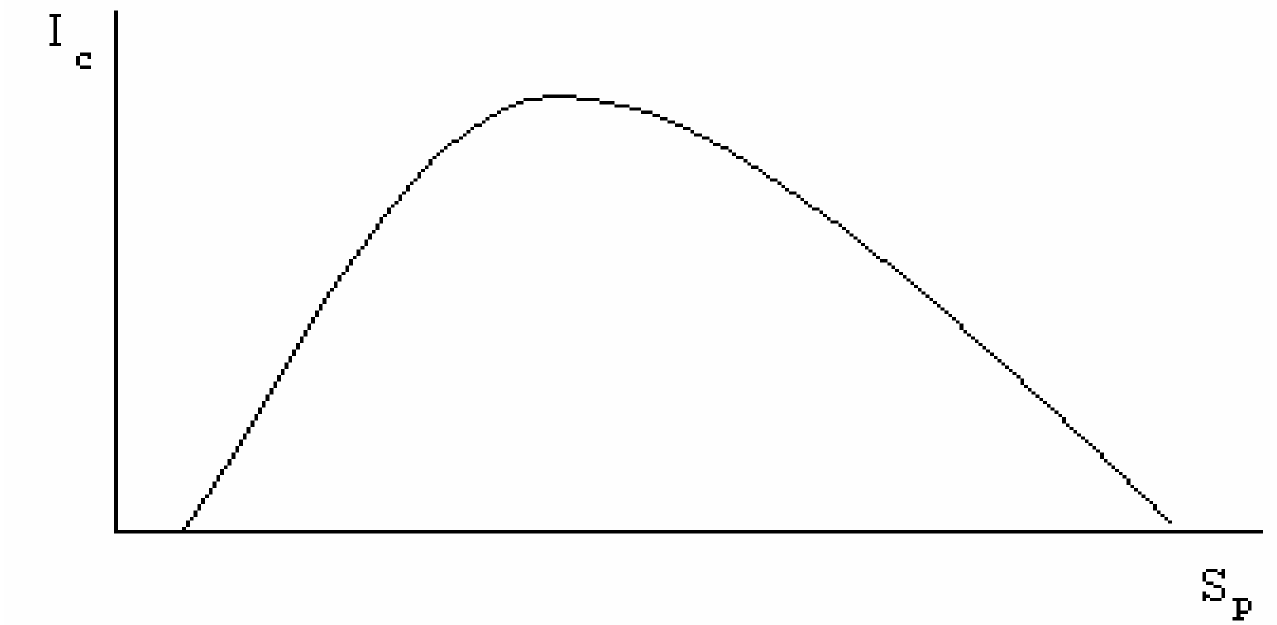


Рис. 2.9. Зависимость количества семантической информации, воспринимаемой потребителем, от его тезауруса $l_c = f(S_p)$

Рассмотрим два предельных случая, когда количество семантической информации I_c равно 0:

1)при S_p пользователь не воспринимает, не понимает поступающую информацию;

2)при S_p пользователь все знает, и поступающая информация ему не нужна.

Максимальное количество семантической информации I_c потребитель приобретает при согласовании ее смыслового содержания S со своим тезаурусом S_p ($S_p = S_{p\ opt}$), когда поступающая информация понятна пользователю и несет ему ранее не известные (отсутствующие в его тезаурусе) сведения. Следовательно, количество семантической информации в сообщении, количество новых знаний, получаемых пользователем, является величиной относительной. Одно и то же сообщение может иметь смысловое содержание для компетентного пользователя и быть бессмысленным (семантический шум) для пользователя некомпетентного. При оценке семантического (содержательного) количества информации необходимо стремиться к согласованию величин S с S_p.

Относительной мерой количества семантической информации может служить коэффициент содержательности, C, который определяется как отношение количества семантической информации к ее объему. Необходимость количественной оценки семантической (смысловой) информации, к которой относится и научно-техническая информация, обусловила появление теории семантической информации и меры ее количественной оценки, разработанной на основе *концепции разнообразия*, а не на шенноновской концепции снятия неопределенности. Однако в целом проблема оценки объемов семантической информации в сфере научно-технической информации остается пока нерешенной. Основными показателями оценки объема и качества семантической информации, выдаваемой на запрос специалиста или группы специалистов, является полнота, точность.

7) структурно-топологическая мера информации

Так, для простейшего графа, состоящего из двух вершин и соединяющего их ребра, нет возможности типологически отличить его вершины друг от друга. Считается, что количество информации такого графа равно нулю. Однако если в этом графе ориентировать ребро, то вершины его уже можно различать: одна из них будет начальной (ребро графа выходит из нее), вторая окажется конечной (ребро графа входит в нее). Информационное содержание такого ориентированного графа будет уже отлично от нуля; чем больше в графе отличающихся друг от друга вершин, тем большее количество информации он содержит.

8) алгоритмическая мера информации

В 1965 г. академик А.Н. Колмогоров ввел принципиально новое алгоритмическое определение понятие количества информации.

Алгоритм – некоторая система правил, предписаний, задающая вычислительный процесс, программу решения той или иной задачи.

Тогда с учетом этого определения по А.Н. Колмогорову количество информации определяется следующим образом.

Количество информации равно минимальной длине программы (сложность), позволяющей однозначно преобразовать один объект (множество) в другой объект (множество).

Так, если дано два одинаковых объекта (последовательность букв а,а,а,а) и необходимо выполнить программу преобразования объектов, то длина такой программы будет равна нулю в силу отсутствия различий между содержаниями объектов (объекты одинаковы). Чем больше различаются два объекта между собой, тем более сложной (длинной) оказывается программа перехода от одного объекта к другому.

Программа измеряет степень тождества (или степень различия) двух объектов, выражает эту степень количеством команд, инструкций, которые необходимо реализовать, выполнив в определенном порядке систему операций, переводящих один объект в другой.

Рассмотренный подход позволяет на базе алгоритмического определения понятия количества информации построить саму теорию вероятности. Так случайными событиями (последовательностями) в «информационном» построении теории вероятностей считаются такие, которые не содержат информации друг о друге. «Не случайными» последовательностями считаются те последовательности, в которых, наблюдается достаточно много закономерностей, т.е. общих свойств элементов последовательности. Например, упомянутая последовательность, все буквы которой одинаковы, не является случайной; можно сказать, что первая буква содержит всю информацию обо всех других элементах последовательности.

9) прагматическая мера информации

Прагматическая мера информации определяет полезность информации (ценность) для достижения пользователем поставленной цели, является величиной относительной, обусловленной особенностями использования этой информации в той или иной системе. Целесообразно в качестве единиц измерения ценности информации применять такие же, как и для измерения целевой функции.

В экономической системе прагматические свойства (ценность) информации можно определить приростом экономического эффекта функционирования, достигнутым благодаря использованию этой информации для управления системой: $Ц_I(У) = ΔЭ = Э(У{:}I) – Э(У)$, где $Ц_I(У)$ – ценность информационного сообщения i для системы управления $У$; $Э(У)$ – априорный ожидаемый экономический эффект функционирования системы управления $У$; $Э(У{:}I)$ – ожидаемый эффект функционирования системы $У$ при условии, что для управления будет использована информация, содержащаяся в сообщении I. В табл. 2.3 представлены для сопоставления введенные меры информации.

Таблица 2.3

Меры информации

Мера информации	Единицы измерения	Примеры (для компьютерной области)
Синтаксическая: шенноновский подход, компьютерный подход	Степень уменьшения неопределенности. Единицы представления информации	Вероятность события. Бит, байт, кбайт и т.д.
Семантическая	Тезаурус. Экономические показатели	Пакет прикладных программ, персональный компьютер, компьютерные сети и т.д. Рентабельность, Производительность, Коэффициент амортизация и т.д.
Прагматическая	Ценность использования	Емкость памяти, производительность компьютера, скорость передачи данных и т.д., денежное выражение время обработки информации и принятия решения решений

2.5. Количественная оценка информации

В качестве основной характеристики сообщения теория информации принимает величину, называемую *количеством информации*. Это понятие не затрагивает смысла и важности передаваемого сообщения, а связано со степенью его неопределенности.

Количество информации в дискретном сообщении. Энтропия.

Предположим, что источник сообщений может в каждый момент времени случайным образом принять одно из конечного множества возможных состояний. Такой источник называют дискретным источником сообщений. При этом принято говорить, что различные состояния реализуются вследствие выбора их источника.

Каждому состоянию источника U ставится в соответствие условное обозначение в виде знака. Совокупность знаков $u_1,u_2,:,u_i,:,u_N$, соответствующих всем N возможным состояниям источника, называют его *алфавитом*, а количество состояний N *объемом алфавита*. Формирование таким источником сообщений сводится к выбору им некоторого состояния u_i и выдачи соответствующего знака. Таким образом, под элементарным *дискретным сообщением*, выдаваемым источником, будем понимать символ u_i, при этом в течение некоторого времени T источник может выдать дискретное сообщение в виде последовательности элементарных дискретных сообщений, представляющей сбой набор символов u_i (например, u_5, u_1, u_3), каждый из которых имеет длительность t_i секунд. В общем случае необязательно одинаковую для различных i. Такая модель источника сообщений соответствует реальной ситуации, имеющей место в телеграфии и передаче данных (t_i = const). Отдельные состояния источника могут выбираться им чаще, другие реже. Поэтому в общем случае он хранится дискретным ансамблем U, т.е. полной совокупностью состояний с вероятностями их появления, составляющими в сумме 1.

$$U = \begin{pmatrix} u_1 & u_2 & ...u_i & ...u_N \\ P(u_1) & P(u_2) & ...P(u_i) & ...P(u_N) \end{pmatrix}, \sum_{i=1}^{N} P(u_i) = 1, \qquad (2.1)$$

где $P(u_i)$ – это вероятность выбора источником состояния u_i. При выдаче источником сообщений в виде последовательности элементарных дискретных сообщений, полным вероятностным описанием является вероятность совместного появления набора различных символов u_i в момент t_1, t_2,...,t_n, где n – длина последовательности

$$P\left(u_i^{t_1}, u_j^{t_2}, \ldots u_k^{t_1}, \ldots u_5^{t_n}\right).$$

Располагая такими сведениями об источнике, можно вычислить вероятность любого отрезка сообщения длиной меньше n. Если функция $P\left(u_i^{t_1}, \ldots u_5^{t_k}\right)$ не меняется во времени, если она равна $P\left(u_i^{t_{1+\tau}}, u_j^{t_{2+\tau}}, \ldots, u_5^{t_{n+\tau}}\right)$ при любых τ, то источник называется стационарным. Если при определении вероятностных характеристик стационарного источника усреднение по ансамблю можно заменить усреднением по времени, то такой источник называется эргодическим. Вероятностные свойства эргодического источника можно оценить, рассматривая лишь одну его достаточно длинную реализацию.

В каждом элементарном сообщении содержится для его получателя определенная информация – совокупность сведений о состоянии дискретного источника сообщения. Определяя количественную меру этой информации, совершенно не обязательно учитывать ее смыслового содержания, а так же ее значения для конкретного получателя. Очевидно, что при отсутствии сведений о состоянии источника имеется неопределенность относительно того, какое сообщение u_i из числа возможных им выбрано, а при наличии этих сведений данная неопределенность полностью исчезает. Естественно: количество информации, содержащейся в дискретном сообщении, измерять величиной исчезнувшей неопределенности.

Введем меру этой неопределенности, которую можно рассматривать и как меру количественной информации. Мера должна удовлетворять ряды естественных условий, одним из них является необходимость ее монотонного возрастания с увеличением возможности выбора, т.е. объема алфавита источника N. Кроме того, желательно, чтобы вводимая мера обладала *свойством аддитивности,* заключающемся в следующем: если 2 независимых источника с объемами алфавита N и M рассматривать, как один источник, одновременно реализующий пары состояний n_i и m_i то в соответствии с принципом аддитивности полагают, что неопределенность объединенного источника равна сумме неопределенностей исходных источников. Поскольку объемы алфавита объединенного источника $= N \cdot M$, то искомая функция при равной вероятности состояний источников должна удовлетворять условию $f(N \cdot M) = f(N) + f(M)$.

Можно математически строго показать, что единственной функцией, при перемножении аргументов которой значение функций складываются, является логарифмическая функция. Поэтому перечисленные требования выполняются, если в качестве меры неопределенности источника с равновероятными состояниями и характеризующего его ансамбля U принять логарифм объема алфавита источника

$$H(U) = \log N . \tag{2.2}$$

Легко видеть, что:

1) с ростом N величина $H(U)$ монотонно возрастает;

2) в случае если объем алфавита источника N равен 1, т.е. когда неопределенность отсутствует,

$$H(U) = \log 1 = 0;$$

3) величина $H(U)$ обладает свойством аддитивности, поскольку

$$\log(N \cdot U) = \log(N) + \log(M).$$

Впервые данная мера была предложена Хартли в 1928 г. Основание логарифма в (2.2) не имеет принципиального значения и определяет только масштаб или единицу количества информации.

При заданных глубине и длине числа количество чисел, которое можно представить, $N = q^n$. Величина N неудобна для оценки информационной емкости. Хартли ввел аддитивную двоичную логарифмическую меру, позволяющую вычислять количество информации в двоичных единицах – битах:

$$I = \log 2N = n \log 2q.$$

При $n = 1$, $q = 2$ $I = \log_2 \cdot 2 = 1$ *бит*. Это и есть единица информации по Хартли.

Следовательно, 1 бит информации соответствует одному элементарному событию, которое может произойти или не произойти. Такая мера количества информации удобна тем, что она обеспечивает возможность оперировать мерой, как числом. Количество информации при этом эквивалентно количеству двоичных символов 0 или 1. При наличии нескольких источников информации общее количество информации:

$$I(q_1, q_2, ..., q_k) = I(q_1) + I(q_2) + ... + I(q_k),$$

где $I(q_k)$ – количество информации от источника k.

Логарифмическая мера информации позволяет измерять количество информации и используется на практике.

Чаще всего в качестве основания используют число 2, при этом единица количества информации называется двоичной единицей или битом, и представляет собой информацию, содержащуюся в одном дискретном сообщении источника равновероятных сообщений с объемом алфавита равным двум. При выборе в (2.2) основания логарифма равным 10 получаем десятичную единицу, называемую битом. Иногда используют натуральную единицу количества информации, называемую натом, при этом основание логарифма в (2.2) равно e=2,7.

Рассматриваемая мера количества информации может иметь лишь ограниченное применение, поскольку предполагает равную вероятность выбора источником любого из возможных его состояний. В более общем случае, когда вероятности различных состояний источника не одинаковы степень неопределенности конкретного состояния зависит не только от объема алфавита источника, но и от вероятности этого состояния. В такой ситуации количество информации, содержащееся в одном дискретном сообщении u_k целесообразно определить как функцию вероятности появления этого сообщения $P(u_k)$ и характеризовать величиной

$$i(u_k) = -\log P(u_k) = \log \frac{1}{P(u_k)}. \qquad (2.3)$$

Основание логарифма в (2.3) выбирается из тех же соображений что и в (2.2). Знак минус в первом равенстве (2.3) необходим для того, чтобы количество информации $i(u_k)$ было неотрицательным числом т.к. всегда $P(u_k) \le 1$. Очевидно что, так же как и мера $H(U)$, определяемая (2.2), величина $i(u_k)$ обладает *свойством аддитивности*.

И в случае достоверного сообщения, когда $P(u_k)=1$, $i(u_k)=0$. Однако теперь количество информации, содержащееся в дискретном сообщении зависит от степени неожиданности этого сообщения, характеризуемой вероятностью его появления. Количество информации в сообщении тем больше, чем оно более неожиданно. Если источник выдает последовательность зависимых между собой *элементарных сообщений*, то наличие предшествующих сообщений может изменить вероятность последующего, а, следовательно, и количество информации в нем. Оно должно определяться по условной вероятности $P(u_k/u_{k-1},u_{k-2},...)$ выдачи сообщений u_k при известных предшествующих сообщений $u_{k-1},u_{k-2},...$, тогда количество информации

$$i(u_k/u_{k-1},u_{k-2},...)=-\log P(u_k/u_{k-1},u_{k-2},...). \quad (2.4)$$

Определения (2.3) и (2.4) количества информации являются случайной величиной, поскольку сами сообщения являются случайными. Его распределение вероятностей определяется распределением вероятностей сообщений в данном ансамбле. Для цифровой характеристики всего ансамбля или источника сообщения используется математическое ожидание количества информации в отдельных сообщениях, *называемых энтропией*.

$$H(U)=M\left\{\log\frac{1}{P(u_i)}\right\}=\sum_{i=1}^{N}P(u_i)\cdot\log\left(\frac{1}{P(u_i)}\right). \quad (2.5)$$

Чем больше энтропия источника, тем больше степень неожиданности выдаваемых им сообщений в среднем, т.е. тем более неопределенным является ожидание сообщений. Впервые мера (2.5) была предложена Клодом Шенноном в его фундаментальной работе «Математические основы теории связи», опубликованной в 1948 г., в которой были заложены основы современной ТИ. Предполагающая мера была названа энтропией не случайно. Дело в том, что вид формулы (2.5) совпадает с полученным ранее Больцманом результатом выражения для энтропии термодинамической системы. Рассмотрим взаимосвязь меры Шеннона с мерой Хартли. Если в источнике может быть реализовано h равновероятных состояний, то вероятность каждого из них $P_i=\dfrac{1}{N}$, с учетом этого меру неопределенности источника

Хартли $H(U) = \log(N) = -\log\frac{1}{N} = -\log P_i$ можно трактовать, как *количество информации*, приходящей на одно дискретное сообщение (поскольку все сообщения источника равновероятные количества информации в каждом из них равны). В то же время, энтропия по Шеннону – это среднее количество информации, содержащееся в одном из не равновероятных состояний. Она позволяет учесть статистические свойства источника информации. Наряду с рассмотренными мерами Хартли и Шеннона существуют и другие подходы к определению количества информации. Наиболее интересной, наиболее новой явилась информационная концепция Колмогорова. Ее основным тезисом является то, что на основании определения энтропии (2.5) количество информации связывается с вероятностью наступления P_i.. Это выглядит достаточно разумно, т.к. понятие вероятности имеет смысл лишь в связи с массовыми явлениями количества единиц информации в единичном акте и представляющих интерес в связи с данным исходом, оказывается выраженным через вероятности массовых явлений.

Эта внутренняя противоречивость содержания информации, опирающейся на концепцию выбора, продолжится на базе общей теории алгоритма. Согласно алгоритмическому подходу, энтропия $H(u,z)$ – есть мнимая длина, записанная в виде последовательности 0 и 1 программы, которая позволяет построить объект u, имея в своем распоряжении объект z. Тогда основные понятия теории информации могут быть обоснованы без обращения к теории вероятности, причем так, что понятие энтропия и количество информации оказываются, строго приемлемыми к индивидуальным объектам. Однако Шенноновская мера интересна не сама по себе, а как основание встроенной теории, позволяющей изменить и расширить существующие предположения о возможностях в технике связи.

2.6. Статистическая мера информации

Статистическая мера информации позволяет осуществлять синтаксическую оценку информации, которая оперирует с обезличенной информацией, не выражающей смыслового отношения к объекту.

О*бъем данных* в сообщении измеряется количеством символов (разрядов) в этом сообщении. В различных системах счисления один разряд имеет различный вес и, соответственно, меняется единица измерения данных:

– в двоичной системе счисления единица измерения – бит *(bit – binary digit* – двоичный разряд);

– в десятичной системе счисления единица измерения – дит (десятичный разряд);

– в современных ЭВМ наряду с минимальной единицей измерения данных «бит» широко используется укрупненная единица измерения «байт», равная 8 бит.

Так сообщение в двоичной системе в виде восьмиразрядного двоичного кода 1011 1011 имеет объем данных $V_д$ = 8 бит. Сообщение в десятичной системе в виде шестиразрядного числа 275903 имеет объем данных $V_д$ = 6 бит.

Статистические меры оценки количества информации на синтаксическом уровне основаны на вероятностно-энтропийном подходе, состоящем в том, что информация рассматривается как сообщение об исходе случайных событий, реализации случайных величин и функций, а количество информации определяется в зависимости от априорных вероятностей событий, величин, функций, явлений.

События x_i, представляющие дискретные состояния измененной случайной величины, рассматриваются как возможные (вероятностные) исходы некоторого опыта, составляющие полную группу событий: $p_1+p_2+\ldots+p_i+\ldots+p_k=1$.

В простейшем случае эти события несовместимы, составляют полную группу, в которой обязательно реализуется одно из событий и имеет место условие:

$$\sum_{i=1}^{k} p(x_i) = p(x_1) + p(x_2) + \ldots + p(k_k) = 1.$$

Рассматриваемые вероятности изменяются во времени, в силу чего статистические характеристики (среднее значение и дисперсия) становятся переменными величинами. Процессы, описываемые этими величинами, называются нестационарными.

Энтропийный подход Шеннона к оценке количества информации основан на том, что различные символы в осмысленном сообщении появляются с различной вероятностью. Например, в русских текстах чаще всего встречается буква «О». Наиболее часто в английском языке встречается буква «Е», а наиболее редко буква «Z».

Если под испытанием понимать чтение очередной буквы, независимо от ее фонетических характеристик (ударная, безударная, короткая гласная, произносимые после мягких, но после твердых согласных и т.п.), то вероятность появления буквы «О» равна p_o= 0,09. Чаще появляется только пробел между словами, $p_{пр}$= 0,17. А вероятность появления буквы «Ф» минимальна и равна $p_ф$= 0,002. Для бессмысленного сообщения, получаемого, например, путем случайного нажатия клавиш печатающего устройства, вероятность появления букв будет примерно одинаковой. Таким образом, в синтаксическом смысле содержание (осмысленность) сообщения может быть определено с помощью соответствующей функции, зависящей от вероятности появления символов в данном сообщении.

Будем считать, что в осмысленном сообщении (в синтаксическом смысле) символы используемого алфавита упорядочены, а в бессмысленном – полностью хаотичны. При этом вполне очевидно, что пример осмысленного, в синтаксическом смысле сообщения, формируемого из ограниченного алфавита, представляющего десятичную, двоичную или другую подобную систему счисления, не достаточно нагляден. Таким образом, задача состоит в том, чтобы отличить порядок от хаоса. Функция, позволяющая решить эту задачу, была найдена в XIX веке при исследовании задач термодинамики. Она получила название энтропии и связана с оценкой неопределенности каждой ситуации. В различных областях знаний энтропия означает:

– в термодинамике – степень устойчивости состояний вещества, определяемой логарифмом числа микросостояний, реализующих данное макросостояние физической системы;

– в математике – степень неопределенности ситуации или задачи;

– информатике – способность источника отдавать информацию.

Так, в соответствии со вторым законом термодинамики (Больцмана), энтропия системы из N молекул выражается формулой $H = -\frac{1}{N}\sum_{i=1}^{N} n_i \ln \frac{n_i}{N}$, где N – общее число молекул в данном объеме; n_i – количество молекул со скоростями в интервале $v_i + \Delta v$. Здесь n_i / N – есть частота события, следовательно, вероятность того, что молекулы имеют скорость $v_i + \Delta v$, будет равна $p_i = n_i / N$.

Впоследствии было установлено, что энтропия, имеет универсальный характер и записывается в виде $H = -\sum_{i=1}^{N} p_i \ln p_i$. Применительно к задачам теории передачи сигналов (называемой в ряде литературных источников теорией информации) p_i – вероятность появления при очередном испытании i-го символа из алфавита языка, а N – объем алфавита. Полагая, $\partial H / \partial p_i = 0$, находим $\log_a p_i = -1$, то есть, функция H достигает экстремума при $p_i = 1/a$. Равенство всех p_i, возможно, как уже говорилось только для белого шума – бессмысленного сообщения. Кроме того, так как $\sum p_i = 1$, то, a=N. Итак, окончательная формула для энтропии имеет следующий вид $H = -\sum_{i=1}^{N} p_i \log_N p_i$. Так как $p_i < 1$, $N > 1$, то энтропия всегда положительна. Можно доказать, что выражение $p \log_N p$ стремится к нулю при $p \to 1$ и при $p \to 0$. Поэтому для сообщения из одного символа, когда p_1=1, p_2=…= p_N=0, мы получаем H=0. Эти рассуждения показывают, что H=$f(p)$, и функция эта выпуклая. С другой стороны, из этих рассуждений, очевидно, что любое отклонение от равновероятного состояния, независимо от числа событий, уменьшает энтропию события появления символа. Для белого шума H=1. Таким образом, белому шуму соответствует максимум энтропии, и чем ближе значение H к единице, тем больше оснований считать сообщение синтаксически бессмысленным. Тогда можно утверждать, что для априорно равновероятных событий количество информации, которую мы получаем в результате совершения этих событий, уменьшается с увеличением вероятности любого из событий.

Наши знания о некотором объекте или системе α есть информация, имеющая определенную энтропию $H(\alpha)$. При получении некоторого сообщения β о системе мы получаем возможность делать некоторые умозаключения или предсказания о характере информации, находящейся в нашем распоряжении, которые вносят элемент упорядоченности в массив информации. Следовательно, энтропия имеющейся информации после получения составит $H_\beta(\alpha)$. Количество информации $I_\beta(\alpha)$, содержащейся в сообщении β, определяется как приращение энтропии с обратным знаком: $I_\beta(\alpha) = H(\alpha) - H_\beta(\alpha)$.

Для случая, когда все состояния системы равновероятны, т.е. их вероятности равны $p_i = 1|N$, энтропия определяется соотношением

$$H(\alpha) = -\sum_{1}^{N} \frac{1}{N} \log \frac{1}{N} = \log N .$$

Логарифмическая мера количества информации, предложенная Р. Хартли, была получена на основе следующих положений. Число сообщений N, которое можно получить, комбинируя m символов алфавита по n элементов в сообщении равно $N=m^n$. С учетом этого выражения, число сообщений N, и соответствующее количество информации находятся в экспоненциальной зависимости от количества элементов в сообщении. В силу этого N нельзя непосредственно использовать, как меру количества информации и Р. Хартли предложил в качестве меры количества информации применять логарифм – числа возможных последовательностей символов: $I = \log N = \log m^n = n \cdot \log m$. Основание логарифма зависит от выбранной единицы количества информации. Такое представление меры количества информации соответствует требованию аддитивности. Оно отражает экспоненциальный характер зависимости количества возможных кодовых комбинаций от количества символов в исходном алфавите, а также согласуется с основным психофизиологическим законом Вебера-Фехтнера $S=K \cdot \log E$ и почти совпадает с рассмотренной выше классической формулой Больцмана для энтропии в статистической термодинамике $H_T = k \cdot \log w$. В этой формуле S – восприятие, K – некоторая константа, зависящая от характера проводимого опыта, E – возбуждение, H_T – термодинамическая энтропия; k – константа; w – вероятность данного состояния системы. Такое совпадение объясняется общностью природы рассматриваемых явлений.

При передаче информации по каналу связи рассматриваемое частное сообщение выбирают из определенного количества возможных сообщений. Так, если по каналу связи передается n-разрядное сообщение, использующее m различных символов, то количество всевозможных кодовых комбинаций будет $N = m^n$, и при равновероятности появления любой из них количество информации в полученном сообщении, будет $I = \log N = n \log m$ (формула Хартли).

Если в качестве основания логарифма принять m, то $I = n$, и количество информации (при условии полного априорного незнания абонентом содержания сообщения) будет равно объему данных $I= V_д$, полученных по каналу связи. Для неравновероятных состояний системы всегда количество информации меньше его объема и равно n.

Коэффициент (степень) информативности Y (лаконичность) сообщения определяется отношением количества информации к объему данных.

С увеличением информативности Y уменьшаются объемы работы по преобразованию информации (данных) в системе. Для повышения информативности разрабатываются специальные методы оптимального кодирования информации.

2.7. Динамическая энтропия

Этим понятием удобно пользоваться в случае, когда целью извлечения информации является обращение в ноль неопределенности ситуации. В ходе распознавания образов, диагноза болезней, расследования преступлений энтропия ситуации изменяется во времени. Изменение обусловливается поступлением в определенное время дополнительной информации, уменьшающей (положительная) или увеличивающей (отрицательная (дезинформация)) неопределенность ситуации.

Такие ситуации можно представить множеством отношений между наблюдаемым следствием $\{a_i\}$ и раскрываемой по этому следствию причиной $\{b_j\}$. Возможная связь $a_i \to b_j$ имеет вероятность p_{ij}, причем: $p_{ij}(t)=0$ – если нет связи между следствием a_i, и причиной b_j, вызвавшей ее; $p_{ij}(t)=1$ – имеется достоверная (однозначная и полная) связь. Тогда для множества следствий и причин энтропия будет определяться выражением: $H(t) = -\sum_{i=1}^{N(t)}\sum_{j=1}^{M(t)} p_{ij}(t) * \log p_{ij}(t)$, где $N(t)$ *и* $M(t)$ – общее количество следствий и причин в момент времени t. При поступлении дополнительной информации за единичный интервал времени происходит изменение количества следствий, причин и вероятность отношений между ними. В результате энтропия ситуации в момент времени $t+1$ получает новое выражение:

$$H(t+1) = -\sum_{i=1}^{N(t+1)}\sum_{j=1}^{M(t+1)} p_{ij}(t+1) * \log p_{ij}(t+1).$$

Мерой информации, влияющей на энтропию, будет разность $I = H(t) - H(t+1)$, которая может быть положительной или отрицательной.

2.8. Семантическая количественная мера информации

Вычислительные машины обрабатывают и преобразуют информацию разного содержания – от числовых данных до сочинения музыки и стихов. Вся эта информация изображается соответствующими символами. Оценка содержания разнохарактерной информации – весьма сложная проблема.

Среди семантических мер наиболее распространены содержательность, логическое количество, целесообразность и существенность информации.

Содержательность события i выражается через функцию меры $m(i)$ – содержательности его отрицания. Оценка содержательности основана на математической логике, в которой логические функции истинности $m(i)$ и ложности $m(o)$ имеют формальное сходство с функциями вероятностей события $p(i)$ и антисобытия $q(i)$ в теории вероятностей.

Как и вероятность, содержательность события изменяется в пределах $0 \le m \le 1$.

Логическое количество информации I_{nf}, сходное со статистическим количеством информации, вычисляется по выражению:

$$I_{nf} = \log 2[1/m(i)] = -\log 2m(o),$$

Отличие статистической оценки от логической состоит в том, что в первом случае учитываются вероятности реализации тех или иных событий, что приближает к оценке смысла информации.

Если информация используется в системах управления, то ее полезность целесообразно оценивать по тому эффекту, который она оказывает на результат управления.

2.9. Количественная мера целесообразности информации

Мера целесообразности информации определяется как изменение вероятности достижения цели при получении дополнительной информации. Полученная информация может быть пустой, т. е. не изменять вероятности достижения цели, и в этом случае ее мера равна нулю. В других случаях полученная информация может изменять положение дела в худшую сторону, т. е. уменьшить вероятность достижения цели, и тогда она будет дезинформацией, измеряющейся отрицательным значением количества информации. Наконец, в благоприятном случае получается добротная информация, которая увеличивает вероятность достижения цели и измеряется положительной величиной количества информации.

Мера целесообразности в общем виде может быть аналитически выражена в виде соотношения:

$$I_{цел} = \log_2 p_1 - \log_2 p_0 = \log_2 \frac{p_1}{p_0},$$

где p_0 и p_1 – начальная (до получения информации) и конечная (после получения информации) вероятности достижения цели.

Следует различать: существенность самого события; существенность времени совершения события или его наблюдения (рано–поздно–момент); существенность координаты совершения события.

Измерение некоторого параметра X можно характеризовать несколькими функциями величины x: вероятностью $p(x)$, погрешностью измерения $e(x)$ и существенностью $c(x)$. Каждой из этих функций можно поставить в соответствие определенную меру информации.

Мерой Хартли оценивается функция погрешности e при фиксированных значениях функции вероятности ($p=const$) и существенности ($c=const$).

Мерой Шеннона оценивается функция вероятности ($p=var$) при фиксированных значениях функций погрешности ($e=const$) и существенности ($c=const$).

Мера существенности относится к ситуации с фиксированными функциями погрешности ($e=const$) и вероятности ($p=const$).

Контрольные вопросы

1. В каких формах можно представить общие свойства информации?
2. Дать определения синтактики, семантики, знака, денотата, концепта, прагматики.
3. Привести основные потребительские показатели свойств информации, определяющих ее важность и эффективность использования.
4. Дать определение понятий следующих качеств информации: выдачи (своевременность, актуальность, полнота, релевантность, толерантность); обработки (глубина, достоверность, адекватность);

защищенности (целостность физическая, целостность логическая, доверие, безопасность).

5. Представить структуру системы показателей свойств информации.

6. Дать определение понятиям следующих свойств информации: атрибутивности, прагматичности, динамичности.

7. Какими показателями можно характеризовать синтаксическую адекватность, дать их определение?

8. Какими мерами могут быть измерены атрибутивные свойства информации?

9. Привести характеристики динамических свойств информации и дать их определения.

10. Привести характеристики прагматических свойств информации и дать их определения.

11. Дать определение важности информации и привести показатели ее оценки.

12. Что характеризуют показатели смысла и новизны, содержательность (релевантность), релевантность, репрезентативность (адекватность), достаточность (полнота), пертинентность, ценность информации?

13. Дать определение информационного кадастра объекта.

14. Привести выражения для оценки меры полноты информации (через коэффициент полноты).

15. Как определяется коэффициент важности элемента информации?

16. Как решается задача определения количества и меры информации в структурной теории, статистической теории, семантической теории информации?

17. Что представляет собой структурная мера информации?

18. Представить математическое выражение меры количества информации, как сочетание видов элементов.

19. Представить математическое выражение меры количества информации, как перестановку видов элементов.

20. Представить математическое выражение меры количества информации, как размещение видов элементов.

21. Что представляет аддитивная мера оценки количества информации по Хартли?

22. Что представляет собой статистическая мера информации?

23. Что называется энтропией?

24. Дать определения семантики, знака, слова, языка, сигнала.

25. Перечислить и дать краткую характеристику семантической мере информации.

26. Раскройте понятие динамической энтропии и содержания меры ее измерения.

27. Сущность динамической энтропии.

28. Раскрыть содержание подходов определения содержательности события.

29. Как определить логическое количество информации?

30. В чем отличие статистической оценки от логической?

31. Раскрыть сущность количественных мер целесообразности информации.

ГЛАВА 3. КЛАССИФИКАЦИЯ ИНФОРМАЦИИ

3.1. Основные понятия системной классификации информации

Рассмотренные ранее подходы к определению понятий и свойств информации дают возможность ее классификации, обеспечивающей ее поиск.

Классификация – система распределения понятий (предметов или отношений), объектов по классам на основании общих признаков, свойств, присущих одним понятиям (предметам или отношениям), объектам и отличающихся от других понятий (предметов и отношений), объектов.

Объект – любой предмет, процесс, явление материального или нематериального мира.

В классификации каждый класс имеет постоянное, определенное место относительно других классов. В основе классификации лежит деление понятий (предметов, отношений, объектов) на группы, которые носят относительно устойчивый характер.

Признак, по которому производится распределение понятий (предметов или отношений), объектов называется *классификационным признаком (основанием деления)*. Каждый признак имеет определенные значения («*значение классификационного признака»)*, которые позволяют установить сходство или различие объектов (по соответствующим свойствам). Аналогом такой интерпретации реквизитов классификации в моделировании служат понятия «показатель свойства объекта» и его «шкала».

Примеры признаков классификации: 1) возраст человека, со значениями: до 20 лет, от 20 до 30 лет, свыше 30 лет; 2) возраст до 20 лет, возраст от 20 до 30 лет, возраст свыше 30 лет

Род – непосредственно подчиняющее понятие классификации.

Вид – непосредственно подчиненное понятие по отношению к роду (родовому понятию) классификации.

Это родовидовые (иерархические) отношения.

Сильная иерархия – видовое понятие имеет только одно предшествующее (родовое) понятие.

Слабая иерархия – видовое понятие имеет два и более предшествующих (родовых) понятия.

Отношения соподчиненности – отношения, которыми связаны подклассы одного и того же класса.

Назначение классификации: 1) выявление общих свойств информационного объекта (по определению); 2) разработка правил (алгоритмов) и процедур обработки информации; 3) осуществление информационного поиска; 4) группирование объектов, выделение определенных классов, характеризующихся рядом общих свойств и др.

С учетом рассмотренного назначения можно дать следующее определение.

Классификация объектов – процедура группировки на качественном уровне, направленная на выделение однородных свойств для образования выделенных классов объектов.

Информационные объекты – выделенные классы информации, представляющие объект исследования.

Пример информационных объектов, характеризующихся общими свойствами (реквизитами): информация о студентах (информационный объект «Студент»), преподавателях (информационный объект «Преподаватель»), факультетах (информационный объект «Факультет») и т.п.

Реквизит – логически неделимый информационный элемент, описывающий определенное свойство объекта, процесса, явления и т.п., определяемое информационными параметрами.

Реквизиты представляются:

– числовыми данными (вес, стоимость, размер, год и др.);

– лингвистически переменными (цвет, марка машины, фамилия, имя, отчество, адрес проживания и др.).

В ходе классификации обеспечивается разработка правил (алгоритмов) и процедур обработки информации, в том числе информационного поиска информации различных форм, имеющих реквизиты. Алгоритм обработки информационных объектов позволяет получить информацию об объемах. Такие алгоритмы различаются целями, видами обрабатываемой информации, способами реализации. Так, алгоритм обработки информационных объектов библиотечного фонда позволяет получить информацию обо всех книгах по определенной тематике, об авторах, абонентах и т.д.

Информационный поиск – процесс коммуникаций между человеком и массивом информации, представляющий совокупность логических и технических операций, реализуемых с целью нахождения документированной информации, фактов, данных, релевантных запросу потребителя.

Информационный поиск реализуется с помощью информационно-поисковых систем и связан с использованием семантики (смысла) информации, документа, данных.

Информационно-поисковая система (ИПС) – совокупность взаимосвязанных методов и средств, предназначенных для хранения и поиска документированной информации, отдельных фактов и данных. ИПС включает:

1) логико-семантический аппарат, состоящий из информационно-поискового языка (одного или нескольких), правил его использования в процессе обработки документированной информации;

2) информационно-поисковый массив объектов;

3) средства реализации хранения и поиска объектов информационно-поискового массива;

4) пользователи и обслуживающий персонал, взаимодействующие с системой.

Основным элементом логико-семантического аппарата ИПС является информационно-поисковый язык (ИПЯ).

Информационно-поисковый язык (ИПЯ) – специально созданный искусственный язык, предназначенный для выражения содержания документированной информации, запросов или описания фактов с целью их последующего поиска.

От качества ИПЯ зависит эффективность информационного поиска. Необходимость создания ИПЯ для выражения смыслового содержания документированной информации, запросов и фактов с целью их поиска обусловлена сложностью естественного языка:

1) неоднозначностью слов (синонимия);

2) многозначностью слов (омонимия);

3) неформализованностью связей слов;

4) сложностью определения значений слов в содержании свернутых информационных документов по контекстуальным отношениям этих слов.

При выборе требований (принципов) и методов классификации информации необходимо учитывать особенности, виды и требования к формированию ИПЯ.

Принципы (*требования*) *системной классификации* информации:

– полнота охвата объектов рассматриваемой области;

– однозначность реквизитов, описывающих только один смысл, одно толкование и однозначную запись информации, понятия;

– достаточная семантическая сила – способность отражать с необходимой полнотой и точностью смысловое содержание документов и запросов предметной области;

– открытость – обеспечение возможности включения новых объектов, корректировки информации и ИПЯ.

3.2. Классификация информации по различным признакам

1) виды информации

Рассмотрим вариант классификации информации по следующим признакам: сфере возникновения; способу передачи и восприятия; общественному назначению (табл. 3.1).

Элементарная информация – информация, возникающая в неживой природе.

Биологическая информация – информация, возникающая в мире животных и растений.

Социальная информация – информация, передающаяся в человеческом обществе в процессе коммуникации между людьми, представленная в форме знаков, понятных членам общества, и способная изменять уровень их знаний о внешнем мире, т.е. изменять состояние их тезаурусов.

Таблица 3.1

Классификация видов информации

<table>
<tr><td rowspan="2"></td><td colspan="17">Классификационные признаки</td></tr>
<tr><td colspan="4">Сфера возникновения</td><td colspan="5">Способ передачи и воспроизведения</td><td colspan="8">Общественное назначение</td></tr>
<tr><td rowspan="3">Виды информации</td><td>Элементарная</td><td>Биологическая</td><td colspan="2">Социальная</td><td rowspan="3">Визуальная</td><td rowspan="3">Аудиальная</td><td rowspan="3">Тактильная</td><td rowspan="3">Вкусовая</td><td rowspan="3">Машинная</td><td>Массовая</td><td colspan="6">Специальная</td><td rowspan="3">Личная</td></tr>
<tr><td colspan="2" rowspan="2">Эстетическая</td><td colspan="2">Семантическая</td><td rowspan="2">Военно-политическая</td><td rowspan="2">Обыденная</td><td rowspan="2">Научно-популярная</td><td rowspan="2">Научная</td><td rowspan="2">Техническая</td><td rowspan="2">Производственная</td><td rowspan="2">Управленческая</td></tr>
<tr><td>Первичная</td><td>Вторичная</td></tr>
</table>

Виды социальной информации: массовая информация – предназначена для всех членов общества, независимо от их положения и рода занятий; публицистическая; обыденная; эстетическая; специальная информация – предназначена целевым социальным группам и подразделяется на научную, техническую, технологическую, планово-экономическую.

Эстетическая информация – информация, образованная сочетаниями информации живой и неживой природы в виде света, цвета, тени, звуков, запахов.

Семантическая информация – результат познания человеческим обществом законов природы, общества и мышления.

Первичная информация – семантическая информация, возникающая непосредственно в ходе или в результате научной, производственной или общественно-политической деятельности людей.

Вторичная информация – семантическая информация, возникающая как результат аналитико-синтетической переработки первичной информации.

Машинная информация – информация, воспринимаемая ЭВТ.

Научная информация – информация, возникающая в сфере науки в результате изучения закономерностей окружающей природы, общественного развития и мышления.

В свою очередь, научная информация классифицируется по областям науки, техническая – по отраслям хозяйствования (машиностроение, приборостроение, транспорт, строительство, сельское хозяйство и др.).

Производственная информация – информация о характере производственных процессов (конструкторская, технологическая, планово-экономическая, эксплуатационная и др.).

Рассмотренный вариант классификации информации отражает разнообразие ее содержания и форм представления. Формальное представление информации, по сути есть отражение тех или иных знаков в сознании человека, познающего реальную действительность. Изучением рассматриваемых аспектов занимается семиотика.

Разнообразие – упорядоченная последовательность, обладающая определенной структурой.

Семиотика – наука, изучающая природу, виды и функции знаков.

Знаковая система – форма передачи информации.

Знак – какой-либо материальный предмет или явление, которым обозначают (представляют) другие предметы и явления в процессах коммуникации и познания.

Рассмотренная классификация видов информации является вариантом, не претендующим на полноту, однозначность и законченность. Она позволяет продемонстрировать разнообразие видов информации и обосновать необходимость ее системного представления (классификации) для удобства пользования и организации информационного поиска.

2) виды системной классификации документированной информации

С учетом существующих видов ИПЯ, рассматриваемых в ряде известных работ, в том числе Макаровой Н.В., предлагается выделить следующие виды систем классификации объектов: классификационный, включающий иерархическую, фасетную, алфавитно-предметную классификации; дескрипторный, включающий дескрипторный словарь, информационно-поисковый тезаурус, указатели связи; предкоординированный, включающий иерархическую, алфавитно-предметную, фасетную классификации.

Каждый из этих видов отличается способами применения классификационных признаков.

3) классификационная система

Система классификационного типа представляет системную классификацию понятий, отражающую смысловые отношения между понятиями. В систематических классификациях лексические единицы задаются системой понятий, выражаемых словами, словосочетаниями и предложениями. Классификационная система может быть представлена такими видами классификации как иерархической, фасетной, алфавитно-предметной.

Иерархическая (перечисленная) система классификации используется для систематизации объектов информации и их поиска и строится по следующим правилам:

– исходное множество элементов составляет 0-й уровень и делится по выбранному классификационному признаку на классы (группировки), образующие 1-й уровень;

– каждый класс 1-го уровня по своим классификационным признакам делится на подклассы, образующие 2-й уровень;

– каждый класс 2-го уровня аналогично делится на группы, образующие 3-й уровень, и т.д.

Основным обобщенным (системным) требованием к процедуре построения структуры классификации является первоочередное определение цели классификации, предполагающей определение (выделение, выявление, формирование) классификационных признаков (свойств объединения объектов в классы).

Кроме системного требования к процедуре формирования иерархической системы классификации к ней предъявляют следующие требования:

– большое внимание уделять выбору классификационных признаков;

– каждый объект на любом уровне должен быть отнесен к одному классу, который характеризуется конкретным значением выбранного классификационного признака;

– для последующей группировки в каждом новом классе необходимо задать свои классификационные признаки и их значения.

Выбор классификационных признаков зависит от семантического содержания того класса объектов, для которого необходима группировка на последующем уровне иерархии.

Количество уровней классификации, соответствующее числу признаков, выбранных в качестве основания деления, характеризует *глубину классификации*.

Достоинствами иерархической системы классификации являются простота построения, использование независимых классификационных признаков в различных ветвях иерархической структуры.

К *недостаткам* иерархической системы классификации можно отнести жесткую структуру, усложняющую внесение изменений за счет необходимости перераспределения всех классификационных группировок, невозможность группировки объектов по заранее не предусмотренным сочетаниям признаков, а также сложность автоматизированной реализации поиска информации.

4) фасетная система классификации

Фасетная система классификации – системная классификация, создаваемая на основе признаков классификации *фасетов* (facet – рамка), не зависящих как друг от друга, так и от семантического содержания классифицируемого объекта.

Фасеты – признаки классификации, каждый из которых содержит совокупность однородных значений данного классификационного признака, располагающихся в произвольном порядке, хотя предпочтительнее их упорядочение.

Примером фасетной классификации служит *цвет,* содержащий значения: красный, белый, зеленый, черный, желтый. Фасет *специ-*

альность содержит названия специальностей. Фасет *образование* содержит значения: среднее, среднее специальное, высшее.

Принципы фасетной классификации:

– разделение одного и того же множества предметов и явлений по разным основаниям (выделение нескольких иерархических «деревьев» для одного и того же множества категорий, фасетов);

– образование сложных индексов путем соединения простых, фиксируемых в иерархических «деревьях».

В отличие от иерархической, в фасетной классификации кроме основного деления на ряд классов, осуществляется второе разбиение всего множества предметов и явлений по категориям или фасетам.

Фасетный класс – множество предметов (понятий), сгруппированных на основе одного фасетного признака, присущего всем предметам (понятиям) данного множества.

В каждом фасетном классе понятие определяется двумя координатами – признаком отраслевого класса и признаком категории. Фасетная система классификации, может быть представлена табл. 3.2.

Названия столбцов соответствуют выделенным классификационным признакам (фасетам), обозначенным $Ф_1$, $Ф_2$, ..., $Ф_i$, ..., $Ф_n$. Например, цвет, размер одежды, вес и т.д. Произведена нумерация строк таблицы. В каждой клетке таблицы хранится конкретное значение фасета. Например, фасет *цвет,* обозначенный $Ф_3$ содержит значения: красный, белый, зеленый, черный, желтый.

Основные этапы построения фасетной системы классификации:

1) каждому объекту присваивается (задается) конкретная группировка фасетов структурной формулой, в которой отражается их порядок следования: $Ks = (Ф_1, Ф_2, Ф_\text{i}, ..., Ф_n)$, где $Ф_i$ – i-й фасет, а n – количество фасетов. При этом могут быть использованы не все фасеты;

2) строится фасетная система классификации без повторения значений, используемых в различных фасетах.

Модификация системы осуществляется путем внесения изменений в конкретные значения любого фасета.

Достоинства фасетной системы классификации:

– возможность создания большой емкости классификации, т.е. использования большого числа признаков классификации и их значений для создания группировок;

– позволяет производить многоаспектное и глубокое индексирование содержания документированной информации (документов);

– индексы классификации короче, чем у иерархических классификаций;

– возможность простой модификации, исправлений и дополнений всей системы классификации без изменения структуры существующих группировок.

Таблица 3.2

Фасетная система классификации

Значения фасетов	Фасеты				
	$Ф_1$ (размер)	$Ф_2$ (рост)	Ф3 (цвет)	$Ф_i$	$Ф_n$
1	48	152	Синий	…	…
2	11	164	Хаки	…	…
…	…	…	…	…	…
К	56	190	Черный	…	…

Индексирование документированной информации (документа) – включает определение его темы или предмета и выражение главной темы или предмета на ИПЯ системы, т.е. формирование поискового образа документа.

Недостатком фасетной системы классификации являются: сложность ее построения, обусловленная необходимостью учета всего многообразия классификационных признаков, ограниченность по использованию в узких предметных областях, обусловленная большой трудоемкостью разработки, сложность методики индексирования;

Вариант известной фасетной системы классификации для информационного объекта «факультет» может быть представлен табл. 3.3.

Структурную формулу любого класса можно представить в виде:

$$Ks = (\text{Факультет, Возраст, Пол, Дети}).$$

Большое распространение в стране получила полуфасетная классификация – универсальная десятичная классификация (УДК), принятая в качестве государственной системы классификации публикаций по естественным и техническим наукам. Она включает более 130 тысяч рубрик. В УДК выделяются 6 фасетов, из которых 4 являются фасетами содержания и 2 – фасетами формы.

Преимущества УДК: применение в качестве нотации арабских цифр; цифровые индексы выполняют роль посредников для разных естественных языков; постоянно обновляется и исправляется при сохранении единства и стабильности; позволяет реализовать машинную обработку и поиск информации.

Таблица 3.3

Вариант фасетной системы классификации информационного объекта «Факультет»

Фасеты			
Название факультета	Возраст	Пол	Дети
Радиотехнический	До 20 лет	М	Есть
Машиностроительный	20–30 лет	Ж	Нет
Коммерческий	Свыше 30 лет	Ж	Есть
Информационный	До 20 лет	М	Нет
Математический	До 20 лет	Ж,	Нет

В любой стране разработаны и применяются государственные, отраслевые, региональные классификаторы, в том числе отраслей промышленности, оборудования, профессий, единиц измерения, статей затрат и т.д.

Кроме этой классификации нашли применение библиотечно-библиографический классификатор (ББК), десятичный классификатор Дьюин, общая система классификации (ОСК), международная классификация изобретений (МКИ) и др.

Классификатор – систематизированный свод наименований и кодов классификационных группировок.

5) алфавитно-предметная классификация

Алфавитно-предметная классификация (АПК) – это классификация, в которой классы понятий (предметов, фактов, сведений) расположены в алфавитном порядке их имен, предназначенная для узкопредметного поиска документированной информации и сведений, главная тема или предмет которых обозначаются заголовком (рубрикой).

АПК в основном применяются для индексирования книг, периодических изданий и составления каталогов к библиотечным фондам и систематическим каталогам.

Рассмотренные системные классификации характеризуются рядом свойств, обусловливающих низкую эффективность и затрудняющих их использование в информационном поиске, особенно с применением технических средств. К таким свойствам относятся:

– предварительная координация (связь) слов и словосочетаний в рубрике;

– трудность обновления и дополнения;

– практическая невозможность полной и детальной разработки схемы классификации;

– трудоемкость использования при индексировании.

Эти недостатки обусловили необходимость создания других классификаций, не имеющих указанных недостатков и позволяющих вести информационный поиск с помощью технических средств.

6) дескрипторная система классификации

Для организации поиска информации, для ведения тезаурусов (словарей) эффективно используется дескрипторная (описательная) система классификации, язык которой приближается к естественному языку описания информационных объектов.

Тезаурус представляет совокупность знаков, между которыми установлены связи в соответствии с их значениями. Тезаурус является моделью накопленных знаний.

Дескрипторная система классификации (от англ. describe – описывать) – классификация, в основе которой лежит алфавитный перечень слов или словосочетаний, обеспечивающий представление содержания документов и запросов в виде некоторого множества слов и словосочетаний естественного языка, являющихся своеобразными их координатами в некотором пространстве предметно-тематических признаков.

Дескриптор – нормализованное слово или словосочетание, выбранное из группы синонимичных или близких по значению, ключевых слов (класса условной эквивалентности) и предназначенное для индексирования документированной информации и запросов.

Суть дескрипторного метода классификации заключается в следующем:

– в качестве лексических единиц отбирается совокупность значимых ключевых слов или словосочетаний, описывающих определенную предметную область или совокупность однородных объектов. Причем среди ключевых слов могут находиться синонимы;

– выбранные ключевые слова и словосочетания подвергаются нормализации, т.е. из совокупности синонимов выбирается один или несколько наиболее употребляемых;

– создается словарь дескрипторов.

Дескрипторный словарь – нормированный словарь, в котором в едином алфавитном ряду приведены все важнейшие ключевые слова, словосочетания и дескрипторы отрасли или области знаний, с соответствующими пометками, отобранные в результате процедуры нормализации.

Для обеспечения полноты отражения содержания документированной информации и запросов с помощью ИПЯ с целью повышения эффективности информационного поиска в дескрипторном словаре для каждого дескриптора указываются его синонимы, а также видовые, родовые и ассоциативные отношения.

Информационно-поисковый тезаурус (ИПТ) – нормированный словарь дескрипторов и ключевых слов с зафиксированными парадигматическими отношениями, предназначенный для координатного индексирования документов и запросов.

Целевое назначение тезауруса определяется предметной областью.

Лингвистический тезаурус обеспечивает помощь авторам в выборе разнообразных слов для выражения одной и той же мысли, а также возможность использования любого множества связных по смыслу слов и словосочетаний для выражения идеи, обозначенной названием понятийной группы.

Научно-информационный тезаурус применяется в деятельности для замены разнообразных слов и словосочетаний, выражающих одну и ту же мысль, одним словом или словосочетанием (дескриптором).

ИПТ отличается от перечисленных и используется для индексирования документированной информации и запросов в системе поиска. Этапы разработки ИПТ:

1) определение тематического охвата ИПТ, определяемого на основе анализа информационных потребностей абонентов, обслуживаемых соответствующей информационной системой;

2) сбор массива лексических единиц;

3) формирование словника на основе выделения лексических единиц в массиве документов или запросов фонда, для которого разрабатывается ИПТ; построение словарных статей и указателей;

4) оформление ИПТ;

5) экспертиза и регистрация ИПТ.

Лексическая единица – последовательность букв, цифр и специальных символов, принятая в данном естественном языке для обозначения определенного понятия. К лексическим единицам относятся слова, термины, комбинации терминов, имена собственные, марки, номенклатурные обозначения, аббревиатуры, общепринятые сокращения, лексически значимые компоненты сложных слов. Лексические единицы в ИПТ делятся на дескрипторы и аскрипторы.

Дескрипторы – лексическая единица ИПТ, предназначенная для использования в поисковых образах документированной информации (запросов).

Аскрипторы (недескриптор) – лексическая единица ИПТ, которая в поисковых образах документированной информации подлежит замене на дескриптор при поиске или обработке информации.

Если лексические единицы относятся к одному классу синонимии, они связываются ссылками «с» и «см», поскольку замена одной из единиц на другую не приводит к изменению смысла контекста. Аскриптор, заменяемый в различных контекстах разными дескрипторами, имеет указание на эти дескрипторы ссылкой «исп а» (используй альтернативно).

Так, например, для искусственных языков:

– *информационно-поисковые языки;*
– *языки международного общения;*
– *языки программирования.*

Родовидовые связи обозначаются ссылками «вр» (выше род) и «нв» (ниже – вид) и устанавливаются между двумя дескрипторами, если объем понятия нижестоящего дескриптора входит в объем понятия вышестоящего.

В качестве объекта дескрипторной классификации можно рассмотреть известный пример с успеваемостью студентов:

➢ выбираются ключевые слова: оценка, экзамен, зачет, преподаватель, студент, семестр, название предмета. Здесь нет синонимов, и поэтому указанные ключевые слова можно использовать как словарь дескрипторов;

➢ в качестве предметной области выбирается учебная деятельность в высшем учебном заведении;

➢ ключевыми словами могут быть «студент», «обучаемый», «учащийся», «преподаватель», «учитель», «педагог», «лектор», «ассистент», «доцент», «профессор», «коллега», «факультет», «подразделение университета», «аудитория», «комната», «лекция», «практическое занятие», занятие» и т.д.;

➢ среди указанных ключевых слов встречаются синонимы: студент, обучаемый, учащийся, преподаватель, учитель, педагог, факультет, подразделение университета и т.д.;

➢ после нормализации словарь дескрипторов будет состоять из следующих слов: студент, преподаватель, лектор, ассистент, доцент, профессор, факультет, аудитория, лекция, практическое занятие и т.д.

Между дескрипторами устанавливаются связи, которые позволяют расширить область поиска информации. Связи могут быть трех видов:

– *синонимические,* рассматривающие некоторую совокупность ключевых слов как синонимы;

– *родовидовые,* отражающие включение некоторого класса объектов в более представительный класс;

– *ассоциативные*, соединяющие дескрипторы, обладающие общими свойствами.

Можно привести примеры: 1) синонимическая связь: студент – учащийся – обучаемый; 2) родовидовая связь: университет – факультет – кафедра; 3) ассоциативная связь: студент – экзамен – профессор – аудитория.

Достоинства дескрипторных систем классификации:

– меньшая, по сравнению с классификационными, трудоемкость разработки;

– возможность осуществлять поиск по любому заранее заданному сочетанию характеристик, входящих в ИПТ;

– возможность автоматизации процесса индексирования документов.

– Недостатки дескрипторных систем классификации:

– необходимость априорной разработки ИПТ, требующая значительных затрат труда и времени, и трудности однозначного применения тезаурусов;

– локальный характер дескрипторных ИПЯ, затрудняющий их использование для обмена информацией между системами с различными ИПЯ;

– фиксированность лексики ИПЯ, ограничивающая отражение в поисковых образах документов новых понятий;

– различная производительность и неадекватность значения термина, выбранного в качестве дескриптора в различных дескрипторных ИПЯ.

Особенностью дескрипторных систем классификации является необходимость использования специальных средств автоматизации информационного поиска документированной информации.

7) решение задачи классификация информации

Решение задачи классификации информации обусловлено необходимостью организации информационного обеспечения деятельности. Ее место в общей системе задач организации информационного обеспечения можно представить табл. 3.4.

Таблица 3.4

Перечень задач организации информационного обеспечения деятельности

Аспекты, определяющие содержание организации деятельности	Задачи организации информационного обеспечения деятельности
1. Внутренняя упорядоченность, взаимодействие частей целого	1.1. Структуризация информации
	1.2. Структуризация информационных потребностей объекта
	1.3. Определение источников информации
2. Совокупность процессов или действий, ведущих к образованию и совершенствованию взаимодействий между частями	2.1. Формирование структуры информационного потока
	2.2. Обоснование содержания обработки информации на всех этапах циркуляции информационного потока
	2.3. Обоснование методов обработки информации
	2.4. Обоснование перечня, содержания и методов решения обеспечивающих процессов
3. Объединение людей, совместно реализующих программу	3.1. Выделение ответственности за организацию информационного обеспечения
	3.2. Распределение прав и обязанностей сотрудников объекта в процессе сбора, обработки и использования информации
	3.3. Обучение сотрудников объекта правилам сбора, обработки и использования информации

Решение задачи структуризации информации предполагает ее системную классификацию для достижения следующих целей: рационализации процессов сбора, обработки, использования, передачи и переработки; обоснования способов представления информации для каждого элемента классификационных структур.

Системная классификация информации в свою очередь предполагает решение двух частных задач: выбора показателей и критериев классификации; обоснования классификационной структуры в соответствии с выбранными показателями.

Следовательно, основой классификации является выбор показателей классификации, который определяется целями самой классификации. Важность этого обстоятельства обусловлена широким пониманием информации, ее содержания и свойств, выбор которых для практической деятельности будет определять, и изменять показатели и результаты классификации. Таким образом, любая классификация всегда будет относительна выбранных видов деятельности. Один и тот же объект может классифицироваться по разным показателям. Подтверждением этого являются рассмотренные ранее вопросы, связанные с определением информации, раскрытием ее содержания и свойств. В этом же вопросе будет продолжено рассмотрение классификации информации с позиций ее прагматических свойств для решения задач и проблем информатики: формирования информационного ресурса общества, необходимого для информационного обеспечения деятельности личности, общества и государства; формирования специфического материализованного сырья, подлежащего обработке по специальным технологиям.

С первой позиции основными показателями классификации должны быть показатели, определяющие удобство использования информации в процессе решения прагматических задач, в том числе практических и научных. Со второй позиции – показатели удобства обработки и хранения информации при использовании современных информационных технологий.

В силу близости содержания информационного ресурса и собственно информации как сырья для обработки, необходимо указанные показатели привести к единой классификации. Очевидно, наиболее целесообразным будет классифицировать информацию исходя из потребностей ее функционального использования человеком, а средства и методы обработки (современные информационные технологии) приспособить к рациональной обработке в соответствии с результатами полученной классификации. Вариант такой известной классификации Герасименко В.А. осуществляется по следующим показателям: по виду источника – сведения об источниках (документальная) и фактах (фактографическая); практике использования; по содержанию, объему и способу оформления и др. (рис. 3.1).

В табл. 3.5. приведена структура классификации видов информации, циркулирующей в организации, без учета предметной ориентации, которая может быть использована в различных условиях, разными потребителями и для различных целей. В основу классификации положено пять наиболее общих показателей: место возникновения, стадия обработки, способ отображения, стабильность, функция управления.

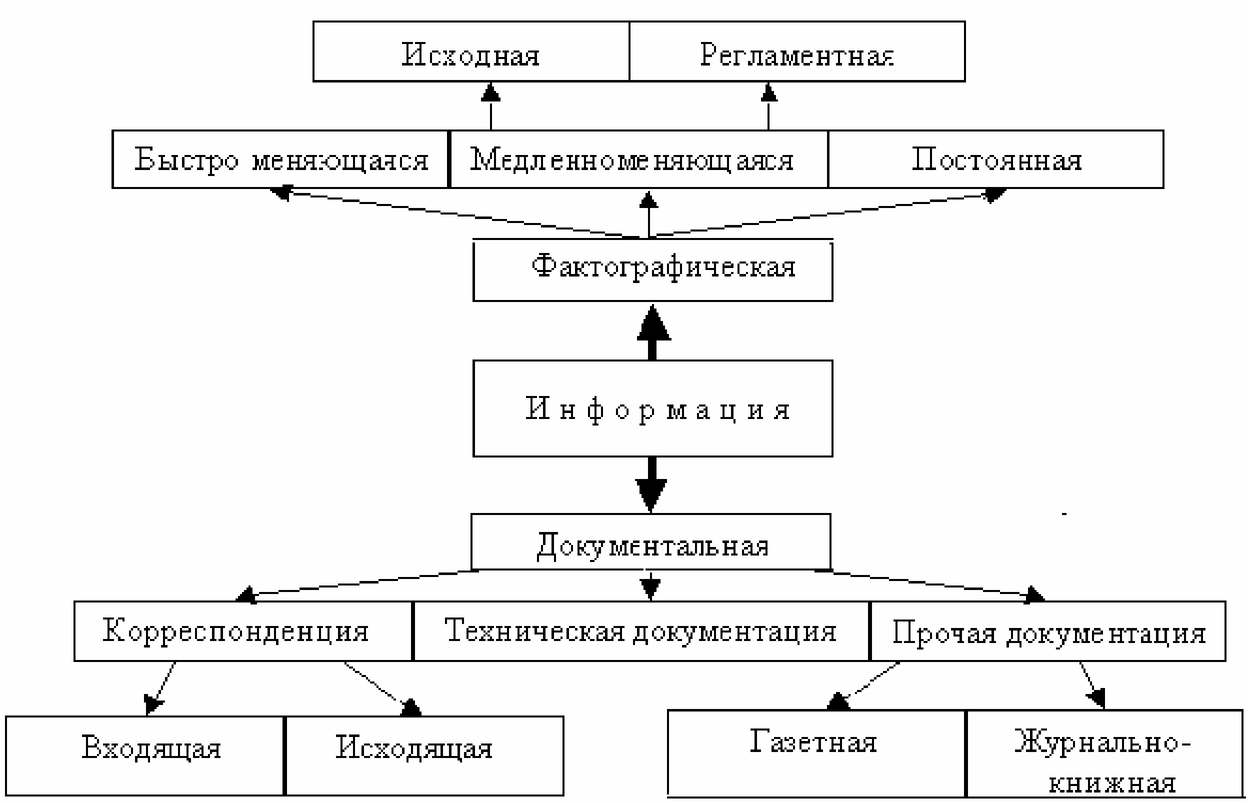


Рис. 3.1. Системная классификация информации

Входная информация – это информация, поступающая в фирму или ее подразделения.

Выходная информация – это информация, поступающая из фирмы в другую фирму, организацию (подразделение).

Например, содержание указа правительства об изменении уровня взимаемых налогов, с одной стороны, внешней информацией, с другой стороны – входной. Сведения организации в налоговую инспекцию о размере отчислений в госбюджет являются, с одной стороны, выходной информацией, с другой стороны – внешней по отношению к налоговой инспекции. *По стадии обработки* информация может быть первичной, вторичной, промежуточной, результатной.

Первичная – информация, которая возникает непосредственно в процессе деятельности объекта и регистрируется на начальной стадии.

Вторичная – информация, которая получается в результате обработки первичной информации и может быть промежуточной и результатной.

Таблица 3.5

Классификация информации, циркулирующей в организации

Информация	Признаки классификации информации				
	По месту возникновения	По стабильности	По стадии обработки	По способу Отражения	По функции управления
	Входная Выходная Внутренняя Внешняя	Переменная Постоянная	Первичная Вторичная Промежуточная Результатная	Текстовая Графическая	Плановая Нормативная Справочная Учетная Оперативная

Одна и та же информация может являться входной для одной организации, а для другой – выходной.

По отношению к объекту управления информация может быть определена как внутренняя, так и внешняя.

Внутренняя информация возникает внутри объекта, *внешняя* информация – за пределами объекта.

Промежуточная информация используется в качестве исходных данных для последующих расчетов.

Результатная информация получается в процессе обработки первичной и промежуточной информации и используется для выработки управленческих решений.

По способу отображения информация подразделяется на текстовую и графическую.

Текстовая информация – это совокупность алфавитных, цифровых и специальных символов, с помощью которых представляется информация на физическом носителе (бумага, изображение на экране дисплея).

Графическая информация – это различного рода графики, диаграммы, схемы, рисунки и т.д.

По стабильности информация может быть переменной (текущей) и постоянной (условно-постоянной). Кроме этого здесь можно говорить о статической и динамической информации.

Статическая информация – числовая, логическая и символьная информация, значения которой не связаны со временем.

Динамическая информация – вся аудиоинформация, существующая только в режиме реального времени.

Эту информацию нельзя «остановить» для подробного изучения, а при изменении масштаба времени ее представления (увеличении или уменьшении), аудиоинформация искажается.

Видеоинформация может быть как статической, так и динамической. Статическая видеоинформация включает текст, рисунки, графики, чертежи, таблицы и др. В свою очередь, рисунки бывают плоские (двухмерные) и объемные (трехмерные). Примерам динамической видеоинформация являются видео-, мульт- и слайдфильмы, представляющие последовательное экспонирование на экране в реальном масштабе времени отдельных кадров в соответствии со сценарием.

Переменная информация отражает фактические количественные и качественные характеристики производственно-хозяйственной деятельности организации.

Она может меняться для каждого случая, как по назначению, так и по количеству.

Постоянная (условно-постоянная) информация – это неизменная и многократно используемая в течение длительного периода времени информация. Постоянная информация может быть справочной, нормативной, плановой.

Постоянная справочная информация включает описание постоянных свойств объекта в виде устойчивых длительное время признаков.

Постоянная нормативная информация представляется данными местных, отраслевых и общегосударственных нормативов. Например, размер налога на прибыль, стандарт на качество продуктов определенного вида, размер минимальной оплаты труда, тарифная сетка оплаты государственным служащим.

Постоянная плановая информация содержит многократно используемые в организации плановые показатели. Например, план выпуска телевизоров, план подготовки специалистов определенной квалификации.

По функциям управления экономическую информацию классифицируют на плановую, нормативно-справочную, учетную и оперативную (текущую).

Плановая информация – информация о параметрах объекта управления на будущий период. На эту информацию идет ориентация всей деятельности фирмы.

Нормативно-справочная информация содержит различные нормативные и справочные данные. Ее обновление происходит достаточно редко.

Нормативно-справочной информацией на предприятии являются: время, предназначенное для изготовления типовой детали (нормы трудоемкости); среднедневная оплата рабочего по разряду; оклад служащего; адрес поставщика или покупателя и т.д.

Учетная информация – информация, которая характеризует деятельность организации за определенный прошлый период времени.

На основании этой информации проводятся корректировка плановой информации, анализ хозяйственной деятельности организации, принятие решения по более эффективному управлению работами и пр.

На практике в качестве учетной информации может выступать информация бухгалтерского учета, статистическая информация и информация оперативного учета.

Оперативная (текущая) информация – информация, используемая в оперативном управлении и характеризующая производственные процессы в текущий (данный) период времени.

К оперативной информации предъявляются серьезные требования по скорости поступления и обработки, а также по степени ее достоверности. От того, насколько быстро и качественно проводится ее обработка, зависит эффективность деятельности.

Основой информатизации всех сфер деятельности личности, общества и государства, а также информатизации правовой, государственной и политической систем, является национальная система правовой информации, аккумулирующая и представляющая потребителю не только нормативную правовую, но и ненормативную правовую информацию (данные МВД, ФСБ, судов, прокуратуры, других ведомств). Она обеспечивает решение практических задач информатизации в целом, создание и использование информационных ресурсов, выработки условий обеспечения ИИБ, механизмов правового регулирования информационных отношений, правового режима информации и информационных ресурсов, собственности, владения и распоряжения информацией и информационными ресурсами, оказания информационных услуг, защиты информации. В формировании национальной системы правовой информации важное место занимает информационная система по законодательству. Ее создание предполагает классификацию правовой информации, которую можно разделить на группы:

1) нормативная правовая информация – содержится в Конституции РФ, конституционных, федеральных и иных законах, Указах Президента, постановлениях Правительства, других нормативных правовых актах;

2) ненормативная правовая информация – возникает в процессе правоприменительной и правоохранительной деятельности государства;

3) доказательственная информация – служит для реализации процесса доказывания по уголовным, гражданским, арбитражным делам (судебные доказательства, криминалистическая, судебно-экспертная и иная информация);

4) отраслевая юридическая информация – специализированная по отраслям права правовая информация (конституционно-правовая, гражданско-правовая, уголовно-правовая и др.);

5) международно-правовая информация – сведения, содержащиеся в международных договорах, соглашениях, конвенциях, выражающие содержание отношений между государствами, народами, международными организациями;

6) научно-юридическая правовая информация – сведения, содержащиеся в юридических монографиях, учебниках, статьях, справочниках, докладах, обзорах и других материалах, не являющиеся официальными изданиями законодательных актов.

Наряду с имеющимся массивом регулируемой правовой информации, имеются массивы ненормативной правовой информации, регулируемые только в ведомственном порядке. К ненормативной правовой информации можно отнести:

– информацию о состоянии законности и правопорядка (о соблюдении прав и свобод человека, о состоянии законности и правопорядка, эффективности прокурорского надзора, о формах и способах защиты прав граждан, о принятых мерах по восстановлению законности, гражданско-правовая информация, административно-правовая информация);

– информация, связанная с раскрытием и расследованием правонарушений (криминалистическая, криминологическая, судебно-экспертная, оперативно-розыскная).

Источниками формирования ресурсов правовой ненормативной информации являются:

– производство и документирование информации в ходе правоприменительной и правоохранительной деятельности;

– поступления информации от граждан, государственных органов и их должностных лиц, органов общественного самоуправления, частных организаций и других субъектов;

– автоматизированные системы, сети, банки и базы данных правовой ненормативной информации;

– производство криминалистической, судебно-экспертной и иной информации в учреждениях и организациях судебной экспертизы и др.

Контрольные вопросы

1) Дать определения понятий: классификация, объект, классификационный признак (основание деления), классификация объектов, информационные объекты, реквизит.

2) Перечислите формы представления реквизитов.

3) Дать определение понятия информационного поиска, релевантности информации и документа, информационно-поисковой системы.

4) Какие взаимосвязанные компоненты включает информационно-поисковая система?

5) Дать определение понятия информационно-поисковый язык (ИПЯ).

6) Перечислите принципы (требования) системной классификации информации.

7) По каким признакам можно классифицировать виды информации?

8) Дать определения элементарной информации, биологической информации, социальной информации, эстетической информации, семантической информации, первичной информации, вторичной информации, машинной информации, научной информации, производственной информации.

9) Что такое семиотика, знак?

10) Назовите виды и содержание систем классификации документированной информации.

11) Что представляет общее содержание системы классификационного типа и виды ее представления?

12) Назначение, порядок построения, требования к процедуре построения структуры, достоинства и недостатки иерархической (перечисленной) системы классификации.

13) Назначение, основные понятия, принципы построения, достоинства и недостатки фасетной системы классификации.

14) Основные этапы построения фасетной системы классификации.

15) Назначение, краткая характеристика, достоинства полуфасетной классификации – универсальная десятичная классификация (УДК).

16) Назначение и применение алфавитно-предметной классификации.

17) Дать определения основных понятий дескрипторной системы классификации: тезаурус, дескрипторная система классификации, дескриптор, дескрипторный словарь, логическая единица, дескриптор, аскриптор, информационно-поисковый тезаурус (ИПТ).

18) В чем назначение информационно-поискового тезауруса и каковы этапы его разработки?

19) Приведите краткую характеристику связей между дескрипторами.

20) В чем достоинства и недостатки дескрипторных систем классификации?

21) В каких видах может проявляться любая информация, отнесенная к области абстрактных категорий?

22) Дать определения информации, данных, сообщения и раскрыть связи между этими понятиями.

23) Привести существующие структуры информации и дать их краткие характеристики.

24) Какова форма представления кодированной информации?

25) Что является предметом изучения теории информации?

26) Раскрыть общее содержание понятия сигнала и его характеристик.

27) Привести основные показатели классификации сигналов.

28) Дать определения: статический сигнал, динамический сигнал, непрерывный сигнал, дискретный сигнал, детерминированный сигнал, случайный сигнал.

29) Что является основой классификации информации?

30) Дать определения: входная информация, выходная информация, внутренняя, вторичная, промежуточная, результатная, графическая информация, переменная, постоянная (условно-постоянная), нормативная, плановая, постоянная справочная, постоянная нормативная, постоянная плановая (текущая), нормативно-справочная, оперативная (текущая).

ГЛАВА 4. ОБЩАЯ ХАРАКТЕРИСТИКА ПРОЦЕССА СБОРА, ПЕРЕДАЧИ, ОБРАБОТКИ И НАКОПЛЕНИЯ ИНФОРМАЦИИ

4.1. Сбор информации

Сбор информации является одним из важнейших этапов информационного процесса и в широком смысле означает процесс комплектования системы информационного обеспечения массивами первичных и вторичных документированных источников информации.

Необходимо напомнить, что к первичным документированным источникам информации относятся опубликованные, неопубликованные и непубликуемые документы, содержащие исходную информацию, к вторичным – документы, полученные в результате аналитико-синтетической переработки одного или нескольких первичных источников, например информационные издания (библиографические, реферативные, обзорные).

Применительно к информационным системам, созданным с применением ЭВТ, используется понятие сбора данных.

Сбор данных (data collection) – процесс идентификации и получения данных от различных источников, группирования полученных данных и представления их в форме, необходимой для ввода в ЭВМ.

Сбор информации связывает систему информационного обеспечения с внешней средой. Эффективность процесса сбора информации (информационного массива) оценивается показателями полноты, точности, оперативности, релевантности, стоимости, трудоемкости.

Полнота – количественная мера содержания в массиве всех пертинентных документов (информации), существующих на данный момент времени с точки зрения всех пользователей системы.

Пертинентность – соответствие содержания документа (информации) информационным потребностям пользователей.

Точность – количественная мера содержания в информационном массиве (системе) только пертинентных документов (информации).

Этот показатель характеризует внутреннее состояние процесса сбора, его способность удовлетворять информационные запросы независимо от времени на поиски информации.

Оперативность – способность процесса сбора выполнить задачу в минимально возможное время.

Стоимость – способность процесса сбора минимизировать затраты ресурса на единицу массива информации.

Трудоемкость – способность процесса сбора минимизировать трудозатраты на единицу массива информации.

В рамках процесса сбора осуществляется структуризация информации, информационных потребностей объекта (пользователя) и выбор источника информации. Задача структуризации информации представляет системную классификацию информации по показателям удобства ее использования человеком в процессе решения практических задач и удобства обработки и хранения с использованием современных средств и методов. Такая классификация была рассмотрена при исследовании видов информации. Решение задачи структуризации информационных потребностей субъекта (пользователя) связано с формированием информационного кадастра, представляющего организованную совокупность всех данных, необходимых и достаточных для целенаправленного информационного обеспечения деятельности современного объекта (рис. 4.1).

Рис. 4.1. Схема процесса формирования информационного кадастра объекта

Выбор источника информации должен осуществляться исходя из следующих требований: полноты информационного кадастра, достоверности информации, актуальности информации, релевантности информации, информативности, толерантности информации, функциональной направленности, легальности поставляемой информации, регулярности поступающей информации, минимизации затрат ресурсов на поддержание информационного кадастра.

Постановка задачи оптимального выбора источников информации заключается в выборе из всего имеющегося потенциального множества такой совокупности, которая при минимальных расходах средств и трудозатрат обеспечила бы удовлетворительное регулярное поступление необходимой информации, отвечающей перечисленным выше требованиям.

Решение таких задач осуществляется известными методами многокритериальной оптимизации.

Многокритериальная оптимизация (multi-criterion optimization) – метод решения задач поиска лучшего (оптимального) результата, удовлетворяющего нескольким несводимым друг к другу критериям.

4.2. Подготовка и обработка информации

1) кодирование как процесс подготовки информации в цифровом виде

Этап подготовки информации, обрабатываемой информационными системами информации, связан с известным процессом формирования структуры информационного потока, представляемого в цифровом виде. Для этого используется кодирование информации, сигналов, знаков, содержание, которого было рассмотрено в предыдущих главах, в рамках изучения вопроса представления данных в цифровых автоматах.

В рамках рассматриваемого вопроса представления информации, в теории информации и кодирования, представленной в библиографии известными работами отечественных и зарубежных авторов, способы ее представления, как уже отмечалось, связаны с кодированием, и в частности, со способами представления кодов. С учетом рассмотренных ранее общих вопросов кодирования, а также известных положений теории информации и кодирования, способы представления кодов основываются на применении теорий *соединений*, *алгебраических преобразований и геометрических построений*. Коды, как известно, могут быть представлены формулами, геометрическими фигурами, таблицами, графами, многочленами, матрицами и т.д.

Пример использования теории соединений при формировании кодов заключается в том, что количество комбинаций определяется выбранным методом образования кода, числом качественных признаков и общим числом элементов кода. Можно задать код в виде формулы *размещения*, где кодовые слова представляют собой комбинации, различающиеся как самими элементами, так и их порядком:

$$A^n_m = m(m–1)(m–2)\ldots(m – n+1),$$

где m- число качественных признаков (алфавита) кода; n – количество элементов кодового слова

С учетом этого выражения максимальное количество размещений будет при $n = m – 1$. Можно задать код, в котором кодовые слова представляют соединения, отличающиеся только порядком входящих в них элементов. Для этого используется, формула перестановок:

$$p_m = 1*2*3*\ldots\ m = m!,$$

где m – число качественных признаков (алфавит) кода.

Если код представляет собой соединения, отличающиеся только самими элементами, то он задается в виде формулы сочетаний:

$$C^n_m = m(m–1)(m–2)\ldots(m – n+1) / n! = A^n_m / P_n.$$

Максимальное число сочетаний получается *при* $n = m/2$ (для четных m) и $n = (m+1)/2$ (для нечетных m).

Пример. Для алфавита a, b, c, $m = 2$, а $C^n_m = 3$ кодовые слова имеют вид ab, ac, bc.

Представление кода в виде многочлена для любой системы счисления с основанием P соответствует выражению:

$$Q = \sum_{i=1}^{m} a_i P^{i-l} = a_m P^{m-1} + a_{m-1} P^{m-2} + \ldots + a_2 P^2 + a_1 P^0.$$

4.3. Передача информации

1) место процесса передачи информации в системе формирования информационного потока

Передача информации в широком смысле рассматривается как процесс и методы формирования и циркуляции информационного потока, который в общем, виде представляет движение структурированной информации в некоторой среде данных. Вопросы структурирования и представления информации рассмотрены в предыдущих параграфах, поэтому рассмотрим вопросы формирования среды циркуляции информационных потоков. Информационные потоки на объекте делятся на входные, внутренние и выходные. В канале телекоммуникации они могут быть разделены на односторонние и двухсторонние. Циркуляцией информационных потоков называется факт регулярного их движения между различными объектами или между различными элементами одного и того же объекта. Основные процессы, обеспечивающие циркуляцию информационных потоков, можно представить в виде схемы (рис. 4.2).

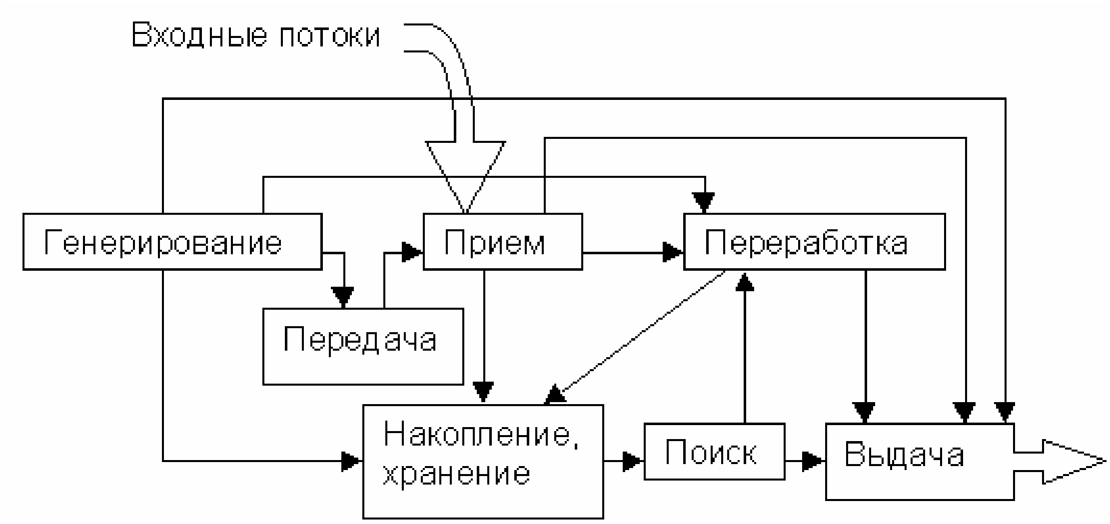


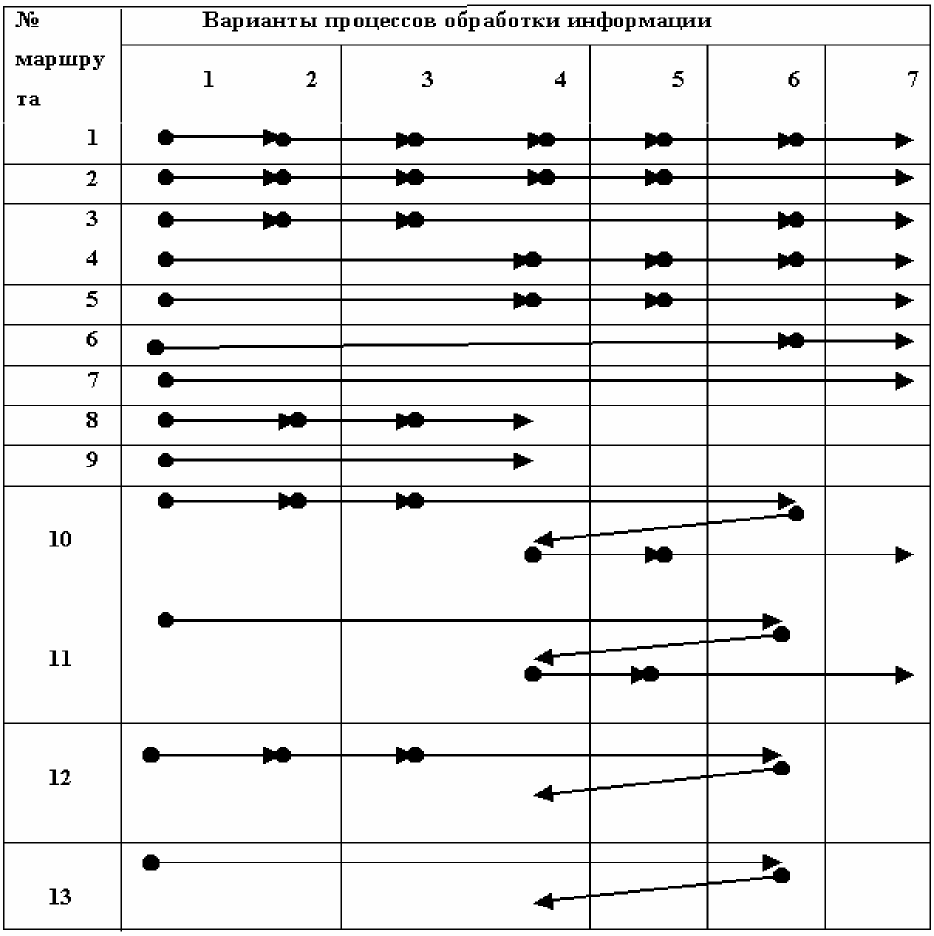
Рис. 4.2. Схема циркуляции информационных потоков

Из этой схемы очевидны различные варианты маршрутов циркуляции информации, определяемые количеством соответствующих процессов (табл. 4.1).

Важное место здесь принадлежит процессу передачи информации, который заключается в ее транспортировке от места генерации (от источника) к местам хранения, обработки или использования (потребитель).

Таблица 4.1

Варианты маршрутов циркуляции информации



2) системы передачи информации

В существующих информационных системах различных классов в зависимости от видов используемых носителей информации и средств обработки можно выделить: устную передачу при непосредственном общении; передачу бумажных носителей с помощью фельдъегерско-почтовой связи; передачу машиночитаемых носителей (магнитных карт, перфокарт, перфолент, магнитных дисков и лент) с помощью фельдъегерско-почтовой связи; передачу в виде различных электрических сигналов по каналам телекоммуникаций, в том числе по автоматизированным каналам связи.

Несмотря на широкое использование технических средств телекоммуникаций, продолжает использоваться фельдъегерско-почтовая связь, при которой передачу информации можно представить в виде общей схемы (рис. 4.3).

В современных информационных телекоммуникационных системах передачу информации по каналам связи принято представлять схемой (рис. 4.4). Основные принципы и методы ее реализации отражены в теории связи и передачи сигналов, материалы которой представлены в библиографии книги.

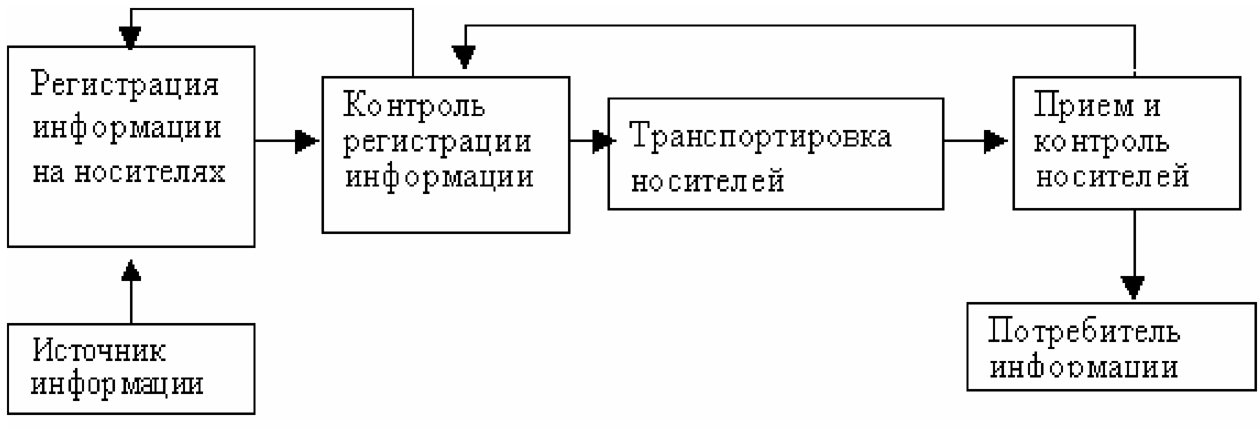


Рис. 4.3. Общая схема процесса передачи информации по каналам фельдъегерско-почтовой связи

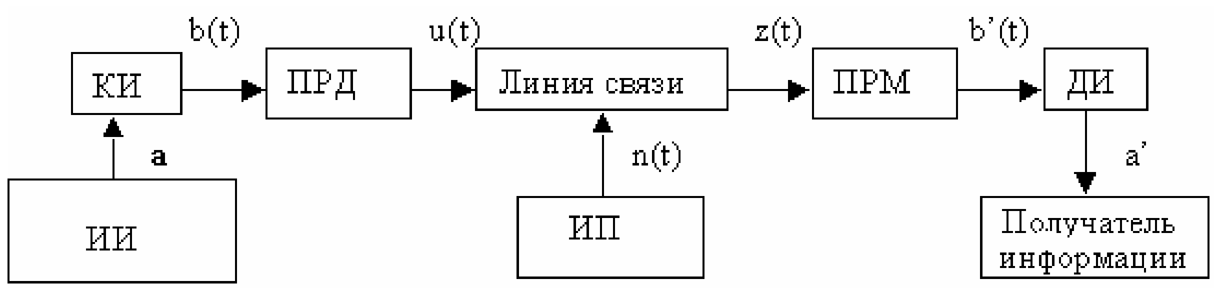


Рис. 4.4. Общая схема системы передачи информации по каналам телекоммуникаций (связи)

Источником информации (ИИ) и ее получателем могут быть как человек, так и различные технические устройства (средства связи, ЭВТ и др.).

С помощью кодера устройства источника (КИ) информация, имеющая любую физическую природу (изображение, звук и т.п.), преобразуется в первичный электрический сигнал b(t). Для непрерывной информации, например речевого сообщения, эта операция сводится к преобразованию звукового давления в пропорционально изменяющийся электрический ток микрофона, который в каждый момент отсчета (времени) можно представить конечным числом сигналов, соответствующих отдельным знакам алфавита источника.

В телеграфии последовательность элементов сообщения (букв алфавита) большого объема заменяются (кодируются) символами (буквами) другого алфавита меньшего объема. В дальнейшем с помощью технического устройства последовательность кодовых символов преобразуется в последовательность электрических сигналов.

Процесс преобразования букв сообщения в сигналы преследует несколько целей. Первой целью является преобразование информации в систему символов, обеспечивающих простоту и надежность аппаратной реализации информационных устройств, и их эффективность:

– простоту технических средств распознавания элементарных символов сообщения;

– снижение избыточности символов, требующихся на букву сообщения;

– минимальное время передачи или минимальный объем запоминающих устройств хранения информации;

– простоту выполнения арифметических и логических действий с хранимой информацией.

Рассмотренное кодирование при отсутствии помех в канале связи дает выигрыш во времени передачи или объеме запоминающего устройства, т.е. повышает эффективность системы. Оно получило название эффективного или оптимального кодирования.

Второй, не менее важной целью кодирования, является обеспечение заданной достоверности сообщения при передаче или хранении информации путем дополнительного внесения избыточности, но уже по простым алгоритмам и с учетом интенсивности и статистических закономерностей помех в канале связи. Такое кодирование называется помехоустойчивым.

Выбор соответствующих кодирующих и декодирующих устройств зависит от статистических свойств источника сообщений, уровня и характера помех в канале связи. В передатчике (ПРД) первичный электрический сигнал преобразуется во вторичный $u(t)$, пригодный для передачи по соответствующему каналу (линии) связи. Такое преобразование осуществляется с помощью модулятора.

Преобразование сообщения в сигнал должно быть обратимым. Это позволит по выходному сигналу восстановить входной первичный сигнал, т.е. получить всю информацию, содержащуюся в переданном сообщении. В противном случае часть информации будет потеряна.

Линия связи – среда, используемая для передачи сигналов от передатчика к приемнику.

В системах электросвязи такими линиями являются кабели, волноводы. В системах радиосвязи линиями являются пространства, в которых распространяются электромагнитные волны от передатчика к приемнику. При передаче по линии связи на сигналы от источника (ИП) могут накладываться помехи $n(t)$, в результате чего сигналы искажаются.

Приемное устройство, в составе приемника (ПРМ) и декодирующего устройства информации (ДИ), обрабатывает принятый сигнал $z(t) = s(t) + n(t)$ и восстанавливает по нему передаваемое сообщение a', адекватное сообщению источника информации a.

Система связи – совокупность технических средств передачи сообщений от источника к потребителю, включающая передающие (КИ, ПРД) и приемное (ДИ, ПРМ) устройства и линию связи.

По виду передаваемых сообщений различают системы связи, предназначенные для передачи речи (телефонная), текста (телеграфная), неподвижных изображений (фототелеграфная), изображений (телевизионная); сигналов радиовещания, передачи данных, видеотекста, телетекста, конференц-связи, телеизмерения и телеуправления и др.

По количеству передаваемых сообщений по одной линии связи системы делятся на одноканальные и многоканальные.

Канал связи – совокупность технических средств (ПРД, линия связи, ПРМ), обеспечивающих передачу сигналов от источника (КИ) до получателя сигнала (ДИ).

Классификация каналов связи может осуществляться по ряду показателей.

По виду сигналов, поступающих на вход канала и снимаемых с его выхода – дискретные, непрерывные, дискретно-непрерывные.

По способу передачи информации между объектами каналы бывают:

– дуплексные, обеспечивающие возможность передачи информации как от объекта А к Б, так и от Б к А;

– симплексные, обеспечивающие передачу информации только в одном направлении.

По структуре связи источника с адресатами канала связи:

– последовательные – однофидерные линии связи проходят через каждый адресат;

– радиальные – каждый адресат соединен с объектом отдельной однофидерной линией;

– древовидные – однофидерные линии непосредственно не соединены с объектом, а подключаются к нему через отдельную линию.

По виду линий связи каналы делятся на проводные, радиоканалы, оптические, гидроакустические и др.

4.4. Хранение и накопление информации

1) подготовка информации к хранению

Хранение информации (information storage) – это ее запись во вспомогательные запоминающие устройства на различных носителях для последующего использования.

Хранение является одной из основных операций, осуществляемых над информацией, и главным способом обеспечения ее доступности в течение определенного промежутка времени. Основное содержание процесса хранения (holding storage) и накопления информации состоит в создании, записи, пополнении и поддержании информационных массивов и баз данных (рис. 4.5).

В результате реализации такого алгоритма, документ, независимо от формы представления, поступивший в информационную систему, подвергается обработке и после этого отправляется в хранилище (базу данных), где он помещается на соответствующую «полку» в зависимости от принятой системы хранения. Результаты обработки передаются в каталог.

Этап хранения информации может быть представлен на следующих уровнях: внешнем, концептуальном, (логическом), внутреннем, физическом.

Внешний уровень отражает содержательность информации и предлагает способы (виды) представления данных пользователю в ходе реализации их хранения.

Концептуальный уровень определяет порядок организации информационных массивов и способы хранения информации (файлы, массивы, распределенное хранение, сосредоточенное и др.).

Внутренний уровень представляет организацию хранения информационных массивов в системе ее обработки и определяется разработчиком.

Физический уровень хранения означает реализацию хранения информации на конкретных физических носителях.

Способы организации хранения информации связаны с ее поиском – операцией, предполагающей извлечение хранимой информации.

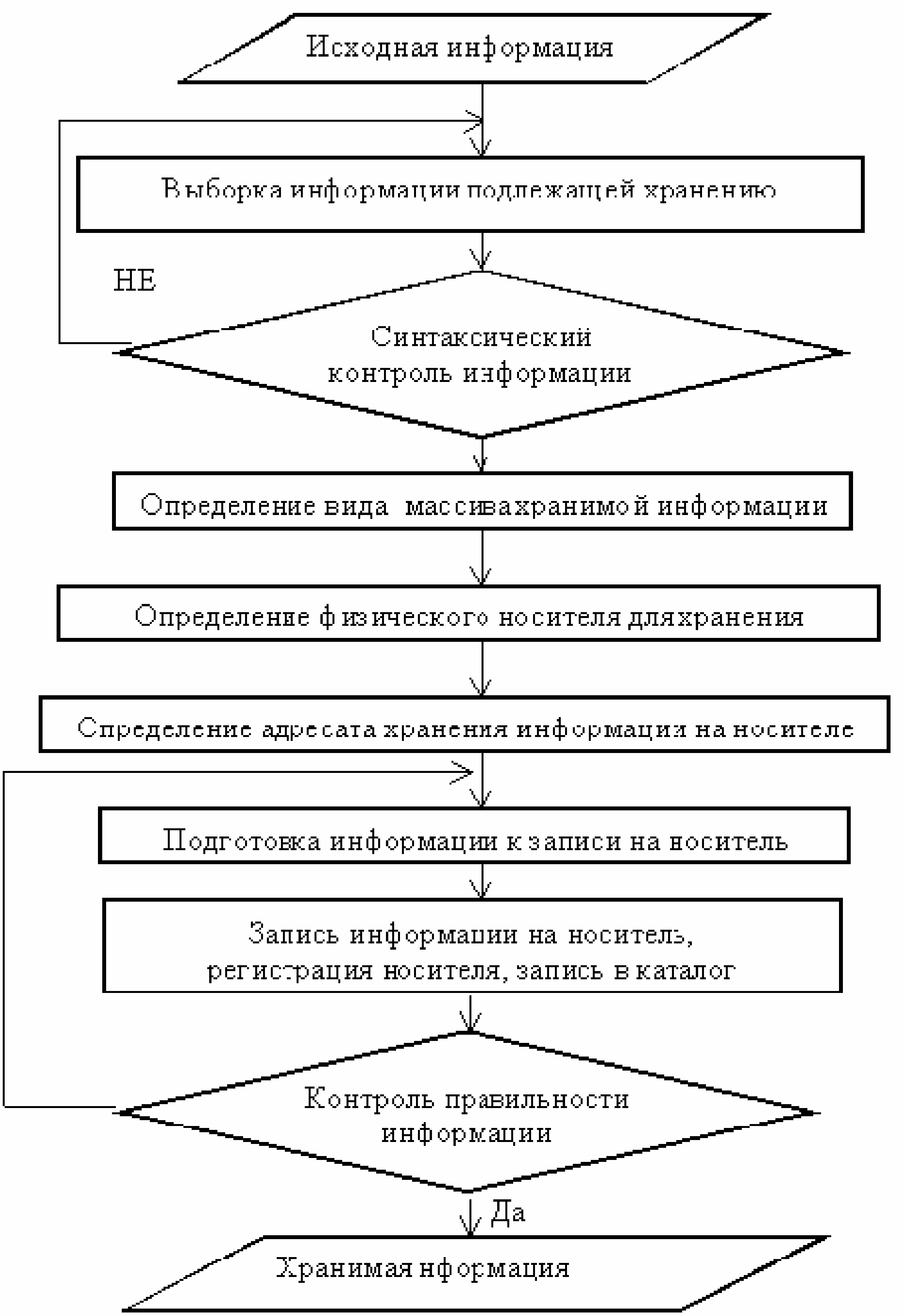


Рис. 4.5. Алгоритм процесса подготовки информации к хранению

2) поиск, обработка и организация массивов информации

Хранение и поиск информации являются не только операциями над ней, но и предполагают использование методов осуществления этих операций. Информация запоминается так, чтобы ее можно было отыскать для дальнейшего использования. Возможность поиска закладывается во время организации процесса запоминания. Для этого используют методы маркирования запоминаемой информации, обеспечивающие поиск и последующий доступ к ней. Эти методы применяются для работы с файлами, графическими базами данных и т.д.

Маркер (mark, marker) – метка на носителе информации, обозначающая начало или конец данных либо их части (блока). В современных носителях информации используются маркеры: адреса (адресный маркер) – код или физическая метка на дорожке диска, указывающие на начало адреса сектора; группы – маркер, указывающий начало или конец группы данных; дорожки (начала оборота) – отверстие на нижнем диске пакета магнитных дисков, указывающие физическое начало каждой дорожки пакета; защиты – прямоугольный вырез на носителе (картонном пакете, конверте, магнитном диске), разрешающий выполнение любых операций над данными: запись, чтение, обновление, удаление и др.; конца файла – метка, используемая для указания окончания считывания последней записи файла; ленты (ленточный маркер) – управляющая запись или физическая метка на магнитной ленте, обозначающая признак начала или конца блока данных или файла; сегмента – специальная метка, записываемая на магнитной ленте для отделения одного сегмента набора данных от другого сегмента.

Хранение информации в ЭВМ связано как с процессом ее арифметической обработки, так и с принципами организации информационных массивов, поиска, обновления, представления информации и др.

Важным этапом автоматизированного этапа хранения является организация информационных массивов.

Массив (англ. array) – упорядоченное множество данных.

Информационный массив – система хранения информации, включающая представление данных и связей между ними, т.е. принципы их организации.

С учетом этого рассматриваются *линейные* и *многомерные* структуры организации информационных массивов. В свою очередь линейная структура данных может быть представлена в виде: строк, одномерных массивов; стеков, очередей, деков и др.

Строка – представление данных в виде элементов, располагающихся по признаку непосредственного следования, т.е. по мере поступления данных в ЭВМ.

Одномерный массив – представление данных, отдельные элементы которых имеют индексы, т.е. поставленные им в соответствие целые числа, рассматриваемые как номер элемента массива.

Индекс обеспечивает поиск и идентификацию элементов, а, следовательно, и доступ к заданному элементу, что облегчает его поиск по сравнению с поиском в строке.

Идентификация – процесс отождествления объекта с одним из известных.

Стек – структура данных, учитывающая динамику процесса ввода и вывода информации, использующая линейный принцип организации, реализующий процедуру обслуживания «последним пришел – первым ушел».

В стеке первым удаляется последний поступивший элемент.

Очередь – структура организации данных, при которой для обработки информации выбирается элемент, поступивший ранее всех других.

Дека – структура организации данных, одновременно сочетающая рассмотренные виды.

Нелинейные структуры хранения данных – структуры использующие многомерные конструкции (многомерные массивы) следующих видов: *деревья, графы, сети*.

Элемент многомерного массива определяется индексом, состоящим из набора чисел. Формой представления является матрица, каждый элемент которой определяется набором индексов требуемого элемента массива. Так, в двухмерном массиве, элемент обозначается двумя индексами, а в трехмерном – тремя.

Массивы по своей структуре близки к файлам и отличаются от последних двумя основными признаками: каждый элемент массива может быть явно обозначен, и к нему имеется прямой доступ; число элементов массива определятся при его описании. Организация расположения данных в многомерном массиве может быть представлена в виде логических структур информационных массивов. В этих массивах структуры данных компонуются в виде записей, располагающихся различным образом. С учетом этого выделяют следующие основные структуры информационных массивов: последовательную, цепную, ветвящуюся, списковую.

В *последовательной структуре* информационного массива записи располагаются последовательно, нахождение требуемой записи осуществляется путем просмотра всех предшествующих. Включение новой записи в информационный массив требует смещения всех записей, начиная с той, которая включается. Обновление информационных массивов при последовательной структуре требует перезаписи всего массива.

В *цепной структуре* информационные массивы располагаются произвольно. Для логической связи отдельных записей необходима их адресация, т.е. каждая предыдущая запись логически связанного информационного массива должна содержать адрес расположения последующей записи. Если с определенного уровня, т.е. признака, значения в записях повторяются в различных сочетаниях, то в целях экономии памяти возможен переход от цепной структуры к ветвящейся.

В *ветвящейся структуре* информационного массива сначала размещается запись, отображающая признак объекта с небольшим числом значений, далее эти значения повторяются в записях в различных сочетаниях. Это обеспечивает возможность перехода от некоторой основной записи к другим в зависимости от запроса, не повторяя основную запись. Чтобы устранить повторяющиеся записи и соответствующие им поля из памяти, их удаляют из основного массива и объединяют в дополнительный, небольшой информационный массив. В таком массиве записи упорядочиваются по какому-то признаку без повторений. Тогда в основном массиве вместо удаленного информационного поля указываются адреса записей, которые размещены в дополнительном массиве. Данная структура является удобной при реорганизации информационной базы, поскольку повторяющиеся записи легко могут быть заменены, так как они хранятся в дополнительном массиве, основной же массив подвергается при этом малому изменению. Однако эта структура требует дополнительного объема памяти.

Списковая структура информационных массивов характеризуется наличием списка, который содержит набор данных, определяющих логический порядок организации информационного массива. Список включает имя поля данных и адреса полей. В памяти ЭВМ элементы списка физически разнесены, но связаны друг с другом логически за счет адресных ссылок. Поле данных в зависимости от характера хранимой информации может представляться двоичным разрядом, словом фиксированной либо переменной длины, а также набором отдельных слов.

Формализовано список может быть представлен в виде таблицы, в которой именам списка и полям данных сопоставлены адреса. Адреса выбираются произвольно по мере наличия свободных мест в запоминающем устройстве. В случае необходимости повторений какой-либо информации рекомендуется многократно обращаться по адресу, который может входить в несколько списков, т.е. применить механизм многократных адресных ссылок. Списковая структура с механизмом адресных ссылок может быть представлена в виде *графа древовидной структуры*. В нем каждый элемент списка включает в себя маркерное поле, поле данных и адресное поле. Маркерное поле предупреждает, имеется ли ссылка на другой список либо такая ссылка отсутствует. В зависимости от этого в маркерном поле ставится знак минус или плюс.

Списки могут отображать и более сложные структуры, чем древовидные. Они могут быть показаны *ориентированными графами* с полями, в которых возможна ссылка вперед и назад. Возникает так называемый симметричный список, и появляется возможность движения в структуре данных в разных направлениях. Рассмотренные списковые структуры информационных массивов отличаются следующими особенностями: обладают высокой логической простотой; относительно большим временем доступа, обусловленным адресным обращением к данным, при котором к каждому элементу списка необходимо иметь ссылку; значительным возрастанием объема памяти запоминающего устройства, по сравнению с последовательной структурой организации информационных массивов, обусловленным адресным обращением к данным.

С учетом рассмотренных структур формирования информационных массивов можно представить ряд способов организации массивов в ЗУ ЭВТ (рис. 4.6).

Рис. 4.6. Способы организации массивов информации

3) ***организация хранения и накопления массивов информации на физических носителях***

На физическом уровне любые записи информационного поля представляются в виде двоичных символов. Обращение к памяти большого объема требует и большой длины адреса. Если память имеет емкость 2*n* слов, то требуются для поиска этих слов *n*-разрядные адреса.

В микропроцессорах восьмиразрядные слова дают возможность обращаться к 256 ячейкам памяти, что оказывается недостаточно для хранения информации в автоматизированных системах. Поэтому переходят к страничной организации памяти, если непосредственно обращение к любой ячейке невозможно. В этом случае выбирают область памяти 2*n* слов и называют страницей, обращение к которой осуществляется командой, содержащей *n*-разрядное адресное поле. В микропроцессорах обычно используют страницы размером в 256 слов в запоминающем устройстве ЭВТ.

Принципы адресации, объемы памяти, количественные характеристики зависят от функционального назначения запоминающих устройств, которые по уровням функциональной иерархии можно разделить на сверхоперативное, оперативное, постоянное, полупостоянное, внешнее, буферное.

Хранение информации осуществляется на специальных носителях. Исторически наиболее распространенным носителем информации была бумага, которая, однако, непригодна в обычных (не специальных) условиях для длительного хранения информации. На бумагу оказывают вредное воздействие температурные условия: либо разбухает, либо ломается, способна к возгоранию.

Для ЭВТ по материалу изготовления различают следующие машинные носители: бумажные, металлические, пластмассовые, комбинированные и др.

По принципу воздействия и возможности изменения структуры выделяют магнитные, полупроводниковые, диэлектрические, перфорационные, оптические и др.

По методу считывания различают контактные, магнитные, электрические, оптические. Особое значение при построении информационного обеспечения имеют характеристики доступа к информации, записанной на носителе. Выделяют носители прямого и последовательного доступа. Пригодность носителя для хранения информации оценивается следующими параметрами: временем доступа, емкостью памяти и плотностью записи.

Таким образом, можно заключить, что хранение информации (data holding) представляет процесс передачи информации во времени, связанный с обеспечением неизменности состояния материального носителя.

Контрольные вопросы

1. Перечислите основные этапы информационного процесса.
2. Дать определения процесса сбора информации (данных) и показателей его эффективности.
3. Дать определения полноты, точности, оперативности, точности, стоимости, трудоемкости сбора информации, пертинентности документов (информации).
4. Какими требованиями определяется выбор источника информации?

5. Дать определения: полноты информационного кадастра; достоверности информации; актуальности информации; релевантности информации; информативности; толерантности информации; функциональной направленности; легальности поставляемой информации; регулярности поступающей информации; минимизации затрат ресурса на поддержание информационного кадастра.

6. Представить структурно-логическую схему процесса формирования информационного кадастра объекта.

7. Дать определение понятия многокритериальной оптимизации.

8. Дать определения понятий: код, кодирование, кодовое слово, первичный алфавит, вторичный алфавит, декодирование, неравномерные (некомплектные) коды, равномерные (комплектные).

9. Какие виды полей данных применяются для представления информации в ЭВТ?

10. Какова структура поля упакованного формата представления чисел в ЭВТ и для каких операций над ними он используется?

11. Дайте характеристику форм представления (записи) числа в разрядной сетке ЭВТ.

12. Какова структура поля распакованного формата представления чисел в ЭВТ и для каких операций над ними он используется?

13. Дайте краткую характеристику кода ASCII (American Standard Code for Information Interchange).

14. Дать определение циркуляции информационных потоков и структуру основных процессов, обеспечивающих эту циркуляцию.

15. Дать определение линий связи и привести примеры их реализации.

16. Привести определение понятия систем связи и показатели их классификации.

17. Дать определение канала связи, привести перечень показателей их классификации.

18. Привести примеры классификации каналов связи по виду сигналов, по способу передачи информации между объектами, по виду линий связи.

19. Дать определение понятия хранения информации и перечислить основные этапы процесса хранения.

20. Представить алгоритм процесса подготовки информации к хранению.

ГЛАВА 5. ОБРАБОТКА ДАННЫХ В СИСТЕМАХ ПЕРЕДАЧИ ИНФОРМАЦИИ

5.1. Сообщения и сигналы в системах передачи данных

В основе функционирования современных систем передачи данных лежит использование дискретных сообщений и сигналов.

Сообщение – формализованные данные, полученные от источника информации и представленные в виде наборов знаков.

Данные становятся информацией в момент их использования, а информацией становятся те сообщения, которые снимают неопределенность, существующую до их поступления. С учетом этого последовательность структурных преобразований информации может быть различной в различных информационных системах.

Кодированная информация имеет форму совокупности чисел, или цифровую форму, основанную на применении какой-либо системы счисления или кодирования.

Сигнал − физический процесс отображающий (несущий) передаваемое сообщение, являющееся формой представления информации в виде речи, текстов, изображения, информации данных, графиков и др.

С учетом знаковой формы сообщение – это совокупность знаков, содержащих ту или иную информацию.

5.2. Виды и характеристики носителей информации и сигналов

Основным *видами сообщений* являются: телеграфная передача, телефонный разговор (телефонная передача), движущееся изображение.

Телеграфная передача – сообщение, представляющее последовательность отдельных знаков (букв, цифр), составляющих текст телеграмм.

Телефонный разговор (передача) – сообщение, представляющее непрерывно изменяющееся во времени звуковое давление, отражающее содержание, интонацию, тембр, ритм и другие характеристики голоса человека.

Движущееся изображение – сообщение, представляющее изменение элемента яркости во времени, несущего информацию.

Форма представления информации − знаки (символы), в том числе: слово, фраза, жест, рисунок, электрический сигнал, форма колебаний, алфавит, математические знаки и другие.

Форма сигнала – физический процесс, изменяющийся в соответствии с переносимым сообщением.

С учетом рассмотренных понятий можно формализовать содержание понятия информации (*И*), его связи с составляющими компонентами – сообщением ($C_{ооб}$), знаками ($З_{нi}$), носителями информации (*НИ*), сигналом ($C_{игн}$), (5.1), а также представить соответствующую схему образования сигнала, рис. 5.1.

$$И \xrightarrow{Зн_i} \left\{C_{общ} = \{Зн_i\} i = \overline{1, N}\right\} \xrightarrow{НИ} \left\{C_{сигн}\right\} \qquad (5.1)$$

Физический оригинал	МАТЕМАТИЧЕСКИЕ МОДЕЛИ			Физическая модель
Материя, явления	Аппарат восприятия	Аппарат Квантования	Аппарат кодирования	Материя сигналов
□⇒	⇒□⇒	⇒□⇒	⇒□⇒	⇒□
Наблюдаемое явление	Начальная информа-ция	Квантованная информация	Кодированная информа-ция	Используемые сигналы

Рис. 5.1. Схема образования сигнала

Носитель информации *(носитель записи информации) − материальное средство (носитель, среда) регистрации, накопления, хранения информации и обмена ею между людьми или машинами.*

Форма представления носителя *информации*: *материальный носитель, физический процесс (звук, электрические волны, электрический ток, напряжение и т.п.).*

Виды носителей информации: человеко-ориентированные; машинные.

К человеко-ориентированным относятся: *бумажные носители рукописей (рукописная информация)*; *бумажная печатная информация*; *бумажная с изобразительной информацией.*

Машинные носители информации классифицируются:

1. По среде накопления информации: с непрерывной средой (магнитные ленты, магнитные диски и т.д.); с дискретной средой (ферритовые сердечники, перфоленты, перфокарты);
2. По материальной основе: бумажные; пластиковые; гибридные;
3. По форме представления информации: перфоносители; магнитные носители; фото, видео, печатные носители;
4. По возможности считывания: машинно-считываемые; человеко-машинно-считываемые;
5. По принципу кратности нанесения записи информации: одноразовые; многоразовые; стирающиеся; нестирающиеся.

При выборе соответствующих видов машинных носителей могут быть использованы показатели целесообразности применения машинных носителей информации: габаритные размеры, плотность записи, временные характеристики (температура записи, считывание, поиска), срок службы, простота записи и считывания, стоимость, юридическая сила машинных носителей информации.

Переходя к рассмотрению основных видов сигналов и их характеристик, необходимо отметить, что теория передачи (обработки сигналов) является частью теории информации. Предметом ее изучения являются оптимальные методы передачи (обработки) сообщений, которые передаются с помощью сигналов, которые представляются сообщением о любом изменении начального состояния объекта, которое может вызвать реакцию человека или прибора. При этом сигналы имеют следующие особенности:

– обладают определенными физическими свойствами;

– различаются друг от друга по природе возникновения, физической сущности (зрительные, звуковые, электрические, радиосигналы и др.);

– способны вызывать другие сигналы (электрический – звуковой, световой – электрический, электрический – магнитный); взаимосвязаны в пространстве и времени (электромагнитные поля и волны, звуковое кино).

С учетом этих особенностей формирование сигнала представляет процесс изменения параметров физического носителя по закону передаваемых сообщений (модуляция).

Основными параметрам сигналов являются:

– длительность сигнала, *Tc* – интервал, в пределах которого он существует;

– ширина частотного спектра, Fc – скорость изменения сигнала внутри интервала его существования;

– средняя мощность Pc;

1) динамический диапазон – энергетическая характеристика сигнала – отношение средней мощности сигнала к средней мощности помехи, $D = log\ P_c/P_n$;

2) объем сигнала – обобщенная характеристика, показывающая условия, которые должен обеспечивать канал связи для качественной передачи сигналов, $V_c = T_c\ F_c\ D_c$.

С учетом этих характеристик можно выделить виды сигналов, а также классифицировать их по следующим показателям:

– по положению во времени и пространстве – динамические, статистические;

– по структуре сообщения – непрерывные, дискретные;

– по характеру изменения процесса (явления) – детерминированные, случайные.

Статический сигнал – сигнал, отображающий устойчивое изменение состояния объекта.

Динамический сигнал – сигнал, отображающий непрерывные изменения состояния объекта либо процесса при переходе из одного устойчивого состояния в другое.

Детерминированный сигнал – сигнал, описывающий физические процессы детерминированными соотношениями.

Случайный сигнал – сигнал, описывающий физические процессы вероятностными математическими выражениями.

Непрерывный сигнал (сообщение)(непрерывный по состоянию, аналоговый) – сигнал, принимающий в конечном интервале амплитуд любое бесконечное количество значений в некотором интервале времени t.

5.3. Спектры сигналов

Спектры сигналов – это характеристики сигналов, которые отражают их частотные свойства, представляемые в виде частоты функций спектра, и определяют преобразование Фурье временной формы.

Спектр сигналов определяет параметры технических устройств, в том числе аппаратуры для образования каналов передачи сигналов, цифровых автоматов и других средств обработки информации. Характеристика спектра определяется шириной полосы. Для периодического сигнала $U_x(t)$ спектр определяется множеством величин (5.2):

$$A_k = \frac{2}{T}\int_{-T/2}^{T/2} U_x(t)e^{-jk\Omega t}dt, \quad \text{k} = \pm 1, \pm 2, \pm 3, \ldots - \text{период сигнала}, \quad (5.2)$$

$$\text{или } A_k = a_k - jb_k,$$

где $a_k = \frac{2}{T}\int_{-T/2}^{T/2} U_ч(t)\cos k\Omega t dt$; $b_k = \int_{-T/2}^{T/2} U_x(t)\sin k\Omega t dt$; $A_k = |A_k|e^{\varphi_k}$ – комплексный коэффициент ряда Фурье; $|A| = \sqrt{a^2 + b^2}$ – модуль величины комплексного коэффициента ряда Фурье; $\varphi_k = -arktg(b_k / a_k)$ – фаза k-й гармонической составляющей; T – период сигнала; $\Omega = 2\pi/T$ – основная круговая частота.

С учетом рассмотренных выражений описание спектра сигнала, нетрудно видеть, что спектр периодической функции является дискретным. Его ширина определяется полосой положительных частот, модуль величины комплексного коэффициента имеет значительную величину. Для непериодической функции спектр сигнала имеет непрерывный характер.

5.4. Кодирование и квантование сигналов

Переход от непрерывного (аналогового) представления сигнала к дискретному (цифровому) дает существенное преимущество при его обработке (передаче, хранении, переработке). Такой переход связан с квантованием (дискретизацией) сигнала по времени и по уровню. Рассмотрим вопросы, связанные с кодированием и квантованием сигналов.

1) квантование сигналов

Дискретный (квантованный) сигнал (дискретный по состоянию) – сигнал, принимающий в конечном интервале амплитуд ограниченное количество значений и представляющий функцию зависимости амплитуды (мощности) $U(t)$, принимающую только определенные дискретные значения Un (1,0 или др.).

Дискретный сигнал по времени − сигнал (сообщение), задаваемое не на всей оси времени, а в определенные моменты времен, t_n.

Разновидностями рассмотренных сигналов являются: непрерывные сигналы по состоянию и по времени; непрерывные по состоянию и дискретные по времени; дискретные по состоянию и непрерывные по времени; дискретные по времени и по состоянию.

Непрерывный сигнал по состоянию и по времени (рис. 5.2) представляет непрерывную функцию $x(t)$, принимающую конечные или бесконечные значения от x_{min} до x_{max}, и непрерывного аргумента t, изменяющегося в промежутке от $-T$ до T.

Непрерывный по состоянию и дискретный по времени (рис. 5.3) сигнал представляет непрерывную функцию $x(t)$, определенную на дискретном значении аргумента t_i, $i = 0, \pm 1, \pm 2, \ldots$; $t \in [-T, T]$. Функция $x(t)$, может принимать любые значения на отрезке $[x_{min}, x_{max}]$.

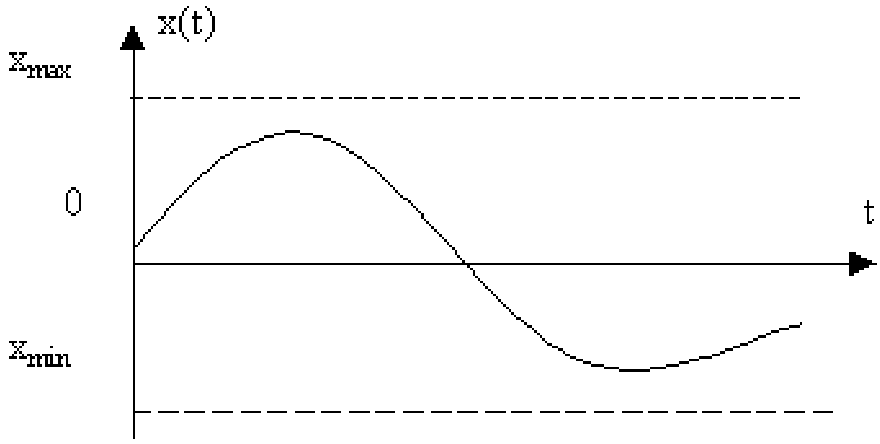


Рис. 5.2. Непрерывный сигнал по состоянию и по времени

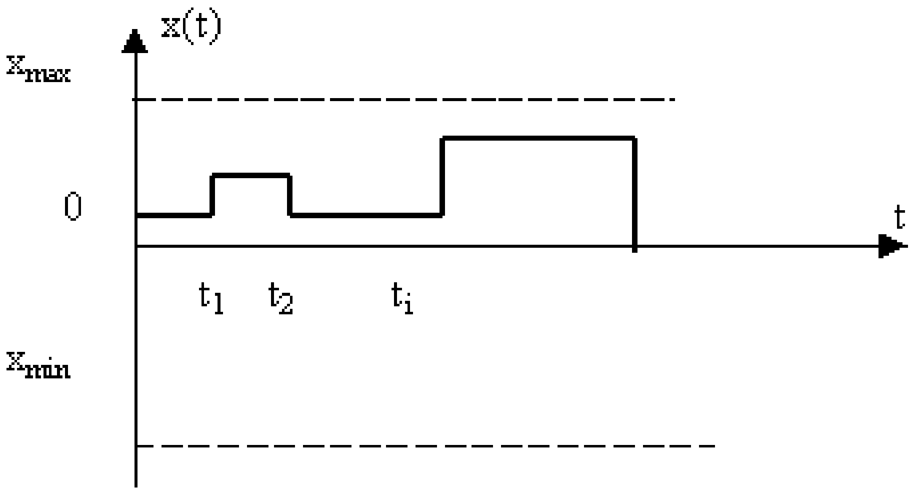


Рис. 5.3. Непрерывный по состоянию и дискретный по времени сигнал

На рис. 5.4 показан дискретный по состоянию и непрерывный по времени сигнал.

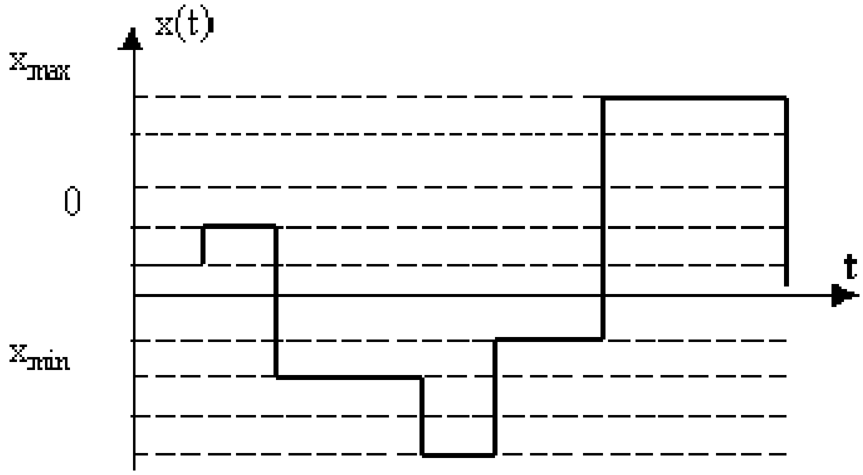


Рис. 5.4. Дискретный по состоянию и непрерывный по времени сигнал

Он представляет дискретную функцию $x(t)$, определенную на непрерывном значении аргумента времени t, который может быть любым на интервале $[-T,T]$. Значения функции представляют дискретный конечный или бесконечный ряд чисел $x_1, x_2, x_3,\ldots, x_k$, каждому из которых можно поставить в соответствие интервал, внутри которого других чисел данного ряда нет.

Дискретный по состоянию и по времени сигнал, рис. 5.5. представляет дискретную функцию $x(t)$, образующую дискретный ряд чисел $x_1, x_2, x_3,\ldots, x_k$ на дискретном ряду значений аргумента времени $-t_i,\ldots, -t_2, -t_1, t_0, t_1, t_2, \ldots, t_i,\ldots$на интервале $[-T,T]$.

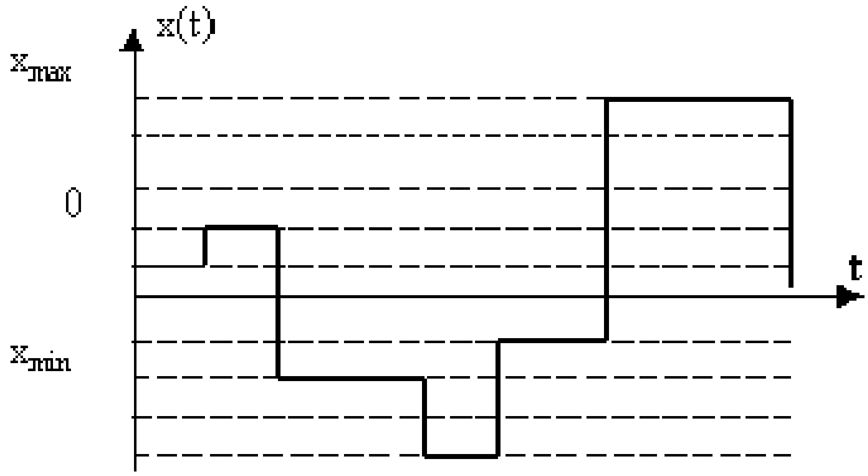


Рис. 5.5. Дискретный по времени и по состоянию сигнал

Квантование (дискретизация) сигнала по времени и по уровню − переход от аналогового представления к цифровому.

Дискретизация сигнала $x(t)$ − преобразование сигнала непрерывного аргумента времени в сигнал $x(t_i)$ дискретного аргумента времени, t_i, представляющее операцию отображения $x(t) \rightarrow x(t_i)$. В результате исходная функция $x(t)$ заменяется совокупностью отдельных значений $x(t_i)$.

По известным значениям функции $x(t_i)$ на приеме восстанавливается исходная функция $x(t)$ с некоторой погрешностью. Полученная в результате восстановления (интерполяции) функция называется *воспроизводящей* $V(t)$, представляет взвешенную сумму некоторого ряда функций $f(t - t_k)$, (5.3):

$$V(t) \sim U(t) = \sum_{\forall k} a_k f(t - t_k), \tag{5.3}$$

где $a_k = \varphi \{x(t_i), x(t_{i-1})\ldots\}$, т.е. зависит от значений функции, описывающей дискретный сигнал.

Выбор шага дискретизации исходного сигнала, $\Delta_{Ti} = t_i - t_{i-1}$, осуществляется так, чтобы по полученным дискретным отчетам $x(t_i)$, можно было получить воспроизводящую функцию $V(t)$, обеспечивающую с заданной точностью воспроизведение исходной $x(t)$.

Признаки классификации методов дискретизации:

– регулярность отсчетов (равномерный, неравномерный, случайный, адаптивный, с кратными интервалами, с некратными интервалами);

– критерий оценки точности (максимальный, среднеквадратичный, интегральный, вероятностно-зональный);

– базисная функция (ряд Фурье, ряд Котельникова, степенные полиномы, гипергеометрические, полином Чебышева, полином Лежандра, функции Уолша, Хаара, Радемахера);

– принципам приближения (интерполяция, экстраполяция, комбинирование).

Для систем передачи и хранения информации справедливо определение кодирования в широком смысле.

2) кодирование сигналов

С учетом известного содержания *кодирования,* как процесса преобразования сообщения (информации) в комбинацию различных символов или соответствующих им сигналов, *кодирование сигналов* рассматривается как процесс представления сообщений в форме, удобной для передачи по каналу телекоммуникации и хранения в определенной среде.

Для непрерывных сообщений, представляемых числом знаков, стремящихся к бесконечности, практически невозможно создать сигнал передачи. Если непрерывные сообщения представить дискретными частями соответствующего алфавита в однозначно определенные моменты отсчета, то такое сообщение можно передать конечным числом сигналов, соответствующих буквам алфавита источника.

В рамках данной главы можно выделить следующие цели кодирования:

1) согласование свойств источника сообщений со свойствами канала связи (по Шеннону);

2) обеспечение заданной достоверности передачи или хранения информации путем внесения избыточности с учетом интенсивности и статистических закономерностей помехи в канале связи;

3) повышение помехоустойчивости, сообщений и сигналов за счёт защиты от искажений (обеспечение помехозащищенного кодирования, обеспечение логической и физической целостности информации);

4) сжатие входной информации.

5.5. Основные виды и способы обработки аналоговой и цифровой информации

Любая информация, отнесенная к области абстрактных категорий, может проявляться только в материально-энергетической форме и в ходе обработки представляется в формализованном виде *данных*. Рассмотрим вопросы, связанные с обработкой данных более подробно. Дадим определение основным понятиям, связанным с обработкой информации.

1) основные понятия обработки информации

Обработка информации:

- решение задач, связанных с преобразованием информации, независимо от их функционального назначения;
- процесс, который осуществляется при помощи устройств, машин, осуществляющих аналоговые или цифровые преобразования поступающих величин, функций.

Обработка связана с циклом обращения информации, включающем воспроизведение, подготовку, нормализацию, квантование, кодирование, модуляцию сигналов, рис. 5.6.

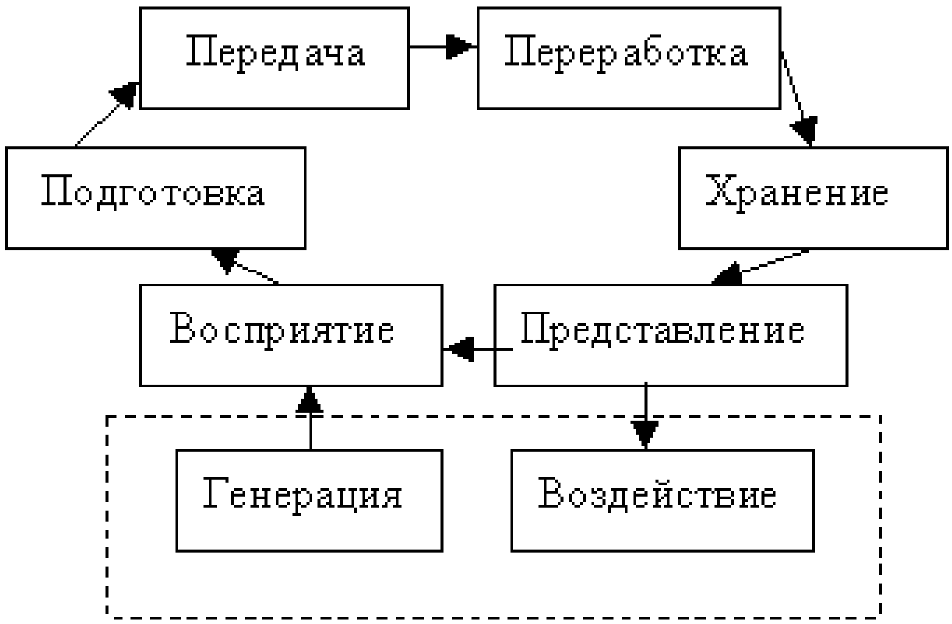


Рис. 5.6. Циклы обращения информации

Воспроизведение − формирование образа, опознавание, оценка объекта и отделение информации от шума, в результате которых образуется сигнал в форме удобной для передачи и преобразования.

Передача – перенос на расстояние с помощью различных сигналов (сигналов различной природы, в том числе механических, гидравлических, пневмонических, акустических, оптических, электрических, электромагнитных и других).

Представление − это обработка информации, связанная с человеком и представляет условное изображение форм сигналов для адекватного воздействия на человека.

Воздействие сигналов − это явление, событие при котором сигнал несущий информацию, производит регулирование и защиту действия и вызывает изменение в объекте.

Содержание рассмотренных циклов обращения информации позволяет увидеть их связь с содержанием и формой представления и структурой обрабатываемой информации.

2) основные виды и способы обработки данных

Данные – информация, представленная в формализованном виде, пригодном для автоматической обработки.

С учетом ранее рассмотренного содержания понятия «сообщение», можно утверждать, что в определенном смысле оно адекватно содержанию понятия «данные».

Виды данных: изображения, звук, аудио, видео, мультимедиа.

Обработка данных − процесс выполнения последовательности операций над данными.

В настоящее время выделяют следующие способы обработки данных: 1) обработка данных прикладными процессами; 2) распределенная обработка данных (выполняемая нескольким системами и процессами); 3) обработка данных в реальном масштабе времени; 4) фоновая обработка данных (без внешнего воздействия из вне); 5) обработка в режиме разделения времени (поочередно); 6) обработка документов: изображений, речи, сигналов, списков, текстов и других.

Обработка документов − процесс создания и преобразования документов, при котором решаются задачи обработки: подбор, структурирование информации, включение разнородной информации, передача информации и др.

Обработка изображения − процесс анализа, преобразования и интерпретации изображения.

Обработка речи − процесс анализа, преобразования и синтеза речи.

Обработка сигналов − процесс преобразования сигналов из аналоговой в цифровую и обратно, в том числе фильтрация, модуляция, демодуляция, разложение на гармоники, выделение из шума.

Обработка сигналов может осуществляться следующими методами − аналоговыми, цифровыми (дискретными). С учетом этого можно дать следующее определение обработки сигналов.

Обработка сигналов − анализ, упорядоченность элементов с использованием методов программирования процессов.

Обработка текстов − процесс ввода, хранения, редактирования, форматирования и печати текстов.

3) обработка аналоговой и цифровой информации

Систему обработки информации можно представить следующей структурой, рис. 5.7.

Источник сообщения − это устройство, осуществляющее выбор из некоторого множества возможных сообщений.

Сигнал – физический процесс, отображающий (несущий) сообщение. Сигнал является элементом некоторого множества, выбираемого с определенной вероятностью.

Множество, на котором задана вероятностная мера, называемая *ансамблем сигнала*.

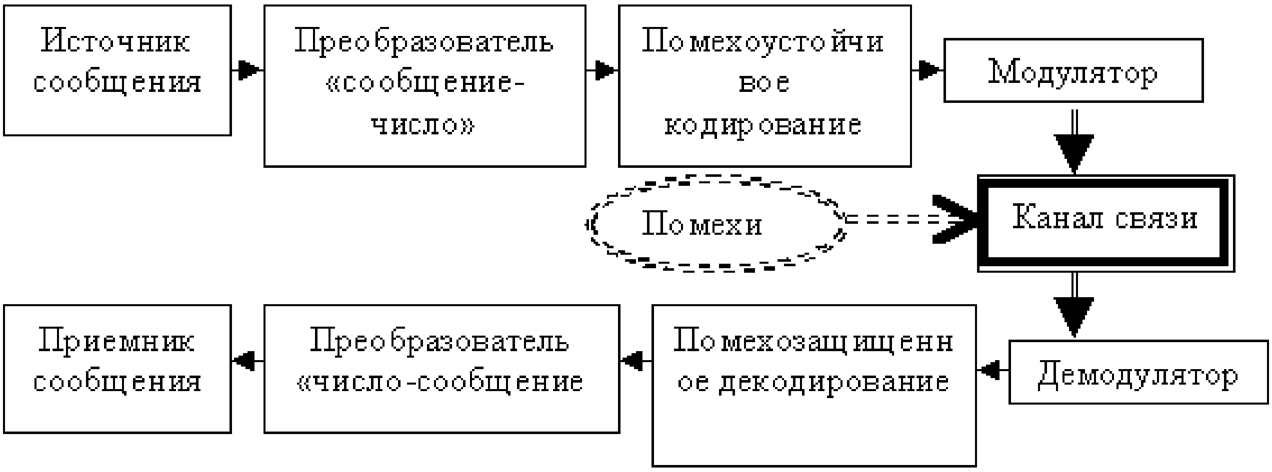


Рис. 5.7. Система обработки передачи информации

Сигналы, основанные на двоичной системе счисления, кодирующие символы алфавита группой двоичных символов.

Ансамбль сигналов (сообщений) может быть конечным (дискретным) или бесконечным. Ансамбль $\{X(t)\}$ функций t является случайным процессом. Каждая входящая функция $X_p(t)$ называется выборочной функцией или реализацией процесса. Наличие реализаций позволяет сигналу переносить информацию.

Канал связи (линия связи) – совокупность средств, включающих систему передачи информации, представляющую телефонный, телеграфный, канал передачи данных или другой канал и среду (воздух, металл, магнитную ленту и т.д.), обеспечивающих поступление сигналов от передающего устройства к приемному.

Передаваемые в канале связи сигналы могут быть подвержены искажению шумами. Сигналы на выходе линии связи могут отличаться от переданных вследствие затухания, искажения и воздействия помех, которые представляют любые мешающие возмущения, как внешние, так и внутренние, вызывающие отклонение принятых сигналов от переданных.

Источник информации – сообщение, состоящее из слов (букв и цифр), т.е. текст, записанный с помощью некоторого алфавита.

Кодирующее устройство – преобразователь сообщения в сигналы, удобные для передач по каналу связи (электрические сигналы, импульсы и т.п. с ограничением по мощности и полосе) и обладающие определенной помехозащищенностью.

Операцию восстановления сообщения по принятому сигналу называют *декодированием*.

Декодер восстанавливает комбинации кода и выдает информацию получателю.

Рассмотренная система обработки информации может быть представлена известной структурной схемой одноканальной системы передачи информации, рис. 5.8. В ней информация поступает в систему в форме сообщений.

Сообщение – совокупность знаков или первичных сигналов, содержащих информацию. Различают дискретные и непрерывные сообщения.

Источник сообщений образует совокупность источника информации ИИ и первичного преобразователя ПП (датчика, человека-оператора и т.д.), воспринимающего информацию о его состоянии или протекающем в нем процессе.

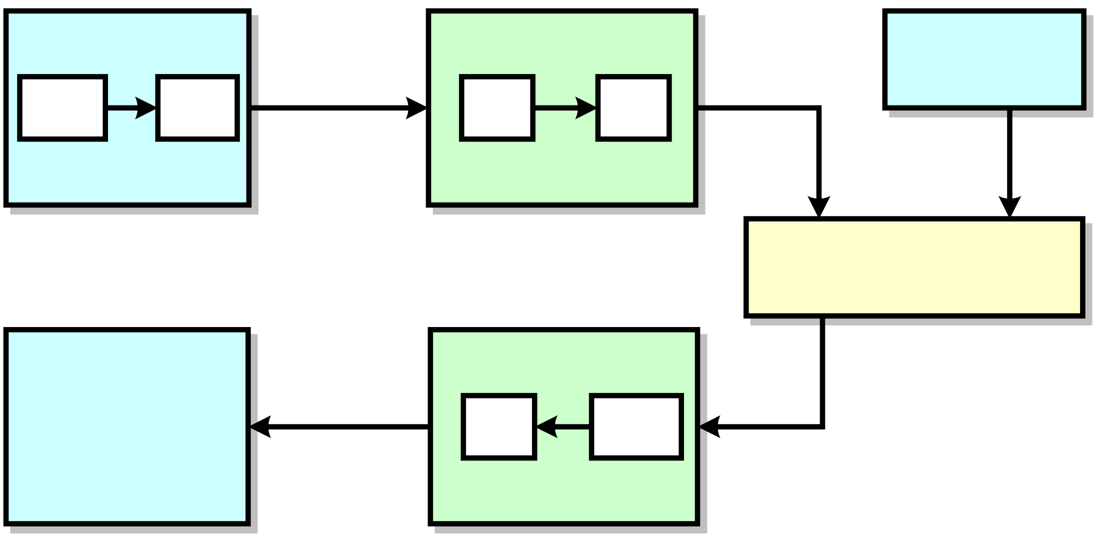


Рис. 5.8. Структурная схема одноканальной системы передачи информации

Дискретные сообщения формируются в результате последовательной выдачи источником отдельных элементов – *знаков*.

Множество различных знаков представляют *алфавит источника сообщения*, а число знаков – *объем алфавита*.

Непрерывные сообщения не разделены на элементы. Они описываются функциями времени, принимающими непрерывное множество значений (речь, телевизионное изображение).

Для передачи сообщения по каналу связи ему ставят в соответствие определенный сигнал.

Как правило, прибегают к операции представления исходных знаков в другом алфавите с меньшим числом знаков, называемых *символами*. При обозначении этой операции используется тот же термин «кодирование», рассматриваемый в узком смысле.

Устройство, выполняющее такую операцию, называют кодирующим или *кодером К*. Так как алфавит символов меньше алфавита знаков, то каждому знаку соответствует некоторая последовательность символов, которую называют *кодовой комбинацией*. Число символов в кодовой комбинации называют ее *знатностью*, число ненулевых символов – *весом*.

Для операции сопоставления символов со знаками исходного алфавита используют термин «декодирование». Техническая реализация ее осуществляется декодирующим устройством или *декодером ДК*.

Передающее устройство осуществляет преобразование непрерывных сообщений или знаков в сигналы, удобные для прохождения по линии связи. При этом один или несколько параметров выбранного носителя изменяют в соответствии с передаваемой информацией. Такой процесс называют *модуляцией*, и осуществляется модулятором *М*. Обратное преобразование сигналов в символы производится демодулятором *ДМ*.

Помехами называют любые мешающие возмущения, как внешние, так и внутренние, вызывающие отклонение принятых сигналов от переданных.

Из смеси сигнала и помехи *приемное устройство* выделяет сигнал и посредством декодера восстанавливает сообщение, которое в общем случае может отличаться от посланного. Меру соответствия принятого сообщения посланному называют *верностью передачи*.

Принятое сообщение с выхода системы связи поступает к абоненту-получателю, которому была адресована исходная информация.

Типовая структура обработки информации, может быть представлена рис. 5.9.

В тракте информация от источника кодируется (преобразуется) не избыточным, как правило, двоичным кодом, затем корректирующим, избыточным (помехоустойчивым) кодом, модулируется и передается по каналу связи.

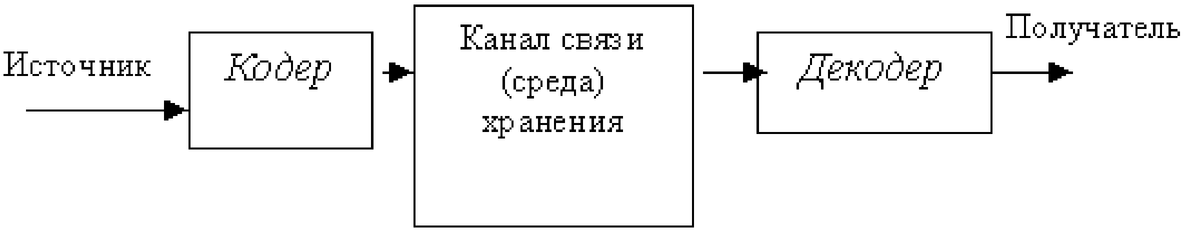


Рис. 5.9. Структурная схема типового тракта обработки информации

На приеме сигнал из канала связи демодулируется (преобразуется в цифровую форму – последовательность, как правило, двоичных сигналов), декодируется с исправлением ошибок и преобразуется в форму, удобную для восприятия потребителю.

Канал связи характеризуется определенной пропускной способностью. Источник информации обладает определенной скоростью. При этом, если скорость информации источника меньше пропускной способности канала связи, то возможно осуществление избыточное, помехозащищенное кодирование и декодирование информации с минимальной вероятностью ошибки. Помехозащищенное кодирование сигналов, передаваемых по каналам, связи обеспечивает защиту физической и логической целостности исходной информации от воздействия помех, которые делятся на флуктуационные, гармонические и импульсные.

Флуктуационная помеха – случайно изменяющиеся напряжение или ток в электронных элементах средств обработки информации.

К таким помехам (шумам) относятся: тепловые шумы средств обработки информации, изменения в среде распространения радиоволн и другие.

Гармоническая помеха – паразитное проникновение различных гармонических напряжений, колебаний генераторного оборудования средств обработки информации, промышленных электротехнических средств, умышленных помех в тракты обработки информации.

Импульсная помеха – помеха, пиковое значение которой соизмеримо или выше величины полезного сигнала.

Виды импульсных помех: помехи естественного происхождения (атмосферные разряды); промышленные (от искрообразующих элементов двигателей, электросварки, линий электропередач и т.п.); помехи, возникающие в средствах и элементах обработки информации.

Флуктуационные и гармонические помехи действуют обычно непрерывно во времени, а импульсные – в течение короткого промежутка времени, и являются сосредоточенными во времени.

В процессе действия помех информация в канале связи искажается и отличается от переданной. Возникновение ошибок – процесс случайный и его предсказание может быть осуществлено на основе статистических данных.

При этом вероятность ошибок может не зависеть от значения передаваемого символа 0 или 1, т.е. среди неправильных символов одинаково часто будут встречаться как искаженные 0, так и 1. Модель такого канала называется симметричным двоичным каналом, рис. 5.10.

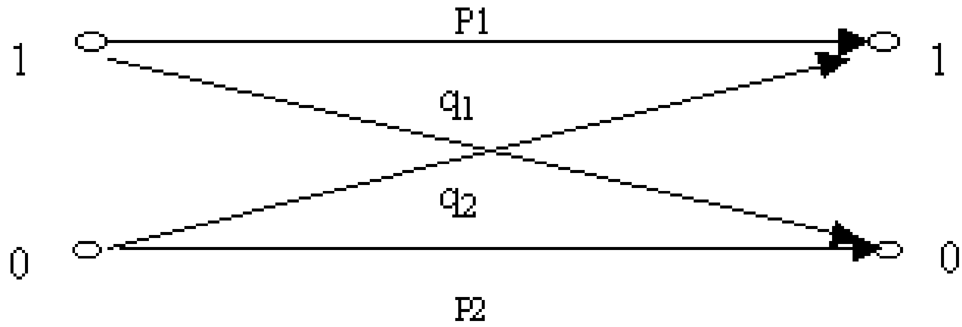


Рис. 5.10. Модель симметричного двоичного канала

В рассматриваемом канале $P_1 = P_2$ – вероятности неискаженной передачи, $q_1 = 1 - P_1$, $q_2 = 1 - P_2$ – вероятности трансформации (искажения) символов.

Существуют другие модели каналов. Так, если условия передачи ряда последовательностей символов не зависят один от другого, то такой канал называется «каналом без памяти». Если условия передачи символа зависят от предыдущих состояний, то такой канал называется «каналом с памятью».

При передаче сигналов в канале связи ошибки символов могут группируются в пачки (пакеты) – последовательности символов, начинающийся и заканчивающийся ошибочно принятыми символами. Внутри пачки могут быть и правильно принятые символы.

При взаимонезависимых ошибках, вероятность искажения каждого символа P_e постоянная. Вероятность безошибочного приема одного символа – ($1\text{-}P_e$), а вероятность распределения ошибок при передаче последовательности из n символов равна:

$$P = \sum_{i=0}^{n} C_n^{\,i} P_e^{\,i} (1\text{-}P_e)^{n-i}, \tag{5.4}$$

где i – кратность ошибок в последовательности n, $C_n^{\,i}$ – число сочетаний из n по i, $C_n^{\,i} = n!/i!(n\text{-}i)!$.

В реальных каналах связи ошибки, как правило, зависимы, и приведенное выражение является приближенным.

Скорость передачи последовательности дискретных сообщений длительностью Т по каналу связи без помех определяется выражением (5.5):

$$\lim_{T\to\infty} I/T = V\ бит/с, \tag{5.5}$$

Тогда пропускная способность канала связи определяется предельным значением скорости и равна:

$$C = V_{max} = \lim_{T\to\infty} I_{max}/T\ \ бит/с. \tag{5.6}$$

Пропускная способность канала – максимальная скорость передачи информации на символ в единицу времени при условии, что канал связи без помех согласован с источником информации:

Для максимального количества информации в сообщении при равной вероятности состояний источника информации пропускная способность канала определяется выражением (5.7):

$$C = V_{max} = \lim_{T\to\infty} n\log_2 m/T\ бит/с. \tag{5.7}$$

Отсюда, скорость передачи сообщений в канале связи зависит от статистических свойств сообщения и параметров канала связи, которые определяются соотношением полезного сигнала к помехе.

Для эффективного использования канала связи необходимо, чтобы скорость источника приближалась к пропускной способности канала ($V \to C$). При $V>C$ – информация будет передана с искажениями. При обратном неравенстве – без искажений.

Неискаженная передача сигналов по каналу связи осуществляется с учетом следующих положений теории передачи К. Шеннона:

1. Если скорость передачи сообщений, достаточно близка к пропускной способности канала C, то всегда существует способ кодиро-

вания, который обеспечит передачу сообщений со скоростью близкой к пропускной способности канала при малой вероятности ошибки;

2. Если к входу канала подключен источник сообщений с энтропией символа, равной пропускной способности канала связи, то считается, что источник согласован с каналом;

3. Если энтропия источника меньше пропускной способности канала (при неравновероятности состояний источника сообщений), то источник не согласован с каналом связи (не полная загрузка канала);

4. Для согласования значения энтропии источника с пропускной способностью канала связи осуществляется на основе статистического кодирования.

Пример. Содержание статистического кодирования демонстрируется примером двух последовательностей сообщений:

1 - 1111000100110100000111011001О1

2 - 000000010000000001100000000000

Для независимого появления символов 1 и 0, в первой последовательности их вероятности равны, $P_1 = P_0 = 0{,}5$, а для второй – $P_1 = 0{,}1$, $P_0 = 0{,}9$. С учетом этого, энтропия первой последовательности равна $H_1 = -2\times0{,}5\times\log_2 0{,}5 = 1$, а второй – $H_2 = -0{,}1\times\log_2 0{,}1 - 0{,}9\times\log_2 0{,}9 \approx 0{,}5$. Отсюда, количество информации на символ во второй последовательности в два раза меньше, чем в первой. Для бинарного канала связи с пропускной способностью $C=\log_2 m$ дв.ед./символ первая последовательность будет согласована с каналом ($H = 1$), а при передаче второй последовательности пропускная способность двоичного канала на символ в два раза больше энтропии источника, т.е. канал недогружен и несогласован с источником ($C>H_2$). Статистическое кодирование позволяет повысить энтропию передаваемых сообщений.

Для канала связи с помехами, сообщения могут искажаться и на приеме, что уменьшает достоверность передаваемой информации. При этом вероятность безошибочного приема будет меньше 1. Количество принимаемой информации, I, уменьшается на величину полученной неопределенности, вносимой помехами и равно:

$$I = H(i) - Hi(j), \tag{5.8}$$

где $H(i)$ – энтропия источника сообщений, $H_i(j)$ – энтропия принятого сообщения.

Скорость передачи информации по каналу связи с помехами определяется выражением (5.9):

$$V' = \lim_{T\to\infty} I/T = \lim_{T\to\infty} (H_i - H_j(i))/T. \tag{5.9}$$

С учетом рассмотренных характеристик источника информации и канала связи, *кодирование информации,* по *Шеннону* представляется следующим содержанием: если энтропия источника информации H_i не превышает пропускной способности канала ($H_i \leq C$), то существует *код,* обеспечивающий передачу информации через канал с помехами со сколь угодно малой частотой ошибок или сколь угодно малой недостоверностью.

Для канала связи с финитной (конечной) полосой пропускания и ограниченной средней мощностью аналогового сигнала пропускная способность равна (5.10):

$$C = F_m log_2(1+W_c/W_ш), \qquad (5.10)$$

где F_m – полоса пропускания канала связи (Гц); W_c – средняя мощность сигнала; $W_ш$ – средняя мощность помех с нормальным законом распределения амплитуд в полосе частот канала. При $W_c >> W_ш$, пропускная способность канала будет равна максимальной скорости передачи информации (5.11):

$$V_{max} = C = F_m log_2 W_c/W_ш. \qquad (5.11)$$

Таким образом, можно заключить, что для передачи информации по каналу связи с помехами с максимальной скоростью, приближающейся к пропускной способности канала связи, используются оптимальные методы избыточного кодирования источника сообщений.

5.6. Модуляция сигналов, принципы построения, работы и характеристики устройств обработки данных

Рассмотрение содержания вопросов модуляции сигналов будет осуществляться в рамках рассмотрения очередного этапа обработки информации в типовом тракте. Рассмотрим вопросы, связанные с модуляцией и кодированием.

1) модуляция сигнала

Модуляция сигнала (носителя) – это процесс:

- нанесения информации на материальный носитель путем изменения некоторых носителей параметров физических процессов, состояний, соединений, комбинаций элементов;

- изменения параметров физических процессов комбинаций или импульсных последовательностей;
- изменения, какого-либо одного или нескольких параметров носителей информации (информационных параметров).

Демодуляция − обратная операция восстановления величин, вызвавших изменение параметров при модуляции.

С учетом содержания понятия модуляции и известных характеристик сигналов, выделяют следующие виды модуляции: прямая, амплитудная (АМ), частотная (ЧМ), фазовая (ФМ), частотно-импульсная (ЧИМ), амплитудно-импульсная (АИМ), временно-импульсная (ВИМ), счетно-импульсная (СИМ), импульсно-кодовая (ИКМ) и др.

Аналогово-цифровой преобразователь, расположенный в передатчике, также называют *кодером* (encoder, coder).

Цифро-аналоговый преобразователь, расположенный в приемнике, называют *декодером* (decoder).

Слово *кодек* образовано из сочетания слов «кодер/декодер».

В добавление к вышеперечисленным основным методам преобразования существуют и более сложные кодеки, которые в зависимости о методов, положенных в основу их разработки, обозначаются:

ADM Adaptive DM – Адаптивная ДМ;

ADPCM Adaptive DPCM – Адаптивная ДИКМ (АДИКМ);

PCM Adaptive PCM – Адаптивная ИКМ;

CDM Continuous DM – Непрерывная ДМ;

DCDM Digitally controlled DM – Управляемая цифровым способом ДМ;

LDM Linear (nonadaptive) DM – Линейная (неадаптивная) ДМ;

LPC Linear predictive codec(s) – Кодек (кодеки) с линейным предсказанием;

CELP Code excited linear – Возбуждаемое кодом кодирование;

predictive coding – с линейным предсказанием;

RELP – Residual excited vocoders – Вокодеры, возбуждаемые остаточным сигналом;

VQ Vector quantization, subband coding, vocoder (s) – Вокодер (вокодеры) с векторным квантованием и субполосным кодированием.

Рассмотрим более подробно основные методы преобразования сигналов.

2) импульсно-кодовая модуляция

Импульсно-кодовая модуляция сигнала электросвязи – преобразование сигнала, при котором сигнал путем дискретизации, квантования отсчётов этого сигнала и их кодированием преобразуется в цифровой сигнал электросвязи.

Это наиболее распространённый вид аналого-цифрового преобразования (АЦП) сигнала. Структурная схема АЦП, представляющая кодек ИКМ, показана на рис. 5.11.

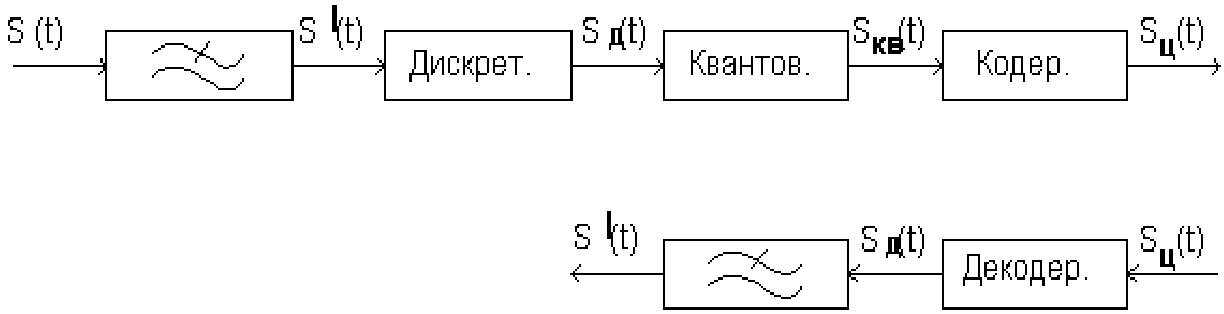


Рис. 5.11. Структурная схема ИКМ кодека

Фильтр нижних частот (ФНЧ) ограничивает спектр аналогового сигнала, обеспечивая выполнение условий теоремы В.А. Котельникова (5.10), (5.11). Так, для речевого сигнала ФНЧ имеет верхнюю частоту среза, равную 3,4 кГц. Для обеспечения требования к затуханию фильтров трактов передачи и приёма, частота дискретизации для речевых сигналов, согласно международным рекомендациям (МККТТ G.711 и G.712), равна 8 кГц. При дискретизации образуется сигнал АИМ первого рода, амплитуда импульса которого на интервале длительности импульса изменяется по закону непрерывного сигнала.

Для уменьшения погрешности амплитудного квантования значение отсчёта в процессе квантования должно оставаться постоянным. Поэтому при импульсно-кодовой модуляции сигналы АИМ-1 преобразуются в сигналы АИМ-2. Для этого аналоговый АИМ сигнал $S_\partial(t)$ поступает на амплитудный квантователь.

Выбор числа уровней равномерного квантования осуществляется с учетом диапазона изменения входных сигналов. Так, динамический уровень телефонного сигнала является случайной величиной, которая зависит от времени передачи и от источника (различных абонентов).

Известно, что динамические уровни телефонных сигналов подчинены нормальному закону с плотностью распределения

$$W(p_\text{c}) = \frac{1}{\sigma_\text{c}\sqrt{2\pi}} e^{-\frac{(p_\text{c}-p_\text{o})^2}{2\sigma_\text{c}^2}}, \qquad (5.12)$$

где p_o – среднее значение динамического уровня телефонного сигнала (дБ), σ_c – его среднеквадратичное отклонение (дБ).

Принято, чтобы при минимальном уровне средней мощности речевого сигнала защищённость от шумов квантования была не менее 22÷25 дБ, тогда требуемое число разрядов равно $n = 10 \div 11$, которое соответствует числу уровней квантования $N = 1024 \div 2048$.

Использование равномерного (линейного) квантования для передачи телефонных сигналов не является оптимальным по следующим причинам:

– распределение амплитуд телефонного сигнала не является равномерным: малые амплитуды сигнала более вероятны, чем большие. В этом случае лучше сделать ошибку квантования меньшей для более вероятных амплитуд за счёт увеличения ошибки для менее вероятных амплитуд;

– телефонные сигналы имеют широкий динамический диапазон (порядка 40 ÷ 45 дБ), в пределах которого должно обеспечиваться примерно постоянное соотношение сигнал-шум квантования. При равномерном квантовании указанное соотношение для слабых сигналов будет практически на 40 дБ хуже, чем для сильных;

– число разрядов, а следовательно, и тактовая частота цифрового сигнала $F_m = n \cdot F_д$, должны иметь большие значения, что потребует увеличения пропускной способности линии связи.

Для выравнивания величины $a_{ш\ кв}$ при изменении уровня сигнала в широких пределах, а также уменьшения разрядности кода применяют неравномерное квантование, при котором шаг квантования имеет минимальное значение для слабых сигналов и увеличивается с увеличением уровня сигнала.

Неравномерное квантование может быть осуществлено различными способами:

– сжатием: динамического диапазона сигнала перед равномерным квантованием с последующим компенсирующим расширением его после декодирования (аналоговое компандирование);

– непосредственно в кодирующем устройстве, то есть посредством использования нелинейного кодирования;

– с помощью соответствующего цифрового преобразования сигнала, формируемого на выходе линейного кодера (цифровое компандирование).

Компандирование – совокупность операций сжатия динамического диапазона компрессором и расширения его экспандером.

Недостатком аналогового компандирования является то, что очень сложно получить с большой точностью взаимообратные амплитудные характеристики компрессора и экспандера, вследствие чего нелинейность суммарной амплитудной характеристики приводит к нелинейным искажениям передаваемых сигналов.

На практике применяют два последних способа, так как нелинейные кодеки и цифровые компандеры обеспечивают лучшую стабильность характеристики канала и лучше его параметры при несогласованной работе.

В настоящее время применяются два приблизительно равноценных закона компандирования: A и μ. В Европе и странах СНГ, где применяется аппаратура ИКМ-30, принят A – закон компандирования.

Применение неравномерного квантования позволяет повысить помехозащищённость слабых сигналов на 26-33 дБ и снизить при прочих равных условиях число разрядов в кодовой группе до восьми.

Таким образом, в системах телефонии сигналы речи, факсимиле или другие сигналы данных, модулированные в полосе тональных частот, ограничены верхней частотой f_m = 3,4 кГц. Для преобразования такого аналогового сигнала в цифровой ИКМ поток данных применяется дискретизация с частотой f_s = 8000 отсчетов в секунду. Каждый отсчет фиксируется на одном из 256 уровней квантования. Количество уровней квантования составляет 8 информационных битов (2^8 = 256). Исходя из этого, канал тональной частоты, дискретизированный с частотой 8000 отсчетов в секунду и требующий 8 битов на отсчет, имеет скорость передачи 64 кбит/с.

Аналоговые сигналы *цветного телевидения* вещательного качества имеют ширину полосы видеочастот около 5 МГц. Для обычного ИКМ кодирования этих видеосигналов используется частота дискретизации f_s = 10^6 отсчетов в секунду и применяется схема кодирования с 9 битами на отсчет. Таким образом, результирующая скорость передачи составляет 90 Мбит/с. Большинство телевизионных изображений сильно коррелированны, и это может быть использовано для *снижения* скорости передачи. Можно предсказать цвет и яркость любого элемента изображения, основываясь на значениях их параметров для соседних элементов, которые уже имели место.

В представленной библиографии описываются методы цифровой обработки сигналов (ЦОС) с применением техники предсказания для цифрового цветного телевидения вещательного качества, требующие большие скорости передачи (от 10 до 45 Мбит/с). Для радиопередачи видеоконференций используются сжатые сигналы изображения со скоростью от 20 до 200 кбит/с.

3) особенности дифференциальной ИКМ

Дифференциальная импульсно-кодовая модуляция (ДИКМ) сигнала – это импульсно-кодовая модуляция сигнала, при которой в цифровой сигнал преобразуется разность между текущими и предсказанными значениями сигнала.

Предсказанное значение сигнала – это ожидаемое значение сигнала, полученное экстраполяцией функции времени, описывающей изменение представляющего параметра сигнала.

Классификация разновидностей ДИКМ приведена на рис. 5.12, а принцип формирования этого цифрового сигнала – на рис. 5.13.

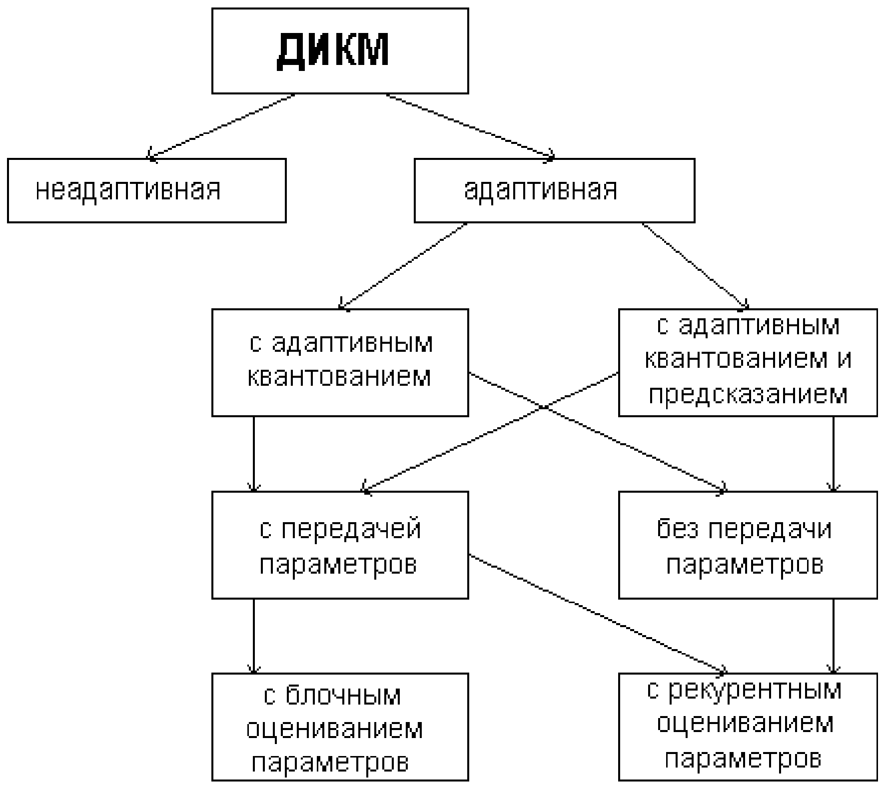


Рис. 5.12. Классификация разновидностей ДИКМ

Квантование разности позволяет уменьшить число уровней квантования разностей отсчётов по сравнению с необходимым числом уровней при квантовании самих отсчётов. Это позволяет уменьшить разрядность кода, и, следовательно, число информационных символов, передаваемых по тракту связи в единицу времени.

Известны два варианта структурных схем ДИКМ:

1) с формированием разностного сигнала в аналоговой форме;

2) формированием разностного сигнала в цифровой форме.

Простейшая структурная схема, в которой разностный сигнал получается в аналоговой форме, приведена на рис. 5.14 (*а*).

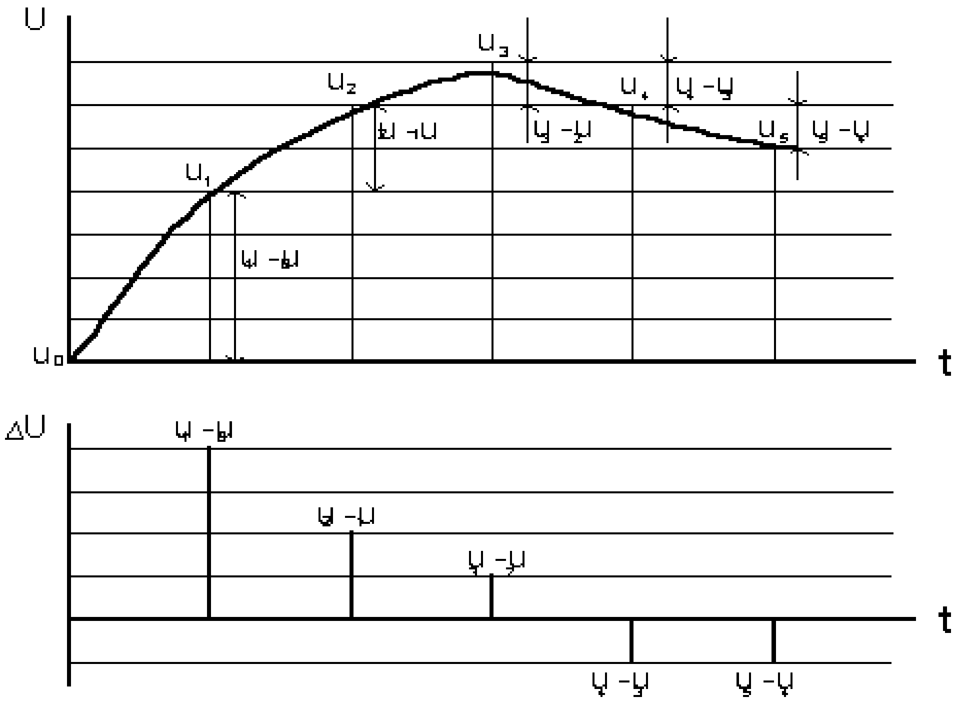


Рис. 5.13. Принцип формирования цифрового сигнала при ДИКМ

В этом варианте входной аналоговый сигнал задерживается на время, равное периоду дискретизации и вычитается из исходного незадержанного сигнала. Полученный разностный сигнал подвергается обычному ИКМ преобразованию и передаётся по тракту связи. На приёмной стороне после декодирования восстанавливается последовательность квантованных разностей, из которых в результате последовательного суммирования восстанавливается последовательность квантованных отсчётов. ФНЧ восстанавливает исходный аналоговый сигнал.

На рис. 5.14 (*б*) приведена схема формирования сигнала ДИКМ, в которой разностный сигнал получается после обработки цифровых сигналов. Здесь входной аналоговый сигнал, как и при ИКМ, преобразуется в цифровую форму. Последовательность кодовых комбинаций поступает на схему арифметического вычитания непосредственно и через устройство задержки, задерживающее полученный цифровой сигнал на время, равное дискретизации. При этом первым подаётся старший разряд кодового слова, а последним – младший, в результате образуется разность кодированных отсчётов, представляющая сигнал ДИКМ.

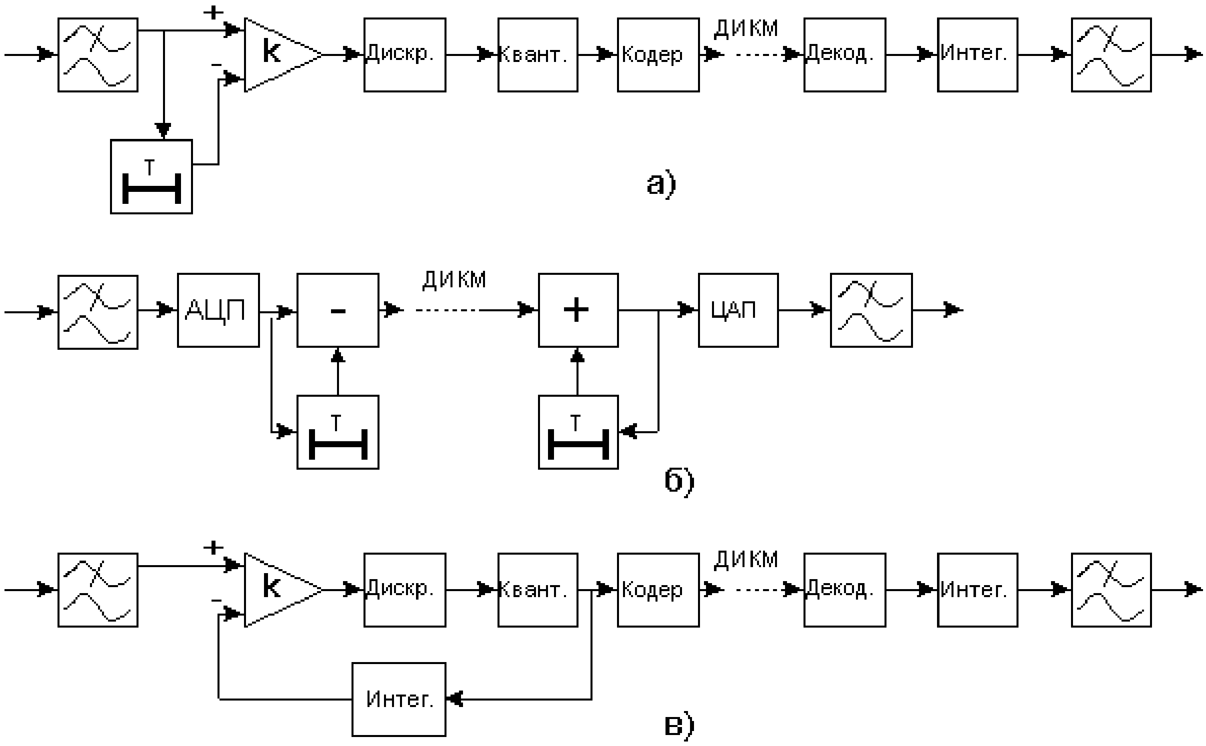


Рис. 5.14. Схемы формирования сигнала ДИКМ

На приеме изменяется порядок следования разрядов в кодовых группах, то есть, сначала получают младший разряд, а последним – старший. Затем полученный цифровой сигнал подаётся на цифровой сумматор, на второй вход которого подаётся этот же сигнал задержанный на период дискретизации. На выходе арифметического сумматора получаются кодовые комбинации, соответствующие квантованным отсчетам сигнала.

В рассмотренных схемах формирования сигналов ДИКМ возникают большие, чем при ИКМ искажения, связанные с процессом формирования цифрового сигнала. Так, в схеме рис. 5.14 (а), появляются специфические нелинейные искажения, которые возникают из-за различия крутизны сигнала на передаче и приёме, что приводит к различию форм переданного и принятого сигнала.

При получении сигнала ДИКМ (рис. 5.14 (б)), погрешность квантования разности соседних отсчётов равна двум шагам квантования, так как при квантовании каждого из отсчётов (задержанного и не задержанного) максимальная погрешность равна одному шагу квантования, что в первом приближении увеличивает уровень мощности шумов квантования (примерно на 6 дБ).

В рассмотренных схемах искажения возникают непосредственно в процессе формирования цифрового сигнала, передаваемого по тракту связи. Эти искажения можно уменьшить, если непрерывно сравнивать исходный сигнал и сигнал с выхода модулятора с тем, чтобы минимизировать искажения, связанные с дискретизацией и квантованием. Для этого необходимо иметь на стороне передачи демодулятор. Так, в схеме, приведённой на рис. 5.14 (в), оценивается не разность между соседними отсчётами передаваемого сигнала, а разность между значением данного отсчёта и квантованным значением предыдущего отсчёта, полученного из цифрового сигнала, поступившего в тракт передачи. Разность между этими сигналами дискретизируется, квантуется и кодируется для последующей передачи по цифровому тракту передачи. В этой схеме происходит как бы отслеживание за передаваемым сигналом.

Ещё больший эффект дают дифференциальные кодеки, в основе которых лежат схемы предикативных квантователей.

Предикативное квантование – квантование ошибки предсказания, то есть разности между исходными и предсказанными значениями сигнала.

Принцип кодирования с предсказанием состоит в следующем. Как и для ИКМ, в каждой точке опробования формируется значение отсчёта (квантованное и неквантованное) исходного аналогового сигнала. На основе вероятностных свойств и конкретных значений в нескольких предыдущих моментах опробования этого аналогового сигнала осуществляется предсказание, которое заключается в выработке наиболее вероятного значения отсчёта в данной точке опробования. Сравнивая это предсказанное значение с действительным значением отсчёта сигнала, можно оценить насколько предсказание оказалось правильным и получить ошибку (поправку) предсказания. Передавая по тракту связи эту поправку (вместо полного значения отсчёта) и осуществляя на приёме аналогичное предсказание, восстанавливают значение выборки аналогового сигнала.

Система ДИКМ с адаптацией предсказателя и квантователя называется адаптивной ДИКМ (АДИКМ). В настоящее время существуют международные рекомендации G.721, определяющие алгоритм АДИКМ на скорость передачи 32 кбит/с.

Преимущества ДИКМ по отношению к ИКМ заключаются в том, что соседние отсчёты дискретизируемого аналогового сигнала с большой вероятностью мало отличаются друг от друга, что особенно характерно для речевых сигналов. Это свойство и даёт возможность уменьшить разрядность кода, отображающего передаваемые разности отсчётов. Следовательно, если период дискретизации Тд взять меньше периода дискретизации, определенного В.А. Котельниковым (Тд = 1/2Fв), то различие между соседними отсчётами аналогового сигнала будет меньше, а значит, при использовании ДИКМ будет и меньше разрядность кода. Отсюда следует, что, уменьшая период дискретизации, можно добиться, чтобы разность между соседними отсчётами стала достаточно малой, а именно такой, чтобы её можно было передавать либо символом «1», если разность двух соседних отсчётов больше выбранного шага квантования, либо символом «0», если разность меньше шага квантования. В этом случае передача сигнала может осуществляться одноразрядным кодом, при котором можно передавать сведения только о знаке приращения. Такой метод формирования цифрового сигнала называется классической линейной дельта-модуляцией, в отличие от других её разновидностей.

4) дельта-модуляция

Дельта-модуляция (ДМ) является частным случаем ДИКМ, когда квантование остатка предсказания осуществляется на два уровня, а частота дискретизации равна скорости передачи.

Дельта-модуляция – это дифференциальная импульсно-кодовая модуляция сигнала, при которой разность между текущими и предсказанными значениями этого сигнала квантуются с использованием только двух уровней квантования сигналов.

Структурная схема кодера линейной дельта-модуляции приведена на рис. 5.15. Кодер состоит из $ФНЧ_{пер}$, компаратора (К), триггера задержки на такт (Т), преобразователя полярности (ПП) и интегратора (Инт).

Принцип работы кодера. Через $ФНЧ_{пер}$ входной непрерывный сигнал $U(t)$ поступает на компаратор, где сравнивается с аппроксимирующим сигналом $U^*(t)$. В результате сравнения, в каждый момент входного и аппроксимирующего напряжений на выходе компаратора формируется двоичное напряжение, логический уровень которого зависит от знака разности между напряжениями $U(t)$ и $U^*(t)$. Если в анализируемый момент $U(t) > U^*(t)$, то на выходе компаратора устанавливается потенциал, соответствующий уровню логической «1». Если $U(t) < U^*(t)$, то на выходе компаратора формируется потенциал, соответствующий уровню логического "0". С приходом на синхронизирующий вход D-триггера тактового импульса двоичный сигнал с выхода компаратора записывается в триггер T, где хранится в течение одного такта. В результате формируется выходной дельта- модулированный сигнал $U_{дм}(t)$, который подается в линию и ответвляется в цепь обратной связи на преобразователь полярности.

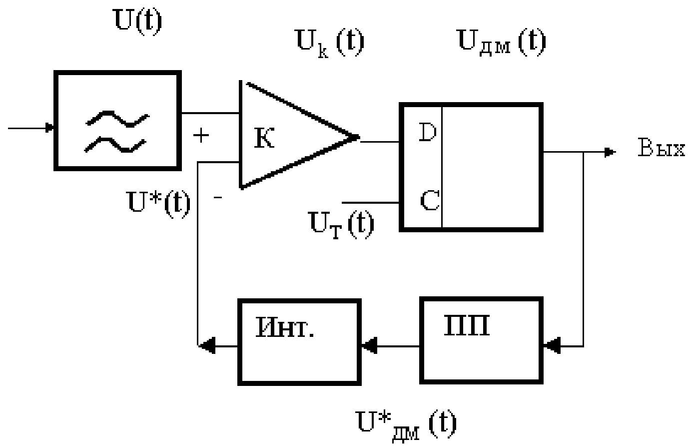


Рис. 5.15. Структурная схема кодера линейной дельта-модуляции

Преобразователь полярности преобразует однополярный ДМ-сигнал $U_{дм}(t)$ в двухполярный сигнал $U_{дм}{}^*(t)$ с постоянными уровнями напряжения $+V$, управляющими интегратором. ПП не изменяет информационного содержания импульсной последовательности ДМ-сигнала и необходим только для исключения из него постоянной составляющей, так как на вход кодера воздействует сигнал, как правило, не содержащий постоянной составляющей, например, речевой. В результате воздействия на интегратор двухполярной последовательности импульсов напряжения $U_{дм}{}^*(\mathrm{t})$ на его выходе формируется аппроксимирующее напряжение $U^*(t)$. За время одного тактового интервала T_m аппроксимирующее напряжение изменяется (увеличивается или уменьшается) на величину Δ, которая называется шагом квантования.

Декодер линейной дельта-модуляции (рис. 5.16) содержит все функциональные узлы обратной связи кодера, а также фильтр нижних частот ФНЧ$_{пр}$, аналогичный ФНЧ$_{пер}$.

Декодером на приеме служит интегратор, после которого аппроксимирующие напряжение $U^*(t)$ (рис. 5.14. в) сглаживается фильтром ФНЧ$_{пр}$ и таким образом восстанавливается переданный аналоговый сигнал $\hat{U}(t)$. При резком изменении крутизны входного сигнала аппроксирующее напряжение $U^*(t)$ с одинаковым шагом квантования не успевает отслеживать изменение входного сигнала. На этих участках возникают шумы перегрузки по крутизне. При их отсутствии искажения обусловлены только процессом квантования. Ошибки квантования часто называют искажениями *(шумами) дробления.*

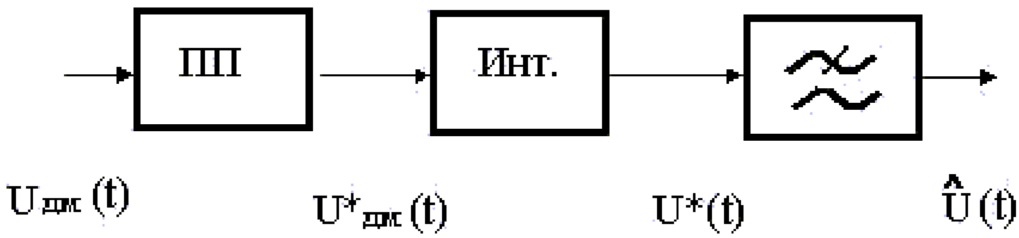


Рис. 5.16. Структурная схема декодера линейной дельта-модуляции

Принципиальное отличие ДМ от ИКМ состоит в том, что при ДМ имеют место шумы перегрузки по крутизне (имеет место ограничение допустимой скорости изменения исходного сигнала), а при ИКМ имеют место шумы ограничения по амплитуде (необходимо предусматривать амплитудное ограничение сигнала).

Структурная схема дельта-кодека с инерционной компрессией включает классический дельта-модулятор с одинарным интегратором, к которому дополнительно подключены амплитудно-импульсный модулятор (М), анализатор пачек символов (АПС), слоговый интегратор (Сл. инт.) и сумматор. Принцип работы схемы: амплитуда импульсов на входе интегратора изменяется в соответствии со структурой цифрового потока, получающейся на выходе дельта-модулятора. С увеличением уровня или частоты входного сигнала $U(t)$ увеличивается число элементов в пачках символов дельта-2.

Тактовая частота в системе передачи с ДМ может быть существенно уменьшена, если во избежание перегрузки шаг квантования менять с изменением крутизны аналогового сигнала или при постоянном шуме квантования уменьшить динамический диапазон сигнала. Это обеспечивается применением ДМ с компандированием.

Виды методов компандирования: аналоговые, цифровые, мгновенные и инерционные (слоговые).

Аналоговое компандирование непрерывных сигналов в дельта-кодеках применяют в исключительных случаях, так как очень сложно обеспечить идентичность характеристик компрессора и экспандера. Его применяют в тех системах, где не требуются каналы высокого качества (для организации каналов служебной связи). Наибольшее применение в современных цифровых системах передачи получило цифровое компандирование.

При мгновенном компандировании сигнал управления шагом квантования формируется и изменяется в течение одного или нескольких тактов, что позволяет с большой скоростью отслеживать изменение входного сигнала. Поэтому мгновенное компандирование применяют при аналого-цифровых преобразованиях широкополосных сигналов, например, телевизионных.

При преобразовании речевых или медленно изменяющихся сигналов применяют, как правило, инерционное компандирование. В этом случае время формирования сигнала управления выбирается исходя из интервала корреляции речевого сигнала, который не превышает средней длительности слога (примерно 2-3 мс). Поэтому постоянная времени слогового фильтра (слогового интегратора) должна быть равной 10-15 мс, причем даже при значительном разбросе значений постоянной времени (от 5 до 40 мс) качественные характеристики восстановленного декодером сигнала остаются практически как у модулированного сигнала $U_{дм}(t)$ (то есть возрастает их плотность).

Применение компандирования приводит к уменьшению динамического диапазона изменения входного сигнала при постоянном шаге квантования для линейной ДМ. Уменьшение зависит от количества разрядов регистра в анализаторе пачек символов. Применение компандирования позволяет снизить скорость передачи цифрового сигнала по сравнению с ЛДМ в 5 ... 15 раз, в зависимости от динамического диапазона изменения входного сигнала.

5) сравнительная оценка различных методов обработки сигналов

Зависимость соотношения сигнал-шум, представляющего минимальную защищенность от шумов квантования, от значения тактовой частоты информационного сигнала для дельта-модуляции (ДМ) на пороге перегрузки и ИКМ, показана на рис. 5.17, а зависимость соотношения сигнал-шум от уровня входного сигнала для различных видов модуляции показана на рис. 5.18.

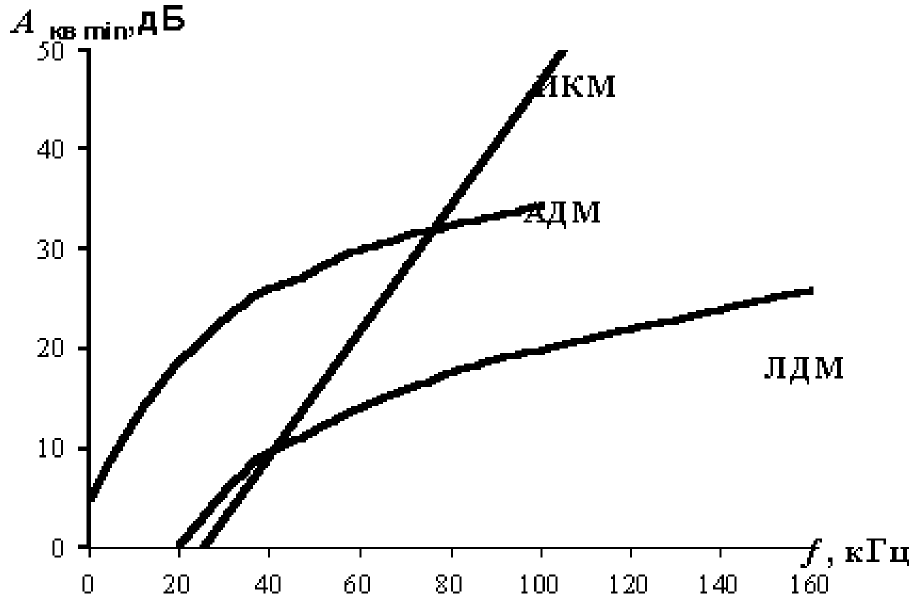


Рис. 5.17. Зависимость минимальной защищенности по шумам квантования от тактовой частоты информационного сигнала для различных видов модуляции

При ИКМ одно и то же увеличение отношения сигнал-шум квантования приводит к меньшему возрастанию скорости передачи, чем при дельта-модуляции. При низких скоростях передачи (до 32 кГц) ДМ обеспечивает лучшее, чем ИКМ отношение $P_c/P_{кв}$. В отдельных случаях для передачи телефонных сигналов целесообразно использовать дельта-модуляцию. При кодировании только одного телефонного сигнала дельта-кодер обладает таким важным преимуществом как отсутствие цикловой синхронизации, что значительно упрощает оборудование. АДИКМ по сравнению с ИКМ позволяет снизить скорость передачи до 24-32 кбит/с без заметного ухудшения качества звучания сигнала. Качество передачи при этом слабо зависит от конкретного источника информации.

Из рис. 5.18 видно, что соотношения сигнал-шум для АДИКМ-32 лучше, чем для АДМ-48 при слабом уровне входного сигнала ($P_{вх} < -28$ *дБ*).

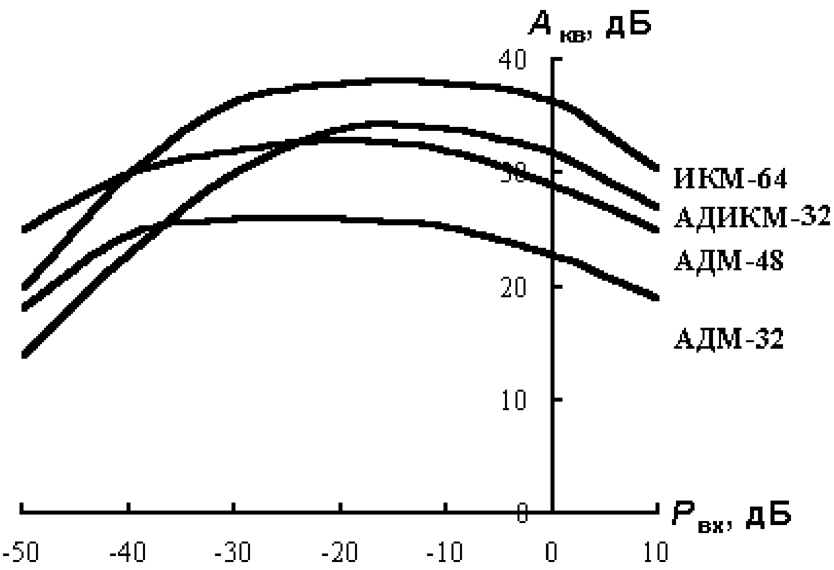


Рис. 5.18. Зависимость защищенности по шумам квантования от уровня входного информационного сигнала для различных видов модуляции

Внедрение АДИКМ в современные цифровые системы передачи стало возможным с появлением быстродействующих цифровых БИС.

Вывод: применение дифференциальных методов кодирования при аналого-цифровом преобразовании позволяет снизить скорость передаваемых цифровых сигналов при сохранении примерно такого же качества передачи информации, как и при импульсно-кодовой модуляции. При передаче сигналов, верхняя частота которых относительно низка, дифференциальные методы обеспечивают большую величину защищенности от шумов квантования.

6) цифро-аналоговое цифро-цифровое преобразование сигналов

Цифровые данные, представляющие информационные биты физически передаются в виде аналоговых или цифровых электрических сигналов. При работе с аналоговыми сигналами для передачи закодированных данных используется аналоговый несущий сигнал синусоидальной формы, а при работе с цифровыми сигналами – двухуровневый дискретный сигнал.

Аналоговые сигналы менее чувствительны к искажению, обусловленному затуханием в передающей среде, однако кодирование и декодирование данных проще осуществляется для цифровых сигналов. Здесь можно говорить об аналоговом и цифровом кодировании сигналов.

Цифро-аналоговое преобразование сигналов (аналоговое кодирование) применяется при передаче цифровых данных по телефонным (аналоговым) линиям связи, преобладающим в региональных и глобальных телекоммуникационных сетях (ТКС). Это преобразование изначально ориентировано на передачу акустических сигналов (речи). Перед передачей цифровые данные, поступающие обычно из ЭВМ, преобразуются в аналоговую форму с помощью модулятора-демодулятора (модема), обеспечивающего цифро-аналоговый интерфейс.

Различают три способа преобразования цифровых данных в аналоговую форму или три метода модуляции:

– *амплитудная модуляция* – изменение амплитуды несущей синусоидальных колебаний в соответствии с последовательностью передаваемых информационных битов. Например, при передаче единицы амплитуда колебаний устанавливается большой, а при передаче нуля – малой либо сигнал несущей вообще отсутствует;

– *частотная модуляция* – под действием модулирующих сигналов (передаваемых информационных битов) меняется только частота несущей синусоидальных колебаний. Например, при передаче нуля – низкая, а при передаче единицы – высокая;

– *фазовая модуляция* – в соответствии с последовательностью передаваемых информационных битов изменяется только фаза несущей синусоидальных колебаний: при переходе от сигнала 1 к сигналу 0 или наоборот фаза меняется на 180°.

Передающий модем преобразует (модулирует) сигнал несущей синусоидальных колебаний (амплитуду, частоту или фазу) так, чтобы он мог нести модулирующий сигнал, т.е. цифровые данные от ЭВМ. Обратное преобразование (демодуляция) осуществляется принимающим модемом.

В соответствии с применяемым методом модуляции различают модемы с амплитудной, частотной и фазовой модуляцией. Наиболее часто применяется частотная и амплитудная модуляции.

Аналоговый способ передачи цифровых данных обеспечивает широкополосную передачу путем использования в одном канале сигналов различных несущих частот. Это обеспечивает взаимодействие большого количества абонентов (каждая пара абонентов работает на своей частоте).

Цифровое кодирование (цифро-цифровое преобразование сигналов) цифровых данных выполняется напрямую, путем изменения уровней сигналов, несущих информацию.

Например, в ЭВМ цифровые данные представляются сигналами уровней: 5В – для кода 1 и 0,2 В – для кода 0, при передаче этих данных в линию связи уровни сигналов преобразуются соответственно в +12 В и в -12 В). Такое кодирование осуществляется, в частности, с помощью асинхронных последовательных адаптеров (например RS-232-C) при передаче цифровых данных от одного компьютера к другому на небольшие (десятки и сотни метров) расстояния. Цифровой способ передачи является узкополосным. При этом цифровые данные передаются в их естественном виде на единой частоте.

7) синхронизация элементов технических средств обработки данных

Важным требованием к построению технических устройств обработки данных является наличие в них синхронизации, обеспечивающей синхронную работу приемника и передатчика, при которой приемник осуществляет выборку поступающих информационных битов (т.е. замер уровня сигнала в линии связи) строго в моменты их прихода. Синхросигналы настраивают приемник на передаваемое сообщение еще до его прихода и поддерживают синхронизацию приемника с приходящими битами данных. Различают синхронную передачу, асинхронную передачу и передачу с автоподстройкой.

Синхронная передача отличается наличием дополнительной линии связи (кроме основной, по которой передаются данные) для передачи синхронизирующих импульсов (СИ) стабильной частоты. Каждый СИ подстраивает приемник. Выдача битов данных в линию связи передатчиком и выборка информационных сигналов приемником производятся в моменты появления СИ. Недостаток данного способа синхронизации – необходимость дополнительной линии связи.

При *асинхронной* передаче и передаче с *автоподстройкой* не требуется дополнительная линия связи.

При *асинхронной передаче* передача данных осуществляется небольшими блоками фиксированной длины (обычно байтами). Синхронизация приемника достигается тем, что перед каждым передаваемым байтом посылается дополнительный бит – старт-бит, а после переданного байта – еще один дополнительный бит – стоп-бит. Для синхронизации используется старт-бит. Этот способ синхронизации используется только в системах с низкими скоростями передачи данных.

Передача с автоподстройкой применяется в современных высокоскоростных системах передачи данных. Синхронизация обеспечивается на основе использования самосинхронизирующих кодов (СК), рис. 5.19.

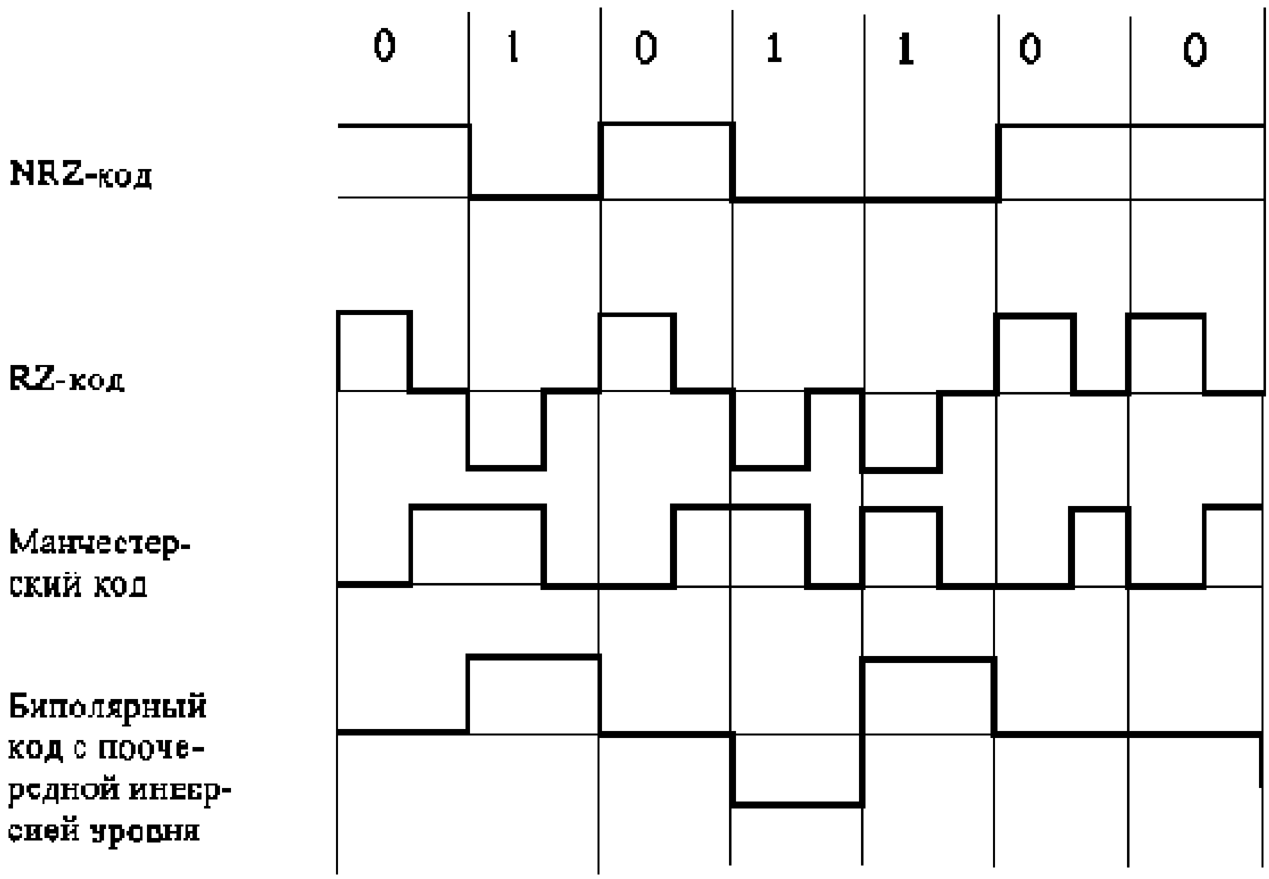


Рис. 5.19. Схемы кодирования сообщения с помощью самосинхронизирующих кодов

На основе кодирования передаваемых данных с помощью СК обеспечиваются регулярные и частые изменения (переходы) уровней сигнала в канале. Каждый переход уровня сигнала от высокого к низкому или наоборот используется для подстройки приемника. Лучшими считаются такие СК, которые обеспечивают переход уровня сигнала не менее одного раза в течение интервала времени, необходимого на прием одного информационного бита. Чем чаще переходы уровня сигнала, тем надежнее осуществляется синхронизация приемника и увереннее производится идентификация принимаемых битов данных.

Наиболее часто применяются самосинхронизирующие коды: NRZ-код (код без возвращения к нулю), RZ-код (код с возвращением к нулю), манчестерский код, биполярный код с поочередной инверсией уровня (код AMI). На рис. 5.19 представлен пример такого кодирования сообщения 0101100.

NRZ-код отличается простотой кодирования и низкой стоимостью при его реализации. Однако при передаче серий одноименных битов (единиц или нулей) уровень сигнала остается неизменным для каждой серии, что существенно снижает качество синхронизации и надежность распознавания принимаемых битов (может произойти рассогласование таймера приемника по отношению к поступающему сигналу и несвоевременный опрос линии). Для этого кода имеют место соотношения $V_1 \leq 2V_2$; $V_{1..max} = 2V_2$, *где*: V_1 – скорость изменения уровня сигналов в линии связи; V_2 – пропускная способность линии связи (бит/с), *RZ-код* отличается тем, что за время передачи одного информационного бита уровень сигнала меняется дважды независимо от того, передаются ли серии одноименных битов или поочередно изменяющихся битов. Этот код обладает хорошими свойствами синхронизации, но стоимость его реализации довольно высокая, так как необходимо обеспечить соотношение $V_1 = 2\ V_2$.

Биполярный код обладает хорошими синхронизирующими свойствами при передаче серий единиц. При передаче нулей синхронизация отсутствует. Сравнительно прост в реализации. Для этого кода $V_1 \leq V_2$; $V_1, MAX = V_2$.

Манчестерский код обеспечивает изменение уровня сигнала при представлении каждого бита, а при передаче серий одноименных битов – двойное изменение. Обладает хорошими синхронизирующими свойствами. Применяется в технике записи информации на магнитных лентах, при передаче информации по коаксиальным и оптоволоконным линиям. Соотношение скоростей для этого кода $V_1 \leq 2V_2$; $V_{1..max} = 2V_2$.

Таким образом, применение различных методов обработки сигналов, а именно модуляции и кодирования, и позволяет в настоящее время преобразовывать и передавать данные, представляемые в различной форме: аналоговой и цифровой.

Контрольные вопросы

1. Дать определение понятий «обработка информации», «данные», «обработка данных».
2. Раскрыть содержание циклов обращения информации.
3. Каково содержание понятия «сообщение» и соотношение понятий «данные» и «информация»?
4. Какие существуют способы обработки данных?
5. Дать определение понятия «сигнал», «телеграфная передача», «телефонный разговор (передача)», «движущееся изображение?
6. Какие существуют виды носителей информации?
7. Перечислить особенности и параметры сигналов и основания их классификации.
8. Дать определение понятий: «статический сигнал», «динамический сигнал», «детерминированный сигнал», «случайный сигнал», «непрерывный сигнал (сообщение)», «дискретный (квантованный) сигнал».
9. В чем состоит основное назначение квантования (дискретизации) сигнала?
10. Дать определение понятий: «источник сообщения», «канал связи (среда)», «источник информации», «кодирующее устройство».
11. Какие компоненты включает типовая структура обработки информации?
12. Что такое помеха, и какие виды помех существуют?
13. От чего зависит пропускная способность канала связи и каково ее содержание?
14. Привести выражение для расчета пропускной способности канала связи с финитной (конечной) полосой пропускания и ограниченной средней мощностью аналогового сигнала.
15. Раскрыть содержание понятий «кодирование», «декодирование», «код».
16. Дать определения: первичный алфавит, вторичный алфавит и привести закон преобразования символов первичного алфавита во вторичный
17. Раскрыть содержание целей кодирования.
18. Дать определение понятий: «регистрационное кодирование», «позиционное кодирование», «побуквенное кодирование», «порядковое кодирование», «пословное кодирование», «серийно-порядковое кодирование».
19. Раскрыть содержание понятий «модуляция сигнала (носителя)», «демодуляция сигнала»
20. Какие виды модуляции существуют?

ГЛАВА 6. КОЛИЧЕСТВЕННЫЕ ИНФОРМАЦИОННЫЕ ХАРАКТЕРИСТИКИ ИСТОЧНИКОВ СООБЩЕНИЙ И КАНАЛОВ

6.1. Энтропии вероятностных схем. Аксиомы Хинчина и Фаддеева

Исходным понятием для построения рассматриваемой теории является понятие вероятностной схемы. Пусть (Ω, M, P) – вероятностное пространство, $\{a_i\}$ – полная группа попарно несовместимых событий.

Определение 6.1. Пара $A = (\{a_i\},(p(a_i))$ *называется вероятностной схемой.*

Известно, что вероятностная схема *A дискретна*, если число событий $\{a_i\}$ не более чем счётно. В этом случае *A* будем записывать в виде:

$$A = \begin{pmatrix} a_1, \ldots\ldots, a_k, \ldots \\ p(a_1), \ldots\ldots, p(a_k), \ldots \end{pmatrix}, \tag{6.1}$$

где a_k – исход вероятностной схемы, $p(a_k)$ – вероятность исхода,

$$\sum_k p(a_k) = 1, \; p(a_k) \geq 0.$$

Если число событий $\{a_k\}$ более чем счётно, то вероятностная схема *A* называется непрерывной. Тогда её задают, описывая множество возможных исходов и указывая с помощью плотности распределения вероятностей $p(x)$ вероятностное распределение на исходах. На протяжении данного курса будем, как правило, рассматривать конечные вероятностные схемы.

Важной характеристикой схемы является энтропия – средняя мера неравновероятности исходов схемы. *Энтропия* и близкое понятие *информация* по-разному определялись рядом авторов. Приведём некоторые из этих определений.

Согласно Р. Хартли количество информации, содержащееся в сообщении, должно удовлетворять двум требованиям:

1. Пусть сообщение T_1, записанное в алфавите A_1, $|A_1| = n_1$, имеет длину l_1, а сообщение T_2, записанное в алфавите A_2, $|A_2| = n_2$ имеет длину l_2. Тогда, если выполнено $n_1^{l_1} = n_2^{l_2}$, то сообщения T_1 и T_2 несут одинаковое количество информации.

2. Количество информации, содержащееся в сообщении, пропорционально его длине.

Из этих положений Хартли получил формулу для количества информации в сообщении:

$$I = \log n^l, \tag{6.2}$$

где n – мощность алфавита, l – длина сообщения.

К. Шеннон рассматривал порождение сообщений в условиях вероятностной схемы:

$$A = \begin{pmatrix} a_1, \ldots, a_i, \ldots \\ p(a_1), \ldots, p(a_i), \ldots \end{pmatrix}, \sum_{i=1}^{n} p_i = 1, 0 \le p_i \le 1.$$

Относительно количества информации, содержащегося в сообщении T длины l, составленном по схеме независимых испытаний, высказываются следующие требования:

1. Пустое сообщение не содержит информации.

2. Количество информации, содержащееся в сообщении, пропорционально его длине.

По Шеннону, количество информации, содержащееся в сообщении $T = a_{i1}, \ldots a_{il}$ равно:

$$I(T) = l * I, \tag{6.3}$$

где $I = -\sum_{i=1}^{n} p_i \log p_i$ – энтропия вероятностной схемы, или количество информации, приходящееся на один символ порождённого сообщения.

А. Я. Хинчин к определению энтропии вероятностной схемы подошёл с аксиоматических позиций. В его работе установлено, что энтропия конечной вероятностной схемы однозначно определяется с точностью до постоянного множителя при задании системы аксиом.

Аксиомы Хинчина:

Аксиома 1.1.1 $H(p_1,...,p_n)$ – ненулевая непрерывная по переменным $p_1,...,p_n$ в симплексе $0 \le p_i \le 1,\ p_1 + ... + p_n = 1$.

Аксиома 1.1.2. $H(p_1, p_2,..., p_n)$ симметрична по переменным $p_1, p_2,..., p_n$

Аксиома 1.1.3. $H(p_1, p_2,..., p_n, 0) = H(p_1, p_2,..., p_n)$.

Аксиома 1.1.4. $H(q_{1l},..., q_{1m}; q_{2l},..., q_{2m};...; q_{nl},..., q_{nm}) =$

$$= H(p_1,..., p_n) + \sum p_l H\left(\frac{q_{l1}}{p_l},...,\frac{q_{lm}}{p_l}\right)\ p_i = q_{i1} + ... + q_{im} (i = 1,..., n).$$

Аксиома 1.1.5. $H(p_1,..., p_n) = H\left(\frac{1}{n},...,\frac{1}{n}\right)$.

Система аксиом Фаддеева эквивалентна системе аксиом Хинчина и позволяет получить тот же результат.

Аксиомы Фаддеева :

Аксиома1.1.1' $H(p, 1-p)$ непрерывна при $0 \le p_i \le 1$ и положительна хотя бы в одной точке.

Аксиома 1.1.2'. $H(p_1, p_2,..., p_n)$ симметрична по $p_1,..., p_n$.

Аксиома 1.1.3'. При $n \ge 2$

$$H(p_1,..., p_{n-1}; q_1, q_2) = H(p_i,..., p_n) + p_n H\left(\frac{q_1}{p_n},\frac{q_2}{p_n}\right),$$

где $p_n = q_1 + q_2$.

Разница в этих системах аксиом заключается в том, что аксиома 1.1.5 *(экстремальность функции энтропии)* заменяется требованием положительности энтропии в одной точке. Аксиомы Хинчина 1.1.3 и 1.1.4 заменяются аксиомой 1.1.3'. Аксиома 1.1.3' естественна, так как *неопределённость* схемы

$$A = \begin{pmatrix} a_1,......, a_{n-1}; b_1, b_2 \\ p_1,..., p_{n-1}; q_1, q_2 \end{pmatrix}$$

отличается от неопределённости схемы

$$A = \begin{pmatrix} a_1,......, a_{n-1}, a_n \\ p_1,..., p_{n-1}, p_n \end{pmatrix}$$

на неопределённость, происходящую от подразделения a_n на два подсобытия b_1, b_2 с условными вероятностями $\frac{q_1}{p_n}, \frac{q_2}{p_n}$. Эта неопределённость должна быть преодолена только в случае, если реализуется событие a_n, вероятность которого равна p_n.

Если рассматривать энтропию как количественную меру неопределённости в реализации вероятностной схемы, то последняя аксиома естественна.

Установим равносильность аксиом 1.1.2, 1.1.3, 1.1.4 и 1.1.2', 1.1.3', следуя упомянутой работе Фаддеева.

Лемма 1. Из аксиом 1.1.2, 1.1.3, 1.1.4 *следует, что* $H(1,0) = 0$.

Доказательство.

Подсчитаем $H\left(\frac{1}{2},\frac{1}{2},0,0\right)$ двумя способами:

1. По аксиоме 1.1.3:

$$H\left(\frac{1}{2},\frac{1}{2},0,0\right) = H\left(\frac{1}{2},\frac{1}{2}\right).$$

2. По аксиоме 1.1.2:

$$H\left(\frac{1}{2},\frac{1}{2},0,0\right) = H\left(\frac{1}{2},0,\frac{1}{2},0\right).$$

Применим аксиому 1.1.4:

$$H\left(\frac{1}{2},0,\frac{1}{2},0\right) = H\left(\frac{1}{2},\frac{1}{2}\right) + \frac{1}{2}H(1,0) + \frac{1}{2}H(1,0) = H\left(\frac{1}{2},\frac{1}{2}\right) + \frac{1}{2}H(1,0).$$

Отсюда следует, что

$$H(1,0) = 0.$$

Лемма 2. Из *аксиом* 1.1.2, 1.1.3, 1.1.4 *следует аксиома* 1.1.3'.

Доказательство. Рассмотрим набор $r_1,...,r_{2n}$, $r_i \geq 0$ $\sum_{i=1}^{2n} r_i = 1$, среди r_i есть $n-1$ нуль. Из аксиом 1.1.2 и 1.1.3 следует, что $H(r_1,...,r_{2n})$ перестановками аргументов можно представить в виде:

$$H(p_i,0;p_2,0;...;p_{n-1},0;q_i,q_2).$$

Применим аксиому 1.1.4:

$$H(p_i,0;...;p_{n-1},0;q_i,q_2) = H(p_1,...;p_{n-1},p_n) + p_1 H(1,0) + ...$$

$$... + p_{n-1}H(1,0) + p_n H\left(\frac{q_1}{p_n},\frac{q_2}{p_n}\right) = H(p_1,...;p_{n-1},p_n) +$$

$$+ p_n H\left(\frac{q_1}{p_n},\frac{q_2}{p_n}\right).$$

Лемма 3. Из *аксиом* 1.1.2', 1.1.3' *следует, что* $H(1,0)=0$.

Доказательство. Подсчитаем двумя способами $H\left(\frac{1}{2},\frac{1}{2},0\right)$.

$$H\left(\frac{1}{2},\frac{1}{2},0\right) = H\left(\frac{1}{2},\frac{1}{2}\right) + \frac{1}{2}H(1,0).$$

$$H\left(\frac{1}{2},\frac{1}{2},0\right) = H\left(0,\frac{1}{2},\frac{1}{2}\right) = H(0,1) + H\left(\frac{1}{2},\frac{1}{2}\right).$$

Следовательно, $H(1,0)=0$.

Лемма 4. *Из аксиом* 1.1.2', 1.1.3' *следует аксиома* 1.1.3.

Доказательство.

$$H(p_1,...,p_{n-1};p_n,0) = H(p_1,...,p_{n-1};p_n) + p_n H(1,0) = H(p_1...,p_{n-1},p_n).$$

Лемма 5. *Из аксиомы* 1.1.3' *следует*

$$H(p_1,...,p_{n-1};q_1,...q_m) = H(p_1,...,p_{n-1},p_n) + p_n H\left(\frac{q_1}{p_n},\frac{q_m}{p_n}\right).$$

Здесь $m \geq 2$, $p_n = q_1 + ... + q_m$.

Доказательство. При $m=2$ утверждение леммы совпадает с аксиомой 1.1.3'. При $m \geq 3$ лемма доказывается индукцией по m.

Лемма 6. *Из аксиом* 1.1.2', 1.1.3' *следует*

$$H(q_{i1},...,p_{im1};q_{21},...,q_{2m2};q_{n1},...,q_{nmn}) = H(p_1,...,p_n) + \sum_{i=1}^{n} p_i H\left(\frac{q_{i1}}{p_i},...,\frac{q_{imi}}{p_i}\right).$$

Здесь $p_i = q_{i1} + ... + q_{imi}$.

Доказательство. Нужно n раз применить лемму 1.1.5 к каждой группе аргументов, что возможно в силу симметрии. Приняв $m_1 = m_2 = ... = m_n = m$, получим, что из аксиом 1.1.2', 1.1.3' следует аксиома 1.1.4.

Установлена равносильность групп аксиом 1.1.2, 1.1.3, 1.1.4 и 1.1.2', 1.1.3'.

Перейдём к доказательству *теоремы единственности функции энтропии.* Введём определение:

$$F(n) = H\left(\frac{1}{n},...,\frac{1}{n}\right), \text{ при } n \geq 2 \text{ и } F(1) = 0.$$

Применяя лемму 6 к случаю $m_1 = m_2 = ... = m_n = m$ $q_{ij} = \dfrac{1}{mn}$, получим

$$F(mn) = F(m) + F(n). \tag{6.4}$$

Применим лемму 1.1.5 к $H\left(\dfrac{1}{n},...,\dfrac{1}{n}\right)$:

$$H\left(\frac{1}{n},...,\frac{1}{n}\right) = H\left(\frac{1}{n},\frac{n-1}{n}\right) + \frac{n-1}{n} H\left(\frac{1}{n-1},...,\frac{1}{n-1}\right).$$

Обозначим $\eta_n = H\left(\dfrac{1}{n},\dfrac{n-1}{n}\right) = F(n) - \dfrac{n-1}{n} F(n-1)$.

Лемма 7. *При* $n \to \infty$

$$\mu_n = \frac{F(n)}{n} \to 0 \text{ и } \lambda_n = F(n) - F(n-1) \to 0.$$

Доказательство. В силу непрерывности $H(p,1-p)$ при $p\in[0,1]$

$$\eta_n = H\left(\frac{1}{n},\frac{n-1}{n}\right)\to H(0,1)=0,\ \text{при}\ n\to\infty$$

Далее,

$$n\eta_n = nF(n)-(n-1)F(n-1).$$

Отсюда, складывая эти соотношения для $k=1,2,...,n$, имеем:

$$nF(n)=\sum_{k=1}^{n} k\eta_k$$

Получаем

$$\frac{F(n)}{n}=\frac{1}{n^2}\sum_{k=1}^{n} k\eta_k=\frac{n+1}{2n}\frac{2}{n(n+1)}\sum_{k=1}^{n} k\eta_k.$$

Однако $\frac{2}{n(n+1)}\sum_{k=1}^{n} k\eta_k$ – среднее арифметическое первых $\frac{n(n+1)}{2}$ членов стремящейся к нулю последовательности η_1, η_2, η_2, η_3, η_3, $\eta_3,\ldots$

Следовательно,

$$\frac{2}{n(n+1)}\sum_{k=1}^{n} k\eta_k \to 0 \text{ и } \mu_n=\frac{F(n)}{n}\to 0,\ \text{при n}\to\infty.$$

Далее,

$$\lambda_n = F(n)-F(n-1)=\eta_n-\frac{1}{n}F(n-1)\to 0.$$

Лемма доказана.

В силу формулы (6.4) функция $F(n)$ будет определена при всех натуральных n, как только зададим значения функции F на простых числах.

Именно, если $n=p_1^{\alpha_1},...,p_S^{\alpha_{S1}}$ где $p_1,...,p_1$ – простые, то

$$F(n)=\alpha_1 F(p_1)+...+\alpha_S F(p_S).$$

Положим,

$$F(p) = c_p \log p, \tag{6.5}$$

где p – простое число.

Тогда

$$F(n) = \alpha_1 c_{p1} \log p_1 + \ldots + \alpha_S p_S \log p_S.$$

Лемма 8. *Среди чисел* $c_p (p = 2,3,5,7,\ldots)$ *существует наибольшее.*

Доказательство. Допустим противное и покажем, что это предположение противоречит предположению о непрерывности $H(P,1-p)$ при $p=0$. Действительно, если в последовательности $c_p (p = 2,3,5,7,\ldots)$ нет наибольшего числа, то можно построить бесконечную последовательность простых чисел $p_1, p_2, \ldots, p_i$ так, что $p_1 = 2$ p_i – есть наименьшее простое число, такое, что $c_{p_i} > c_{p_{i-1}}$. Из способа построения этой последовательности ясно, что как только простое число q меньше p_i, то $c_q \le c_{pi}$.

Пусть $i > 1$ и $q_1^{\alpha_1}, \ldots, q_s^{\alpha_s}$ есть каноническое разложение числа $p_i - 1$.

Рассмотрим

$$\lambda_{p_i} = F(p_i) - F(p_i - 1) = F(p_i) - \frac{F(p_i)}{\log p_i} \log(p_i - 1) + c_{pi} \log(p_i - 1) -$$

$$- F(p_i - 1) = \frac{F(p_i)}{p_i} \frac{p_i}{\log p_i} \log \frac{p_i}{p_i - 1} + \sum_{j=1}^{s} \alpha_j (c_{pi} - c_{qi}) \log q_j,$$

Ясно, что

$$\sum_{j=1}^{s} \alpha_j (c_{pi} - c_{qi}) \log q_j \ge (c_{pi} - c_2) \log 2 \ge (c_{p2} - c_2) \log 2,$$

так как $q_i < p_i$, $c_{qj} \le c_{pi}$, и в силу чётности $p_i - 1$ среди q_j присутствует число 2 с ненулевым показателем. Далее, при $i \to \infty$

$$\frac{F(p_i)}{p_i} \to 0, \ \frac{p_i}{\log p_i} \log \frac{p_i}{p_i - 1} \to 0, \text{ и } \lambda_{p_i} \to 0.$$

Следовательно, $(c_{p2} - c_2) \log 2 \le 0$, что невозможно.

Таким же образом устанавливается, что среди c_p существует наименьшее.

Лемма 9. *Функция* $F(n) = c\log n$, *где c – постоянное.*

Доказательство. Достаточно установить, что все c_p равны друг другу. Допустим, что найдётся такое простое число p', что $c_{p'} > c_2$. Обозначим через p то простое число, для которого c_p принимает наибольшее значение. Тогда $c_p > c_2$ и $p > 2$. Пусть m – натуральное число и $q_1^{\alpha_1},...,q_s^{\alpha_s} = p^m - 1$. Рассмотрим:

$$\lambda_{p^m} = F(p^m) - F(p^m - 1) =$$

$$= F(p^m) - \frac{F(p^m)}{\log p^m}\log(p^m - 1) + c_p \log(p^m - 1) - F(p^m - 1) =$$

$$= \frac{F(p^m)}{p^m}\frac{p^m}{\log p^m}\log\frac{p^m}{p^m - 1} + \sum_{j=1}^{s}\alpha_j(c_p - c_{qi})\log q_j \geq$$

$$\geq \frac{F(p^m)}{p^m}\frac{p^m}{\log p^m}\log\frac{p^m}{p^m - 1} + (c_p - c_2)\log 2.$$

При переходе к пределу при $m \to \infty$ получим $(c_p - c_2)\log 2 \leq 0$, что невозможно. Таким образом, для всех p' имеет место $(c_{p'} \leq c_2)$. Совершенно таким же образом устанавливается, что $(c_{p'} \leq c_2)$ при всех p'. Лемма доказана.

С учётом доказанных лемм 7, 8, 9 в системе аксиом Фаддеева может быть доказана:

Теорема 1.1. *Справедливо представление функции* $H(p_1, p_2,..., p_n)$:

$$H(p_1, p_2,..., p_n) = c\sum_{i=1}^{n} p_i \log\frac{1}{p_i}(c > 0). \qquad (6.6)$$

Доказательство. Рассмотрим случай $n = 2$.

Пусть $p = \frac{r}{s}$ при натуральных r и s. Применим лемму 6 к $H\left(\frac{1}{s},...,\frac{1}{s}\right)$, соединив аргументы в две группы из r и $s - r$ элементов. Получим:

$$H\left(\frac{1}{s},..,\frac{1}{s}\right) = H\left(\frac{r}{s},1-\frac{1}{s}\right) + \frac{r}{s}H\left(\frac{1}{r},...,\frac{1}{r}\right) + \frac{s-r}{s}H\left(\frac{1}{s-r},...,\frac{1}{s-r}\right).$$

Отсюда

$$H(p,1-p) = F(s) - pF(r) - (1-p)F(s-r) =$$
$$= c\log s - pc\log r - (1-p)c\log(s-r) =$$
$$= c\left(p\log\frac{s}{r} + (1-p)\log\frac{s}{s-r}\right) = c\left(p\log\frac{1}{p} + (1-p)\log\frac{1}{1-p}\right)$$

В силу непрерывности $H(p,1-p)$ полученная формула может быть распространена и на иррациональные значения p. Постоянная $C > 0$ в силу положительности $H(p,1-p)$, хотя бы в одной точке.

Переход к общему случаю осуществляется методом математической индукции на основании аксиомы 1.1.3'.

Выводы этой фундаментальной теоремы Д. К. Фаддеева о представлении функции энтропии для дискретной вероятностной схемы позволят в дальнейшем характеризовать производительность источника информации, оценивать возможности сжатия информации, получить формулу для вычисления пропускной способности дискретного канала связи.

6.2. Свойства энтропии. Энтропия при непрерывном сообщении

При равновероятности знаков алфавита $P_i = 1/m$ и из общей формулы получаем

$$H = -\sum_{i=1}^{m} P_i \log P_i = -\sum_{i=1}^{m} \frac{1}{m} \log\frac{1}{m} = -\left(m\frac{1}{m}\right)(-\log m) = \log m$$

В этом случае энтропия определяется исключительно числом знаков m алфавита и по существу является характеристикой только алфавита.

Если же буквы неравновероятны, то алфавит можно рассматривать как дискретную случайную величину, заданную статистическим распределением частот n_i (или вероятностей $P_i = n_i / n$):

Знаки x_i	x_1	x_2	...	x_m
Частоты n_i	n_1	n_2	...	n_m

Такие распределения получают обычно на основе статистического анализа конкретных типов сообщений (например, русских или английских текстов, численных значений и т.п.).

Поэтому, хотя формально в выражении для энтропии входят только характеристики алфавита, оно отражает статистические свойства некоторой совокупности сообщений.

На основании выражения

$$H=-\sum_{i=1}^{m} P_i \log P_i=\sum_{i=1}^{m} P_i \log \frac{1}{P_i}$$

величину $\log 1/P_i$ можно рассматривать как *частную энтропию*, характеризующую информативность знака x_i, а энтропию H – как среднее значение частных энтропий.

При малых P_i частная энтропия велика, а с приближением P_i к единице она стремится к нулю (рис. 6.1).

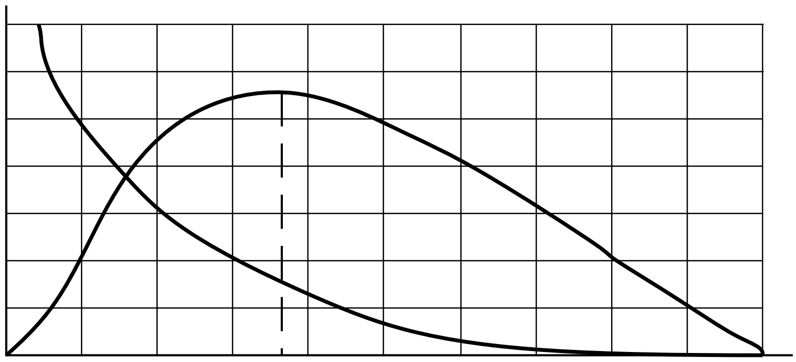


Рис. 6.1. Графики функций $\log 1/P_i$ и $-P_i \cdot \log P_i$

Функция $-P_i \cdot \log P_i$ отражает вклад знака x_i в энтропию H. Как видно, при $P_i=1$ эта функция равна нулю, затем возрастает до своего максимума и при уменьшении P_i стремится к нулю.

Из условия

$$\frac{\partial}{\partial P_i}(-P_i \log P_i)=-\log P_i-\log e=-\log P_i e=0$$

находим $P_i\ e=1$,где e – основание натурального логарифма. Таким образом, функция $-P_i \log P_i$ при $P_i=1/e=0{,}37$ имеет максимум $\frac{1}{e}\log e=0{,}531.$

Энтропия H – величина вещественная, неотрицательная и ограниченная, т.е. $H \geq 0$ (это свойство следует из того, что такими же качествами обладают все ее слагаемые $P_i \log 1/P_i$).

Энтропия равна нулю, если сообщение известно заранее (в этом случае каждый элемент сообщения замещается некоторым знаком с вероятностью, равной единице, а вероятности остальных знаков равны нулю).

Можно также показать, что энтропия максимальна, если все знаки алфавита равновероятны, т.е. $H_{\max} = \log m$.

Пример 1: Распределение знаков алфавита имеет вид $p(x_1) = 0{,}1$ $p(x_2) = 0{,}1$ $p(x_3) = 0{,}1$ $p(x_4) = 0{,}7$. Определить число знаков другого алфавита, у которого все знаки равновероятны, а энтропия такая же, как и у заданного алфавита.

Особый интерес представляют *бинарные сообщения*, использующие двузнаковый алфавит (0,1). Так как при $m = 2$ вероятности знаков алфавита $P_1 \neq P_2 = 1$, то можно положить $P_1 = P$ и $P_2 = 1 - P$. Тогда энтропия определяется соотношением

$$H = -P_1 \log P_1 - P_2 \log P_2 = -P \log P - (1-P)\log(1-P),$$

график которого показан на рис. 6.2.

Как видно, энтропия бинарных сообщений достигает максимального значения, равного 1 *биту*, при P = 0,5, и ее график симметричен относительно этого значения.

Пример 2. Сравнить неопределенность, приходящуюся на одну букву источника информации (алфавита русского языка), характеризуемого ансамблем, представленным в таблице, с неопределенностью, которая была бы у того же источника при равновероятном использовании букв (табл. 6.1).

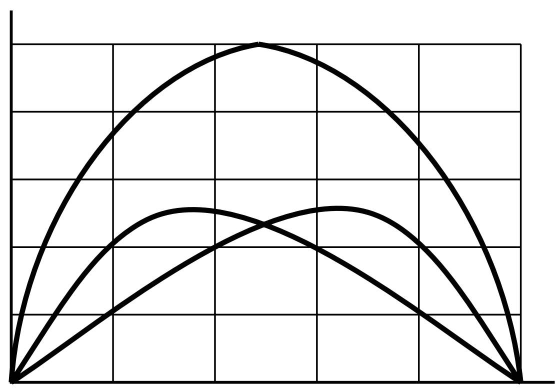


Рис. 6.2. График энтропии двоичных сообщений 1 и ее составляющих 2: –(1–P)log(1–P) и 3: –Plog$P$

Таблица 6.1

Статистические вероятности появления букв русского алфавита

Буква	Вероят-ность	Бу-ква	Веро-ят-ность	Буква	Веро-ят-ность	Буква	Вероят-ность
а	0,064	й	0,010	т	0,056	ы	0,016
б	0,015	к	0,029	у	0,021	э	0,003
в	0,039	л	0,036	ф	0,02	ю	0,007
г	0,014	м	0,026	х	0,09	я	0,019
д	0,026	н	0,056	ц	0,04	пробел	0,143
е,ё	0,074	о	0,096	ч	0,013		
ж	0,008	п	0,024	ш	0,006		
з	0,015	р	0,041	ш	0,003		
и	0,064	с	0,047	ъ,ь	0,015		

Решение: При одинаковых вероятностях появления всех m = 32 букв алфавита неопределенность, приходящаяся на одну букву, характеризует энтропия

$$H = \log m = \log 32 = 5 \text{ бит}$$

Энтропию источника, характеризуемого заданным таблицей ансамблем

$$H = -\sum_{i=1}^{m} p_i \log p_i = -0{,}064 \log 0{,}064 \log 0{,}064 - 0{,}015 \log 0{,}015 - \ldots -$$

$$- 0{,}143 \log 0{,}143 \approx 4{,}43 \textit{бит}$$

Таким образом, неравномерность распределения вероятностей использования букв снижает энтропию источника с 5 до 4,42 бит

Пример 3: Заданы ансамбли X и Y двух дискретных величин x и y:

Случайные величины x_i	0,5	0,7	0,9	0,3
Вероятности их появления	0,25	0,25	0,25	0,25
Случайные величины y_j	5	10	15	8
Вероятности их появления	0,25	0,25	0,25	0,25

Сравнить их энтропии.

Решение: Энтропия не зависит от конкретных значений случайной величины.

Так как вероятности их появления у обоих величин одинаковы, то

$$H(X) = H(Y) = -\sum_{i=1}^{m} p_i \log p_i = -4(0{,}25\log 0{,}25) = -4(1/4\log 1/4)$$
$$= \log 6 = 2\textit{бит}$$

1) энтропия при непрерывном сообщении

Так как при непрерывном сообщении оно не разделено на элементы, то вместо конечного алфавита необходимо рассматривать бесконечное множество возможных состояний элементов, определяемое непрерывным распределением плотности вероятностей $P(x)$.

Для обобщения формулы Шеннона разобьем интервал возможных состояний на равные непересекающиеся отрезки Δx и рассмотрим множество дискретных состояний x_1, x_2, ... , x_m с вероятностями $P_i = p(x_i)\Delta x$ $(i = 1,2,...,m)$. Тогда

$$H = -\sum_{i=1}^{m} p(xi)\Delta x \log p(xi)\Delta x = -\sum_{i=1}^{m} p(xi)\Delta x \log p(xi) =$$
$$= -\sum_{i=1}^{m} p(xi)\Delta x \log \Delta x.$$

В пределе при $\Delta x \to 0$ с учетом соотношения

$$\int_{-\infty}^{\infty} p(x)dx = 1$$

получим

$$H(x) = -\int_{-\infty}^{\infty} p(x)\log p(x)dx - \log\Delta x$$

Первое слагаемое в этой сумме, называемое *приведенной энтропией*, целиком определяет информативность сообщений, обусловленных статистикой состояний их элементов.

Величина $\log\Delta x$ зависит только от выбранного интервала Δx, определяющего точность квантования состояний, и при $\Delta x = const$ она постоянна.

Итак, энтропия и количество информации зависят от распределения плотности вероятностей $p(x)$.

В теории информации большое значение имеет решение вопроса о том, при каком распределении обеспечивается максимальная энтропия $H(x)$.

Можно показать, что при заданной дисперсии

$$\sigma^2 = \int_{-\infty}^{\infty} x^2 p(x)dx = const$$

наибольшей информативностью сообщение обладает только тогда, когда состояния элементов распределены по нормальному закону:

$$p(x) = \frac{1}{\sigma\sqrt{\pi}} e^{-\frac{x^2}{2\sigma^2}}$$

Так как дисперсия определяет среднюю мощность сигнала, то отсюда следуют практически важные выводы.

Передача наибольшего количества информации при заданной мощности сигнала (или наиболее экономичная передача данного количества информации) достигается при такой обработке сигнала, которая приближает распределение к нормальному. В то же время, приписывая нормальное распределение помехе, обеспечивают ее наибольшую «информативность», т.е. учитывают ее пагубное воздействие на прохождение сигналов в самом худшем случае.

Найдем значение энтропии, когда состояния элементов распределены по нормальному закону:

$$H = -\frac{1}{e\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} \log\left[\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}\right] dx - \log\Delta x =$$

$$= \log(\sigma\sqrt{2\pi}\,\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} dx + \frac{\log e}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} dx - \log\Delta x) =$$

$$= \log(\frac{\sigma}{\Delta x}\sqrt{2\pi e})$$

Найдем значение энтропии, когда состояния элементов распределены внутри интервала их существования $a \le x \le b$ по равномерному закону, т.е

$$x) = \begin{cases} \dfrac{1}{b-a} & при \quad a \le x \le b \\ 0 \quad при \quad x < a, x > b \end{cases}$$

$$H_p(x) = -\frac{1}{b-a}\int_a^b \log\left(\frac{1}{b-a} - \Delta x\right)dx = \log\frac{b-a}{\Delta x}$$

Дисперсия равномерного распределения $\sigma_p^2 = \dfrac{(b-a)^2}{12}$, поэтому $b - a = 2\sqrt{3}\sigma_p$. С учетом этого можно записать

$$H_p(x) = \log\left(\frac{\sigma_p}{\Delta x}2\sqrt{3}\right)$$

Сравнивая между собой сообщения с равномерным и нормальным распределением вероятностей при условии $H_н(x) = H_p(x)$, получаем

$$\sigma_p^2 = \frac{e}{6}\sigma^2 \approx 1{,}42\sigma^2$$

Это значит, что при одинаковой информативности сообщений средняя мощность сигналов для равномерного распределения их амплитуд должна быть на 42 % больше, чем при нормальном распределении.

Пример 4: Определить энтропию непрерывной случайной величины, распределенной по экспоненциальному закону с плотностью вероятности:

$$p(x) = \begin{cases} ce^{-cx}, x \le 0 \\ 0, x < 0 \end{cases}$$

Пример 5: Найдите энтропию случайной величины, распределенной по закону с плотностью вероятности

$$p(x) = \begin{cases} 0, x \le 0 \\ x^2, 0 < x \le 1 \\ 1, x > 1 \end{cases}$$

Пример 6: При организации мешающего воздействия при передаче информации можно использовать источник шума с нормальным распределением плотности и источник, имеющий в некотором интервале равномерную плотность распределения. Определить, какой источник шума применять экономичнее и каков при этом выигрыш в мощности.

Решение: Очевидно, что сравнение источников следует проводить из условия обеспечения равенства энтропий, когда каждый источник вносит одинаковое мешающее воздействие при передаче информации.

Как было показано выше, значение энтропии, когда состояния элементов распределены по нормальному закону:

$$H_{,}(x) = \log\left(\frac{\sigma}{\Delta x}\sqrt{2\pi e}\right) = \log(\sigma_{д}\sqrt{2\pi e}) \ ,$$

где $\Delta x = 1Ом$, а $\dfrac{\sigma}{\Delta x} = \sigma_{д}$, т.е. $\sigma_{г}^{2}$ – дисперсия, характеризующая мощность, выделяемую на резистор с сопротивлением 1*Ом*.

Для равномерного распределения:

$$H_{p}(x) = \log\frac{b-a}{\Delta x}$$

Так как дисперсия равномерного распределения

$$\sigma_{h}^{2} = \frac{(b-a)^{2}}{12}, \qquad то \quad b - a = 2\sqrt{3}\sigma_{p}, \text{ и, следовательно}$$

$$H_{p}(x) = \log\left(\frac{\sigma_{p}}{\Delta x}2\sqrt{3}\right) = \log\left(\sigma_{pГ}2\sqrt{3}\right),$$

где $\sigma_{pГ} = \dfrac{\sigma_{p}}{\Delta x}$, $\Delta x = 1Ом$

Так как

$$H_{Н}(x) = H_{p}(x), \qquad то \qquad \log(\sigma_{Г}\sqrt{2\pi e}) = \log(\sigma_{pГ}2\sqrt{3}),$$

$$2\pi e\sigma_{Г}^{2} = 12\sigma_{pГ},$$

$$\sigma_{pГ}^{2} = \frac{\pi e}{6}\sigma_{Г}^{2} \approx 1{,}42\sigma_{Г}^{2}$$

Поэтому следует выбирать источник шума с нормальным распределением плотности, т.к. при той же неопределенности, вносимой им в канал связи, можно выиграть в мощности 42 %.

Условная энтропия. Предполагалось, что все элементы сообщения независимы, т.е. появление каждого данного элемента никак не связано с предшествующими элементами. Рассмотрим теперь два ансамбля

$$X = (x_1, x_2, ..., x_r)$$
$$Y = (y_1, y_2, ..., y_r),$$

которые определяются не только собственными вероятностями $p(x_i)$ и $p(y_i)$, но и условными вероятностями $p_{xi}(y_i)$, $p_{yi}(x_i)$, где $i = 1,2,...,r$; $j = 1,2,...,s$.

Систему двух случайных величин *X*, *Y* можно изобразить случайной точкой на плоскости. Событие, состоящее в попадании случайной точки (*X*,*Y*) в область *D*, принято обозначать в виде $(X, Y) \subset D$. Закон распределения системы двух дискретных случайных величин может быть задан с помощью таблицы, где P_{ij} – вероятность события, заключающегося в одновременном выполнении равенства $X = x_i$, $Y = y_i$. При этом

$$\sum_{i=1}^{r} \sum_{j=1}^{s} P_{ij} = 1 .$$

Закон распределения системы непрерывных случайных величин (*X*,*Y*) задают при помощи функции плотности вероятности *p*(*x*,*y*).

Вероятность попадания случайной точки (*X*,*Y*) в область *D* определяется равенством

$$P[(X,Y) \subset D] = \iint_D p(x,y)dxdy.$$

X \ Y	y_1	y_2	. . .	y_s
x_1	P_{11}	P_{12}	. . .	P_{1s}
x_2	P_{21}	P_{22}	. . .	P_{2s}
:	:	:	:	:
x_r	P_{r1}	P_{r2}	. . .	P_{rs}

Функция плотности вероятности обладает следующими свойствами:

1) $p(x, y) \geq 0$

2) $\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} p(x, y)\,dxdy = 1$

Если все случайные точки (X,Y) принадлежат области *D*, то $\iint\limits_{D} p(x, y)\,dxdy = 1$.

Предположим, что приведена серия опытов, в результате которых каждый раз наблюдалась двумерная случайная величина (*X*,*Y*). Условимся, исход каждого опыта изображать точкой на декартовой плоскости.

Может оказаться, что изображающие точки хаотично расположатся на плоскости (рис. 6.3):

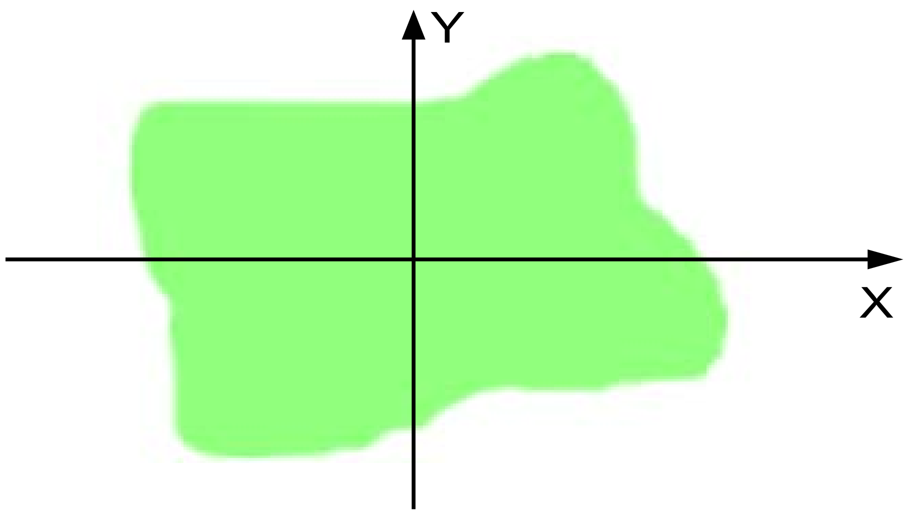


Рис. 6.3. Расположение точек хаотично

Говорят, что при этом рассматриваемые величины некоррелированы, т.е. между ними нет устойчивой связи в вероятном смысле. При этом, случайные величины *X* и *Y* называют независимыми, т.к. вероятность одной из них принять значение, лежащее в любом промежутке области ее значений, не зависит от того, какое значение приняла другая величина.

Однако возможен случай, когда изображающие точки в среднем расположены вдоль некоторой прямой так, что в каждом отдельном испытании величины x_i и y_j имеют чаще всего одинаковый знак. Это наводит на мысль о том, что между x_i и y_j есть статистическая связь, называемая *коррекцией* (рис. 6.4).

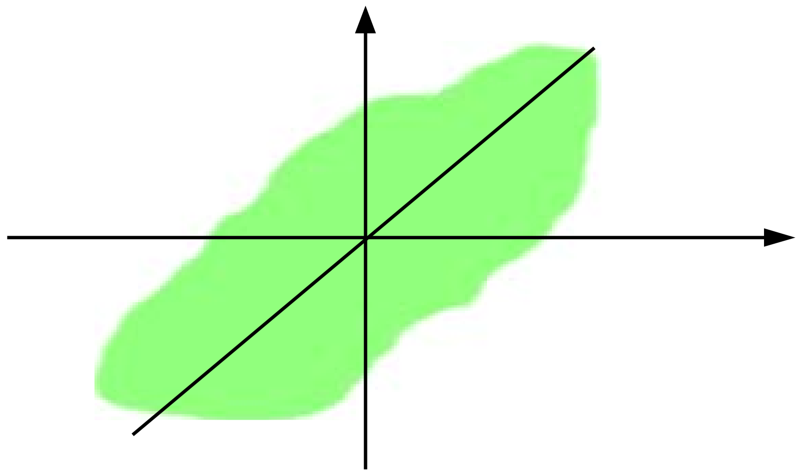


Рис. 6.4. Изображающие точки в среднем расположены вдоль некоторой прямой

Пример 7: Задано распределение вероятностей дискретной двумерной случайной величины:

X \ Y	4	5
3	0,17	0,10
10	0,13	0,30
12	0,25	0,05

Найти законы распределения составляющих X и Y.

Решение:

1) Сложив вероятности «по строкам», получим вероятности возможных значений X:

$$P(3) = 0{,}17 + 0{,}10 = 0{,}27$$
$$P(3) = 0{,}14 + 0{,}30 = 0{,}43$$
$$P(12) = 0{,}25 + 0{,}05 = 0{,}30.$$

Запишем закон распределения составляющей X:

X	3	10	12
$P(x_i)$	00,27	0,43	0,30

Контроль: $0{,}27 + 0{,}43 + 0{,}30 = 1$

2) Сложив вероятности «по столбцам», аналогично найдем распределение составляющей *Y*:

Y	4	5
$P(y_j)$	0,55	0,45

Контроль: $0{,}55 + 0{,}45 = 1$

Условным распределением составляющей X при $Y = y_j$ (j сохраняет одно и то же значение при всех возможных значениях X) называют совокупность условных вероятностей

$$P_{yj}(x_1), P_{yj}(x_2), ..., P_{yj}(x_r)$$

Аналогично определяется условное распределение *Y*.

Условные вероятности составляющих *X* и *Y* вычисляют соответственно по формулам:

$$P_{xi}(y_j) = \frac{P(x_i, y_j)}{P(x_i)}$$

$$P_{yj}(x_i) = \frac{P(x_i, y_j)}{P(y_j)}$$

Для контроля вычислений целесообразно убедиться, что сумма вероятностей условного распределения равна единице.

Пример 8: Задана дискретная двумерная случайная величина (*X*,*Y*):

X \ Y	$y_1 = 0{,}4$	$y_2 = 0{,}8$
$x_1 = 2$	0,15	0,05
$x_2 = 5$	0,30	0,12
$x_3 = 8$	0,35	0,03

Найти:

1) безусловные законы распределения составляющих;

2) условный закон распределения составляющей *X* при условии, что составляющая *Y* приняла значение $y_1 = 0{,}4$;

3) условный закон распределения составляющей *Y* при условии, что составляющая *X* приняла значение $x_2 = 5$.

Решение:

1) Сложив вероятности «по строкам», напишем закон распределения *X*.

X	2	5	8
$P(x)$	0,20	0,42	0,38

2) Сложив вероятности «по столбцам», найдем закон распределения *Y*.

Y	0,4	0,8
$P(y)$	0,80	0,20

3) Найдем условные вероятности возможных значений *X* при условии, что составляющая *Y* приняла значение $y_1 = 0{,}4$

$$P_{y1}(x_1) = \frac{P(x_1, y_1)}{P(y_1)} = \frac{0{,}15}{0{,}80} = \frac{3}{16},$$

$$P_{y1}(x_2) = \frac{P(x_2, y_1)}{P(y_1)} = \frac{0{,}30}{0{,}80} = \frac{3}{8},$$

$$P_{y1}(x_3) = \frac{P(x_3, y_1)}{P(y_1)} = \frac{0{,}35}{0{,}80} = \frac{7}{16}$$

Напишем искомый условный закон распределения *X*:

X	2	5	8
$P_{y1}(x_i)$	3/16	3/8	7/16

Контроль: 3/16 + 3/8 + 7/16 = 1

Аналогично найдем условный закон распределения *Y*:

Y	0,4	0,8
$P_{x2}(y_j)$	5/7	2/7

Контроль: 5/7 + 2/7 = 1.

Так как условная вероятность события y_j при условии выполнения события x_i принимается по определению

$$P_{x_i}(y_j) = \frac{P(x_i, y_j)}{P(x_i)},$$

то вероятность совместного появления совокупности состояний

$$P(x_i, y_i) = P(x_i) P_{xi}(y_i)$$

Аналогично, условимся вероятность события x_i при условии выполнения события y_j:

$$P(x_i, y_j) = P(y_j) P_{yi}(x_i)$$

Поэтому общую энтропию зависимых ансамблей X и Y определяют по формуле Шеннона:

$$H(X,Y) = -\sum_{i=1}^{r}\sum_{j=1}^{s} P(x_i, y_j)\log P(x_i, y_j) =$$

$$= \sum_{i=1}^{r}\sum_{j=1}^{s} P(x_i)P_{x_i}(y_j)\log[P(x_i)P_{x_i}(y_j)] =$$

$$= \sum_{i=1}^{r} P(x_i)\log P(x_i)\sum_{j=1}^{s} P_{x_i}(y_j) - \sum_{i=1}^{r} P(x_i)\sum_{j=1}^{s} P_{x_i}(y_j)\log P_{x_i}(y_j).$$

С учетом соотношения $\sum_{j=1}^{s} P_{x_i}(y_j) = 1$ получают

$$H(X,Y) = H(X) + H_X(Y),$$

где $H(X)$ – энтропия ансамбля X;

$H_X(Y)$ – условная энтропия ансамбля Y при условии, что сообщение ансамбля X известны:

$$H_X(Y) = -\sum_{i=1}^{r} P(x_i)\sum_{j=1}^{s} P_{x_i}(y_j)\log P_{x_i}(y_j)$$

Для независимых событий X и Y: $P_{xi}(y_i) = P(y_j)$ и поэтому $H_X(Y) = H(Y)$ и, следовательно,

$$H(X,Y) = H(X) + H(Y).$$

Если X и Y полностью зависимы, т.е. при появлении x_i неизбежно следует y_j, то $P(x_i, y_j)$ равна единице при $i = j$ и нулю при $i \neq j$. Поэтому $H_X(Y) = 0$, и, следовательно,

$$H(X,Y) = H(X),$$

т.е. при полной зависимости двух ансамблей один из них не вносит никакой информации.

Полученное выражение для условной энтропии

$$H_X(Y) = -\sum_{i=1}^{r} P(x_i) \sum_{j=1}^{s} P_{x_i}(y_j) \log P_{x_i}(y_j)$$

можно использовать и как информативную характеристику одного ансамбля X, элементы которого взаимно зависимы. Положив $Y = X$, получим

$$H' = -\sum_{i=1}^{r} P(x_i) \sum_{j=1}^{s} P_{x_i}(x_j) \log P_{x_i}(x_j).$$

Например, алфавит состоит из двух элементов 0 и 1. Если эти элементы равновероятны, то количество информации, приходящееся на один элемент сообщения: $H_0 = \log m = \log 2 = 1 бит$

Если же, например, $P(0) = s$, а $P(1) = j$, то

$$H = -\sum_{i=1}^{m} P_i \log P_i = -P(0)\log P(0) - P(1)\log P(1) =$$

$$= -\left(\frac{3}{4}\log\frac{3}{4} + \frac{1}{4}\log\frac{1}{4}\right) = 0{,}815$$

В случае же взаимной зависимости элементов, определяемой, например, условными вероятностями $P_0(0) = 2/3$; $P_0(1) = 1/3$; $P_1(0) = 1$; $P_1(1) = 0$, то условная энтропия

$$H' = -P(0)[P_0(0)\log P_0(0) + P_0(1)\log P_0(1)] - P(1)[P_1(0)\log P_1(0) +$$
$$+ P_1(1)\log P_1(1)] = -\frac{3}{4}\left(\frac{2}{3}\log\frac{2}{3} + \frac{1}{3}\log\frac{1}{3}\right) = 0{,}685a$$

Легко показать, что энтропия при взаимно зависимых элементах всегда меньше, чем при независимых, т.е. $H' < H$.

Пример 9: Закон распределения вероятностей системы, объединяющей зависимые источники информации *X* и *Y*, задан с помощью таблицы:

X \ Y	y_1	y_2	y_3
x_1	0,4	0,1	0
x_2	0	0,2	0,1
x_3	0	0	0,2

Определить энтропии $H(X)$, $H(Y)$, $H_X(Y)$, $H(X,Y)$.

Решение:

1. Вычислим безусловные вероятности $P(x_i)$ и $P(y_j)$ системы:

а) сложив вероятности «по строкам», получим вероятности возможных значений *X*:

$$P(x_1) = 0{,}5$$
$$P(x_2) = 0{,}3$$
$$P(x_3) = 0{,}2$$

б) сложив вероятности «по столбцам», получим вероятности возможных значений *Y*:

$$P(y_1) = 0{,}4$$
$$P(y_2) = 0{,}3$$
$$P(y_3) = 0{,}3$$

2. Энтропия источника информации *X*:

$$H(X) = -\sum_{i=1}^{r} P(x_i)\log P(x_i) = -(0{,}5\log 0{,}5 + 0{,}3\log 0{,}3 +$$
$$+ 0{,}2\log 0{,}2) = 1{,}485\text{бит}$$

3. Энтропия источника информации *Y*:

$$H(Y) = -\sum_{j=1}^{S} P(y_j)\log P(y_j) = -(0{,}4\log 0{,}4 + 0{,}3\log 0{,}3 +$$
$$+ 0{,}3\log 0{,}3) = 1{,}57\,бит$$

4. Условная энтропия источника информации *Y* при условии, что сообщения источника *X* известны:

$$H(Y) = -\sum_{i=1}^{r} P(x_i)\sum_{j=1}^{s} P_{x_i}(y_j)\log P_{x_i}(y_j).$$

Так как условная вероятность события y_j при условии выполнения события x_i принимается по определению

$$P_{x_i}(y_j) = \frac{P(x_i, y_j)}{P(x_i)},$$

поэтому найдем условные вероятности возможных значений *Y* при условии, что составляющая *X* приняла значение x_1:

$$P_{x_1}(y_1) = \frac{P(x_1, y_1)}{P(x_1)} = \frac{0{,}4}{0{,}5} = 0{,}8$$

$$P_{x_1}(y_2) = \frac{P(x_1, y_2)}{P(x_1)} = \frac{0{,}1}{0{,}5} = 0{,}2$$

$$P_{x_1}(y_3) = \frac{P(x_1, y_3)}{P(x_1)} = \frac{0}{0{,}5} = 0$$

Для x_2:

$$P_{x_2}(y_1) = \frac{P(x_2, y_1)}{P(x_2)} = \frac{0}{0{,}3} = 0$$

$$P_{x_2}(y_2) = \frac{P(x_2, y_2)}{P(x_2)} = \frac{0{,}2}{0{,}3} = 0{,}67$$

$$P_{x_2}(y_3) = \frac{P(x_2, y_3)}{P(x_2)} = \frac{0{,}1}{0{,}3} = 0{,}33$$

Для x_3:

$$P_{x_3}(y_1) = \frac{P(x_3, y_1)}{P(x_3)} = \frac{0}{0{,}2} = 0$$

$$P_{x_3}(y_2) = \frac{P(x_3, y_2)}{P(x_3)} = \frac{0}{0{,}2} = 0$$

$$P_{x_3}(y_3) = \frac{P(x_3, y_3)}{P(x_3)} = \frac{0{,}2}{0{,}2} = 1$$

Поэтому:

$$H_X(Y) = -[0{,}5(0{,}8\log 0{,}8 + 0{,}2\log 0{,}2) +$$
$$+ 0{,}3(0{,}67\log 0{,}67 + 0{,}33\log 0{,}33) + 0{,}2(1\log 1)] = 0{,}635$$

5. Аналогично, условная энтропия источника информации X при условии, что сообщения источника Y известны:

$$H_Y(X) = -\sum_{j=1}^{s} P(y_j) \sum_{i=1}^{r} P_{y_j}(x_i) \log P_{y_j}(x_i).$$

$$P_{y_j}(x_i) = \frac{P(x_i, y_j)}{P(y_j)};$$

Для y_1:

$$P_{y_1}(x_1) = \frac{P(x_1, y_1)}{P(y_1)} = \frac{0{,}4}{0{,}4} = 1$$

$$P_{y_1}(x_2) = \frac{P(x_2, y_1)}{P(y_1)} = \frac{0}{0{,}4} = 0$$

$$P_{y_1}(x_3) = \frac{P(x_3, y_1)}{P(y_1)} = \frac{0}{0{,}4} = 0$$

Для y_2:

$$P_{y_2}(x_1) = \frac{P(x_1, y_2)}{P(y_2)} = \frac{0{,}1}{0{,}3} = 0{,}33$$

$$P_{y_2}(x_2) = \frac{P(x_2, y_2)}{P(y_2)} = \frac{0{,}2}{0{,}3} = 0{,}67$$

$$P_{y_2}(x_3) = \frac{P(x_3, y_2)}{P(y_2)} = \frac{0}{0{,}3} = 0$$

Для y_3:

$$P_{y_3}(x_1) = \frac{P(x_1, y_3)}{P(y_3)} = \frac{0}{0{,}3} = 0$$

$$P_{y_3}(x_2) = \frac{P(x_2, y_3)}{P(y_3)} = \frac{0{,}1}{0{,}3} = 0{,}33$$

$$P_{y_3}(x_3) = \frac{P(x_3, y_3)}{P(y_3)} = \frac{0{,}2}{0{,}3} = 0{,}67$$

$$H_Y(X) = -[0{,}4(1\log 1) + 0{,}3(0{,}33\log 0{,}33 + 0{,}67\log 0{,}67) +$$
$$+ 0{,}3(0{,}33\log 0{,}33 + 0{,}67\log 0{,}67)] \approx 0{,}55 бит.$$

6. Общая энтропия зависимых источников информации X и Y:

$$H(X,Y) = -\sum_{i=1}^{r}\sum_{j=1}^{s} P(x_i, y_j)\log P(x_i, y_j) =$$
$$= -(0{,}4\log 0{,}4 + 0{,}1\log 0{,}1 + 0{,}2\log 0{,}2 + 0{,}1\log 0{,}1 + 0{,}2\log 0{,}2) =$$
$$= 0{,}529 + 0{,}332 + 0{,}464 + 0{,}332 + 0{,}464 = 2{,}12 \text{ бит}.$$

Проверим результат по формуле:

$$H(X,Y) = H(X) + H_X(Y) = 1{,}485 + 0{,}635 = 2{,}12 бит$$
$$H(X,Y) = H(Y) + H_Y(X) = 1{,}57 + 0{,}55 = 2{,}12 бит$$

Пример 10: Известны энтропии двух зависимых источников $H(X) = 5 бит$; $H(H) = 10 бит$. Определить, в каких пределах будет изменяться условная энтропия $H_X(Y)$ в максимально возможных пределах.

Решение: Уяснению соотношений между рассматриваемыми энтропиями источников информации способствует их графическое отображение.

При отсутствии взаимосвязи между источниками информации (рис. 6.5), $H_X(Y) = H(Y) = 10 бит$, $H_Y(X) = H(X) = 5 бит$, $H(X,Y) = H(X) + H(Y) = 5 + 10 = 15 бит$.

Т.е., когда источники независимы $H_X(Y) = H(Y) = 10 бит$, и поэтому принимают максимальное значение.

При полной зависимости двух источников один из них не вносит никакой информации, т.к. при появлении x_i неизбежно следует y_j, т.е. $P(x_i, y_j)$ равно единице при $i = j$ и нулю при $i \neq j$.

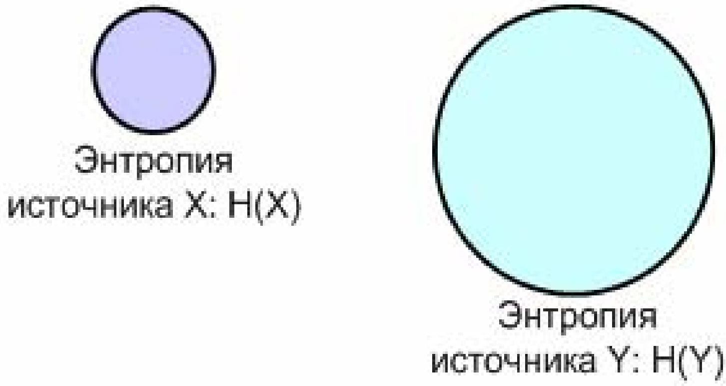


Рис. 6.5. Энтропии источников

По мере увеличения взаимосвязи источников $H_X(Y)$ и $H_Y(X)$ будут уменьшаться, рис. 6.6:

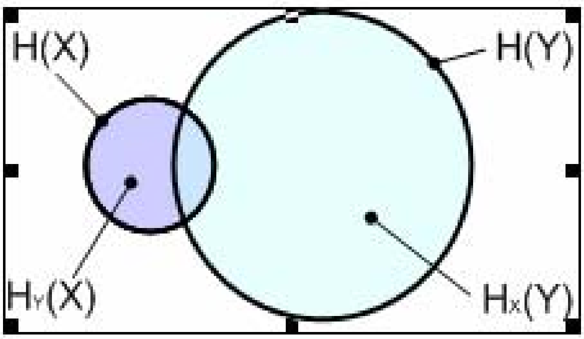


Рис. 6.6. Энтропии при взаимосвязи источников

Поэтому $H_Y(X) = 0$ и, следовательно, $H(X,Y) = H_X(Y)$, рис. 6.7.

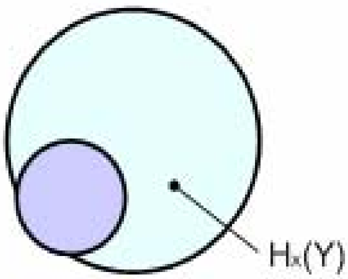


Рис. 6.7. Полная зависимость источников (один из них не вносит никакой информации)

При этом $H_X(Y) = H(Y) - H(X) = 10 - 5 = 5\textit{бит}$. Поэтому $H_X(Y)$ будет изменяться от $10\textit{бит}$ до $5\textit{бит}$ при максимально возможном изменении $H_Y(X)$ от $5\textit{бит}$ до $0\textit{бит}$.

Пример 11: Определите $H(X)$ и $H_X(Y)$, если $P(x_1, y_1) = 0{,}3$; $P(x_1, y_2) = 0{,}2$; $P(x_3, y_2) = 0{,}25$; $P(x_3, y_3) = 0{,}1$.

Пример 12: Определите $H(X)$, $H(Y)$, $H(X,Y)$, если $P(x_1, y_1) = 0{,}2$; $P(x_2, y_1) = 0{,}4$; $P(x_2, y_2) = 0{,}25$; $P(x_2, y_3) = 0{,}15$.

2) взаимная энтропия

Пусть ансамбли X и Y относятся к передаваемым и принимаемым сообщениям. Различия между X и Y обуславливаются искажениями сигналов и при отсутствии помех $H(X) = H(Y)$. Воздействие помех характеризуется условной энтропией $H_Y(X)$, так что получаемое потребителем количество информации на один элемент сообщения:

$$E(X,Y) = H(X) - H_Y(X).$$

Величину $E(X,Y)$ называют *взаимной энтропией.*

Если X и Y независимы, т.е. шумы в канале приводят к полному искажению сообщения, то $H_Y(X) = H(X)$ и $E(X,Y) = 0$.

Если X и Y полностью зависимы, т.е. шумы в канале отсутствуют, то $H_Y(X) = 0$ и $E(X,Y) = H(Y)$.

Так как $H_Y(X) = H(X,Y) - H(Y)$, то

$$E(X,Y) = H(X) + H(Y) - H(X,Y),$$

или

$$E(X,Y) = \sum_{i=1}^{r}\sum_{j=1}^{s} P(x_i, y_j)\log\frac{P_{y_j}(x_i)}{P(x_i)}.$$

Пример13: Определите $H(X)$ и $E(X,Y)$, если $P(x_1, y_1) = 0{,}3$; $P(x_1, y_2) = 0{,}2$; $P(x_2, y_3) = 0{,}1$; $P(x_3, y_2) = 0{,}1$; $P(x_3, y_3) = 0{,}25$ P(x_3,y_3) = = 0,25.

3) избыточность сообщений

Чем больше энтропия, тем большее количество информации содержит в среднем каждый элемент сообщения.

Пусть энтропии двух источников сообщений $H_1 < H_2$, а количество информации, получаемое от них одинаковое, т.е. $I = n_1 H_1 = n_2 H_2$, где n_1 и n_2 – длина сообщения от первого и второго источников.

Обозначим

$$\mu = \frac{n_2}{n_1} = \frac{H_1}{H_2}$$

При передаче одинакового количества информации сообщение тем длиннее, чем меньше его энтропия.

Величина μ, называемая *коэффициентом сжатия*, характеризует степень укорочения сообщения при переходе к кодированию состояний элементов, характеризующихся большей энтропией.

При этом доля излишних элементов оценивается *коэффициентом избыточности*:

$$r = \frac{H_2 - H_1}{H_2} = 1 - \frac{H_1}{H_2} = 1 - \mu .$$

Русский алфавит, включая пропуски между словами, содержит 32 элемента (см. пример), следовательно, при одинаковых вероятностях появления всех 32 элементов алфавита, неопределенность, приходящаяся на один элемент, составляет

$$H_0 = \log 32 = 5 бит$$

Анализ показывает, что с учетом неравномерного появления различных букв алфавита $P = 4{,}42 бит$, а с учетом зависимости двухбуквенных сочетаний $H' = 3{,}52 бит$, т.е.

$$H' < H < H_0 .$$

Обычно применяют три коэффициента избыточности:

1) частная избыточность, обусловленная взаимосвязью

$$r' = 1 - H'/H ;$$

2) частная избыточность, зависящая от распределения

$$r'' = 1 - H/H_0 ;$$

3) полная избыточность

$$r_0 = 1 - H' / H_0 .$$

Эти три величины связаны зависимостью

$$r_0 = r' + r'' - r' r'' .$$

Вследствие зависимости между сочетаниями, содержащими две и больше букв, а также смысловой зависимости между словами, избыточность русского языка (как и других европейских языков) превышает 50 % ($r_0 = 1 - H' / H_0 = 1 - 3{,}52 / 5 = 0{,}30$).

Избыточность устраняют построением оптимальных кодов, которые укорачивают сообщения по сравнению с равномерными кодами. В то же время избыточность играет и положительную роль, т.к. благодаря ей сообщения менее уязвимы со стороны помех. Это обстоятельство используют при помехоустойчивом кодировании.

6.3. Условная энтропия и взаимная информация

В технике связи очень часто представляет интерес выявление количества информации, содержащегося в одном ансамбле сообщений *и* с *объемом алфавита N* относительно другого в общем случае зависящего от него ансамбля *Z*, с объемом алфавита *M*. В качестве *U* и *Z* можно рассматривать, например ансамбль сообщений *U* и сигналов *Z*, с помощью которых передают сообщения *Z*. Для определения такой информационной характеристики введем *понятие условной энтропии*, которое будем обозначать $H(U/Z)$, определяющей среднее количество информации, даваемое сообщением ансамбля *и* при условии, что сообщение ансамбля *Z* уже известно. Если оба ансамбля имеют независимые элементы, то мера неопределенности $H(U/Z)$ находится усреднением по всем значениям Z_j средней неопределенности элементов ансамбля $H(U/Z_i)$ при данном Z_j. Последнее находится аналогично энтропии $H(u)$ заменой безусловных вероятностей

$P(U_k)$ (появление сообщения U_k) на условные вероятности появления U_k при условии Z_j $P(U_k / Z_j)$.

$$H(u/z) = \sum_{j=1}^{M} P(z_j) \cdot \underbrace{\sum_{j=1}^{N} P(u_k / z_j) \cdot \log \frac{1}{P(u_k / z_j)}}_{H(u/z_j)}. \qquad (6.7)$$

По теореме умножения вероятности имеем:

$$P(z_j) \cdot P(u_k / z_j) = P(u_k, z_j), \qquad (6.8)$$

где $P(u_k, z_j)$ – вероятность совместного появления сообщений u_k и z_j.

С учетом выражения (6.8), выражение (6.7) можно переписать в виде

$$H(u/z) = \sum_{j=1}^{M} \sum_{k=1}^{N} P(u_k, z_j) \cdot \log(P(u_k, z_j)/z_j). \qquad (6.9)$$

Возможно также другое представление (6.8) и (6.8)

$$H(u/z) = M\left\{\log \frac{1}{P(u_k / z_j)}\right\} = -\sum_{j=1}^{M} \sum_{k=1}^{N} P(u_k, z_j) \cdot \log(P(u_k, z_j), \quad (6.9а)$$

где $M\{\}$ – символ математического ожидания. Условная энтропия удовлетворяет неравенству

$$0 \le H(u/z) \le H(U), \qquad (6.10),$$

причем $H(u/z) = 0$, когда по реализации ансамбля Z можно точно установить реализацию ансамбля U (канал без помех). $H(u/z) \le H(U)$, когда ансамбли U и Z независимы и знание реализации Z ничего не говорит о реализации U. В общем случае $H(u/z) < H(U)$ и знание реализации Z снижает первоначальную неопределенность U. На основании этого можно ввести информационную характеристику двух ансамблей U и Z, называемую *взаимной информацией* между U и Z или *количеством информации*, содержащимся в Z относительно U, которая определяется, как

$$I(u,z) = H(U) - H(u/z), \qquad (6.11)$$

Взаимная информация измеряется в тех же единицах что и энтропия, например в битах. Величина $I(u,z)$ показывает, сколько в среднем бит информации о реализации ансамбля *u* дает наблюдение о реализации ансамбля *z*. Подставляя в (6.11), имеем:

$$I(u,z) = -\sum_{k=1}^{N} P(u_k) \cdot \log P(u_k) + \sum_{k=1}^{N}\sum_{j=1}^{M} P(u_k, z_j) \cdot \log \frac{P(u_k, z_j)}{P(z_j)}.$$

Учитывая, что $P(u_k) = \sum_{j=1}^{M} P(u_k, z_j)$, последнее выражение можно записать в виде

$$\begin{aligned} I(u,z) &= -\sum_{k=1}^{N}\sum_{j=1}^{M} P(u_k, z_j) \cdot \log P(u_k) + \\ &+ \sum_{k=1}^{N}\sum_{j=1}^{M} P(u_k, z_j) \cdot \log \frac{P(u_k, z_j)}{P(z_j)} = \\ &= \sum_{k=1}^{N}\sum_{j=1}^{M} P(u_k, z_j) \cdot \log \frac{P(u_k, z_j)}{P(u_k) \cdot P(z_j)}. \end{aligned} \tag{6.12}$$

Взаимная информация обладает следующими свойствами:

1) $I(u,z) \geq 0$ (6.13),

причем равенство имеет место только в том случае, когда *u* и *z* независимы между собой. Это следует из определения (6.11) и неравенства (6.10).

2) $I(u,z) = I(z,u)$ (6.14),

т.е *z* содержит столько же информации относительно *u*, сколько *u* содержит относительно *z*, это свойство вытекает из симметрии (6.12). Поэтому можно так же записать

$I(u,z) = H(Z) - H(z/u)$ (6.15).

$$\left.\begin{aligned} I(u,z) &\leq H(U) \quad (6.16) \\ I(u,z) &\leq H(Z) \quad (6.17) \end{aligned}\right\}.$$

Причем равенство имеет место, когда по реализации z можно точно восстановить реализацию u или наоборот. Это следует из (6.10) и (6.11). (6.16) вытекает из (6.14) и (6.17).

Полагая $Z = U$ и учитывая, что $H(u/u) = 0$, получаем $I(u,u) = H(U)$

Это позволяет интерпретировать энтропию источника как его собственную информацию ансамбля U о самом себе. Пусть U ансамбль дискретных сообщений, а Z ансамбль дискретных сигналов, в которые преобразуется сообщение U, тогда $I(u,z) = H(U)$ только в том случае когда преобразование U в Z обратимо, т.е. однозначно. При необратимом преобразовании $H(u/z) < H(U)$ и разность $H(U) - I(u,z) = H(u/z)$ называют потерей информации или ненадежностью преобразования U в Z. Таким образом, информация не теряется только при обратимых преобразованиях, величина $H(z/u) = H(Z) - I(u,z)$ называется энтропией шума преобразования или ложной информацией, создаваемой при образовании.

1) источники информации, энтропия источников, дискретный источник без памяти

Из определения энтропии $H(U)$ дискретного источника, характеризуемого ансамблем U следует, что величина $H(U)$ зависит от распределения вероятности ансамбля (свойство 2 энтропии). Энтропия максимальна только в том случае, когда все сообщения источника равновероятны. При генерировании источником последовательности сообщений существует еще один фактор, оказывающий влияние на величину энтропии, а именно наличия или отсутствия у источника памяти. Источник дискретных сообщений называется *источником с памятью*, если вероятность выдачи им очередного элементарного сообщения u_k зависит от того, какое (или какие) элементарные сообщения были выданы ранее. Иначе говоря, сообщения источника с памятью являются зависимыми. Стационарный источник независимых сообщений называется *источником без памяти*.

В качестве примера источника дискретных сообщений с памятью можно привести источник связанного русского текста, в качестве элементарных сообщений которого рассматриваются отдельные буквы русского алфавита. Наряду с тем, что различными являются вероятности появления разных букв в тексте (а чаще чем ъ) имеет место зависимость вероятности появления каждой буквы, от того какая буква ей предшествовала. Так сочетание букв «ар» может встретиться чаще чем «аъ».

В тоже время если передачи подлежат расчеты ЭВМ и в качестве элементарных сообщений выступают отдельные цифры, то есть основания в общем случае такой источник информации считать источником без памяти. Количество информации, содержащееся в одном элементарном сообщении источника с памятью, определяется с помощью условных вероятностей.

Следовательно, и энтропия такого источника, определяемая на основе (6.3.3) так же будет равна условной энтропии $H(u/z)$. Сообщения ансамбля U при условии, что ему предшествовало сообщение (или несколько сообщений) ансамбля Z. В соответствии со свойством условной энтропии (6.3.11) для случая зависимых ансамблей U и Z всегда, $H(u/z) < H(U)$ т.е. энтропия источника с памятью всегда меньше энтропии источника независимых сообщений. Таким образом, энтропия дискретного источника максимальна в том случае, когда выполняются 2 условия:

1) все сообщения источника независимы (источник без памяти).

2) все сообщения источника равновероятны.

Невыполнение любого из этих требований уменьшает энтропию источника и является причиной избыточности. *Избыточностью* источника дискретных сообщений с энтропией H_u и объемом алфавита N называется величина

$$\mu = \frac{H_{\max} - H_u}{H_{\max}} = 1 - \frac{H_u}{H_{\max}} = 1 - \frac{H_u}{\log N}, \tag{6.18}$$

где $H_{\max}$ – максимально возможное значение энтропии при данном объеме алфавита, оно достигается при выполнении условий 1) и 2)

$$H_{\max} = \log N .$$

Избыточность показывает, какая доля максимально возможной при заданном объеме алфавита неопределенности (энтропии) не используется источником. В частности избыточность современного английского текста составляет 50 %, избыточность русского текста 70 %.

Производительность источника дискретных сообщений. Скорость передачи информации.

Обычно источники передают сообщения, с некоторой скоростью, затрачивая в среднем время Т на передачу одного сообщения.

Производительностью источника $H'(U)$ назовем суммарную энтропию сообщений, переданных за единицу времени

$$H'(U) = \frac{1}{T} \cdot H(U). \tag{6.19}$$

Производительность измеряется в битах на секунду. Если сообщение может быть представлено в виде последовательности элементарных дискретных сообщений u_k источника с энтропией $H(U)$ следующих со скоростью $\upsilon_c = \dfrac{1}{T}$ элементов в секунду, то

$$H'(U) = \upsilon_c \cdot H(U). \tag{6.20}$$

Аналогичным образом, т.е. разделив формулы (6.10-6.17) на T и обозначая $H'(u/z) = \dfrac{1}{T} \cdot H(u/z)$, $I'(u/z) = \dfrac{1}{T} \cdot I(u,z)$, получим соответственные равенства для условных энтропии и количества информации, рассчитанных на одно сообщение a в единицу времени. Величина $I'(u,z)$ называется *скоростью передачи информации* от U к Z или наоборот. Если, например U ансамбль сигналов на входе дискретного канала, а Z ансамбль сигналов на его выходе то скорость передачи информации по каналу

$$I'(u/z) = H'(U) - H'(u/z) = H'(Z) - H'(z/u). \tag{6.21}$$

Это соотношение наглядно иллюстрируется на рисунке 6.7.

Здесь $H'(U)$ производительность источника передаваемого сигнала U, а $H'(Z)$ «производительность» канала, т.е. полная собственная информация в принятом сигнале за единицу времени. Величина $H'(u/z)$ представляет собой *потерю информации или ненадежность канала* в единицу времени, а $H'(z/u)$ скорость создания ложной, посторонней информации в канале, не имеющей отношение к U, и обусловленная присутствующими в канале помехами. По определению Шеннона ненадежность канала является энтропией входа, когда выход известен, т.е. ее можно считать мерой средней неопределенности принятого сигнала. Величина же $H'(z/u)$ есть энтропия выхода, когда вход известен и служит мерой средней неопределенности передаваемого сигнала. Соотношение между $H'(u/z)$ и $H'(z/u)$ зависит от свойств канала. Так, например, при передаче звукового сигнала по каналу с узкой полосой пропускания, недостаточной для качественного воспроизведения сигнала и с низким уровнем помех.

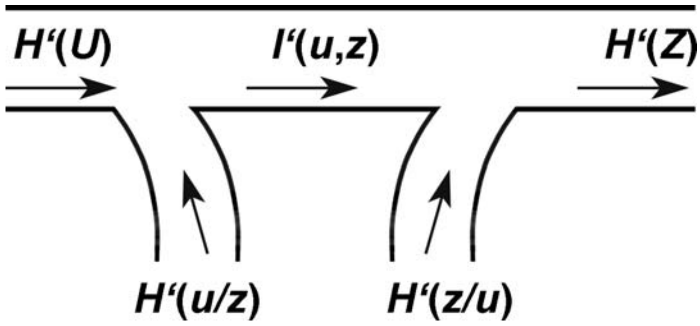


Рис. 6.7 Соотношение скоростей передачи сигналов по каналу

Теряется часть полезной информации, но почти не получается бесполезной, в этом случае $H'(u/z) >> H'(z/u)$. Если же сигнал воспроизводится качественно, но при этом прослушиваются наводки от соседнего радиоканала, то это означает, что почти не теряя полезной информации мы получили много лишней мешающей информации и $H'(u/z) << H'(z/u)$.

2) пропускная способность дискретного канала

В любой системе связи по каналу передается информация, скорость ее передачи определяется выражением (6.15), как видно из него эта скорость зависит не только от свойств самого канала, но и от подаваемого на его вход сигнала, и поэтому не может характеризовать канал как средство передачи информации. Попытаемся найти объективную характеристику способности канала передавать информацию. Рассмотрим дискретный канал, через который в единицу времени передается υ_k символов источника с объемом алфавита M. При передаче каждого символа в среднем по каналу проходит количество информации $I(u/z) = H(U) - H(u/z) = H(Z) - H(z/u)$, где U и Z ансамбли сообщений на входе и выходе канала. Из четырех фигурирующих здесь энтропий лишь $H(U)$ – собственная информация источника передаваемых символов, определяется источником входного сигнала и не зависит от свойств канала. Остальные три энтропии в общем случае зависят как от свойств источника, так и от канала. Представим себе, что на вход канала можно подавать символы от различных источников характеризуемых различными распределениями вероятностей $P(U)$ при одних и тех же значениях υ_k и M. Для каждого такого источника количество информации, переданной по каналу, принимает свое значение. Очевидно, существует какой-то источник входного сигнала с некоторым распределением $P(U)$ для которого величина $I(U,Z)$ максимальна. Максимальное количество переданной информации, взятое по всевозможным источникам входного сигнала, характеризует сам канал, и называется пропускной способностью канала в расчете на один символ.

$$C_{символ} = \max_{P(U)} I(u, z) \ \ бит/символ, \qquad (6.22, а)$$

где максимизация производится по всем возможным многомерным (т.е. учитывающим и статистическую взаимозависимость последовательно выдаваемых элементарных сообщений) распределением вероятностей $P(U)$). Обычно определяют пропускную способность в расчете на единицу времени.

$$C_{символ} = \max_{P(U)} I'(u, z) \ \ бит/символ, \qquad (6.22, б)$$

которую и называют просто *пропускной способностью канала*. Пропускная способность канала удовлетворяет системе неравенств

$$0 \le C \le \upsilon_k \cdot \log M, \qquad (6.23)$$

причем C=0 при независимых входе и выходе канала, т.е. $H(U/Z)=H(U)$ (обрыв канала или сильные помехи). Ограниченное значение

$$C = \upsilon_k \cdot \log M \qquad (6.23, а)$$

наблюдается в том случае, когда помех в канале нет $H(U/Z)=H(Z/U)=0$, при этом $H(U)=H(Z)=I(U,Z)$ если учесть что при заданном M $H_{\max}(U) = \log M$ (см. свойство 2 энтропии). Таким образом, пропускная способность дискретного канала без шума определяется равенством (6.23, *б*) при наличии шума

$$C < \upsilon_k \cdot \log M. \qquad (6.23, б)$$

В качестве примера вычислим пропускную способность дискретного симметричного канала без памяти. В таком канале каждый переданный кодовый символ может быть принят ошибочно с фиксированной вероятностью P и правильно с вероятностью 1-P, при чем в случае ошибки вместо переданного символа u_k может быть с равной вероятностью принят любой другой символ. Таким образом, вероятность того, что принят символ Z_j при условии, что передавался символ U_k, равна

$$P(z_j/u_k) = \begin{cases} \dfrac{P}{N-1} & при \quad j \neq k \\ 1-P & при \quad j = k \end{cases}, \tag{6.24}$$

где N – объем алфавита источника.

Термин без памяти означает, что вероятность ошибки в канале не зависит от того, какие символы передавались ранее и как они были приняты. В соответствии с (6.22, а) $C_{символ} = \max\limits_{P(U)} H(Z) - H(z/u)$. В соответствии с (6.17) и (6.24) имеем

$$\begin{aligned} H(z/u) &= \sum_{k=1}^{N} P(u_k) \cdot \sum_{j=1}^{N} P(z_j/u_k) \cdot \log \frac{1}{P(z_j/u_k)} = \\ &= \sum_{k=1}^{N} P(u_k) \cdot (1-P) \log \frac{1}{1-P} + \sum_{k=1}^{N} P(u_k) \cdot \sum_{j=1,\, j\neq k}^{N} \frac{P}{N-1} \cdot \log \frac{N-1}{P} = \\ &= (1-P) \log \frac{1}{1-P} \sum_{k=1}^{N} P(u_k) + \frac{P}{N-1} \log \frac{N-1}{P} \cdot N-1 \cdot \underbrace{\sum_{k=1}^{N} P(u_k)}_{1} = \\ &= -(1-P)\log(1-P) - P\log\frac{P}{N-1}, \end{aligned} \tag{6.24, а}$$

итак $C_{символ} = \max\limits_{P(U)} \left[H(Z) + (1-P)\log(1-P) + P\log\frac{P}{N-1} \right]$.

В правой части от $P(U)$ зависит только $H(Z)$ следует минимизировать только ее. В соответствии со *вторым свойством энтропии* максимальное значение $H(Z) = \log N$ и реализуется оно тогда когда все принятые сигналы Z_j равновероятные и независимые. Легко убедиться, что это условие выполняется, когда $P(u_k) = \frac{1}{N}$,

$$P(Z_j) = \sum_{k=1}^{N} P(U_j) \cdot P\left(Z_j \big/ U_k \right) = \sum_{k=1}^{N} \frac{1}{N} \cdot P\left(Z_j \big/ U_k \right) = \frac{1}{N} \cdot \sum_{k=1}^{N} P\left(Z_j \big/ U_k \right).$$

При каждом значении j для одного из слагаемых этой суммы соответствующего $P(z_j / u_k) = 1 - P$, при $j=k$, а для всех остальных $P(z_j / u_k) = \frac{P}{N-1}$. При этом

$$H(Z) = \log N \text{ и } C_{символ} = \log N + (1-P)\log(1-P) + P\log\frac{P}{N-1}. \quad (6.25)$$

Отсюда пропускная способность канала

$$C = \upsilon_k \left[\log N + (1-P)\log(1-P) + P\log\frac{P}{N-1} \right]. \quad (6.25, а)$$

Для двоичного симметричного канала N=2 и (6.25, *а*) принимает вид

$$C = \upsilon_k \left[1 + (1-P)\log(1-P) + P\log P \right]. \quad (6.26)$$

Зависимость, $\frac{C}{\upsilon} = f(P)$ показана на рисунке 6.8.

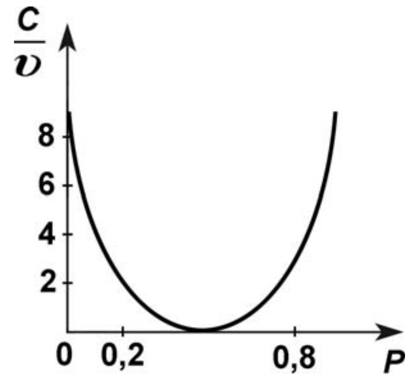


Рис. 6.8. Общий вид зависимости $f(P)$

То, что пропускная способность при P=1 в двоичном канале такая же, что и при P=0 (канал без шумов) объясняется тем, что при

P=1 достаточно все выходные символы инвертировать, чтобы правильно восстановить исходный сигнал.

Контрольные вопросы

1. Раскрыть содержание базового понятия энтропии в теории передачи сообщений.
2. В чем различие существующих подходов в определении энтропии и информации при описании процесса передачи сообщений?
3. Представить вероятностную математическую модель порождения сообщений.
4. Как определяется относительное количество информации, содержащейся в сообщении?
5. Сформулировать аксиомы Хинчина.
6. Сформулировать систему аксиом Фаддеева.
7. В чем различие аксиом Хинчина и аксиом Фаддеева?
8. Сформулировать и раскрыть содержание свойств энтропии.
9. Как оценить энтропию при непрерывном сообщении?
10. Что такое приведенная энтропия?
11. Раскрыть содержание условной энтропии.
12. Как рассчитывается энтропия источника информации?
13. Представить выражения для определения условной энтропии источника информации Y при условии, что сообщения источника X известны.
14. Как определяется общая энтропия зависимых источников информации X и Y?
15. Как определяется взаимная энтропия источников информации?
16. Что такое избыточность сообщений?
17. Как определяется коэффициент сжатия и коэффициент избыточности информации?
18. Раскрыть содержание понятий условной энтропии и взаимной информации.
19. В чем сущность определения энтропии дискретного источника информации?
20. При каких условиях энтропия дискретного источника максимальна?
21. Дать определение и представить выражения для определения избыточности источника дискретных сообщений.

22. Что такое производительность источника сообщений и как она определяется?

23. Как определяется соотношение скоростей передачи сигналов по каналу связи?

24. Как определить пропускную способность дискретного канала связи?

ГЛАВА 7. КОДИРОВАНИЕ ИНФОРМАЦИИ

7.1. Постановка задачи кодирования. Теоремы Шеннона об источниках информации

Теорема Шеннона – это одна из основных теорем теории информации о передаче сигналов по каналам связи при наличии помех, приводящих к искажениям. Пусть надлежит передать последовательность символов, появляющихся с определёнными вероятностями, причём имеется некоторая вероятность того, что передаваемый символ в процессе передачи будет искажён. Простейший способ, позволяющий надёжно восстановить исходную последовательность по получаемой, состоит в том, чтобы каждый передаваемый символ повторять большое число (N) раз. Однако это приведёт к уменьшению скорости передачи в N раз, т. е. сделает её близкой к нулю.

Шеннон также утверждает, что можно указать такое, зависящее только от рассматриваемых вероятностей положительное число v, что при сколько угодно малом $e > 0$ существуют способы передачи со скоростью $v'(v' < v)$, сколь угодно близкой к v, дающие возможность восстанавливать исходную последовательность с вероятностью ошибки, меньшей e. В то же время при скорости передачи v', большей v, это уже невозможно. Упомянутые способы передачи используют надлежащие «помехоустойчивые» коды. Критическая скорость v определяется из соотношения $Hv = C$, где H – энтропия источника на символ, C – ёмкость канала в двоичных единицах в секунду.

Теоремы Шеннона затрагивают проблему эффективного *символьного кодирования*.

Первая теорема декларирует возможность создания системы эффективного кодирования дискретных сообщений, у которой среднее число двоичных символов на один символ сообщения асимптотически стремится к энтропии источника сообщений (в отсутствии помех).

Вторая теорема Шеннона гласит, что при наличии помех в канале всегда можно найти такую систему кодирования, при которой сообщения будут переданы с заданной достоверностью.

Теория кодирования информации является одним из разделов теоретической информатики. К основным задачам, решаемым в данном разделе, необходимо отнести следующие:

– разработка принципов *наиболее экономичного* кодирования информации;

– согласование параметров передаваемой информации с особенностями канала связи;

– разработка приемов, обеспечивающих *надежность* передачи информации по каналам связи, т.е. отсутствие потерь информации.

Две последние задачи связаны с процессами передачи информации. Первая же задача – кодирование информации – касается не только передачи, но и обработки, и хранения информации, т.е. охватывает широкий круг проблем; частным их решением будет представление информации в компьютере.

Для представления дискретных сообщений используется некоторый алфавит. Однако однозначное соответствие между содержащейся в сообщении информацией и его алфавитом отсутствует.

В целом ряде практических приложений возникает необходимость перевода сообщения хода из одного алфавита к другому, причем, такое преобразование не должно приводить к потере информации.

Введем ряд определений.

Будем считать, что источник представляет информацию в форме дискретного сообщения, используя для этого алфавит, который в дальнейшем условимся называть *первичным.* Далее это сообщение попадает в устройство, преобразующее и представляющее его в другом алфавите – этот алфавит назовем *вторичным.*

Код – (1) правило, описывающее соответствие знаков или их сочетаний первичного алфавита знакам или их сочетаниям вторичного алфавита;

(2) набор знаков вторичного алфавита, используемый для представления знаков или их сочетаний первичного алфавита.

Кодирование – перевод информации, представленной сообщением в первичном алфавите, в последовательность кодов.

Декодирование – операция, обратная кодированию, т.е. восстановление информации в первичном алфавите по полученной последовательности кодов.

Кодер – устройство, обеспечивающее выполнение операции кодирования.

Декодер – устройство, производящее декодирование.

Операции кодирования и декодирования называются *обратимыми***,** если их последовательное применение обеспечивает возврат к исходной информации без каких-либо ее потерь.

Примером обратимого кодирования является представление знаков в телеграфном коде и их восстановление после передачи. Примером кодирования необратимого может служить перевод с одного естественного языка на другой – обратный перевод, вообще говоря, не восстанавливает исходного текста. Безусловно, для практических задач, связанных со знаковым представлением информации, возможность восстановления информации по ее коду является необходимым условием применения кода, поэтому в дальнейшем изложении ограничим себя рассмотрением только обратимого кодирования.

Кодирование предшествует передаче и хранению информации. При этом, как указывалось ранее, хранение связано с фиксацией некоторого состояния носителя информации, а передача – с изменением состояния с течением времени (т.е. процессом). Эти состояния или сигналы будем называть *элементарными сигналами* – именно их совокупность и составляет вторичный алфавит.

Не обсуждая технических сторон передачи и хранения сообщения (т.е. того, каким образом фактически реализованы передача-прием последовательности сигналов или фиксация состояний), попробуем дать математическую постановку задачи кодирования.

Пусть первичный алфавит A состоит из N знаков со средней информацией на знак $I^{(A)}$, а вторичный алфавит B – из M знаков со средней информацией на знак $I^{(B)}$. Пусть также исходное сообщение, представленное в первичном алфавите, содержит n знаков, а закодированное сообщение – m знаков. Если исходное сообщение содержит $I_{st}(A)$ информации, а закодированное – $I_{fin}(B)$, то условие обратимости кодирования, т.е. *неисчезновения* информации при кодировании, очевидно, может быть записано следующим образом:

$$I_{st}(A) \le I_{fin}(B),$$

смысл которого в том, что *операция обратимого кодирования может увеличить количество информации в сообщении, но не может его уменьшить*. Однако каждая из величин в данном неравенстве может быть заменена произведением числа знаков на среднее информационное содержание знака, т.е.:

$$n \cdot I^{(A)} \le m \cdot I^{(B)}$$

или

$$I^{(A)} \le \frac{m}{n} \cdot I^{(B)}.$$

Отношение *m/n,* очевидно, характеризует среднее число знаков вторичного алфавита, которое приходится использовать для кодирования одного знака первичного алфавита – будем называть его *длиной кода* или *длиной кодовой цепочки* и обозначим *К(A,B).* Следовательно

$$K(A,B) \geq \frac{I^{(A)}}{I^{(B)}}. \tag{7.1}$$

Обычно *N*>*M* и *I*(*A*)>*I*(*B*), откуда *К*(*A*,*B*) > 1, т.е. один знак первичного алфавита представляется несколькими знаками вторичного. Поскольку способов построения кодов при фиксированных алфавитах *А* и *В* существует множество, возникает проблема выбора (или построения) наилучшего варианта – будем называть его *оптимальным кодом*. Выгодность кода при передаче и хранении информации – это экономический фактор, так как более эффективный код позволяет затратить на передачу сообщения меньше энергии, а также времени и, соответственно, меньше занимать линию связи; при хранении используется меньше площади поверхности (объема) носителя. При этом следует сознавать, что выгодность кода не идентична временной выгодности всей цепочки кодирование-передача-декодирование; возможна ситуация, когда за использование эффективного кода при передаче придется расплачиваться тем, что операции кодирования и декодирования будут занимать больше времени и иных ресурсов (например, места в памяти технического устройства, если эти операции производятся с его помощью).

Как следует из (7.1), минимально возможным значением средней длины кода будет:

$$K^{\min}(A,B) \geq \frac{I^{(A)}}{I^{(B)}}. \tag{7.2}$$

Данное выражение следует воспринимать как соотношение оценочного характера, устанавливающее нижний предел длины кода, однако, из него неясно, в какой степени в реальных схемах кодирования возможно приближение *К*(*A*,*B*) к $K^{\min}(A,B)$. По этой причине для теории кодирования и теории связи важнейшее значение имеют две теоремы, доказанные Шенноном. Первая – ее мы сейчас рассмотрим – затрагивает ситуацию с кодированием при отсутствии помех, искажающих сообщение. Первая теорема Шеннона, которая называется *основной теоремой о кодировании при отсутствии помех*, формулируется следующим образом:

При отсутствии помех всегда возможен такой вариант кодирования сообщения, при котором среднее число знаков кода, приходящихся на один знак первичного алфавита, будет сколь угодно близко к отношению средних информации на знак первичного и вторичного алфавитов.

Приведенное утверждение является теоремой и, следовательно, должно доказываться. Для интересующихся именно доказательной стороной можно обратиться к книге А. М. Яглома и И. М. Яглома. Чрезвычайно важно, что теорема открывает принципиальную возможность оптимального кодирования, т.е. построения кода со средней длиной $K^{\min}(A,B)$. Однако необходимо сознавать, что из самой теоремы никоим образом не следует, как такое кодирование осуществить практически – для этого должны привлекаться какие-то дополнительные соображения, что и станет предметом нашего последующего обсуждения.

Имеются два пути сокращения $K^{\min}(A,B)$:

– уменьшение числителя – это возможно, если при кодировании учесть различие частот появления разных знаков в сообщении, корреляции двухбуквенные, трехбуквенные и т.п. (в п.6.3. было показано, что $I_0 > I_1 > \mathrm{I}_2 > \ldots > I_\infty$);

– увеличение знаменателя – для этого необходимо применить такой способ кодирования, при котором появление знаков вторичного алфавита было бы равновероятным, т.е. $I^{(B)} = \log_2 M$.

В частной ситуации, рассмотренной подробно К. Шенноном, при кодировании сообщения в первичном алфавите учитывается различная вероятность появления знаков, однако их корреляции не отслеживаются – источники подобных сообщений называются *источниками без памяти*. Если при этом обеспечена равная вероятность появления знаков вторичного алфавита, то, как следует из (3.2), для минимальной средней длины кода оказывается справедливым соотношение:

$$K^{\min}(A,B) \geq \frac{I_1^{(A)}}{\log M}. \tag{7.3}$$

В качестве меры превышения $K(A,B)$ над $K^{\min}(A,B)$ можно ввести *относительную избыточность кода* ($Q(A,B)$:

$$Q(A,B) = \frac{K(A,B) - K^{\min}(A,B)}{K^{\min}(A,B)} = \frac{K(A,B)}{K^{\min}(A,B)} - 1 = \frac{K(A,B) \cdot I^{(B)}}{I^{(A)}} \quad (7.4)$$

Данная величина показывает, насколько операция кодирования увеличила длину исходного сообщения. Очевидно, $Q(A,B) \to 0$ при $K(A,B) \to K^{\min}(A,B)$. Следовательно, решение проблемы оптимизации кода состоит в нахождении таких схем кодирования, которые обеспечили бы приближение средней длины кода к значению $K^{\min}(A,B)$, равному отношению средних информации на знак первичного и вторичного алфавитов. Легко показать, что чем меньше $Q(A,B)$, тем $I_{fin}(B)$ ближе к $I_{st}(A)$, т.е. возникает меньше информации, связанной с кодированием, более выгодным оказывается код и более эффективной операция кодирования.

Используя понятие избыточности кода, можно построить иную формулировку теоремы Шеннона:

При отсутствии помех всегда возможен такой вариант кодирования сообщения, при котором избыточность кода будет сколь угодно близкой к нулю.

Наиболее важной для практики оказывается ситуация, когда $M = 2$, т.е. для представления кодов в линии связи используется лишь два типа сигналов – технически это наиболее просто реализуемый вариант (например, существование напряжения в проводе (будем называть это *импульсом*) или его отсутствие (пауза); наличие или отсутствие отверстия на перфокарте или намагниченной области на дискете); подобное кодирование называется *двоичным*. Знаки двоичного алфавита принято обозначать «0» и «1», но нужно воспринимать их как буквы, а не цифры. Удобство двоичных кодов и в том, что при равных длительностях и вероятностях каждый элементарный сигнал (0 или 1) несет в себе 1 бит информации ($\log_2 M = 1$); тогда из (7.3)

$$K^{\min}(A,2) = I_1^{(A)},$$

и первая теорема Шеннона получает следующую интерпретацию:

При отсутствии помех средняя длина двоичного кода может быть сколь угодно близкой к средней информации, приходящейся на знак первичного алфавита.

Для двоичных сообщений источника без памяти при кодировании знаками равной вероятности получаем:

$$Q(A,2)=\frac{K(A,2)}{I_1^{(A)}}=1. \qquad (7.5)$$

При декодировании двоичных сообщений возникает проблема выделения из потока сигналов (последовательности импульсов и пауз) кодовых слов (групп элементарных сигналов), соответствующих отдельным знакам первичного алфавита. При этом приемное устройство фиксирует *интенсивность* и *длительность* сигналов, а также может соотносить некоторую последовательность сигналов с эталонной (*таблицей кодов*).

Возможны следующие особенности вторичного алфавита, используемого при кодировании:

– элементарные сигналы (0 и 1) могут иметь одинаковые длительности ($\tau_0 = \tau_1$) или разные ($\tau_0 \neq \tau_1$);

– длина кода может быть одинаковой для всех знаков первичного алфавита (в этом случае код называется *равномерным*) или же коды разных знаков первичного алфавита могут иметь различную длину (*неравномерный код*);

– коды могут строиться для отдельного знака первичного алфавита (*алфавитное кодирование*) или для их комбинаций (*кодирование блоков, слов*).

Комбинации перечисленных особенностей определяют основу конкретного способа кодирования, однако даже при одинаковой основе возможны различные варианты построения кодов, отличающихся своей эффективностью. Нашей ближайшей задачей будет рассмотрение различных схем кодирования для некоторых основ.

7.2. Марковские и эргодические источники

1) марковские источники

Определение 7.1. Дискретный стационарный источник [A, $\tilde{p}(s)$] называется Марковским источником порядка m, если для любого $l(l \geq m)$ и любой последовательности $c_l = (a_{i1},\ldots,a_{il})$ выполняется:

$$p(a_{il}/a_{il-1},\ldots,a_{i1}) = p(a_{il}/a_{il-1},\ldots,a_{il-m+1}). \qquad (7.6)$$

Из определения следует, что последовательности $\{C_l\}$ являются реализациями конечной стационарной цепи Маркова с глубиной зависимости m. Ниже будем предполагать, что цепь Маркова неразложимая и ациклическая.

Определение 7.2. Величина

$$H^{(k)} = -\sum_{\{C_k\}} p(a_{i1},...,a_{ik}) \log p\left(a_{ik}/a_{ik-1},...,a_{i1}\right) \tag{7.7}$$

называется шаговой энтропией марковского источника порядка k.

Определение 7.3. Величина

$$H_k = -\frac{1}{k}\sum_{\{C_k\}} p(a_{i1},...,a_{ik}) \log p\left(a_{i1},...,a_{ik}\right) \tag{7.8}$$

называется энтропией источника на один знак.

Рассмотрим соотношение:

$$p(a_{i1},...,a_{ik}) = p(a_{i1})\, p(a_{i2}/a_{i1})\, p(a_{i3}/a_{i2}\ a_{i1}) ... p(a_{ik}/a_{ik-1},...,a_{i1}).$$

Логарифмируя его, усредняя по множеству $\{C_k\}$, умножая на $-\dfrac{1}{k}$, получаем связь шаговой энтропии и энтропии «на знак»:

$$H_k = -\frac{1}{k}\sum_{\{C_k\}} p(a_{i1},...,a_{ik}) \log p\left(a_{i1}\right) - \frac{1}{k}\sum_{\{C_k\}} p(a_{i1},...,a_{ik}) \log p\left(a_{i2}/a_{i1}\right) -$$
$$-\frac{1}{k}\sum_{\{C_k\}} p(a_{i1},...,a_{ik}) \log p\left(a_{ik}/a_{ik-1},...,a_{i1}\right) = \frac{1}{k}\sum_{i=1}^{k} H^{(i)}$$

Покажем, что последовательность $H^{(k)}$ $k = 1,2,...$ является невозрастающей, то есть для любого $k \in N$ $H^{(k+1)} \le H^{(k)}$.

Действительно:

$$H^{(k+1)} = -\sum_{\{C_k\}} p(a_{i1},...,a_{ik+1}) \log p\left(a_{ik+1}/a_{ik},...,a_{i1}\right) =$$
$$= H\left(a_{ik+1}/a_{ik},...,a_{i1}\right) \le H\left(a_{ik+1}/a_{ik},...,a_{i2}\right) = H\left(a_{ik}/a_{ik-1},...,a_{i1}\right) = H^{(k)}.$$

Последовательность H_k также является невозрастающей:

$$H_k - H_{k+1} = \frac{(k+1)\sum_{i=1}^{k} H^{(i)} - k\sum_{i=1}^{k} H^{(i)}}{k(k+1)} = \frac{1}{k(k+1)} \sum_{i=1}^{k} (H^{(i)} - H^{(k+1)}) \geq 0.$$

Определение 7.4. Величина $\lim_{k\to\infty} H^{(k)} = \lim_{k\to\infty} H_k = H_\infty \geq 0$ – называется энтропией Марковского источника.

Первый предел существует в силу теоремы Вейерштрасса, второй – как среднее арифметическое членов последовательности $\{H^{(k)}\}$, имеющей конечный предел.

В проведённых рассуждениях зависимость Марковского типа могла распространяться на произвольную глубину m. Рассмотрим некоторые частные случаи.

$m = 0$. Зависимость между отдельными знаками отсутствует. Марковский источник является источником без памяти.

$$H^{(k)} = H_k - H_1 = H_\infty,$$

где $H_1 = -\sum_{i-1}^{n} p_i \log p_i$.

$m = 1$. Появление очередного знака в последовательности зависит от предыдущего знака.

$$H^{(k)} = H^{(2)}, \text{ при } k \geq 2.$$

$$\lim_{k\to\infty} H^{(k)} = H^{(2)} = H_\infty = -\sum_{\{C_k\}} p(a_{i1}a_{i2}) \log p\left(a_{i2}/a_{i1}\right).$$

Энтропия на знак для этого случая представляется выражением:

$$H_k = \frac{1}{k}\left[H^{(1)} + (k-1)H^{(2)}\right].\lim_{k\to\infty} H_k = H^{(2)} = H_\infty.$$

Пусть зависимость между знаками в последовательностях $\{c_k\}$ распространяется на глубину m. Тогда $H^{(k)} = H^{(m)}$, при $k \geq m$.

$$\lim_{k\to\infty} H_k = H^{(m)} = H_\infty = -\sum_{\{C_m\}} p(a_{i1},...,a_{im}) \log p\left(a_{im}/a_{im-1},...,a_{i1}\right). \quad (7.9)$$

$$H_k = \frac{1}{k}\left[\sum_{i=1}^{m-1} H^{(i)} + (k-m+1)H^{(m)}\right] \lim_{k\to\infty} H_k = H^{(m)} = H_\infty.$$

Рассмотрим источник сообщений $[\text{А}, \tilde{p}(\text{s})]$, порождающий символы алфавита А согласно простой, стационарной и эргодической цепи Маркова с конечным числом состояний. Обозначим так же, как и в §2, посредством $C_l = \{c_l\}$ совокупность всевозможных последовательностей длины *l*, порождённых этим источником.

Теорема 1. (Первая теорема Шеннона для марковских источников порядка $m=1$). Для любых $\varepsilon > 0$, $\eta > 0$ существует l_0 такое, что при $l>l_0$ все реализации длины l источника $[\text{А}, \tilde{p}(\text{s})]$ могут быть разбиты на два класса:

$$C_l = C_l' + C_l''.$$

Для любой последовательности $c_l' \in C_l'$ имеет место:

$$\left| \frac{1}{l} \log \frac{1}{p(c_l')} - H_\infty \right| < \eta, \tag{7.10}$$

где $H_\infty = -\sum\limits_{(i,j)} p_i p_{ij} \log p_{ij} = -\sum\limits_{(i,j)} p(ij) \log p_{ij}$ – энтропия источника сообщений. Суммарная вероятность последовательностей из класса C_l'' меньше ε.

Доказательство. Зададимся произвольными $\varepsilon > 0$, $\eta > 0$. Множество всевозможных последовательностей C_l, порождённых рассматриваемым источником, разделим на два класса. Отнесём к первому классу C_l' те и только те последовательности, для которых выполнено неравенство:

$$\left| m_{ij} - l p_i p_{ij} \right| < l\delta, \quad i, j = 1, \ldots, n, \tag{7.11}$$

где m_{ij} – абсолютная частота встречаемости биграммы $(a_i a_j)$. Ко второму классу C_l'' отнесём все последовательности, для которых неравенство (7.11) не выполнено, по крайней мере, для одной пары (ij). Условие (7.11) равносильно выполнению системы неравенств:

$$lP_i P_{ij} - l\delta < m_{lj} < lP_i P_{ij} + lb. \quad i, j = 1, \ldots, n.$$

Эта система эквивалентна следующему представлению абсолютных частот m_{ij} появления биграмм $(a_i\, a_j)$:

$$m_{lj} = lP_i P_{ij} + l\delta\theta_{ij}, \; (\left|\theta_{ij}\right| < 1, i, j = 1, \ldots, n).$$

Тогда вероятность порождения последовательности $c_l^{'} \in C_l^{'}$ равна:

$$p(c_l^{'}) = p_{ij} \prod_{(i,j)} p_{ij}^{lp_i p_{ij} + l\delta\theta_{ij}}$$

Логарифмируя равенство, получаем:

$$\log \frac{1}{p(c_l^{'})} = -\log p_{i_1} - \sum_{(i,j)} (lp_i p_{ij} + l\delta\theta_{ij}) \log p_{ij} =$$
$$= \log \frac{1}{p_{i_1}} + lH_\infty - l\delta \sum_{(i,j)} \theta_{ij} \log p_{ij}$$

Далее, оцениваем сверху модуль разности:

$$\left| \frac{1}{l} \log \frac{1}{p(c_l')} - H_\infty \right| < \frac{1}{l} \log \frac{1}{p_{i_1}} + \delta \sum_{(i,j)} \log \frac{1}{p_{ij}}. \quad (7.12)$$

При $l > l_0$ и достаточно малом δ из неравенства (3.12) следует первое утверждение теоремы:

$$\left| \frac{1}{l} \log \frac{1}{p(c_l^{'})} - H_\infty \right| \langle \eta,$$

где $H_\infty = -\sum_{(i,j)} p_i p_{ij} \log p_{ij}$

Ко второму классу $C_l^{''}$ отнесем последовательности, для которых неравенство (7.12) не выполнено, по крайней мере для одной пары (i,j). Для доказательства второго утверждения теоремы следует оценить сверху сумму

$$\sum_{(i,j)} p(\left| m_{ij} - lp_i p_{ij} \right| > \delta l) . \quad (7.13)$$

Зафиксируем пару (i,j). В силу эргодичности рассматриваемой цепи Маркова для произвольного $\varepsilon' > 0$ при достаточно большом l

будет выполняться закон больших чисел:

$$P\left\{|m_i - lp_i| < \frac{\delta}{2} l\right\} > 1 - \varepsilon'$$

$$P\left\{\left|\frac{m_{ij}}{m_i} - p_{ij}\right| < \frac{\delta}{2}\right\} > 1 - \varepsilon' \tag{7.14}$$

для абсолютных частот встречаемости символов и биаграмм.

Из неравенств (7.14) получаем:

$$P\left\{|m_i - lp_i| < \frac{\delta}{2} l; \left|\frac{m_{ij}}{m_i} - p_{ij}\right| < \frac{\delta}{2}\right\} > 1 - 2\varepsilon'.$$

Учитывая включение случайных событий

$$\left\{\left|\frac{m_{ij}}{m_i} - p_{ij}\right| < \frac{\delta}{2}\right\} = \left\{|m_{ij} - p_{ij} m_i| < \frac{\delta}{2} m_i\right\} \subseteq \left\{|m_{ij} - p_{ij} m_i| < \frac{\delta}{2} l\right\}.$$

получаем оценку

$$P\left\{|m_i - lp_i| < \frac{\delta}{2} l; |m_{ij} - p_{ij} m_i| < \frac{\delta}{2} l\right\} > 1 - 2\varepsilon'.$$

Вероятность дополнительного события оценивается сверху таким образом:

$$P\left\{|m_{ij} - lp_i p_{ij}| \geq \delta l\right\} \leq 2\varepsilon'.$$

Следовательно, для суммы (7.13) получаем:

$$\sum_{c_l'' \in C_l''} p(c_l'') \leq \sum_{(ij)} P\left\{|m_{ij} - lp_i p_{ij}| \geq \delta l\right\} < \varepsilon,$$

чем и завершается доказательство теоремы 1.

2) эргодические источники

При рассмотрении источников без памяти и марковских источников в первой теореме Шеннона была показана информационная устойчивость последовательностей, входящих в класс C_l' Имеется широкий класс источников более общей природы, для которых это свойство устойчивости также имеет место. Это эргодические источники сообщений.

Если дискретный источник рассматривать, как марковский процесс, то среди всех возможных марковских процессов можно выделить группу со свойствами, важными для теории связи. В этот класс входят так называемые «эргодические» процессы, и поэтому соответствующие источники называют эргодическими. Хотя строгое определение эргодического процесса достаточно сложно для восприятия, его основная идея проста. Для эргодического процесса все сгенерированные последовательности обладают одинаковыми статистическими свойствами, то есть, к примеру, частоты встречаемости букв, диаграмм и так далее, оцененные по отдельным последовательностям, сходятся с ростом длины выборок к определенным пределам, не зависящим от последовательности. На самом деле это верно не для всякой последовательности, однако множество последовательностей, для которых это не выполняется, имеет меру ноль (то есть обладает нулевой вероятностью). В общей постановке свойство эргодичности означает статистическую однородность.

Это связано со структурой соответствующих графов. Если граф обладает двумя следующими свойствами, соответствующий процесс будет эргодическим:

1. Граф не состоит из двух изолированных частей A и B, таких, что невозможно перейти с вершин одной части на вершины другой по линиям графа в разрешенном направлении, и с вершин второй – на вершины первой.

2. Назовем замкнутые серии ребер графа, которые можно обойти в разрешенном направлении, «контуром». Длиной контура назовем число ребер в нем. Вторым требуемым свойством является равенство наибольшего общего делителя длин всех контуров на графе единице.

Марковский источник называется *эргодическим,* если вероятность перехода через произвольное (большее некоторого фиксированного числа *m)* число шагов из каждого состояния S- в произвольное состояние s_J больше нуля.

Будем рассматривать квантование с равномерным шагом $\Delta x = const$, т.е. равномерное квантование. В процессе квантования неизбежно возникает ошибка квантования. Последовательность ошибок квантования, возникающая при квантовании процесса с дискретным временем, называется шумом квантования. Обычно шум квантования предполагают стационарным эргодическим случайным процессом.

Чаще всего интерес представляют максимальное значение ошибки квантования, ее среднее значение, равное математическому ожиданию шума и среднеквадратическое отклонение, равное квадратному корню из дисперсии шума (она характеризует мощность шума квантования). Все эти величины зависят от способа округления, применяемого при квантовании, а также от закона распределения мгновенных значений сигнала в пределах шага квантования.

Информационная дивергенция [Informational divergence] – функция, определенная для двух распределений вероятностей и характеризующая степень их близости. Широко используется в задачах теории информации.

Дивергенция (от лат. *divergere* – обнаруживать расхождение) – дифференциальный оператор, отображающий векторное поле на скалярное (то есть операция дифференцирования, в результате применения которой к векторному полю получается скалярное поле), который определяет (для каждой точки), «насколько расходится входящее и исходящее из малой окрестности данной точки поле» (точнее – насколько расходятся входящий и исходящий поток).

Если учесть, что потоку можно приписать алгебраический знак, то нет необходимости учитывать входящий и исходящий потоки по отдельности, всё будет автоматически учтено при суммировании с учетом знака. Поэтому можно дать более короткое определение дивергенции:

дивергенция – это дифференциальный оператор на векторном поле, характеризующий поток данного поля через поверхность малой окрестности каждой внутренней точки области определения поля.

Оператор дивергенции, применённый к полю *F*, обозначают как $divF$ или $\nabla \cdot F$.

Определение дивергенции выглядит следующим образом:

$$divF = \lim_{S \to 0} \frac{\Phi_F}{V},$$

где Φ_F – поток векторного поля *F* через сферическую поверхность площадью *S*, ограничивающую объем *V*. Еще более общим, а потому удобным в применении, является это определение, когда форма области с поверхностью *S* и объемом *V* допускается любой, единственным требованием является ее нахождение внутри сферы радиусом, стремящимся к нулю. Это определение, в отличие от приводимого ниже, не привязано к определенным координатам, например, к декартовым, что может представлять дополнительное удобство в определенных случаях. (Например, если выбирать окрестность в форме куба или параллелепипеда, легко получаются формулы для декартовых координат, приведенные в следующем параграфе).

3) определение в декартовых координатах

Допустим, что векторное поле дифференцируемо в некоторой области. Тогда в трёхмерном декартовом пространстве дивергенция будет определяться выражением

$$divF = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}.$$

Это же выражение можно записать с использованием *оператора набла*

$$divF = \nabla \cdot F .$$

Многомерная, а также двумерная и одномерная, дивергенция определяется в декартовых координатах в пространствах соответствующей размерности совершенно аналогично (в верхней формуле меняется лишь количество слагаемых, а нижняя остается той же, подразумевая оператор набла подходящей размерности).

4) физическая интерпретация

С точки зрения физики, дивергенция векторного поля является показателем того, в какой степени данная точка пространства является источником или стоком этого поля:

$divF > 0$ – точка поля является источником;

$divF < 0$ – точка поля является стоком;

$divF = 0$ – стоков и источников нет, либо они компенсируют друг друга.

Например, если в качестве векторного поля взять совокупность направлений наискорейшего спуска на земной поверхности, то дивергенция покажет местоположение вершин и впадин, причём на вершинах дивергенция будет положительная (направления спуска расходятся от вершин), а на впадинах отрицательная (ко впадинам направления спуска сходятся).

Ещё одним, быть может, несколько схематическим, примером может служить озеро (для простоты – постоянной единичной глубины с горизонтальной скоростью течения воды, не зависящей от глубины, давая, таким образом двумерное векторное поле на двумерном пространстве). Если угодно иметь более реалистическую картину, то можно рассмотреть горизонтальную проекцию скорости, проинтегрированную по вертикальной пространственной координате, что даст ту же картину двумерного векторного поля на двумерном пространстве, причем картина качественно будет для наших целей не сильно отличаться от упрощенной первой, количественно же являться ее обобщением (весьма реалистическим). В такой модели (и в первом, и во втором варианте) родники, бьющие из дна озера будут давать положительную дивергенцию поля скоростей течения, а подводные стоки (пещеры, куда вода утекает) – отрицательную дивергенцию.

Дивергенция вектора плотности тока дает минус скорость накопления заряда в обычной трехмерной физике (так как заряд сохраняется, то есть не исчезает и не появляется, а может только переместиться через границы какого-то объема, чтобы накопиться в нем или уйти из него; а если и возникают или исчезают где-то положительные и отрицательные заряды – то только в равных количествах).

5) свойства

Из обычных правил дифференцирования могут быть получены следующие свойства.

Линейность*:* для любых векторных полей F и G и для всех действительных чисел a и b

$$div(aF + bG) = adiv(F) + bdiv(G).$$

Если φ – скалярное поле, а F – векторное, тогда:

$$div(\varphi F) = grad(\varphi)\cdot F + \varphi div(F) \text{ или}$$
$$\nabla\cdot(\varphi F) = (\nabla\varphi)\cdot F + \varphi(\nabla\cdot F).$$

Свойство, связывающее векторные поля F и G, заданные в трёхмерном пространстве, с ротором:

$$div(F\times G) = rot(F)\cdot G - F\cdot rot(G), \text{ или}$$
$$\nabla\cdot(F\times G) = (\nabla\times F)\cdot G - F\cdot(\nabla\times G).$$

Дивергенция от градиента есть лапласиан:

$$div(grad(\varphi)) = \Delta\varphi.$$

Дивергенция от ротора:

$$div(rot(F)) = 0.$$

7.3. Эффективное кодирование

При кодировании каждая буква исходного алфавита представляется различными последовательностями, состоящими из кодовых букв (цифр).

Если исходный алфавит содержит m букв, то для построения равномерного кода с использованием k кодовых букв необходимо удовлетворить соотношение

$$m \le k_q ,$$

где q – количество элементов в кодовой последовательности.

Поэтому

$$q \ge \frac{\log m}{\log k} = \log_k m .$$

Для построения равномерного кода достаточно пронумеровать буквы исходного алфавита и записать их коды как q – разрядные числа в k-ичной системе счисления.

Например, при двоичном кодировании 32 букв русского алфавита используется $q = \log_2 32 = 5$ разрядов, на чем и основывается телетайпный код.

Кроме двоичных, наибольшее распространение получили восьмеричные коды.

Пусть, например, необходимо закодировать алфавит, состоящий из 64 букв. Для этого потребуется $q = \log_2 64 = 6$ двоичных разряда или $q = \log_8 64 = 2$ восьмеричных разряда. При этом буква с номером 13 при двоичном кодировании получает код 001101, а при восьмеричном кодировании 15.

Общепризнанным в настоящее время является позиционный принцип образования системы счисления. Значение каждого символа (цифры) зависит от его положения – позиции в ряду символов, представляющих число.

Единица каждого следующего разряда больше единицы предыдущего разряда в m раз, где m – основание системы счисления.

Полное число получают, суммируя значения по разрядам:

$$Q = \sum_{i=1}^{l} a_i m^{i-1} = a_l m^{l-1} + a_{l-1} m^{l-2} + \ldots + a_2 m^l + a_1 m^0,$$

где i – номер разряда данного числа; l – количество рядов; a_i – множитель, принимающий любые целочисленные значения в пределах от 0 до m–1 и показывающий, сколько единиц i – ого ряда содержится в числе.

Часто используются двоично-десятичные коды, в которых цифры десятичного номера буквы представляются двоичными кодами. Так, например, для рассматриваемого примера буква с номером 13 кодируется как 0001 0011.

Ясно, что при различной вероятности появления букв исходного алфавита равномерный код является избыточным, т.к. его энтропия (полученная при условии, что все буквы его алфавита равновероятны):

$$q = \log_k m = H_0$$

всегда больше энтропии $H = \log m$ данного алфавита (полученной с учетом неравномерности появления различных букв алфавита, т.е. информационные возможности данного кода используются не полностью).

Например, для телетайпного кода $H_0 = \log_k m = \log_{32} m = 5 бит$, а с учетом неравномерности появления различных букв исходного алфавита

$$H \approx 4{,}35\ бит.$$

Устранение избыточности достигается применением неравномерных кодов, в которых буквы, имеющие наибольшую вероятность, кодируются наиболее короткими кодовыми последовательностями, а более длинные комбинации присваиваются редким буквам.

Если i-я буква, вероятность которой P_i, получает кодовую комбинацию длины q_i, то средняя длина комбинации

$$q_{-p} = \sum_{i=1}^{m} p_i \cdot q_i\ .$$

Считая кодовые буквы равномерными, определяем наибольшую энтропию закодированного алфавита как $q_{cp} \log m$, которая не может быть меньше энтропии исходного алфавита H, т.е.

$$q_{cp} \log m \geq H\ .$$

Отсюда имеем

$$q_{-\text{р}} \geq \frac{\text{н}}{\log m}.$$

При двоичном кодировании (m=2) приходим к соотношению $q_{ср} \geq H$, или

$$\sum_{i=1}^{m} P_i q_i \geq -\sum_{i=1}^{m} P_i \log P_i .$$

Чем ближе значение $q_{ср}$ к энтропии H, тем более эффективно кодирование. В идеальном случае, когда $q_{ср} \approx H$, код называют *эффективным*. Оно устраняет избыточность, сокращает длины сообщений, уменьшая время передачи или объем памяти для их хранения.

При построении неравномерных кодов необходимо обеспечить возможность их однозначной расшифровки. В равномерных кодах такая проблема не возникает, т.к. при расшифровке достаточно кодовую последовательность разделить на группы, каждая из которых состоит из q элементов.

В неравномерных кодах можно использовать разделительный символ между буквами алфавита (так поступают, например, при передаче сообщений с помощью азбуки Морзе). Если же отказаться от разделительных символов, то следует запретить такие кодовые комбинации, начальные части которых уже использованы в качестве самостоятельной комбинации. Например, если 101 означает код какой-то буквы, то нельзя использовать комбинации 1, 10 или 10101.

Практические методы оптимального кодирования просты и основаны на очевидных соображениях. Так, буквы исходного алфавита записывают в порядке убывающей вероятности. Такое упорядоченное множество разбивают так, чтобы суммарные вероятности этих подмножеств были примерно равны. Всем знакам (буквам) верхней половины в качестве первого символа присваивают кодовый элемент 1, а всем нижним – 0. Затем каждое подмножество снова разбивается на два подмножества с соблюдением того же условия равенства вероятностей и с тем же условием присваивания кодовых элементов в качестве второго символа. Такое разбиение продолжается, пока в подмножестве не окажется только по одной букве кодируемого алфавита.

При каждом разбиении буквам верхнего подмножества присваивается кодовый элемент 1, а буквам нижнего подмножества – 0.

Так, при проведении эффективного кодирования ансамбля из восьми знаков, из $q_{ср} = H$, следует, что полученный код является оптимальным. Рассмотренный метод известен, как *метод Шеннона – Фано*.

Пример. Построить код Шеннона – Фано, если известны вероятности: $P(x_1) = 0{,}5$; $P(x_2) = 0{,}25$; $P(x_3) = 0{,}125$; $P(x_4) = 0{,}125$ и провести эффективное кодирование ансамбля из восьми знаков (m = 8) (табл. 7.1).

Решение. При обычном (не учитывающем статистических характеристик) двоичном кодировании с использованием k=2 знаков при построении равномерного кода количество элементов в кодовой последовательности будет $q \geq \log_k m = \log_2 8 = 3$, т.е. для представле-

Таблица 7.1

Эффективное кодирование ансамбля из восьми знаков (код Шеннона – Фано)

Буква (знак) x_i	Вероятность P_i	Кодовые последовательности. Номер разбиения				Длина q_i	p_i q_i	$-p_i \log p_i$
		1	2	3	4			
x_1	0,25	1	1			2	0,5	0,50
x_2	0,25	1	0			2	0,5	0,50
x_3	0,15	0	1	1		3	0,45	0,41
x_4	0,15	0	1	0		3	0,45	0,41
x_5	0,05	0	0	1	1	4	0,2	0,22
x_6	0,05	0	0	1	0	4	0,2	0,22
x_7	0,05	0	0	0	1	4	0,2	0,22
x_8	0,05	0	0	0	0	4	0,2	0,22

ния каждого знака использованного алфавита потребуется три двоичных символа.

$$q_{cp} = \sum_{i=1}^{m} p_i q_i = 2{,}7$$

$$H = -\sum_{i=1}^{m} P_i \log P_i = 2{,}7$$

Методика Шеннона – Фано позволяет построить кодовые комбинации (табл. 7.2), в которых знаки исходного ансамбля, имеющие наибольшую вероятность, кодируются наиболее короткими кодовыми по последовательностями. Таким образом, устраняется избыточность обычного двоичного кодирования, информационные возможности которого используются не полностью.

Так как вероятности знаков представляют собой целочисленные отрицательные степени двойки, то избыточность при кодировании устранена полностью. Среднее число символов на знак в этом случае точно равно энтропии.

Таблица 7.2

Кодовые комбинации исходного ансамбля (код Шеннона – Фано)

Знаки (буквы) x_i	Вероятность Pi	Кодовые комбинации						
		номер разбиения						
		1	2	3	4	5	6	7
x_1	1/2	1						
x_2	1/4	0	1					
x_3	1/8	0	0	1				
x_4	1/16	0	0	0	1			
x_5	1/32	0	0	0	0	1		
x_6	1/64	0	0	0	0	0	1	
x_7	1/128	0	0	0	0	0	0	1
x_8	1/128	0	0	0	0	0	0	0

В общем случае для алфавита из восьми знаков среднее число символов на знак будет меньше трех, но больше энтропии алфавита. Вычислим энтропию алфавита:

$$H = -\sum_{i=1}^{m=8} P(x_i) \log P(x_i) = 1\frac{63}{64}$$

Вычислим среднее число символов на знак:

$$q_{-\text{р}} = \sum_{i=1}^{m=8} P(x_i)q(x_i) = 1\frac{63}{64},$$

где $q(x_i)$ – число символов в кодовой комбинации, соответствующей знаку x_i.

Пример. Определить среднюю длину кодовой комбинации при эффективном кодировании ансамбля из восьми знаков и энтропию алфавита.

1. Средняя длина кодовых комбинаций

$$q_{-\text{р}} = \sum_{i=1}^{m=8} P_i q_i = 2{,}84.$$

2. Энтропия алфавита

$$H = -\sum_{i=1}^{m=8} P_i \log P_i = 2{,}76.$$

При кодировании по методике Шеннона – Фано некоторая избыточность в последовательностях символов, как правило, остается ($q_{ср} > H$) (табл. 7.3).

Таблица 7.3

Избыточность в последовательностях символов (по методике Шеннона – Фано)

Знаки (буквы) x_i	Вероятность P_i	Кодовые комбинации				
		номер разбиения				
		1	2	3	4	5
x_1	0,22	1	1			
x_2	0,20	1	0	1		
x_3	0,16	1	0	0		
x_4	0,16	0	1			
x_5	0,10	0	0	1		
x_6	0,10	0	0	0	1	
x_7	0,04	0	0	0	0	1
x_8	0,02	0	0	0	0	0

Решение:

Эту избыточность можно устранить, если перейти к кодированию достаточно большими блоками.

Пример: Рассмотрим процедуру эффективного кодирования сообщений, образованных с помощью алфавита, состоящего всего из двух знаков x_1 и x_2 с вероятностями появления соответственно

$$P(x_1) = 0{,}9;\; P(x_2) = 0{,}1.$$

Так как вероятности не равны, то последовательность из таких букв будет обладать избыточностью. Однако при побуквенном кодировании мы никакого эффекта не получим. Действительно, на передачу каждой буквы требуется символ либо 1, либо 0, в то время как энтропия равна

$$H = -\sum_{i=1}^{m=2} P_i \log P_i = 0{,}47,$$

т.е. оказывается

$$q_{ср} = \sum_{i=1}^{m} P_i q_i = 1 > H = 0{,}47.$$

При кодировании блоков (табл. 7.4), содержащих по две буквы, получим коды:

Таблица 7.4

Кодирование блоков, содержащих две буквы

Блоки	Вероятности	Кодовые комбинации		
		номер разбиения		
		1	2	3
x_1x_1	0,81	1		
x_1x_2	0,09	0	1	
x_2x_1	0,09	0	0	1
x_2x_2	0,01	0	0	0

Так как знаки статистически не связаны, вероятности блоков определяют, как произведение вероятностей составляющих знаков.

Среднее число символов на блок

$$q_{-\text{р}} = \sum_{i=1}^{m=4} P_i q_i = 1{,}29,$$

а на букву 1,29/2 = 0,645, т.е. приблизилось к H=0,47 и таким образом удалось повысить эффективность кодирования.

Кодирование блоков, содержащих по три знака (табл. 7.5), дает еще больший эффект:

Таблица 7.5

Кодирование блоков из трех знаков

Блоки	Вероят-ность P_i	кодовые комбинации				
		номер разбиения				
		1	2	3	4	5
$x_1x_1x_1$	0,729	1				
$x_2x_1x_1$	0,081	0	1	1		
$x_1x_2x_1$	0,081	0	1	0		
$x_1x_1x_2$	0,081	0	0	1		
$x_2x_2x_1$	0,009	0	0	1	1	
$x_2x_1x_2$	0,009	0	0	0	1	0
$x_1x_2x_2$	0,009	0	0	0	0	1
$x_2x_2x_2$	0,001	0	0	0	0	0

Среднее число символов на блок равно 1,59, а на знак – 0,53, что всего на 12 % больше энтропии.

Следует подчеркнуть, что увеличение эффективности кодирования при укрупнении блоков не связано с учетом все более далеких статистических связей, т.к. нами рассматривались алфавиты с независимыми знаками.

Повышение эффективности определяется лишь тем, что набор вероятностей, получившихся при укрупнении блоков, можно делить на более близкие по суммарным вероятностям подгруппы.

Рассмотренная методика Шеннона – Фано не всегда приводит к однозначному построению кода, т.к. при разбиении на подгруппы можно сделать большей по вероятности как верхнюю, так и нижнюю подгруппы, табл. 7.6.

От указанного недостатка свободна *методика Хаффмена*.

Эта методика гарантирует однозначное построение кода с наименьшим для данного распределения вероятностей средним числом символов на букву.

Таблица 7.6.

Разбиение на подгруппы по методике Шеннона – Фано

Знаки (буквы) x_i	Вероятность P_i	1-е кодовые комбинации					2-е кодовые комбинации				
		номер разбиения					номер разбиения				
		1	2	3	4	5	1	2	3	4	5
x_1	0,22	1	1				1	1			
x_2	0,20	1	0	1			1	0			
x_3	0,16	1	0	0			0	1	1		
x_4	0,16	0	1				0	1	0		
x_5	0,10	0	0	1			0	0	1		
x_6	0,10	0	0	0	1		0	0	0	1	
x_7	0,04	0	0	0	0	1	0	0	0	0	1
x_8	0,02	0	0	0	0	0	0	0	0	0	0

Для двоичного кода методика сводится к следующему. Буквы алфавита сообщений выписывают в основной столбец в порядке убывания вероятностей. Две последние буквы объединяют в одну вспомогательную букву, которой приписывают суммарную вероятность. Вероятности букв, не участвующих в объединении и полученная суммарная вероятность, снова располагаются в порядке убывания вероятностей в дополнительном столбце, а две последние объединяются.

Процесс продолжается до тех пор, пока не получат единственную вспомогательную букву с вероятностью, равной единице.

Используя методику Хаффмана, осуществим эффективное кодирование ансамбля из восьми знаков, табл. 7.7.

Для наглядности построим кодовое дерево (рис.7.1). Из точки, соответствующей вероятности 1, направляем две ветви, причем ветви с большей вероятностью присваиваем символ 1, а с меньшей 0.

Такое последовательное ветвление продолжаем до тех пор, пока не дойдем до вероятности каждой буквы.

Двигаясь по кодовому дереву сверху вниз, можно записать для каждой буквы соответствующую ей кодовую комбинацию.

Средней длиной кодового слова называется величина:

$$t_{cp} = \sum_{i-1}^{n} p(a_i)\lambda_i ,$$

где λ_i – длина кодового слова $\tilde{\alpha}_i$.

Таблица 7.7

Кодирование ансамбля из восьми знаков (по методике Хаффмана)

Знаки	Вероятности	Вспомогательные столбцы							Новая комбинация
		1	2	3	4	5	6	7	
x_1	0,22	0,22	0,22	0,26	0,32	0,42	0,58	1	01
x_2	0,20	0,20	0,20	0,22	0,26	0,32	0,42		00
x_3	0,16	0,16	0,16	0,20	0,22	0,26			111
x_4	0,16	0,16	0,16	0,16	0,20				110
x_5	0,10	0,10	0,10	0,16					100
x_6	0,10	0,10							1011
x_7	0,04	0,06							10101
x_8	0,02								10100

Для равномерного, блокового кода понятие средней длины становится тривиальным, так как не зависит от вероятностного распределения $\tilde{p}(s)$ и совпадает с длиной блока. Иное дело получается, когда рассматривается неравномерный код, для которого величина λ_{cp} меняется в зависимости от выбора длины кодовых слов. Очевидно, что если более вероятным буквам, порожденным источником, ставить в соответствие короткие кодовые слова, то при этом будем получать уменьшение средней длины.

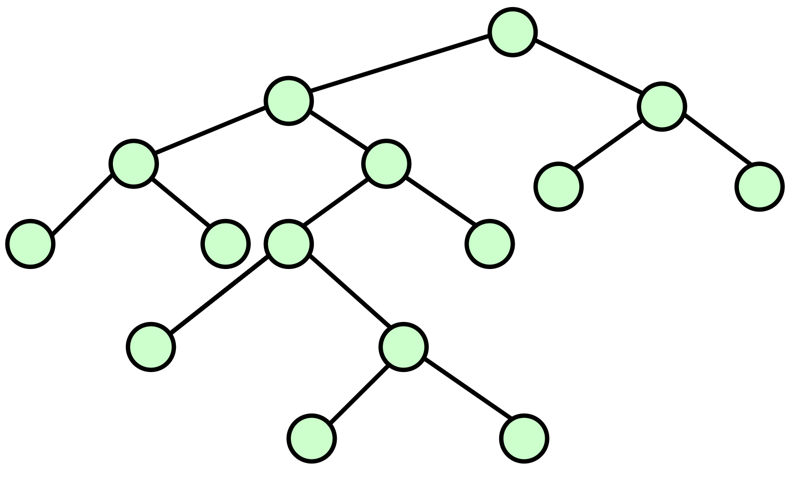


Рис. 7.1. Кодовое дерево

Однозначно декодируемый код $\{\tilde{a}_1 \ldots \tilde{a}_n\}$ называется *оптимальным***,** если он обладает минимально возможной средней длиной во множестве всех кодов, построенных для данного источника. Другими словами, как известно оптимальное кодирование приводит к максимальному уплотнению, сжатию информации.

7.4. Алфавитное неравномерное двоичное кодирование сигналами равной длительности. Префиксные коды

В способах *построения двоичных кодов* знаки первичного алфавита (например, русского) кодируются комбинациями символов двоичного алфавита (т.е. 0 и 1), причем, длина кодов и, соответственно, длительность передачи отдельного кода, могут различаться. Длительности элементарных сигналов при этом одинаковы ($\tau_0 = \tau_1 = \tau$). Очевидно, для передачи информации, в среднем приходящейся на знак первичного алфавита, необходимо время $K(A,2) \cdot \tau$. Таким образом, задачу оптимизации неравномерного кодирования можно сформулировать следующим образом: *построить такую схему кодирования, в которой суммарная длительность кодов при передаче (или суммарное число кодов при хранении) данного сообщения была бы наименьшей*. За счет чего возможна такая оптимизация? Очевидно, суммарная длительность сообщения будет меньше, если применить следующий подход: тем знакам первичного алфавита, которые встречаются в сообщении чаще, присвоить меньшие по длине коды, а тем, относительная частота которых меньше – коды более длинные. Другими словами, коды знаков первичного алфавита, вероятность появления которых в сообщении выше, следует строить из возможно меньшего числа элементарных сигналов, а длинные коды использовать для знаков с малыми вероятностями.

Параллельно должна решаться проблема *различимости кодов*. Представим, что на выходе кодера получена следующая последовательность элементарных сигналов:

00100010000111010101110000110

Каким образом она может быть декодирована? Если бы код был равномерным, приемное устройство просто отсчитывало бы заданное (фиксированное) число элементарных сигналов (например, 5, как в коде Бодо) и интерпретировало их в соответствии с кодовой таблицей. При использовании неравномерного кодирования возможны два подхода к обеспечению различимости кодов.

Первый состоит в использовании специальной комбинации элементарных сигналов, которая интерпретируется декодером как *разделитель знаков*. Второй – в применении *префиксных кодов*. Рассмотрим подробнее каждый из подходов.

Неравномерный код с разделителем

Условимся, что разделителем отдельных кодов букв будет последовательность *00* (признак конца знака), а разделителем слов-слов – *000* (признак конца слова – пробел). Довольно очевидными оказываются следующие правила построения кодов:

- код признака конца знака может быть включен в код буквы, поскольку не существует отдельно (т.е. коды всех букв будут заканчиваться *00*);
- коды букв не должны содержать двух и более нулей подряд в середине (иначе они будут восприниматься как конец знака);
- код буквы (кроме пробела) всегда должен начинаться с *1;*
- разделителю слов (*000*) всегда предшествует признак конца знака; при этом реализуется последовательность *00000* (т.е., если в конце кода встречается комбинация *...000* или *...0000,* они не воспринимаются как разделитель слов); следовательно, коды букв могут оканчиваться на *0* или *00* (до признака конца знака).

В соответствии с перечисленными правилами построим кодовую табл. 7.8 для букв русского алфавита, основываясь на вероятностях появления отдельных букв.

Теперь можно найти среднюю длину кода $K(r,2)$ для данного способа кодирования:

$$K(r,2) = \sum_{j=1}^{32} p_j \cdot k_j = 4{,}964 \qquad (7.15)$$

Поскольку для русского языка $I_1^{(r)} = 4{,}356$ бит, избыточность данного кода, составляет:

$$Q(r,2) = 4{,}964 / 4{,}356 - 1 \approx 0{,}14$$

это означает, что при данном способе кодирования будет передаваться приблизительно на 14 % больше информации, чем содержит исходное сообщение. Аналогичные вычисления для английского языка дают значение $K(e,2) = 4{,}176$, что при $I_1^{(e)} = 4{,}036$ *бит* приводят к избыточности кода $Q(e,2) = 0{,}168$.

Таблица 7.8

Кодовая таблица для букв русского алфавита

Буква	*Код*	$p_i \cdot 10^3$	k_i	*Буква*	*Код*	$p_i \cdot 10^3$	k_i
пробел	000	174	3	я	1011000	18	7
о	100	90	3	ы	1011100	16	7
е	1000	72	4	з	1101000	16	7
а	1100	62	4	ь,ъ	1101100	14	7
и	10000	62	5	б	1110000	14	7
т	10100	53	5	г	1110100	13	7
н	11000	53	5	ч	1111000	12	7
с	11100	45	5	й	1111100	10	7
р	101000	40	6	х	10101000	9	8
в	101100	38	6	ж	10101100	7	8
л	110000	35	6	ю	10110000	6	8
к	110100	28	6	ш	10110100	6	8
м	111000	26	6	ц	10111000	4	8
д	111100	25	6	щ	10111100	3	8
п	1010000	23	7	э	11010000	3	8
у	1010100	21	7	ф	11010100	2	8

Рассмотрев один из вариантов двоичного неравномерного кодирования, попробуем найти ответы на следующие вопросы:

– возможно ли такое кодирование без использования разделителя знаков?

– существует ли наиболее эффективный (оптимальный) способ неравномерного двоичного кодирования?

Суть первой проблемы состоит в нахождении такого варианта кодирования сообщения, при котором последующее выделение из него каждого отдельного знака (т.е. декодирование) оказывается однозначным без специальных указателей разделения знаков. Наиболее простыми и употребимыми кодами такого типа являются так называемые *префиксные коды*, которые удовлетворяют следующему условию (*условию Фано*):

Неравномерный код может быть однозначно декодирован, если никакой из кодов не совпадает с началом (префиксом) какого-либо иного более длинного кода.

В языковедении термин «*префикс*» означает «*приставка*».

Например, если имеется код *110,* то уже не могут использоваться коды *1, 11, 1101, 110101* и пр. Если условие Фано выполняется, то при прочтении (расшифровке) закодированного сообщения путем сопоставления с таблицей кодов всегда можно точно указать, где заканчивается один код и начинается другой.

Пример: Пусть имеется следующая таблица префиксных кодов:

а	л	м	р	у	ы
10	010	00	11	0110	0111

Требуется декодировать сообщение:

00100010000111010101110000110

Декодирование производится циклическим повторением следующих действий:

(a) отрезать от текущего сообщения крайний левый символ, присоединить справа к рабочему кодовому слову;

(b) сравнить рабочее кодовое слово с кодовой таблицей; если совпадения нет, перейти к (a);

(c) декодировать рабочее кодовое слово, очистить его;

(d) проверить, имеются ли еще знаки в сообщении; если «да», перейти к (a). Применение данного алгоритма дает следующие результаты, табл. 7.9.

Таблица 7.9

Декодирование сообщения циклическим повторением

Шаг	Рабочее слово	Текущее время	Распознанный знак	Декодированное сообщение
0	Пусто	**0**0100010000111010101110000110	–	–
1	0	**0**100010000111010101110000110	нет	–
2	**00**	**1**00010000111010101110000110	**м**	м
3	1	**0**0010000111010101110000110	нет	м
4	**10**	**0**010000111010101110000110	**а**	ма
5	0	**0**10000111010101110000110	нет	ма
6	**00**	**1**0000111010101110000110	**м**	мам
…				

Доведя процедуру до конца, получим сообщение: *«мама мыла раму»*.

Таким образом, использование префиксного кодирования позволяет делать сообщение более коротким, поскольку нет необходимости передавать разделители знаков. Однако условие Фано не устанавливает способа формирования префиксного кода и, в частности, наилучшего из возможных. Мы рассмотрим две схемы построения префиксных кодов.

Префиксный код Шеннона–Фано

Данный вариант кодирования был предложен в 1948–1949 гг. независимо Р. Фано и К. Шенноном и по этой причине назван по их именам. Рассмотрим схему на следующем примере: пусть имеется первичный алфавит *A*, состоящий из шести знаков $a_1 ... a_6$ с вероятностями появления в сообщении, соответственно, 0,3; 0,2; 0,2; 0,15; 0,1; 0,05. Расположим эти знаки в табл. 7.10, в порядке убывания вероятностей.

Таблица 7.10

Кодирование на основе префиксного кода Шеннона – Фано

Знак	p_i	Разряды кода				Код
		1	2	3	4	
a_1	0,30	0	0			00
a_2	0,20	0	1			01
a_3	0,20	1	0			10
a_4	0,15	1	1	0		110
a_5	0,10	1	1	1	0	1110
a_6	0,05	1	1	1	1	1111

Разделим знаки на две группы таким образом, чтобы суммы вероятностей в каждой из них были бы приблизительно равными. В нашем примере в первую группу попадут a_1 и a_2 с суммой вероятностей 0,5 – им присвоим первый знак кода «0». Сумма вероятностей для остальных четырех знаков также 0,5; у них первый знак кода будет «1». Продолжим деление каждой из групп на подгруппы по этой же схеме, т.е. так, чтобы суммы вероятностей на каждом шаге в соседних подгруппах были бы возможно более близкими.

Из процедуры построения кодов легко видеть, что они удовлетворяют условию Фано и, следовательно, код является префиксным. Средняя длина кода равна:

$$K(A,2) = 0{,}3\cdot 2 + 0{,}2\cdot 2 + 0{,}2\cdot 2 + 0{,}15\cdot 3 + 0{,}1\cdot 4 + 0{,}05\cdot 4 = 2{,}45$$

$I_1^{(A)} = 2{,}390$ бит. Подставляя указанные значения в (7.5), получаем избыточность кода $Q(A,2) = 0{,}0249$, т.е. около 2,5 %. Однако данный код нельзя считать оптимальным, поскольку вероятности появления 0 и 1 неодинаковы (0,35 и 0,65, соответственно). Применение изложенной схемы построения к русскому алфавиту дает избыточность кода 0,0147.

Префиксный код Хаффмана

Способ *оптимального* префиксного двоичного кодирования был предложен Д. Хаффманом. Построение кодов Хаффмана мы рассмотрим на том же примере. Создадим новый вспомогательный алфавит A_1, объединив два знака с наименьшими вероятностями (a_5 и a_6) и заменив их одним знаком (например, $a^{(1)}$); вероятность нового знака будет равна сумме вероятностей тех, что в него вошли, т.е. 0,15; остальные знаки исходного алфавита включим в новый без изменений; общее число знаков в новом алфавите, очевидно, будет на 1 меньше, чем в исходном.

Аналогичным образом продолжим создавать новые алфавиты, пока в последнем не останется два знака; ясно, что количество таких шагов будет равно $N - 2$, где N – число знаков исходного алфавита (в нашем случае $N = 6$, следовательно, необходимо построить 4 вспомогательных алфавита). В промежуточных алфавитах каждый раз будем переупорядочивать знаки по убыванию вероятностей.

Теперь в обратном направлении проведем процедуру кодирования. Двум знакам последнего алфавита присвоим коды 0 и 1 (которому какой – роли не играет; условимся, что верхний знак будет иметь код 0, а нижний – 1).

В нашем примере знак $a_1^{(4)}$ алфавита $A^{(4)}$, имеющий вероятность 0,6, получит код 0, а $a_2^{(4)}$ с вероятностью 0,4 – код 1. В алфавите $A^{(3)}$ знак $a_1^{(3)}$ получает от $a_2^{(4)}$ его вероятность 0,4 и код (1); коды знаков $a_2^{(3)}$ и $a_3^{(3)}$, происходящие от знака $a_1^{(4)}$ с вероятностью 0,6, будут уже двузначным:

их первой цифрой станет код их «родителя» (т.е. 0), а вторая цифра – как условились – у верхнего 0, у нижнего – 1;

таким образом, $a_2^{(3)}$ будет иметь код 00, а $a_3^{(3)}$ – код 01.

Из процедуры построения кодов вновь видно, что они удовлетворяют условию Фано и, следовательно, не требуют разделителя. Построения представим в виде табл. 7.11.

Таблица 7.11

Способ оптимального префиксного кодирования (префиксный код Хаффмана)

№ знака	*Вероятности*				
	Исходный алфавит	*Промежуточные алфавиты*			
		$A^{(1)}$	$A^{(2)}$	$A^{(3)}$	$A^{(4)}$
1	0,3	0,3	0,3	0,4	0,6
2	0,2	0,2	0,3	0,3	0,4
3	0,2	0,2	0,2	0,3	
4	0,15	0,15	0,2		
5	0,1	0,15			
6	0,05				

Средняя длина кода, как и в предыдущем примере оказывается:
$K(A,2) = 0{,}3 \cdot 2 + 0{,}2 \cdot 2 + 0{,}2 \cdot 2 + 0{,}15 \cdot 3 + 0{,}1 \cdot 4 + 0{,}05 \cdot 4 = 2{,}45$.
Построения представим в виде табл. 7.12.

Таблица 7.12

Процедура кодирования в обратном направлении

№ знака	*Вероятности*									
	Исходный алфавит		*Промежуточные алфавиты*							
			$A^{(1)}$		$A^{(2)}$		$A^{(3)}$		$A^{(4)}$	
1	0,3	**00**	0,3	**00**	0,3	**00**	0,4	**1**	0,6	**0**
2	0,2	**10**	0,2	**10**	0,3	**01**	0,3	**00**	0,4	**1**
3	0,2	**11**	0,2	**11**	0,2	**10**	0,3	**01**		
4	0,15	**010**	0,15	**010**	0,2	**11**				
5	0,1	**0110**	0,15	**011**						
6	0,05	**0111**								

Избыточность снова оказывается равной $Q(A,2) = 0{,}0249$, однако вероятности 0 и 1 сблизились (0,47 и 0,53, соответственно). Более высокая эффективность кодов Хаффмана по сравнению с кодами Шеннона-Фано становится очевидной, если сравнить избыточности кодов для какого-либо естественного языка.

Применение описанного метода для букв русского алфавита порождает коды, представленные в табл. 7.13 (для удобства сопоставления они приведены в формате табл. 7.8).

Таблица 7.13

Применение оптимального префиксного двоичного кодирования для букв русского алфавита

Буква	*Код*	$p_i \cdot 10^3$	k_i	*Буква*	*Код*	$p_i \cdot 10^3$	k_i
пробел	000	174	3	я	001101	18	6
о	111	90	3	ы	010110	16	6
е	0100	72	4	з	010111	16	6
а	0110	62	4	ь,ъ	100001	14	6
и	0111	62	4	б	101100	14	6
т	1001	53	4	г	101101	13	6
н	1010	53	4	ч	110011	12	6
с	1101	45	4	й	0011001	10	7
р	00101	40	5	х	1000000	9	7
	00111	38	5	ж	1000001	7	7
л	01010	35	5	ю	1100101	6	7
к	10001	28	5	ш	00110000	6	8
м	10111	26	5	ц	11001000	4	8
д	11000	25	5	щ	11001001	3	8
п	001000	23	6	э	001100010	3	9
у	001001	21	6	ф	001100011	2	9

Средняя длина кода оказывается равной $K(r,2) = 4{,}395$; избыточность кода $Q(r,2) = 0{,}0090$, т.е. не превышает 1%, что заметно меньше избыточности кода Шеннона-Фано (см. выше).

Код Хаффмана важен в теоретическом отношении, поскольку можно доказать, что он является *самым экономичным* из всех возможных, т.е. *ни для какого метода алфавитного кодирования длина кода не может оказаться меньше, чем код Хаффмана*.

Таким образом, можно заключить, что существует способ построения оптимального неравномерного алфавитного кода. Метод Хаффмана и его модификация – метод адаптивного кодирования *(динамическое кодирование Хаффмана)* – нашли широчайшее применение в программах-архиваторах, программах резервного копирования файлов и дисков, в системах сжатия информации в модемах и факсах.

Неравенство Л. Крафта

Пусть $\lambda_1, \lambda_2, \ldots, \lambda_n$ – набор натуральных чисел. Для того, чтобы существовал префиксный код с длинами кодовых слов $\lambda_1, \lambda_2, \ldots, \lambda_n$, необходимо и достаточно, чтобы выполнялось неравенство

$$\sum_{i=1}^{n} D^{-\ell_i} \leq 1,$$

где D – число концевых вершин полного дерева;

n – максимальный порядок концевых вершин полного дерева.

Для каждого источника существует оптимальный код, поскольку множество префиксных кодов источника с избыточностью, меньшей либо равной 1, непустое и конечно. Один источник может иметь несколько оптимальных кодов с разными наборами длин кодовых слов.

7.5. Линейные коды. Параметры кодов и их границы

Самый большой класс разделимых кодов составляют *линейные коды*, у которых значения проверочных символов определяются в результате проведения линейных операций над определенными информационными символами. Для случая двоичных кодов каждый проверочный символ выбирают таким образом, чтобы его сумма с определенными информационными символами была равна 0. Символ проверочной позиции имеет значение 1, если число единиц информационных разрядов, входящих в данное проверочное равенство, нечетно, и 0, если оно четно. Число проверочных равенств (а следовательно, и число проверочных символов) и номера конкретных информационных разрядов, входящих в каждое из равенств, определяется тем, какие и сколько ошибок должен исправлять или обнаруживать данный код. Проверочные символы могут располагаться на любом месте кодовой комбинации. При декодировании определяется справедливость проверочных равенств. В случае двоичных кодов такое определение сводится к проверкам на четность числа единиц среди символов, входящих в каждое из равенств (включая проверочные). Совокупность проверок дает информацию о том, имеется ли ошибка, а в случае необходимости и о том, на каких позициях символы искажены.

Любой двоичный линейный код является групповым, так как совокупность входящих в него кодовых комбинаций образует группу.

Уточнение понятий линейного и группового кода требует ознакомления с основами линейной алгебры.

Основой математического описания линейных кодов является *линейная алгебра* (теория векторных пространств, теория матриц, теория групп). Кодовые комбинации рассматривают как элементы множества, например, кодовые комбинации двоичного кода принадлежат множеству положительных двоичных чисел.

Множества, для которых определены некоторые алгебраические операции, называют алгебраическими системами. Под алгебраической операцией понимают однозначные сопоставление двум элементам некоторого третьего элемента по определенным правилам. Обычно основную операцию называют сложением (обозначают $a+b=c$) или умножением (обозначают $a \cdot b = c$), а обратную ей – вычитанием или делением, даже, если эти операции проводятся не над числами и не идентичны.

Основные алгебраические системы, использующиеся в теории корректирующих кодов

Группой – множество элементов, в котором определена одна основная операция и выполняются следующие аксиомы:

1. В результате применения операции к любым двум элементам группы образуется элемент этой же группы (требование замкнутости).

2. Для любых трех элементов группы a,b,c удовлетворяется равенство $(a+b)+c = a+(b+c)$, если основная операция – сложение, и равенство $a(bc) = (ab)c$, если основная операция – умножение.

3. В любой группе G_n существует однозначно определенный элемент, удовлетворяющий при всех значениях а из G_n условию $a+0=0+a$, если основная операция – сложение, или условию $a \cdot 1 = 1 \cdot a = a$, если основная операция – умножение. В первом случае этот элемент называют *нулем* и обозначают символом 0, а во втором – *единицей* и обозначают символом 1.

4. Всякий элемент a группы обладает элементом, однозначно определенным уравнением $a+(-a) = -a+a = 0$, если основная операция – сложение, или уравнением $a \cdot a^{-1} = a^{-1} \cdot a = 1$, если основная операция – умножение.

В первом случае этот элемент называют *противоположным* и обозначают (-a), а во втором – *обратным* и обозначают a^{-1}.

Если операция, определенная в группе, коммутативна, то есть справедливо равенство $a+b=b+a$ (для группы по сложению) или равенство $a \cdot b = b \cdot a$ (для группы по умножению), то группу называют *коммутативной* или *абелевой*.

Группу, состоящую из конечного числа элементов, называют *порядком* группы.

Чтобы рассматриваемое нами множество n-разрядных кодовых комбинаций было конечной группой, при выполнении основной операции число разрядов в результирующей кодовой комбинации не должно увеличиваться. Этому условию удовлетворяет операция символического поразрядного сложения по заданному модулю q (q – простое число), при которой цифры одинаковых разрядов элементов группы складываются обычным порядком, а результатом сложения считается остаток от деления полученного числа по модулю q. При рассмотрении двоичных кодов используется операция сложения по модулю 2. Результатом сложения цифр данного разряда является 0, если сумма единиц в нем четная, и 1, если сумма единиц в нем нечетная, например

$$\oplus\begin{array}{r} 1011101 \\ 0111101 \\ 0001110 \\ \hline 1101110 \end{array}$$

Выбранная нами операция коммутативна, поэтому рассматриваемые группы будут абелевыми.

Нулевым элементом является комбинация, состоящая из одних нулей. Противоположным элементом при сложении по модулю 2 будет сам заданный элемент. Следовательно, операция вычитания по модулю 2 тождественна операции сложения.

Пример: Определить, являются ли группами следующие множества кодовых комбинаций:

1) 0001, 0110, 0111, 0011;

2) 0000, 1101, 1110, 0111;

3) 000, 001, 010, 011, 100, 101, 110, 111.

Решение: Первое множество не является группой, так как не содержит нулевого элемента.

Второе множество не является группой, так как не выполняется условие замкнутости, например, сумма по модулю 2 комбинаций 1101 и 1110 дает комбинацию 0011, не принадлежащую исходному множеству.

Третье множество удовлетворяет всем перечисленным условиям и является группой.

Подмножества группы, являющиеся сами по себе группами относительно операции, определенной в группе, называют *подгруппами*. Например, подмножество трехразрядных кодовых комбинаций: 000, 001, 010, 011 образуют подгруппу указанной в примере группы трехразрядных кодовых комбинаций.

Пусть в абелевой группе G_n задана определенная подгруппа A. Если B – любой, не входящий в A элемент из G_n, то суммы (по модулю 2) элементов B с каждым из элементов подгруппы A образуют определенный класс группы G_n по подгруппе A, порождаемый элементом B.

Элемент B, естественно, содержится в этом смежном классе, так как любая подгруппа содержит нулевой элемент. Взяв последовательно некоторые элементы B_j группы, не вошедшие в уже образованные смежные классы, можно разложить всю группу на смежные классы по подгруппе A.

Элементы B_j называют *образующими элементами смежных классов по подгруппам*.

В таблице разложения, иногда называемой *групповой таблицей*, образующие элементы обычно располагают в крайнем левом столбце, причем крайним левым элементом подгруппы является нулевой элемент.

Пример: Разложить группу трехразрядных двоичных кодовых комбинаций по подгруппе двухразрядных кодовых комбинаций.

Решение: Разложение выполняют в соответствии с таблицей:

$A_1=0$	A_2	A_3	A_4
000	001	010	011
B_1	$A_2 \oplus B_1$	$A_3 \oplus B_1$	$A_4 \oplus B_1$
100	101	110	111

Пример: Разложить группу четырехразрядных двоичных кодовых комбинаций по подгруппе двухразрядных кодовых комбинаций.

Решение: Существует много вариантов разложения в зависимости от того, какие элементы выбраны в качестве образующих смежных классов.

Один из вариантов:

$A_1 = 0$	A_2	A_3	A_4
0000	0001	0110	0111
B_1 0100	$A_2 \oplus B_1$ 0101	$A_3 \oplus B_1$ 0110	$A_4 \oplus B_1$ 0111
B_2 1010	$A_2 \oplus B_2$ 1011	$A_3 \oplus B_2$ 1000	$A_4 \oplus B_2$ 1001
B_3 1100	$A_2 \oplus B_3$ 1101	$A_3 \oplus B_3$ 1110	$A_4 \oplus B_3$ 1111

Кольцом называют множество элементов R, на котором определены две операции (сложение и умножение), такие, что:

1) множество R является коммутативной группой по отношению;

2) произведение элементов $a \in R$ и $b \in R$ есть элемент R (замкнутость по отношению и умножению);

3) для любых трех элементов a,b,c из R справедливо равенство $a(bc) = (ab)c$ (ассоциативный закон для умножения);

4) для любых трех элементов a,b,c из R выполняются соотношения $a(b + c) = ab + ac$ и $(b + c)a = ba + ca$ (дистрибутивные законы).

Если для любых двух элементов кольца справедливо соотношение $ab = ba$, то кольцо называют *коммутативным*.

Кольцо может не иметь единичного элемента по умножению и обратных элементов.

Примером кольца может служить множество действительных четных целых чисел относительно обычных операций сложения и умножения.

Полем F называют множество, по крайней мере, двух элементов, в котором определены две операции – сложение и умножение, и выполняются следующие аксиомы:

1) множество элементов образуют коммутативную группу по сложению;

2) множество ненулевых элементов образуют коммутативную группу по умножению;

3) для любых трех элементов множества a,b,c выполняется соотношение (дистрибутивный закон) $a(b+c)=ab+ac$.

Поле F является, следовательно, коммутативным кольцом с единичным элементом по умножению, в котором каждый ненулевой элемент обладает обратным элементом. Примером поля может служить множество всех действительных чисел.

Поле GF(P), состоящее из конечного числа элементов P, называют *конечным полем* или *полем Галуа*. Для любого числа P, являющегося степенью простого числа q, существует поле, насчитывающее p элементов. Например, совокупность чисел по модулю q, если q – простое число, является полем.

Поле не может содержать менее двух элементов, поскольку в нем должны быть по крайней мере единичный элемент относительно операции сложения (0) и единичный элемент относительно операции умножения (1). Поле, включающее только 0 и 1, обозначим GF(2).

Правила сложения и умножения в поле с двумя элементами следующие, рис. 7.2:

+	0	1
0	0	1
1	1	0

×	0	1
0	0	0
1	0	1

Рис. 7.2. Правила сложения и умножения в поле с двумя элементами

Двоичные кодовые комбинации, являющиеся упорядоченными последовательностями из n элементов поля GF(2), рассматриваются в теории кодирования как частный случай последовательностей из n элементов поля GF(P). Такой подход позволяет строить и анализировать коды с основанием, равным степени простого числа. В общем случае суммой кодовых комбинаций A_j и A_i называют комбинацию $A_f = A_i + A_j$, в которой любой символ A_k $(k=1,2,...,n)$ представляет собой сумму k–x символов комбинаций, причем суммирование производится по правилам поля GF(P). При этом вся совокупность n-разрядных кодовых комбинаций оказывается абелевой группой.

В частном случае, когда основанием кода является простое число q, правило сложения в поле GF(q) совпадает с правилом сложения по заданному модулю q.

Линейным кодом называют множество векторов, образующих подпространства векторного пространства всех n-разрядных кодовых комбинаций над полем GF(P).

В случае двоичного кодирования такого подпространства комбинаций над полем GF(2) образует любая совокупность двоичных кодовых комбинаций, являющаяся подгруппой группы всех n-разрядных двоичных кодовых комбинаций. Поэтому любой двоичный линейный код является групповым.

1) корректирующие коды

Корректирующие коды строятся так, что для передачи сообщения используются не все кодовые комбинации m^n, а лишь некоторая часть их (так называемые *разрешенные* кодовые комбинации). Тем самым создается возможность обнаружения и исправления ошибки при неправильном воспроизведении некоторого числа символов. Корректирующие свойства кодов достигаются введением в кодовые комбинации дополнительных (избыточных) символов.

Декодирование состоит в восстановлении сообщения по принимаемым кодовым символам. Устройства, осуществляющие кодирование и декодирование, называют соответственно кодером и декодером. Как правило, кодер и декодер выполняются физически в одном устройстве, называемым кодеком.

Рассмотрим основные принципы построения *корректирующих кодов* или *помехоустойчивого кодирования*.

Напомним, что расстоянием Хэмминга между двумя кодовыми n-последовательностями, b_i и b_j, которое будем далее обозначать $d(i;j)$, является число разрядов, в которых символы этих последовательностей не совпадают.

Говорят, что в канале произошла ошибка кратности q, если в кодовой комбинации q символов приняты ошибочно. Легко видеть, что кратность ошибки есть не что иное, как расстояние Хэмминга между переданной и принятой кодовыми комбинациями, или, иначе, вес вектора ошибки.

Рассматривая все разрешенные кодовые комбинации и определяя кодовые расстояния между каждой парой, можно найти наименьшее из них $d = \min d(i; j)$, где минимум берется по всем парам разрешенных комбинаций. Это минимальное кодовое расстояние является важным параметром кода. Очевидно, что для простого кода $d = 1$.

Обнаруживающая способность кода характеризуется следующей теоремой. Если код имеет $d > 1$ и используется декодирование по методу обнаружения ошибок, то все ошибки кратностью $q < d$ обнаруживаются. Что же касается ошибок кратностью q и d, то одни из них обнаруживаются, а другие нет.

Исправляющая способность кода при этом правиле декодирования определяется следующей теоремой. Если код имеет $d > 2$ и используется декодирование с исправлением ошибок по наименьшему расстоянию, то все ошибки кратностью $q < d/2$ исправляются. Что же касается ошибок большей кратности, то одни из них исправляются, а другие нет.

Задача кодирования состоит в выборе кода, обладающего максимально достижимым d. Впрочем, такая формулировка задачи неполна. Увеличивая длину кода n и сохраняя число кодовых комбинаций M, можно получить сколь угодно большое значение d. Но такое «решение» задачи не представляет интереса, так как с увеличением n уменьшается возможная скорость передачи информации от источника.

Если длина кода n задана, то можно получить любое значение d, не превышающее n, уменьшая число комбинаций M. Поэтому задачу поиска наилучшего кода (в смысле максимального d) следует формулировать так: при заданных M и n найти код длины n, содержащий M комбинаций и имеющий наибольшее возможное d. В общем виде эта задача в теории кодирования не решена, хотя для многих значений n и M ее решения получены.

На первый взгляд помехоустойчивое кодирование реализуется весьма просто. В память кодирующего устройства (кодера) записываются разрешенные кодовые комбинации выбранного кода и правило, по которому с каждым из M сообщений источника сопоставляется одна из таких комбинаций. Данное правило известно и декодеру.

Получив от источника определенное сообщение, кодер отыскивает соответствующую ему комбинацию и посылает ее в канал. В свою очередь, декодер, приняв комбинацию, искаженную помехами, сравнивает ее со всеми *M* комбинациями списка и отыскивает ту из них, которая ближе остальных к принятой.

Однако даже при умеренных значениях *n* такой способ весьма сложный. Покажем это на примере. Пусть выбрана длина кодовой комбинации $n = 100$, а скорость кода примем равной 0,5 (число информационных и проверочных символов равно). Тогда число разрешенных комбинаций кода будет $2^{50} \cdot 10^{15}$. Соответственно размер таблицы будет $100 \cdot 10^{15} = 10^{17}\,бит \cdot 10^{16}\,байт = 10000\text{Тбайт}$.

Таким образом, применение достаточно эффективных (а значит, и достаточно длинных) кодов при табличном методе кодирования и декодирования технически невозможно.

Поэтому основное направление теории помехоустойчивого кодирования заключается в поисках таких классов кодов, для которых кодирование и декодирование осуществляются не перебором таблицы, а с помощью некоторых регулярных правил, определенных алгебраической структурой кодовых комбинаций.

Построение двоичного группового кода. Построение конкретного *корректирующего* кода производят, исходя из требуемого объема кода *Q*, т. е. необходимого числа передаваемых команд или дискретных значений измеряемой величины и статистических данных о наиболее вероятных векторах ошибок в используемом канале связи. *Вектором ошибки* называют *n*-разрядную двоичную последовательность, имеющую единицы в разрядах, подвергшихся искажению, и нули во всех остальных разрядах. Любую искаженную кодовую комбинацию можно рассматривать теперь как сумму (или разность) по модулю 2 исходной разрешенной кодовой комбинации и вектора ошибки.

Исходя из неравенства $2^k - 1 \geq Q$ (нулевая комбинация часто не используется, так как не меняет состояния канала связи), определяем число информационных разрядов *k*, необходимое для передачи заданного числа команд обычным двоичным кодом.

Каждой из $2^k - 1$ ненулевых комбинаций *k*-разрядного безызбыточного кода нам необходимо поставить в соответствие комбинацию из *n* символов. Значения символов в *n*–*k* проверочных разрядах такой комбинации устанавливаются в результате суммирования по модулю 2 значений символов в определенных информационных разрядах.

Поскольку множество 2^k комбинаций информационных символов (включая нулевую) образует подгруппу группы всех *n*-разрядных комбинаций, то и множество 2^k *n*-разрядных комбинаций, полученных по указанному правилу, тоже является подгруппой группы *n*-разрядных кодовых комбинаций. Это множество разрешенных кодовых комбинаций и будет групповым кодом.

Нам надлежит определить число проверочных разрядов и номера информационных разрядов, входящих в каждое из равенств для определения символов в проверочных разрядах.

Разложим группу 2^n всех n-разрядных комбинаций на смежные классы по подгруппе 2^k разрешенных n-разрядных кодовых комбинаций, проверочные разряды в которых еще не заполнены. Помимо самой подгруппы кода в разложении насчитывается $2^{n-k}-1$ смежных классов. Элементы каждого класса представляют собой суммы по модулю 2 комбинаций кода и образующих элементов данного класса. Если за образующие элементы каждого класса принять те наиболее вероятные для заданного канала связи вектора ошибок, которые должны быть исправлены, то в каждом смежном классе сгруппируются кодовые комбинации, получающиеся в результате воздействия на все разрешенные комбинации определенного вектора ошибки. Для исправления любой полученной на выходе канала связи кодовой комбинации теперь достаточно определить, к какому классу смежности она относится. Складывая ее затем (по модулю 2) с образующим элементом этого смежного класса, получаем истинную комбинацию кода.

Ясно, что из общего числа 2^n-1 возможных ошибок групповой код может исправить всего $2^{n-k}-1$ разновидностей ошибок по числу смежных классов.

Чтобы иметь возможность получить информацию о том, к какому смежному классу относится полученная комбинация, каждому соответствующему арифметическим операциям смежному классу должна быть поставлена в соответствие некоторая контрольная последовательность символов, называемая опознавателем (синдромом).

Каждый символ опознавателя определяют в результате проверки на приемной стороне справедливости одного из равенств, которые мы составим для определения значений проверочных символов при кодировании.

Ранее указывалось, что в двоичном линейном коде значения проверочных символов подбирают так, чтобы сумма по модулю 2 всех символов (включая проверочный), входящих в каждое из равенств, равнялась нулю. В таком случае число единиц среди этих символов четное. Поэтому операции определения символов опознавателя называют *проверками на четность*. При отсутствии ошибок в результате всех проверок на четность образуется опознаватель, состоящий из одних нулей. Если проверочное одних нулей. Если проверочное равенство не удовлетворяется, то в соответствующем разряде опознавателя появляется единица. Исправление ошибок возможно лишь при наличии взаимно однозначного соответствия между множеством опознавателей и множеством смежных классов, а следовательно, и множеством подлежащих исправлению векторов ошибок.

Таким образом, количество подлежащих исправлению ошибок является определяющим для выбора числа избыточных символов $n–k$. Их должно быть достаточно для того, чтобы обеспечить необходимое число опознавателей.

Если, например, необходимо исправить все одиночные независимые ошибки, то исправлению подлежат n ошибок:

000…01
000…10
……….
010…00

100…00

Различных ненулевых опознавателей должно быть не менее n. Необходимое число проверочных разрядов, следовательно, должно определяться из соотношения

$$2^{n-k}-1\geq n$$

или

$$2^{n-k}-1\geq C_n^1$$

Если необходимо исправить не только все единичные, но и все двойные независимые ошибки, соответствующее неравенство принимает вид

$$2^{n-k}-1\geq C_n^1+C_n^2$$

В общем случае для исправления всех независимых ошибок кратности до s включительно получаем

$$2^{n-k}-1\geq C_n^1+C_n^2+...+C_n^S$$

Стоит подчеркнуть, что в приведенных соотношениях указывается теоретический предел минимально возможного числа проверочных символов, который далеко не во всех случаях можно реализовать практически. Часто проверочных символов требуется больше, чем следует из соответствующего равенства.

Одна из причин этого выяснится при рассмотрении процесса сопоставления каждой подлежащей исправлению ошибки с ее опознавателем.

2) понятие качества корректирующего кода

Одной из основных характеристик корректирующего кода является *избыточность кода,* указывающая степень удлинения кодовой комбинации для достижения определенной корректирующей способности.

Если на каждые m символов выходной последовательности кодера канала приходится k информационных и $(m-k)$ проверочных, то относительная избыточность кода может быть выражена одним из соотношений:

$$R_m = (m-k)/m \text{ или } R_k = (m-k)/k\,.$$

Величина R_k, изменяющаяся от 0 до ∞, предпочтительнее, так как лучше отвечает смыслу понятия избыточности. Коды, обеспечивающие заданную корректирующую способность при минимально возможной избыточности, называют *оптимальными.*

В связи с нахождением оптимальных кодов оценим, например, наибольшее возможное число Q разрешенных комбинаций m-значного двоичного кода, обладающего способностью исправлять взаимно независимые ошибки кратности до s включительно. Это равносильно отысканию числа комбинаций, кодовое расстояние между которыми не менее $d = 2s+1$.

Общее число различных исправляемых ошибок для каждой разрешающей комбинации составляет

$$\sum_{i=1}^{S} C_m^i\,,$$

где C_m^i – число ошибок кратности i.

Каждая из таких ошибок должна приводить к запрещенной комбинации, относящейся к подмножеству данной разрешенной комбинации. Совместно с этой комбинацией подмножество включает

$$1+\sum_{i=1}^{s} c_m^i$$

комбинаций.

Однозначное декодирование возможно только в том случае, когда названные подмножества не пересекаются. Так как общее число различных комбинаций m-значного двоичного кода составляет 2^m, число разрешенных комбинаций не может превышать

$$2^m \Big/ (1 + \sum_{i=1}^{s} c_m^i)$$

или

$$Q \le 2^m \Big/ \sum_{i=0}^{s} c_m^i .$$

Эта верхняя оценка найдена Хэммингом. Для некоторых конкретных значений кодового расстояния d, соответствующие Q укажем в табл. 7.14:

Таблица 7.14

Верхняя оценка кодового расстояния

d	Q	d	Q
1	2^m	5	$\le \dfrac{2^{m+1}}{m^2+1}$
2	$\le 2^{m-1}$	…	…
3	$\le \dfrac{2^m}{m+1}$	…	…
4	$\le \dfrac{2^{m-1}}{m}$	…	$\le \dfrac{2^m}{1+c_m^1+c_m^2+\ldots+c_m^k}$

Коды, для которых в приведенном соотношении достигается равенство, называют также *плотноупакованными*.

Однако не всегда целесообразно стремиться к использованию кодов, близких к оптимальным. Необходимо учитывать другой, не менее важный показатель качества корректирующего кода – сложность технической реализации процессов кодирования и декодирования.

Если информация должна передаваться по медленно действующей и дорогостоящей линии связи, а кодирующее и декодирующее устройства предполагается выполнить на высоконадежных и быстродействующих элементах, то сложность этих устройств не играет существенной роли. Решающим фактором в этом случае является повышение эффективности пользования линией связи, поэтому желательно применение корректирующих кодов с минимальной избыточностью.

Если же корректирующий код должен быть применен в системе, выполненной на элементах, надежность и быстродействие которых равны или близки надежности и быстродействию элементов кодирующей и декодирующей аппаратуры (например, для повышения достоверности воспроизведения информации с запоминающего устройства ЭВМ), то критерием качества корректирующего кода является надежность системы в целом, то есть с учетом возможных искажений и отказов в устройствах кодирования и декодирования. В этом случае часто более целесообразны коды с большей избыточностью, но обладающие преимуществом простоты технической реализации.

7.6. Составление таблицы опознавателей. Коды Хэмминга. Корректирующие свойства кодов

Начнем для простоты с установления опознавателей для случая исправления одиночных ошибок. Допустим, что необходимо закодировать 15 команд. Тогда требуемое число информационных разрядов равно четырем. Пользуясь соотношением $2^{n-k}-1=n$, определяем общее число разрядов кода, а следовательно, и число ошибок, подлежащих исправлению (n=7). Три избыточных разряда позволяют использовать в качестве опознавателей трехразрядные двоичные последовательности.

В данном случае ненулевые последовательности в принципе могут быть сопоставлены с подлежащими исправлению ошибками в любом порядке. Однако более целесообразно сопоставлять их с ошибками в разрядах, начиная с младшего, в порядке возрастания двоичных чисел (табл. 7.15).

При таком сопоставлении каждый опознаватель представляет собой двоичное число, указывающее номер разряда, в котором произошла ошибка.

Коды, в которых опознаватели устанавливаются по указанному принципу, известны как *коды Хэмминга*.

Таблица 7.15

Исправление одиночных ошибок в коде Хэмминга (7,4)

Векторы ошибок	Опознаватели	Векторы ошибок	Опознаватели
0000001	001	0010000	101
0000010	010	0100000	110
0000100	011	1000000	111
0001000	100		

Возьмем теперь более сложный случай исправления одиночных и двойных независимых ошибок. В качестве опознавателей одиночных ошибок в первом и втором разрядах можно принять, как и ранее, комбинации 0...001 и 0...010.

Однако в качестве опознавателя одиночной ошибки в третьем разряде комбинацию 0...011 взять нельзя. Такая комбинация соответствует ошибке одновременно в первом и во втором разрядах, а она также подлежит исправлению и, следовательно, ей должен соответствовать свой опознаватель 0...011.

В качестве опознавателя одиночной ошибки в третьем разряде можно взять только трехразрядную комбинацию 0...0100, так как множество двухразрядных комбинаций уже исчерпано. Подлежащий исправлению вектор ошибки 0...0101 также можно рассматривать как результат суммарного воздействия двух векторов ошибок 0...0100 и 0...001 и, следовательно, ему должен быть поставлен в соответствие опознаватель, представляющий собой сумму по модулю 2 опознавателей этих ошибок, т.е. 0...0101.

Аналогично находим, что опознавателем вектора ошибки 0...0110 является комбинация 0...0110.

Определяя опознаватель для одиночной ошибки в четвертом разряде, замечаем, что еще не использована одна из трехразрядных комбинаций, а именно 0...0111. Однако, выбирая в качестве опознавателя единичной ошибки в i-м разряде комбинацию с числом разрядов, меньшим i, необходимо убедиться в том, что для всех остальных подлежащих исправлению векторов ошибок, имеющих единицы в i-м и более младших разрядах, получатся опознаватели, отличные от уже использованных. В нашем случае подлежащими исправлению векторами ошибок с единицами в четвертом и более младших разрядах являются:

0...01001, 0...01010, 0...01100.

Если одиночной ошибке в четвертом разряде поставить в соответствие опознаватель 0...0111, то для указанных векторов опознавателями должны были бы быть соответственно

$$\oplus\begin{array}{r}0\ldots0111\\0\ldots0001\\\hline 0\ldots0110\end{array}\qquad\oplus\begin{array}{r}0\ldots0111\\0\ldots0010\\\hline 0\ldots0101\end{array}\qquad\oplus\begin{array}{r}0\ldots0111\\0\ldots0100\\\hline 0\ldots0011\end{array}$$

Однако эти комбинации уже использованы в качестве опознавателей других векторов ошибок, а именно:

0...0110, 0...0101, 0...0011.

Следовательно, во избежание неоднозначности при декодировании в качестве опознавателя одиночной ошибки в четвертом разряде следует взять четырехразрядную комбинацию 1000. Тогда для векторов ошибок

0...01001, 0...01010, 0...01100

опознавателями соответственно будут:

0...01001, 0...01010, 0...01100.

Аналогично можно установить, что в качестве опознавателя одиночной ошибки в пятом разряде может быть выбрана не использованная ранее четырехразрядная комбинация 01111.

Действительно, для всех остальных подлежащих исправлению векторов ошибок с единицей в пятом и более младших разрядах получаем опознаватели, отличающиеся от ранее установленных:

Векторы ошибок	Опознаватели
0...010001	0…01110
0...010010	0…01101
0...010100	0…01011
0...011000	0…00111

Продолжая сопоставление, можно получить таблицу опознавателей для векторов ошибок данного типа с любым числом разрядов. Так как опознаватели векторов ошибок с единицами в нескольких разрядах устанавливаются как суммы по модулю 2 опознавателей одиночных ошибок в этих разрядах, то для определения правил построения кода и составления проверочных равенств достаточно знать только опознаватели одиночных ошибок в каждом из разрядов. Для построения кодов, исправляющих двойные независимые ошибки, таблица таких опознавателей определена с помощью вычислительной машины вплоть до 29-го разряда. Опознаватели одиночных ошибок в первых пятнадцати разрядах приведены в табл. 7.16.

Таблица 7.16

Опознаватели одиночных ошибок

Номер разряда	Опознаватель	Номер разряда	Опознаватель	Номер разряда	Опознаватель
1	00000001	6	00010000	11	01101010
2	00000010	7	00100000	12	10000000
3	00000100	8	00110011	13	10010110
4	00001000	9	01000000	14	10110101
5	00001111	10	01010101	15	11011011

По тому же принципу аналогичные таблицы определены и для ошибок других типов, например для тройных независимых ошибок, пачек ошибок в два и три символа.

1) определение проверочных равенств

Итак, для любого кода, имеющего целью исправлять наиболее вероятные векторы ошибок заданного канала связи (взаимно независимые ошибки или пачки ошибок), можно составить табл. 7.15, опознавателей одиночных ошибок в каждом из разрядов. Пользуясь этой таблицей, нетрудно определить, символы каких разрядов должны входить в каждую из проверок на четность.

Рассмотрим в качестве примера опознаватели для кодов, предназначенных исправлять единичные ошибки (табл. 7.17).

В принципе можно построить код, усекая эту таблицу на любом уровне. Однако из таблицы видно, что оптимальными будут коды

(7, 4), (15, 11), где первое число равно n, а второе k, и другие, которые среди кодов, имеющих одно и то же число проверочных символов, допускают наибольшее число информационных символов.

Таблица 7.17

Опознаватели для кодов единичных ошибок

Номер разря-дов	Опозна-ватель	Номер разрядов	Опозна-ватель	Номер разрядов	Опозна-ватель
1	0001	7	0111	12	1100
2	0010	8	1000	13	1101
3	0011	9	1001	14	1110
4	0100	10	1010	15	1111
5	0101	11	1011	16	10000
6	0110				

Усечем эту таблицу на седьмом разряде и найдем номера разрядов, символы которых должны войти в каждое из проверочных равенств.

Предположим, что в результате первой проверки на четность для младшего разряда опознавателя будет получена единица. Очевидно, это может быть следствием ошибки в одном из разрядов, опознаватели которых в младшем разряде имеют единицу. Следовательно, первое проверочное равенство должно включать символы 1, 3, 5 и 7-го разрядов:

$$a_1 \oplus a_3 \oplus a_5 \oplus a_7 = 0.$$

Единица во втором разряде опознавателя может быть следствием ошибки в разрядах, опознаватели которых имеют единицу во втором разряде. Отсюда второе проверочное равенство должно иметь вид

$$a_2 \oplus a_3 \oplus a_6 \oplus a_7 = 0.$$

Аналогично находим и третье равенство:

$$a_4 \oplus a_5 \oplus a_6 \oplus a_7 = 0.$$

Чтобы эти равенства при отсутствии ошибок удовлетворялись для любых значений информационных символов в кодовой комбинации, в нашем распоряжении имеется три проверочных разряда. Мы должны так выбрать номера этих разрядов, чтобы каждый из них входил только в одно из равенств. Это обеспечит однозначное определение значений символов в проверочных разрядах при кодировании. Указанному условию удовлетворяют разряды, опознаватели которых имеют по одной единице. В рассматриваемом случае это будут первый, второй и четвертый разряды.

Таким образом, для кода (7, 4), исправляющего одиночные ошибки, искомые правила построения кода, т. е. соотношения, реализуемые в процессе кодирования, принимают вид:

$$\begin{aligned} a_1 &= a_3 \oplus a_5 \oplus a_7, \\ a_2 &= a_3 \oplus a_6 \oplus a_7, \\ a_4 &= a_5 \oplus a_6 \oplus a_7. \end{aligned} \tag{7.16}$$

Поскольку построенный код имеет минимальное хэммингово расстояние d_{min}=3, он может использоваться с целью обнаружения единичных и двойных ошибок. Обращаясь к табл. 7.17, легко убедиться, что сумма любых двух опознавателей единичных ошибок дает ненулевой опознаватель, что и является признаком наличия ошибки.

Пример. Построим групповой код объемом 15 слов, способный исправлять единичные и обнаруживать двойные ошибки.

В соответствии с $d_{И0\min} \geq r + s + 1$ код должен обладать минимальным хэмминговым расстоянием, равным 4. Такой код можно построить в два этапа. Сначала строим код заданного объема, способный исправлять единичные ошибки. Это код Хэмминга (7, 4). Затем добавляем еще один проверочный разряд, который обеспечивает четность числа единиц в разрешенных комбинациях.

Таким образом, получаем код (8, 4). В процессе кодирования реализуются соотношения:

$$\begin{aligned} a_1 &= a_3 \oplus a_5 \oplus a_7, \\ a_2 &= a_3 \oplus a_6 \oplus a_7, \\ a_4 &= a_5 \oplus a_6 \oplus a_7. \\ a_8 &= a_1 \oplus a_2 \oplus a_3 \oplus a_4 \oplus a_5 \oplus a_6 \oplus a_7, \end{aligned}$$

Обозначив синдром кода (7,4) через S_1, результат общей проверки на четность – через S_2 ($S_2 = \sum_{i=1}^{8} a_i$), и пренебрегая возможностью возникновения ошибок кратности 3 и выше, запишем алгоритм декодирования:

при $S_1 = 0$ и $S_2 = 0$ ошибок нет;

при $S_1 = 0$ и $S_2 = 1$ ошибка в восьмом разряде;

при $S_1 \neq 0$ и $S_2 = 0$ двойная ошибка (коррекция блокируется, посылается запрос повторной передачи);

при $S_1 \neq 0$ и $S_2 = 1$ одиночная ошибка (осуществляется ее исправление).

Пример. Используя таблицу 7.12, составим правила построения кода (8,2), исправляющего все одиночные и двойные ошибки.

Усекая табл. 7.11 на восьмом разряде, найдем следующие проверочные равенства:

$$
\begin{gathered}
a_1 \oplus a_5 \oplus a_8 = 0, \\
a_2 \oplus a_5 \oplus a_8 = 0, \\
a_3 \oplus a_5 = 0, \\
a_4 \oplus a_5 = 0, \\
a_6 \oplus a_8 = 0, \\
a_7 \oplus a_8 = 0.
\end{gathered}
$$

Соответственно правила построения кода выразим соотношениями

$$a_1 = a_5 \oplus a_8 \tag{7.16, а}$$

$$a_2 = a_5 \oplus a_8 \tag{7.16, б}$$

$$a_3 = a_5 \tag{7.16, в}$$

$$a_4 = a_5 \tag{7.16, г}$$

$$a_6 = a_8 \tag{7.16, д}$$

$$a_7 = a_8 \tag{7.16, е}$$

Отметим, что для построенного кода d_{min}=5, и, следовательно, он может использоваться для обнаружения ошибок кратности от 1 до 4.

Соотношения, отражающие процессы кодирования и декодирования непосредственно с использованием сумматоров по модулю два. Однако декодирующие устройства, построенные таким путем для кодов, предназначенных исправлять многократные ошибки, чрезвычайно громоздки. В этом случае более эффективны другие принципы декодирования.

2) мажоритарное декодирование групповых кодов

Для линейных кодов, рассчитанных на исправление многократных ошибок, часто более простыми оказываются декодирующие устройства, построенные по мажоритарному принципу. Этот метод декодирования называют также принципом голосования или способом декодирования по большинству проверок. В настоящее время известно значительное число кодов, допускающих мажоритарную схему декодирования, а также сформулированы некоторые подходы при конструировании таких кодов.

Мажоритарное декодирование тоже базируется на системе проверочных равенств. Система последовательно может быть разрешена относительно каждой из независимых переменных, причем в силу избыточности это можно сделать не единственным способом.

Любой символ a_i выражается d (минимальное кодовое расстояние) различными независимыми способами в виде линейных комбинаций других символов. При этом может использоваться тривиальная проверка a_i=a_i. Результаты вычислений подаются на соответствующий этому символу мажоритарный элемент. Последний представляет собой схему, имеющую d входов и один выход, на котором появляется единица, когда возбуждается больше половины его входов, и нуль, когда возбуждается число таких входов меньше половины. Если ошибки отсутствуют, то проверочные равенства не нарушаются и на выходе мажоритарного элемента получаем истинное значение символа. Если число проверок $d \geq 2s+1$ и появление ошибки кратности s и менее не приводит к нарушению более s проверок, то правильное решение может быть принято по большинству неискаженных проверок. Чтобы указанное условие выполнялось, любой другой символ a_j ($j \neq i$) не должен входить более чем в одно проверочное равенство. В этом случае мы имеем дело с системой *разделенных проверок*.

Пример. Построим систему разделенных проверок для декодирования информационных символов рассмотренного ранее группового кода (8,2).

Поскольку код рассчитан на исправление любых единичных и двойных ошибок, число проверочных равенств для определения каждого символа должно быть не менее 5. Подставив в равенства (7.16, а) и (7.16, б) значения a_8, полученные из равенств (7.16, д) и (7.16, е) и записав их относительно a_5 совместно с равенствами (7.16, в) и (7.16, г) и тривиальным равенством a_5=a_5, получим следующую систему разделенных проверок для символа a_5:

$$
\begin{gathered}
a_5 = a_6 \oplus a_1, \\
a_5 = a_7 \oplus a_2, \\
a_5 = a_3, \\
a_5 = a_4, \\
a_5 = a_5.
\end{gathered}
$$

Для символа a_8 систему разделенных проверок строим аналогично:

$$
\begin{gathered}
a_8 = a_3 \oplus a_1, \\
a_8 = a_4 \oplus a_2, \\
a_8 = a_6, \\
a_8 = a_7, \\
a_8 = a_8.
\end{gathered}
$$

Матричное представление линейных кодов. *Матрицей* размерности l×n называют упорядоченное множество l×n элементов, расположенных в виде прямоугольной таблицы с l строками и n столбцами:

$$
A \equiv \begin{bmatrix} a_{11}a_{12}.....a_{1n} \\ a_{21}a_{22}....a_{2n} \\ .... \\ a_{l1}a_{l2}.....a_{\mathrm{ln}} \end{bmatrix}
$$

Транспонированной матрицей к матрице A называют матрицу, строками которой являются столбцы, а столбцами строки матрицы A:

$$A^T \equiv \begin{bmatrix} a_{11}a_{12}.....a_{l1} \\ a_{21}a_{22}....a_{l2} \\ .... \\ a_{1n}a_{2n}.....a_{\ln} \end{bmatrix}$$

Матрицу размерности $n \times n$ называют *квадратной матрицей порядка* n. Квадратную матрицу, у которой по одной из диагоналей расположены только единицы, а все остальные элементы равны нулю, называют *единичной матрицей* I. Суммой двух матриц $A \equiv |a_{ij}|$ и $B \equiv |b_{ij}|$ размерности $1 \times n$ называют матрицу размерности $1 \times n$:

$$A + B \equiv |a_{ij}| + |b_{ij}| \equiv |a_{ij} + b_{ij}|.$$

Умножение матрицы $A \equiv |a_{ij}|$ размерности $l \times n$ на скаляр c дает матрицу размерности $l \times n$:

$$cA \equiv c\ |a_{ij}| \equiv |c\ a_{ij}|.$$

Матрицы $A \equiv |a_{ij}|$ размерности $l \times n$ и $B \equiv |b_{jk}|$ размерности $n \times m$ могут быть перемножены, причем элементами c_{ik} матрицы – произведения размерности $l \times m$ являются суммы произведений элементов l-й строки матрицы A на соответствующие элементы k-го столбца матрицы B:

$$c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}.$$

В теории кодирования элементами матрицы являются элементы некоторого поля $GF(P)$, а строки и столбцы матрицы рассматриваются как векторы. Сложение и умножение элементов матриц осуществляется по правилам поля $GF(P)$.

Пример Вычислим произведение матриц с элементами из поля GF (2):

$$M_1 = \begin{bmatrix} 011 \\ 100 \\ 001 \end{bmatrix} \qquad M_2 = \begin{bmatrix} 111 \\ 010 \\ 100 \end{bmatrix}$$

Элементы c_{ik} матрицы произведения $M = M_1M_2$ будут равны:

$$c_{11} = (011)\,(101) = 0 + 0 + 1 = 1$$
$$c_{12} = (011)\,(110) = 0 + 1 + 0 = 1$$
$$c_{13} = (011)\,(100) = 0 + 0 + 0 = 0$$
$$c_{21} = (100)\,(101) = 1 + 0 + 0 = 1$$
$$c_{22} = (100)\,(110) = 1 + 0 + 0 = 1$$
$$c_{23} = (100)\,(100) = 1 + 0 + 0 = 1$$
$$c_{31} = (001)\,(101) = 0 + 0 + 1 = 1$$
$$c_{32} = (001)\,(110) = 0 + 0 + 0 = 0$$
$$c_{33} = (001)\,(100) = 0 + 0 + 0 = 0$$

Следовательно,

$$M = \begin{bmatrix} 110 \\ 111 \\ 100 \end{bmatrix}$$

Зная закон построения кода, определим все множество разрешенных кодовых комбинаций. Расположив их друг под другом, получим матрицу, совокупность строк которой является подпространством векторного пространства n-разрядных кодовых комбинаций (векторов) из элементов поля $GF(P)$. В случае двоичного (n,k)-кода матрица насчитывает n столбцов и 2^k–1 строк (исключая нулевую). Например, для рассмотренного ранее кода (8,2), исправляющего все одиночные и двойные ошибки, матрица имеет вид

$$\begin{bmatrix} a_5a_8a_1a_2a_3a_4a_6a_7 \\ 0\,1\,1\,1\;\;0\,0\,1\,1 \\ 1\,0\,1\,1\;\;1\,1\,0\,0 \\ 1\,1\,0\,0\;\;1\,1\,1\,1 \end{bmatrix}$$

При больших n и k матрица, включающая все векторы кода, слишком громоздка. Однако совокупность векторов, составляющих линейное пространство разрешенных кодовых комбинаций, является линейно зависимой, так как часть векторов может быть представлена в виде линейной комбинации некоторой ограниченной совокупности векторов, называемой *базисом пространства*.

Совокупность векторов V_1, V_2, V_3, ..., V_n называют *линейно зависимой*, когда существуют скаляры $c_1,..c_n$ (не все равные нулю), при которых

$$c_1V_1 + c_2V_2 + \ldots + c_nV_n = 0$$

В приведенной матрице, например, третья строка представляет собой суммы по модулю два первых двух строк.

Для полного определения пространства разрешенных кодовых комбинаций линейного кода достаточно записать в виде матрицы только совокупность линейно независимых векторов. Их число называют *размерностью векторного пространства*.

Среди $2^k - 1$ ненулевых двоичных кодовых комбинаций – векторов их только k. Например, для кода (8,2)

$$M_{8,2} = \begin{bmatrix} a_1a_2a_3a_4a_5a_6a_7a_8 \\ 1\ 1\ 0\ 0\ 0\ 1\ 1\ \ 1 \\ 1\ 1\ 1\ \ 1\ 1\ 0\ 0\ \ 0 \end{bmatrix}$$

Матрицу, составленную из любой совокупности векторов линейного кода, образующей базис пространства, называют *порождающей (образующей)* матрицей кода.

Если порождающая матрица содержит k строк по n элементов поля $GF(q)$, то код называют (n,k)-кодом. В каждой комбинации (n, k)-кода k информационных символов и n–k проверочных. Общее число разрешенных кодовых комбинаций (исключая нулевую) $Q = q^k$-1.

Зная порождающую матрицу кода, легко найти разрешенную кодовую комбинацию, соответствующую любой последовательности A_{ki} из k информационных символов. Она получается в результате умножения вектора A_{ki} на порождающую матрицу $M_{n,k}$:

$$A_{ni} = A_{ki} \cdot M_{n,k}.$$

Найдем, например, разрешенную комбинацию кода (8,2), соответствующую информационным символам a_5=l, a_8 = 1:

$$|11|M_{8,2} = |11| \begin{bmatrix} 1\,1\,0\,0\,0\,1\,1\,1 \\ 1\,1\,1\,1\,1\,0\,0\,0 \end{bmatrix} = 0\;0\,1\,1\,1\,1\,1\,1.$$

Пространство строк матрицы остается неизменным при выполнении следующих элементарных операций над строками: 1) перестановка любых двух строк; 2) умножение любой строки на ненулевой элемент поля; 3) сложение какой-либо строки с произведением другой строки на ненулевой элемент поля, а также при перестановке столбцов.

Если образующая матрица кода M_2 получена из образующей матрицы кода M_1 с помощью элементарных операций над строками, то обе матрицы порождают один и тот же код. Перестановка столбцов образующей матрицы кода приводит к образующей матрице *эквивалентного* кода. Эквивалентные коды весьма близки по своим свойствам. Корректирующая способность таких кодов одинакова.

Для анализа возможностей линейного (n,k)-кода, а также для упрощения процесса кодирования удобно, чтобы порождающая матрица ($M_{n,k}$) состояла из двух матриц: единичной матрицы размерности $k \times k$ и дописываемой справа *матрицы-дополнения* (контрольной подматрицы) размерности $k \cdot (n-k)$, которая соответствует n – k проверочным разрядам:

$$M_{n,k} = \left[I_k P_{k,n-k}\right] = \begin{bmatrix} 1\;0...0\,p_{1,k+1}\,p_{1,k+2}...p_{1,j}...p_{1,n} \\ 0\;1...0\,p_{2,k+1}\,p_{2,k+2}...p_{2,j}...p_{2,n} \\ .... \\ 0\;1...0\,p_{i,k+1}\,p_{i,k+2}...p_{i,j}...p_{i,n} \\ .... \\ 0\;0...1\,p_{k,k+1}\,p_{k,k+2}...p_{k,j}...p_{k,n} \end{bmatrix}. \qquad (7.17)$$

Разрешенные кодовые комбинации кода с такой порождающей матрицей отличаются тем, что первые k символов в них совпадают с исходными информационными, а проверочными оказываются (n–k) последних символов.

Действительно, если умножим вектор-строку $A_{k,i}$= =(a_1,a_2…a_i...a_k) на матрицу $M_{n,k}$= [$I_k P_{k,n-k}$], получим вектор

$$A_{n,i} = (a_1 a_2 ... a_i ... a_k ... a_{k+1} ... a_j ... a_n),$$

где проверочные символы $a_j (k+1 \le j \le n)$ являются линейными комбинациями информационных:

$$a_j = \sum_{i=1}^{k} a_i p_{ij}. \qquad (7.18)$$

Коды, удовлетворяющие этому условию, называют *систематическими*. Для каждого линейного кода существует эквивалентный систематический код.

Информацию о способе построения такого кода содержит матрица-дополнение. Если правила построения кода (уравнения кодирования) известны, то значения символов любой строки матрицы-дополнения получим, применяя эти правила к символам соответствующей строки единичной матрицы.

Пример. Запишем матрицы I_k, $P_{k,n-k}$ и $M_{n,k}$ для двоичного кода (7,4). Единичная матрица на четыре разряда имеет вид

$$I_4 = \begin{bmatrix} 1000 \\ 0100 \\ 0010 \\ 0001 \end{bmatrix}.$$

Один из вариантов матрицы дополнения можно записать, используя соотношения для кода (8,1)

$$P_{4,3} = \begin{bmatrix} 110 \\ 101 \\ 011 \\ 111 \end{bmatrix}.$$

Тогда для двоичного кода Хэмминга имеем:

$$M_{7,4} = \begin{bmatrix} a_3 a_5 a_6 a_7 a_1 a_2 a_4 \\ 1\,0\,0\,0\,1\,1\,0 \\ 0\,1\,0\,0\,1\,0\,1 \\ 0\,0\,1\,0\,0\,1\,1 \\ 0\,0\,0\,1\,1\,1\,1 \end{bmatrix}.$$

Запишем также матрицу для систематического кода (7,4):

$$M_{7,4} = \begin{bmatrix} a_1 a_2 a_3 a_4 a_5 a_6 a_7 \\ 1\,0\,0\,0\,1\,1\,0 \\ 0\,1\,0\,0\,1\,0\,1 \\ 0\,0\,1\,0\,0\,1\,1 \\ 0\,0\,0\,1\,1\,1\,1 \end{bmatrix}.$$

В свою очередь, по заданной матрице-дополнению $P_{k,n\text{-}k}$ можно определить равенства, задающие правила построения кода. Единица в первой строке каждого столбца указывает на то, что в образовании соответствующего столбцу проверочного разряда участвовал первый информационный разряд. Единица в следующей строке любого столбца говорит об участии в образовании проверочного разряда второго информационного разряда и т, д.

Так, как матрица-дополнение содержит всю информацию о правилах построения кода, то систематический код с заданными свойствами можно синтезировать путем построения соответствующей матрицы-дополнения.

Так как минимальное кодовое расстояние d для линейного кода равно минимальному весу его ненулевых векторов, то в матрицу-дополнение должны быть включены такие k строк, которые удовлетворяли бы следующему общему условию: вектор-строка образующей матрицы, получающаяся при суммировании любых l $(1 \le l \le k)$ строк, должна содержать не менее d–l отличных от нуля символов.

Действительно, при выполнении указанного условия любая разрешенная кодовая комбинация, полученная суммированием l строк образующей матрицы, имеет не менее d ненулевых символов, так как l ненулевых символов она всегда содержит в результате суммирования строк единичной матрицы.

Синтезируем таким путем образующую матрицу двоичного систематического кода (7,4) с минимальным кодовым расстоянием d = 3.

В каждой вектор-строке матрицы-дополнения, согласно сформулированному условию (при l=1), должно быть не менее двух единиц. Среди трехразрядных векторов таких имеется четыре: 011, 110, 101, 111.

Эти векторы могут быть сопоставлены со строками единичной матрицы в любом порядке. В результате получим матрицы систематических кодов, эквивалентных коду Хэмминга, например:

$$M_{7,4}=\begin{bmatrix}1\,0\,0\,0 & 0\,1\,1\\ 0\,1\,0\,0 & 1\,1\,1\\ 0\,0\,1\,0 & 1\,0\,1\\ 0\,0\,0\,1 & 1\,1\,0\end{bmatrix}$$

Нетрудно убедиться, что при суммировании нескольких строк такой матрицы (l>1) получим вектор-строку, содержащую не менее d=3 ненулевых символов.

Имея образующую матрицу систематического кода $M_{n,k}=[I_k P_{k,n-k}]$, можно построить так называемую проверочную (контрольную) матрицу H размерности $(n-k)\times n$:

$$H=\left[-P^T{}_{k,n-k}I_{n-k}\right]=\begin{bmatrix}p_{1,k+1}\,p_{2,k+1}\cdots p_{k,k+1}\ -1\quad 0.....0\\ p_{1,k+2}\,p_{2,k+2}\cdots p_{k,k+2}\quad 0-1......0\\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ p_{1,n}\quad p_{2,n}.....p_{k,n}\qquad 0\ \ 0...-1\end{bmatrix}$$

При умножении неискаженного кодового вектора A_{ni} на матрицу, транспонированную к матрице H, получим вектор, все компоненты которого равны нулю:

$$A_{n,i}H^T=\left|a_1,a_2,...,a_k,a_{k+1,},...,a_j,...,a_n\right|\cdot\begin{bmatrix}p_{1,k+1}\cdots p_{1,j}\cdots p_{1,n}\\ p_{2,k+1}\cdots p_{2,j}\cdots p_{2,n}\\ \cdots\cdots\cdots\cdots\cdots\cdots\\ p_{k,k+1}\cdots p_{k,j}\cdots p_{k,n}\\ -1........0........0\\ 0......-1........0\\ 0.........0.....-1\end{bmatrix}=$$

$$=\left|S_{k+1},S_{k+2},...,S_j,...S_n\right|=\left|0,0,..,0,..,0\right|.$$

Каждая компонента S_j является результатом проверки справедливости соответствующего уравнения декодирования:

$$S_j = \sum_{i=1}^{k} a_i P_{ij} - a_j = 0.$$

В общем случае, когда кодовый вектор

$$A_{n,i} = (a_1, a_2, ..., a_i, ..., a_k, a_{k+1}, ..., a_j, ..., a_n)$$

искажен вектором ошибки $\xi_{n,i} = (\xi_1, \xi_2, ..., \xi_i, ..., \xi_k, \xi_{k+1}, ..., \xi_j, ..., \xi_n)$, умножение вектора $(A_{ni} + \xi_{ni})$ на матрицу H^m дает ненулевые компоненты:

$$S_j = \sum_{i=1}^{k} \xi_i P_{ij} - \xi_j = 0.$$

Отсюда видно, что $S_j (k+1 \le j \le n)$ представляют собой символы, зависящие только от вектора ошибки, а вектор $S = (S_{k+1}, S_{k+2}, ..., S_j, ..., S_n)$ является не чем иным, как опознавателем ошибки (синдромом).

Для двоичных кодов (операция сложения тождественна операции вычитания) проверочная матрица имеет вид:

$$H = \left[P^T{}_{k,n-k} I_{n-k} \right] = \begin{bmatrix} p_{1,k+1} p_{2,k+1} ... p_{k,k+1} \ 1\ 0......0 \\ p_{1,k+2} p_{2,k+2} ... p_{k,k+2} 0\ 1.....0 \\ .... \\ p_{1,n} \quad p_{2,n} ..... p_{k,n} \quad 0\ 0....1 \end{bmatrix}$$

Пример. Найдем проверочную матрицу H для кода (7,4) с образующей матрицей M:

$$M_{7,4} = \begin{bmatrix} 1\,0\,0\,0\,1\,1\,0 \\ 0\,1\,0\,0\,1\,0\,1 \\ 0\,0\,1\,0\,0\,1\,1 \\ 0\,0\,0\,1\,1\,1\,1 \end{bmatrix}$$

Определим синдромы в случаях отсутствия и наличия ошибки в кодовом векторе 1100011.

Выполним транспонирование матрицы $P_{4,3}$

$$P_{4,3} = \begin{bmatrix} 1101 \\ 1011 \\ 0111 \end{bmatrix}.$$

Запишем проверочную матрицу:

$$H = \begin{bmatrix} 1101100 \\ 1011010 \\ 0111001 \end{bmatrix}.$$

Умножение на H^m неискаженного кодового вектора 1100011 дает нулевой синдром:

$$[1100011] \begin{bmatrix} 110 \\ 101 \\ 011 \\ 111 \\ 100 \\ 010 \\ 001 \end{bmatrix} = [000].$$

При наличии в кодовом векторе ошибки, например, в 4-м разряде (1101011) получим:

$$[1101011] \begin{bmatrix} 110 \\ 101 \\ 011 \\ 111 \\ 100 \\ 010 \\ 001 \end{bmatrix} = [111].$$

Следовательно, вектор-строка 111 в данном коде является опознавателем (синдромом) ошибки в четвертом разряде. Аналогично можно найти и синдромы других ошибок. Множество всех опознавателей идентично множеству опознавателей кода Хэмминга (7,4), но сопоставлены они конкретным векторам ошибок по-иному, в соответствии с образующей матрицей данного (эквивалентного) кода.

7.7. Циклические коды, БЧХ – коды, код Хэмминга, сверточные коды

1) общие понятия и определения

Любой групповой код (n, k) может быть записан в виде матрицы, включающей k линейно независимых строк по n символов и, наоборот, любая совокупность k линейно независимых n-разрядных кодовых комбинаций может рассматриваться как образующая матрица некоторого группового кода. Среди всего многообразия таких кодов можно выделить коды, у которых строки образующих матриц связаны дополнительным условием цикличности.

Все строки образующей матрицы такого кода могут быть получены циклическим сдвигом одной комбинации, называемой образующей для данного кода. Коды, удовлетворяющие этому условию, получили название *циклических кодов*.

Сдвиг осуществляется справа налево, причем крайний левый символ каждый раз переносится в конец комбинации. Запишем, например, совокупность кодовых комбинаций, получающихся циклическим сдвигом комбинации 001011:

$$G = \begin{bmatrix} 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

Число возможных циклических (n,k)-кодов значительно меньше числа различных групповых (n, k)-кодов.

При описании циклических кодов n-разрядные кодовые комбинации представляются в виде многочленов фиктивной переменной x.

Показатели степени у x соответствуют номерам разрядов (начиная с нулевого), а коэффициентами при x в общем случае являются элементы поля $GF(q)$. При этом наименьшему разряду числа соответствует фиктивная переменная $x^0=1$. Многочлен с коэффициентами из поля $GF(q)$ называют многочленом над полем $GF(q)$. Так как мы ограничиваемся рассмотрением только двоичных кодов, то коэффициентами при x будут только цифры 0 и 1. Иначе говоря, будем оперировать с многочленами над полем $GF(2)$.

Запишем, например, в виде многочлена образующую кодовую комбинацию 01011:

$$G(x) = 0 \cdot x^4 + 1 \cdot x^3 + 0 \cdot x^2 + 1 \cdot x + 1.$$

Поскольку члены с нулевыми коэффициентами при записи многочлена опускаются, образующий многочлен:

$$G(x) = x^3 + x^3 + 1$$

Наибольшую степень x в слагаемом с ненулевым коэффициентом называют *степенью многочлена*. Теперь действия над кодовыми комбинациями сводятся к действиям над многочленами. Суммирование многочленов осуществляется с приведением коэффициентов по модулю два.

Указанный циклический сдвиг некоторого образующего многочлена степени $n–k$ без переноса единицы в конец кодовой комбинации соответствует простому умножению на x. Умножив, например, первую строку матрицы (001011), соответствующую многочлену $g_0(x) = x^3 + x + 1$, на x, получим вторую строку матрицы (010110), соответствующую многочлену $x \bullet g_0(x)$.

Нетрудно убедиться, что кодовая комбинация, получающаяся при сложении этих двух комбинаций, также будет соответствовать результату умножения многочлена $x^3 + x + 1$ на многочлен $x+1$. Действительно,

$$\begin{array}{r} x^3 + 0 + x + 1 \\ x + 1 \\ \hline \end{array}$$

$$\oplus \begin{array}{r} 001011 \\ \hline 010110 \\ \hline 011101 \end{array} \qquad \oplus \begin{array}{r} x^3 + 0 + x + 1 \\ \hline x^4 + 0 + x^2 + x \\ \hline x^4 + x^3 + x^2 + 0 + 1 \end{array}$$

Циклический сдвиг строки матрицы с единицей в старшем (n-м) разряде (слева) равносилен умножению соответствующего строке многочлена на x с одновременным вычитанием из результата многочлена $x^n + 1 = x^n - 1$, т. е. с приведением по модулю $x^n + 1$.

Отсюда ясно, что любая разрешенная кодовая комбинация циклического кода может быть получена в результате умножения образующего многочлена на некоторый другой многочлен с приведением результата по модулю $x^n + 1$. Иными словами, при соответствующем выборе образующего многочлена любой многочлен циклического кода будет делиться на него без остатка.

Ни один многочлен, соответствующий запрещенной кодовой комбинации, на образующий многочлен без остатка не делится. Это свойство позволяет обнаружить ошибку. По виду остатка можно определить и вектор ошибки.

Умножение и деление многочленов весьма просто осуществляется на регистрах сдвига с обратными связями, что и явилось причиной широкого применения циклических кодов.

2) математическое введение к циклическим кодам

Так как каждая разрешенная комбинация n-разрядного циклического кода есть произведение двух многочленов, один из которых является образующим, то эти комбинации можно рассматривать как подмножества всех произведений многочленов степени не выше n–1. Это наталкивает на мысль использовать для построения этих кодов еще одну ветвь теории алгебраических систем, а именно теорию колец.

Как следует из приведенного ранее определения, для образования кольца на множестве n-разрядных кодовых комбинаций необходимо задать две операции: сложение и умножение.

Операция сложения многочленов уже выбрана нами с приведением коэффициентов по модулю два.

Определим теперь операцию умножения. Нетрудно видеть, что операция умножения многочленов по обычным правилам с приведением подобных членов по модулю два может привести к нарушению условия замкнутости. Действительно, в результате умножения могут быть получены многочлены более высокой степени, чем $n-1$, вплоть до $2(n-1)$, а соответствующие им кодовые комбинации будут иметь число разрядов, превышающее n и, следовательно, не относятся к рассматриваемому множеству. Поэтому операция символического умножения задается так:

1) многочлены перемножаются по обычным правилам, но с приведением подобных членов по модулю два;

2) если старшая степень произведения не превышает $n-1$, то оно и является результатом символического умножения;

3) если старшая степень произведения больше или равна n, то многочлен произведения делится на заранее определенный многочлен степени n и результатом символического умножения считается остаток от деления.

Степень остатка не превышает $n-1$, и, следовательно, этот многочлен принадлежит к рассматриваемому множеству k-разрядных кодовых комбинаций.

При анализе циклического сдвига с перенесением единицы в конец кодовой комбинации установлено, что таким многочленом n-й степени является многочлен x^n+1.

Действительно, в результате умножения многочлена степени $n-1$ на x получим

$$G(x)=(x^{n-1}+x^{n-2}+\ldots+x+1)x=x^n+x^{n-1}+\ldots+x$$

Следовательно, чтобы результат умножения и теперь соответствовал кодовой комбинации, образующейся путем циклического сдвига исходной кодовой комбинации, в нем необходимо заменить x^n на 1. Такая замена эквивалентна делению полученного при умножении многочлена на x^n+1 с записью в качестве результата остатка от деления, что обычно называют *взятием остатка* или *приведением по модулю* x^n+1 (сам остаток при этом называют вычетом).

Выделим теперь в нашем кольце подмножество всех многочленов, кратных некоторому многочлену $g(x)$. Такое подмножество называют *идеалом,* а *многочлен* $g(x)$*-порождающим многочленом идеала*.

Количество различных элементов в идеале определяется видом его порождающего многочлена. Если на порождающий многочлен взять 0, то весь идеал будет составлять только этот многочлен, так как умножение его на любой другой многочлен дает 0.

Если за порождающий многочлен принять $1[g(x)=1]$, то в идеал войдут все многочлены кольца. В общем случае число элементов идеала, порожденного простым многочленом степени $n-k$, составляет 2^k.

Теперь становится понятным, что циклический двоичный код в построенном нами кольце n-разрядных двоичных кодовых комбинаций является идеалом.

Остается выяснить, как выбрать многочлен $g(x)$, способный породить циклический код с заданными свойствами.

3) требования, предъявляемые к образующему многочлену

Согласно определению циклического кода все многочлены, соответствующие его кодовым комбинациям, должны делиться на $g(x)$ без остатка. Для этого достаточно, чтобы на $g(x)$ делились без остатка многочлены, составляющие образующую матрицу кода. Последние получаются циклическим сдвигом, что соответствует последовательному умножению $g(x)$ на x с приведением по модулю x^n+1.

Следовательно, в общем случае многочлен $g_i(x)$ является остатком от деления произведения $g(x)\cdot x^i$ на многочлен x^n+1 и может быть записан так:

$$g(x) = g(x)x^i + c(x^n+1),$$

где c =1, если степень $g(x)\,x^i$ превышает n-1; c = 0, если степень $g(x)\,x^i$ не превышает $n-1$.

Отсюда следует, что все многочлены матрицы, а поэтому и все многочлены кода будут делиться на $g(x)$ без остатка только в том случае, если на $g(x)$ будет делиться без остатка многочлен x^n+1.

Таким образом, чтобы $g(x)$ мог породить идеал, а, следовательно, и циклический код, он должен быть делителем многочлена x^n+1.

Поскольку для кольца справедливы все свойства группы, а для идеала – все свойства подгруппы, кольцо можно разложить на смежные классы, называемые в этом случае *классами вычетов по идеалу*.

Первую строку разложения образует идеал, причем нулевой элемент располагается крайним слева. В качестве образующего первого класса вычетов можно выбрать любой многочлен, не принадлежащий идеалу. Остальные элементы данного класса вычетов образуются путем суммирования образующего многочлена с каждым многочленом идеала.

Если многочлен $g(x)$ степени $m=n-k$ является делителем x^n+1, то любой элемент кольца либо делится на $g(x)$ без остатка (тогда он является элементом идеала), либо в результате деления появляется остаток $r(x)$, представляющий собой многочлен степени не выше $m-1$.

Элементы кольца, дающие в остатке один и тот же многочлен $r_i(x)$, относятся к одному классу вычетов. Приняв многочлены $r(x)$ за образующие элементы классов вычетов, разложение кольца по идеалу с образующим многочленом $g(x)$ степени m, где $f(x)$ – произвольный многочлен степени не выше $n-m-1$ (табл. 7.18).

Как отмечалось, групповой код способен исправить столько разновидностей ошибок, сколько различных классов насчитывается в приведенном разложении. Следовательно, корректирующая способность циклического кода будет тем выше, чем больше остатков может быть образовано при делении многочлена, соответствующего искаженной кодовой комбинации, на образующий многочлен кода.

Таблица 7.18

Разложение кольца по идеалу с образующим многочленом $g(x)$ степени m

0	$g(x)$	$x \cdot g(x)$	$(x+1) \cdot g(x)$	…	$f(x) \cdot g(x)$
$r_1(x)$	$g(x) + r_1(x)$	$x \cdot g(x) + r_1(x)$	$(x+1) \cdot g(x) + r_1(x)$	…	$f(x) \cdot g(x) + r_1(x)$
$r_2(x)$	$g(x) + r_2(x)$	$x \cdot g(x) + r_2(x)$	$(x+1) \cdot g(x) + r_2(x)$	…	$f(x) \cdot g(x) + r_2(x)$
…	…	…	…	…	…
…	…	…	…	…	…
$r_n(x)$	$g(x) + r_n(x)$	$x \cdot g(x) + r_n(x)$	$(x+1) \cdot g(x) + r_n(x)$	…	$f(x) \cdot g(x) + r_n(x)$

Наибольшее число остатков, равное $2^m - 1$ (исключая нулевой), может обеспечить только неприводимый (простой) многочлен, который делится сам на себя и не делится ни на какой другой многочлен (кроме 1).

4) выбор образующего многочлена по заданному объему кода и заданной корректирующей способности

По заданному объему кода однозначно определяется число информационных разрядов k. Далее необходимо найти наименьшее n, обеспечивающее обнаружение или исправление ошибок заданной кратности. В случае циклического кода эта проблема сводится к нахождению нужного многочлена $g(x)$.

Начнем рассмотрение с простейшего циклического кода, обнаруживающего все одиночные ошибки.

5) обнаружение одиночных ошибок

Любая принятая по каналу связи кодовая комбинация $h(x)$, возможно содержащая ошибку, может быть представлена в виде суммы по модулю два неискаженной комбинации кода $f(x)$ и вектора ошибки $\xi(x)$:

$$h(x) = f(x) \oplus \xi(x)$$

При делении $h(x)$ на образующий многочлен $g(x)$ остаток, указывающий на наличие ошибки, обнаруживается только в том случае, если многочлен, соответствующий вектору ошибки, не делится на $g(x)$: $f(x)$ – неискаженная комбинация кода и, следовательно, на $g(x)$ делится без остатка.

Вектор одиночной ошибки имеет единицу в искаженном разряде и нули во всех остальных разрядах. Ему соответствует многочлен $\xi(x) = x^i$. Последний не должен делиться на $g(x)$. Среди неприводимых многочленов, входящих в разложении $x^n + 1$, многочленом наименьшей степени, удовлетворяющим указанному условию, является $x + 1$. Остаток от деления любого многочлена на $x + 1$ представляет собой многочлен нулевой степени и может принимать только два значения: 0 или 1. Все кольцо в данном случае состоит из идеала, содержащего многочлены с четным числом членов, и одного класса вычетов, соответствующего единственному остатку, равному 1. Таким образом, при любом числе информационных разрядов необходим только один проверочный разряд. Значение символа этого разряда как раз и обеспечивает четность числа единиц в любой разрешенной кодовой комбинации, а, следовательно, и делимость ее на $x^n + 1$.

Полученный циклический код с проверкой на четность способен обнаруживать не только одиночные ошибки в отдельных разрядах, но и ошибки в любом нечетном числе разрядов.

6) исправление одиночных (код Хэмминга) или обнаружение двойных ошибок

Прежде, чем исправить одиночную ошибку в принятой комбинации из n разрядов, необходимо определить, какой из разрядов был искажен. Это можно сделать только в том случае, если каждой одиночной ошибке в определенном разряде соответствуют свой класс вычетов и свой опознаватель. Так как в циклическом коде опознавателями ошибок являются остатки от деления многочленов ошибок на образующий многочлен кода $g(x)$, то $g(x)$ должно обеспечить требуемое число различных остатков при делении векторов ошибок с единицей в искаженном разряде. Как отмечалось, наибольшее число остатков дает неприводимый многочлен. При степени многочлена $m = n - k$ он может дать $2n^{n-k} - 1$ ненулевых остатков (нулевой остаток является опознавателем безошибочной передачи).

Следовательно, необходимым условием исправления любой одиночной ошибки является выполнение неравенства

$$2^{n-k} - 1 \geq c_n^1 = n,$$

где C_n^1 – общее число разновидностей одиночных ошибок в кодовой комбинации из n символов; отсюда находим степень образующего многочлена кода

$$m = n - k \geq \log_2(n+1)$$

и общее число символов в кодовой комбинации. Наибольшие значения k и n для различных m можно найти пользуясь таблицей:

M	1	2	3	4	5	6	7	8	9	10
N	1	3	7	15	31	63	127	255	511	1023
K	0	1	4	11	26	57	120	247	502	1013

Как указывалось, образующий многочлен $g(x)$ должен быть делителем двучлена $x^n + 1$. Доказано, что любой двучлен типа $2x^{2m-1} + 1 = x^n + 1$ может быть представлен произведением всех неприводимых многочленов, степени которых являются делителями числа m (от 1 до m включительно). Следовательно, для любого m существует по крайней мере один неприводимый многочлен степени m, входящий сомножителем в разложение двучлена $x^n + 1$.

Пользуясь этим свойством, а также имеющимися в ряде книг таблицами многочленов, неприводимых при двоичных коэффициентах, выбрать образующий многочлен при известных n и m несложно. Определив образующий многочлен, необходимо убедиться в том, что он обеспечивает заданное число остатков.

Пример. Выберем образующий многочлен для случая $n = 15$ и $m = 4$.

Двучлен $x^{15} + 1$ можно записать в виде произведения всех неприводимых многочленов, степени которых являются делителями числа 4. Последнее делится на 1, 2, 4.

В таблице неприводимых многочленов находим один многочлен первой степени, а именно $x+1$, один многочлен второй степени $x^2 + x + 1$ и три многочлена четвертой степени: $x^4 + x + 1$, $x^4 + x^3 + 1$, $x^4 + x^3 + x^2 + x + 1$. Перемножив все многочлены, убедимся в справедливости соотношения

$$(x+1)(x^2+x+1)(x^4+x+1)(x^4+x^3+1)(x^4+x^3+x^2+x+1) = x^{15}+1.$$

Один из сомножителей четвертой степени может быть принят за образующий многочлен кода. Возьмем, например, многочлен $x^4 + x^3 + 1$, или в виде двоичной последовательности 11001.

Чтобы убедиться, что каждому вектору ошибки соответствует отличный от других остаток, необходимо поделить каждый из этих векторов на 11001.

Векторы ошибок m младших разрядов имеют вид:

00…000, 00…0010, 00…0100, 00…1000.

Степени соответствующих им многочленов меньше степени образующего многочлена $g(x)$. Поэтому они сами являются остатками при нулевой целой части. Остаток, соответствующий вектору ошибки в следующем старшем разряде, получаем при делении 00...10000 на 11001, т.е.

$$\begin{array}{r|l} \oplus\, 10000 & 11001 \\ \cline{2-2} \underline{11001} & \\ 1001 & \end{array}$$

Аналогично могут быть найдены и остальные остатки. Однако их можно получить проще, деля на $g(x)$ комбинацию в виде единицы с рядом нулей и выписывая все промежуточные остатки:

$$
\begin{array}{l|l}
10000000000\ldots & 11001 \\
\oplus\ 11001 & \\
10010 & \\
\oplus\ 11001 & \\
10110 & \\
\oplus\ 11001 & \\
11110 & \\
\oplus\ 11001 & \\
011100 & \\
\oplus\ 11001 & \\
010100 & \\
\oplus\ 11001 & \\
11010 & \\
\oplus\ 11001 & \\
0011000 & \\
\oplus\ 11001 & \\
0001000 &
\end{array}
$$

Остатки

1001
1011
1111
0111
1110
0101
1010
1101
0011
0110
1100
0001
0010
0100
1000

При последующем делении остатки повторяются.

Таким образом, мы убедились в том, что число различных остатков при выбранном $g(x)$ равно n = 15, и, следовательно, код, образованный таким $g(x)$, способен исправить любую одиночную ошибку. С тем же успехом за образующий многочлен кода мог быть принят и многочлен $x^4 + x + 1$. При этом был бы получен код, эквивалентный выбранному.

Однако использовать для тех же целей многочлен $x^4 + x^3 + x^2 + x + 1$ нельзя. При проверке числа различных остатков обнаруживается, что их у него не 15, а только 5. Действительно

$$
\begin{array}{l|l}
10000.. & 11111 \\
\oplus\ 11111 & \\
11110 & \\
\oplus\ 11111 & \\
0001000 &
\end{array}
$$

Остатки

1111
0001
0010
0100
1000

Это объясняется тем, что многочлен $x^4 + x^3 + x^2 + x + 1$ входит в разложение не только двучлена $x^{15} + 1$, но и двучлена $x^5 + 1$.

Из приведенного примера следует, что в качестве образующего следует выбирать такой неприводимый многочлен $g(x)$ (или произведение таких многочленов), который, являясь делителем двучлена x^n+1, не входит в разложение ни одного двучлена типа $x^\lambda+1$, степень которого λ меньше n. В этом случае говорят, что многочлен $g(x)$ принадлежит показателю степени n.

В табл. 7.19 приведены основные характеристики некоторых кодов, способных исправлять одиночные ошибки или обнаруживать все одиночные и двойные ошибки.

Таблица 7.19

Основные характеристики кодов для исправления одиночных ошибок

Показатель неприводимого многочлена	Образующий многочлен	Число остатков	Длина кода
2	$x^2 + x + 1$	3	3
3	$x^3 + x + 1$	7	7
3	$x^3 + x^2 + 1$	7	7
4	$x^4 + x^3 + 1$	15	15
4	$x^4 + x + 1$	15	15
5	$x^5 + x^2 + 1$	31	31
5	$x^5 + x^3 + 1$	31	31

Это циклические коды Хэмминга для исправления одной ошибки, в которых в отличие от групповых кодов Хэмминга все проверочные разряды размещаются в конце кодовой комбинации.

Эти коды могут быть использованы для обнаружения любых двойных ошибок. Многочлен, соответствующий вектору двойной ошибки, имеет вид $\xi(x) = x^i - x^j$, или $\xi(x) = x^i(x^{j-i} + 1)$ при $j>i$. Так как $j-i<n$, а $g(x)$ не кратен x и принадлежит показателю степени n, то $\xi(x)$ не делится на $g(x)$, что и позволяет обнаружить двойные ошибки.

7) обнаружение ошибок кратности три и ниже

Образующие многочлены кодов, способных обнаруживать одиночные, двойные и тройные ошибки, можно определить, базируясь на следующем указании Хэмминга. Если известен образующий многочлен $p(x^m)$ кода длины n, позволяющего обнаруживать ошибки некоторой кратности z, то образующий многочлен $g(x)$ кода, способного обнаруживать ошибки следующей кратности $(z+1)$, может быть получен умножением многочлена $p(x^m)$ на многочлен $x+1$, что соответствует введению дополнительной проверки на четность. При этом число символов в комбинациях кода за счет добавления еще одного проверочного символа увеличивается до $n+1$.

В табл. 7.20 приведены основные характеристики некоторых кодов, способных обнаруживать ошибки кратности три и менее.

Таблица 7.20

Основные характеристики кодов, способных обнаруживать ошибки кратности три

Показатель неприводимого многочлена	Образующий многочлен	Число информационных символов	Длина кода
3	$(x+1)(x^3 + x + 1)$	4	8
4	$(x+1)(x^4+ x + 1)$	11	16
5	$(x+1)(x^5+ x + 1)$	26	32

8) обнаружение и исправление независимых ошибок произвольной кратности, БЧХ-коды

Важнейшим классом кодов, используемых в каналах, где ошибки в последовательностях символов возникают независимо, являются коды *Боуза-Чоудхури-Хоквингема (БЧХ)*. Учеными доказано, что для любых целых положительных чисел m и $s < n/2$ существует двоичный код этого класса длины $n = 2^m - 1$ с числом проверочных символов не более ms, который способен обнаруживать ошибки кратности $2s$ или исправлять ошибки кратности s. Для понимания теоретических аспектов этих кодов необходимо ознакомиться с рядом известных понятий высшей алгебры.

9) обнаружение и исправление пачек ошибок

Для произвольного линейного блокового (n,k)-кода, рассчитанного на исправление пакетов ошибок длины b или менее, основным соотношением, устанавливающим связь корректирующей способности с числом избыточных символов, является граница Рейджера:

$$n - k \geq 2b$$

При исправлении линейным кодом пакетов длины b или менее с одновременным обнаружением пакетов длины $l \geq b$ или менее требуется по крайней мере $b + l$ проверочных символов.

Из циклических кодов, предназначенных для исправления пакетов ошибок, широко известны коды Бартона, Файра и Рида-Соломона.

Первые две разновидности кодов служат для исправления одного пакета ошибок в блоке. Коды Рида-Соломона способны исправлять несколько пачек ошибок.

Особенности декодирования циклических кодов, исправляющих пакеты ошибок, рассмотрены далее на примере кодов Файра.

10) методы образования циклического кода. Границы

Существует несколько различных способов кодирования. Принципиально наиболее просто комбинации циклического кода можно получить, умножая многочлены $a(x)$, соответствующие комбинациям безизбыточного кода (информационным символам), на образующий многочлен кода $g(x)$. Такой способ легко реализуется. Однако он имеет тот существенный недостаток, что получающиеся в результате умножения комбинации кода не содержат информационные символы в явном виде. После исправления ошибок такие комбинации для выделения информационных символов приходится делить на образующий многочлен кода.

Применительно к циклическим кодам принято (хотя это и не обязательно) отводить под информационные k символов, соответствующих старшим степеням многочлена кода, а под проверочные $n–k$ символов низших разрядов.

Чтобы получить такой систематический код, применяется следующая процедура кодирования.

Многочлен $a(x)$, соответствующий k-разрядной комбинации безизбыточного кода, умножаем на x^m, где $m = n - k$. Степень каждого одночлена, входящего в $a(x)$, увеличивается, что по отношению к комбинации кода означает необходимость приписать со стороны младших разрядов m нулей. Произведение $a(x)x^m$ делим на образующий многочлен $g(x)$. В общем случае при этом получаем некоторое частное $q(x)$ той же степени, что и $a(x)$ и остаток $r(x)$. Последний прибавляем к $a(x)x^m$. В результате получаем многочлен

$$f(x) = a(x)x^m + r(x).$$

Поскольку степень $g(x)$ выбираем равной m, степень остатка $r(x)$ не превышает $m-1$. В комбинации, соответствующей многочлену

$a(x)x^m$, m младших разрядов нулевые, и, следовательно, указанная операция сложения равносильна приписыванию $r(x)$ к $a(x)$ со стороны младших разрядов.

Покажем, что $f(x)$ делится на $g(x)$ без остатка, т. е. является многочленом, соответствующим комбинации кода. Действительно, запишем многочлен $a(x)x^m$ в виде

$$a(x)x^m = q(x)g(x) + r(x).$$

Так как операции сложения и вычитания по модулю два идентичны, $r(x)$ можно перенести влево, тогда

$$a(x)x^m + r(x) = f(x) = q(x)g(x).$$

что и требовалось доказать.

Таким образом, циклический код можно строить, приписывая к каждой комбинации безизбыточного кода остаток от деления соответствующего этой комбинации многочлена на образующий многочлен кода. Для кодов, число информационных символов в которых больше числа проверочных, рассмотренный способ реализуется наиболее просто.

Следует указать еще на один способ кодирования. Так как циклический код является разновидностью группового кода, то его проверочные символы должны выражаться через суммы по модулю два определенных информационных символов.

Равенства для определения проверочных символов могут быть получены путем решения рекуррентных соотношений:

$$a_{i+k} = \sum_{j=0}^{k-1} h_j a_{i+j}, \tag{7.19}$$

где h – двоичные коэффициенты так называемого генераторного многочлена $h(x)$, определяемого так

$$h(x) = (x^n + 1)/g(x) = h_0 + h_1 x + ... + h_k x^k$$

Соотношение (7.9) позволяет по заданной последовательности информационных сигналов $a_0, a_1, ..., a_{k-1}$ вычислить $n-k$ проверочных символов $a_k, a_{k+1}, ..., a_{n-1}$. Проверочные символы, как и ранее, размещаются на местах младших разрядов. При одних и тех же информационных символах комбинации кода, получающиеся таким путем, полностью совпадают с комбинациями, получающимися при использовании предыдущего способа кодирования. Применение данного способа целесообразно для кодов с числом проверочных символов, превышающим число информационных, например для кодов Боуза-Чоудхури-Хоквингема. (БЧХ)

11) матричная запись циклического кода

Полная образующая матрица циклического кода $M_{n,k}$ составляется из двух матриц: единичной I_k (соответствующей k информационным разрядам) и дополнительной $C_{k,n-k}$ (соответствующей проверочным разрядам):

$$M_{n,k} = \| I_k C_{k,n-k} \|$$

Построение матрицы I_k трудностей не представляет. Если образование циклического кода производится на основе решения рекуррентных соотношений, то его дополнительную матрицу можно определить, воспользовавшись правилами, указанными ранее. Однако обычно строки дополнительной матрицы циклического кода $C_{k,n-k}$ определяются путем вычисления многочленов $r(x)$. Для каждой строки матрицы I_k соответствующий $r(x)$ находят делением информационного многочлена $a(x)x^m$ этой строки на образующий многочлен кода $g(x)$.

Дополнительную матрицу можно определить и не строя I_k. Для этого достаточно делить на $g(x)$ комбинацию в виде единицы с рядом нулей и получающиеся остатки записывать в качестве строк дополнительной матрицы. При этом если степень какого-либо $r(x)$ оказывается меньше $n-k-1$, то следующие за этим остатком строки матрицы получают путем циклического сдвига предыдущей строки влево до тех пор, пока степень $r(x)$ не станет равной $n-k-1$. Деление производится до получения k строк дополнительной матрицы.

Пример. Запишем образующую матрицу для циклического кода (15,11) с порождающим многочленом $g(x) = x^4 + x^3 + 1$.

Воспользовавшись результатами ранее проведенного деления, получим соответствующую матрицу $M_{15,11}$. Существует другой способ построения образующей матрицы, базирующийся на основной особенности циклического (n,k)-кода. Он проще описанного, но получающаяся матрица менее удобна. Матричная запись кодов достаточно широко распространена.

$$M_{15,11} = \begin{bmatrix} 0&0&0&0&0&0&0&0&0&0&1&&1&0&0&1 \\ 0&0&0&0&0&0&0&0&0&1&0&&1&0&1&1 \\ 0&0&0&0&0&0&0&0&1&0&0&&1&1&1&1 \\ 0&0&0&0&0&0&0&1&0&0&0&&0&1&1&1 \\ 0&0&0&0&0&0&1&0&0&0&0&&1&1&1&0 \\ 0&0&0&0&0&1&0&0&0&0&0&&0&1&0&1 \\ 0&0&0&0&1&0&0&0&0&0&0&&1&0&1&0 \\ 0&0&0&1&0&0&0&0&0&0&0&&1&1&0&1 \\ 0&0&1&0&0&0&0&0&0&0&0&&0&0&1&1 \\ 0&1&0&0&0&0&0&0&0&0&0&&0&1&1&0 \\ 1&0&0&0&0&0&0&0&0&0&0&&1&1&0&0 \end{bmatrix}$$

12) укороченные циклические коды

Корректирующие возможности циклических кодов определяются степенью m образующего многочлена. В то время как необходимое число информационных символов может быть любым целым числом, возможности в выборе разрядности кода весьма ограничены.

Если, например, необходимо исправить единичные ошибки при k=5, то нельзя взять образующий многочлен третьей степени, поскольку он даст только семь остатков, а общее число разрядов получится равным 8.

Следовательно, необходимо брать многочлен четвертой степени и тогда n=15. Такой код рассчитан на 11 информационных разрядов.

Однако можно построить код минимальной разрядности, заменив в (n,k)-коде j первых информационных символов нулями и исключив их из кодовых комбинаций. Код уже не будет циклическим, поскольку циклический сдвиг одной разрешенной кодовой комбинации не всегда приводит к другой разрешенной комбинации того же кода. Получаемый таким путем линейный (n–j, k–j)-код называют *укороченным циклическим кодом.* Минимальное расстояние этого кода не менее, чем минимальное кодовое расстояние (n,k)-кода, из которого он получен. Матрица укороченного кода получается из образующей матрицы (n,k)-кода исключением j строк и столбцов, соответствующих старшим разрядам. Например, образующая матрица кода (9,5), полученная из матрицы кода (15,11), имеет вид

$$M_{9.5} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & & 1 & 1 & 1 & 0 \end{bmatrix}$$

13) передача информации по каналу связи, пропускная способность канала

Введение понятий энтропии, количества информации, скорости выдачи информации источником, избыточности позволяют характеризовать свойства информационных систем. Однако для сравнения информационных систем только такого описания недостаточно. Обычно нас интересует не только передача данного количества информации, но передача его в возможно более короткий срок; не только хранение определенного количества информации, но хранение с помощью минимальной по объему аппаратуры и т.п.

Пусть количество информации, которое передается по каналу связи за время *T*, равно $I_T = H_T(X) - H_T(X/Y)$. Если передача сообщения длится *T* единиц времени, то скорость передачи информации составит

$$R = \frac{I_T}{T} = \frac{1}{T}[H_T(X) - H_T(X/Y)] = H(X) - H(X/Y).$$

Это количество информации, приходящееся в среднем на одно сообщение. Если в секунду передается *n* сообщений, то скорость передачи будет составлять $R = n[H(X) - H(X/Y)]$.

Пропускная способность канала есть максимально достижимая для данного канала скорость передачи информации:

$$C = \max R = n[H(X) - H(X/Y)]_{\max}. \qquad (7.20)$$

Или максимальное количество информации, передаваемое за единицу времени:

$$C = nI(X,Y)_{\max}$$

Скорость передачи может быть технической или информационной.

Под технической скоростью V_T, называемой также скоростью манипуляции, подразумевается число элементарных сигналов (символов), передаваемых в единицу времени $V_T = 1/\tau$ *бод*.

Информационная скорость или скорость передачи информации, определяется средним количеством информации, которое передается в единицу времени и измеряется (*бит*/*сек*). $R=nH$.

Для равновероятных сообщений составленных из равновероятных взаимно независимых символов $R = \frac{1}{\tau}\log m$.

В случае если символы не равновероятны $R = -\frac{1}{\tau}\sum_i p_i \log p_i$.

В случае если символы имеют разную длительность

$$R = \frac{\sum_i p_i \log p_i}{\sum_i \tau_i p_i}. \qquad (7.21)$$

Выражение для пропускной способности отличается тем, что характеризуется максимальной энтропией

$$C_{\max} = \frac{H_{\max}}{\tau} \text{ бит/сек}$$

Для двоичного кода $C_{\max} = \frac{\log 2}{\tau} = \frac{1}{\tau}$ *бит*/*сек*

Пропускная способность является важнейшей характеристикой каналов связи. Возникает вопрос: какова должна быть пропускная способность канала, чтобы информация от источника *X* к приемнику *Y* поступала без задержек? Ответ на этот вопрос дает 1-ая теорема Шеннона.

14) 1-ая теорема Шеннона

Если имеется источник информации с энтропией $H(x)$ и канал связи с пропускной способностью C, то если $C > H(X)$, то всегда можно закодировать достаточно длинное сообщение таким образом, что оно будет передано без задержек. Если же, напротив, $C < H(X)$, то передача информации без задержек невозможна.

В любом реальном канале всегда присутствуют помехи. Однако, если их уровень настолько мал, что вероятность искажения практически равна нулю, можно условно считать, что все сигналы передаются неискаженными. В этом случае среднее количество информации, переносимое одним символом равно $I(X,Y)=I(X,X)=H(X)$. Максимальное значение $H_{max}=logm$. Следовательно, пропускная способность дискретного канала без помех за единицу времени равна

$$C = n\log m.$$

Реальные каналы характеризуются тем, что на каналы всегда воздействуют помехи. Пропускная способность дискретного канала с помехами вычисляется по формуле

$$C = n\left[h(Y) - H(Y/X)\right]_{\max}.$$

Где средняя, условная энтропия со стороны приемника сигналов

$$H(Y/X) = -\sum_i \sum_j p(x_i, y_j)\log p(y_j / x_i) =$$
$$= -\sum_i p(x_i)\sum_j p(y_j / x_i)\log p(y_j / x_i).$$

А энтропия принимаемых сигналов определяется из условия максимального значения $H(y)= log\ m$.

Пример. Пусть требуется определить пропускную способность бинарного канала связи. При этом с вероятностью p каждый из двоичных сигналов может перейти в противоположный сигнал.

На рис. 7.3. представлена модель передачи бинарных сигналов,

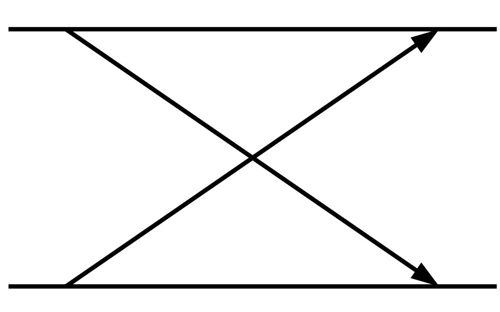


Рис. 7.3. Симметричный канал передачи сигналов
в условиях помех

где $1-p$ вероятность неискаженной передачи сигналов;

p – вероятность искажения сигналов, x_1 и x_2 передаваемые сигналы типа «0» или «1», y_1 и y_2, принимаемые сигналы.

Матрица для нахождения условной вероятности

$$P(y/x) = \begin{array}{c|cc|} & y_1 & y_2 \\ x_1 & P(y_1/x_1) & P(y_2/x_1) \\ x_2 & P(y_1/x_2) & P(y_2/x_2) \end{array} = \begin{vmatrix} 1-p & p \\ p & 1-p \end{vmatrix}$$

Найдем полную условную энтропию системы y относительно x:

$$H(y/x) = -\sum_{i=1} p(x_i) \sum_{j=1} p(y_j/x_i) \log p(y_j/x_i) =$$
$$= -p(x_1)\left[(1-p_n)\log(1-p_n) + p_n \log p_n\right] - p(x_2)$$
$$\left[p_n \log p_n + (1-p_n)\log(1-p_n)\right] =$$
$$= -(p(x_1) + p(x_2))\left[p_n \log p_n + (1-p_n)\log(1-p_n)\right]$$

Откуда $H(y/x) = -p_n \log p_n - (1-p_n)\log(1-p_n)$. $H(y)$ находим из условия максимального значения $H(y) = \log 2 = 1$.

Формула для нахождения пропускной способности бинарного канала связи будет иметь вид

$$C = n\left[1 + p_n \log p_n + (1-p_n)\log(1-p_n)\right]. \tag{7.22}$$

График функции имеет следующий вид (рис. 7.4):

Наибольшее значение эта функция принимает при $p = 0$ (то есть при отсутствии помех) и при $p = 1$ (то есть при негативной передаче). При $p = 1/2$ пропускная способность минимальна.

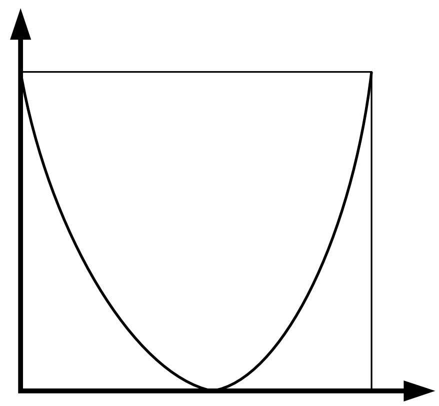


Рис. 7.4. График функции $C = f(p)$

В качестве примера рассмотрим более общий случай передачи по дискретному каналу. Найдем пропускную способность m-ичного канала связи. На рис. 7.5 представлена модель передачи m-ичных сигналов, где $x_1, x_2, \ldots, x_m$ источники информации, $y_1, y_2, \ldots, y_m$ приемники информации.

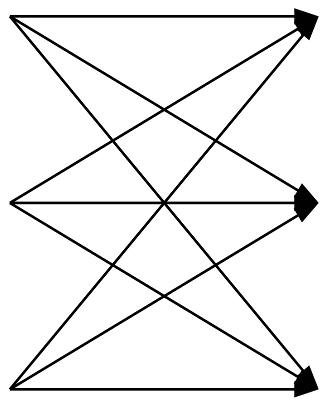


Рис. 7.5. m-ичный канал передачи информации

Вероятность ошибки – p. Вероятность безошибочной передачи сигналов равняется $1 - p$, а в случае ошибки переданный сигнал может с одинаковой вероятностью (равной $\frac{p}{m-1}$) быть воспринятым, как любой из $m - 1$ отличных от него сигналов.

Матрица условных вероятностей имеет вид

$$p(y/x) = \begin{matrix} 1-p & \frac{p}{m-1} & \vdots & \frac{p}{m-1} \\ \frac{p}{m-1} & 1-p & \vdots & \frac{p}{m-1} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{p}{m-1} & \frac{p}{m-1} & \vdots & 1-p \end{matrix}.$$

Полная условная энтропия системы Y относительно X

$$H(y/x) = -1(1-p)\log(1-p) - (m-1)\frac{p}{m-1}\log\frac{p}{m-1}$$

$$H(y) = \log m$$

Формула для нахождения пропускной способности m-ичного канала связи будет иметь вид:

$$C = n\left[\log m + (1-p)\log(1-p) + (m-1)\frac{p}{m-1}\log\frac{p}{m-1}\right] \qquad (7.23)$$

График функции $C(p)$ пропускной способности канала связи при m=4 представлен на рис.7.6.

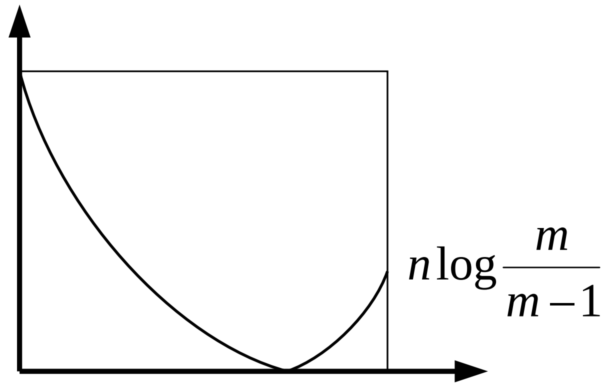


Рис. 7.6. График функции $C(p)$

Эта функция максимальна при $p=0$, при вероятности $p = \frac{m-1}{m} = 0.75$ $C = 0$. При $p=1$,

$$C = n\log\frac{m}{m-1}.$$

Для дискретных каналов с помехами Шеннон сформулировал вторую теорему.

15) 2-я Теорема Шеннона

Пусть имеется источник информации X, энтропия которого в единицу времени равна $H(X)$, и канал с пропускной способностью C. Если $H(X)>C$, то при любом кодировании передача сообщений без задержек и искажений невозможна. Если же $H(X)<C$, то любое достаточно длинное сообщение можно всегда закодировать так, что оно будет передано без задержек и искажений с вероятностью сколь угодно близкой к единице.

nlog

Контрольные вопросы

1. Каково значение и в чем сущность теоремы Шеннона о передаче сигналов по каналам связи при наличии помех, приводящих к искажениям?
2. В чем сущность проблемы эффективного символьного кодирования?
3. Раскрыть содержание первой и второй теорем Шеннона, относящихся к решению задачи передачи сигналов по каналам связи при наличии помех, приводящих к искажениям.
4. Какие задачи должна решать теория кодирования информации?
5. Дать определения понятиям кода, кодирования, декодирования, кодера, декодера, длины кодовой цепочки, оптимального кода?
6. Чем объясняется выгодность кода при передаче и хранении информации?
7. Сформулировать основную теорему о кодировании при отсутствии помех и каковы ее следствия?
8. Каково содержание марковских источников сообщений?
9. Дать определение дискретного стационарного источника, называемого Марковским источником порядка m.
10. В чем сущность первой теоремы Шеннона для марковских источников порядка m=1?
11. Что такое эргодические источники?
12. Какой марковский источник называется эргодическим?
13. Что такое информационная дивергенция, ее определения и свойства?
14. Чем определяется эффективное кодирование?
15. В чем сущность метода Шеннона – Фано, используемого для построение эффективных кодов?
16. Дать представление об алфавитном неравномерном двоичном кодировании сигналов.
17. В чем сущность префиксного кода Шеннона – Фано?
18. В чем сущность префиксного кода Хаффмана?
19. Раскрыть сущность неравенства Крафта при решении задач оптимального кодирования.
20. Дать определение и раскрыть основные особенности линейного кода.
21. Что является основой математического описания линейных кодов?

22. Дать определения основных алгебраических систем, использующихся в теории корректирующих кодов.

23. Дать определение группы, кольца, поля F, поля Галуа GF(P),

24. Что такое линейный код, корректирующие коды?

25. Как определяется расстояние Хэмминга между двумя кодовыми n-последовательностями?

26. Каковы особенности построения двоичного группового кода?

27. Как оценивается качество корректирующего кода?

28. Какой код называют оптимальным?

29. Что такое коды Хэмминга и каковы их основные свойства?

30. Как формируются коды Хэмминга, обеспечивающие исправление ошибок?

31. Как осуществляется мажоритарное декодирование групповых кодов?

32. Раскрыть содержание, назначение и особенности циклических кодов, БЧХ – кодов, сверточных кодов.

33. Как обеспечивается обнаружение одиночных ошибок с помощью кодов?

34. Раскрыть особенности кода Хэмминга, обеспечивающего исправление одиночных или обнаружение двойных ошибок.

35. Как осуществляется обнаружение ошибок кратности три ниже?

36. Как осуществляется обнаружение и исправление независимых ошибок произвольной кратности, БЧХ-кодом?

37. Каковы особенности обнаружения и исправления пачек ошибок?

38. В чем сущность метода образования циклического кода?

39. Как осуществляется построение матричного циклического кода и какова его запись?

40. Что называется укороченным циклическим кодом и как он строится?

41. Как определяется скорость передачи информации по каналу связи?

42. Представить выражение для оценки пропускной способности канала связи и раскрыть его содержание.

43. В чем сущность 1-й теоремы Шеннона для источника информации с заданной энтропией H(x) и каналом связи с заданной пропускной способностью C?

ГЛАВА 8. ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ ТЕОРИИ СВЯЗИ

8.1. Математическая модель канала связи

Канал связи предназначен для транспортировки сообщений. Математическая модель канала связи описывается некоторой совокупностью X_1 элементов x_1 ($x_1 = \{x_{11}, x_{12}, ..., x_{1j}\}$), называемых сигналами на входе канала, совокупностью X_2 элементов x_2 ($x_2 = \{x_{21}, x_{22}, ..., x_{2k}\}$), называемых выходными сигналами, и условными распределениями вероятностей $p_2 = p_2(a_2 \mid x_1)$ в пространстве x_2 выходных сигналов x_2. Если посланный сигнал (сигнал на входе) есть x_1, то с вероятностью $P_2 = P_2(A_2 \mid x_1)$ на выходе канала будет принят сигнал x_2 из некоторого множества $A_2 \subset X_2$ (распределения задают вероятности того или иного искажения посланного сигнала x_1). Совокупность всех возможных сообщений обозначим символом x_0. Предполагается, что каждое из сообщений $x_0 \in X_0$ может поступать с определенной вероятностью. То есть, в пространстве X_0 имеется определенное распределение вероятностей $P_0 = P_0(A_0)$.

Сообщения x_0 не могут быть переданы по каналу связи непосредственно, для их пересылки используются сигналы $x_1 \in X_1$. Кодирование сообщений x_0 в сигналы x_1 описывается при помощи условного распределения вероятностей P_1=$P_1(A_1|x_0)$. Если поступает сообщение x_0, то с вероятностью P_1=$P_1(A_1|x_0)$ будет послан один из сигналов x_1, входящих в множество $A_1 \subset X_1$ (условные распределения $P_1(A_1|x_0)$ учитывают возможные искажения при кодировании сообщений). Аналогичным образом описывается декодирование принимаемых сигналов x_2 в сообщения x_3. Оно задается условным распределением вероятностей P_3=$P_3(A_3|x_2)$ на пространстве X_3 сообщений x_3, принимаемых на выходе канала связи.

На вход канала связи поступает случайное сообщение ξ_0 с заданным распределением вероятностей P_0=$P_0(A_0)$. При его поступлении передается сигнал ξ_1, распределение вероятностей которого задается правилом кодирования P_1=$P_1(A_1|x_0)$:

$$P\{\xi_2 \in A_2 | \xi_0, \xi_1\} = P_2(A_2 | \xi_1)$$

Принятый сигнал ξ_2 декодируется, в результате чего получается сообщение ξ_3:

$$P\{\xi_3 \in A_3 | \xi_0, \xi_1, \xi_2\} = P_3(A_3 | \xi_2)$$

Последовательность $\xi_0 \to \xi_1 \to \xi_2 \to \xi_3$ является марковской. При любых правилах кодирования и декодирования описанного типа имеет место неравенство:

$$I(\xi_0, \xi_3) \le I(\xi_1, \xi_2),$$

где $I(\xi_0, \xi_3)$ – количество информации о ξ_0 в принятом сообщении ξ_3;

$I(\xi_1, \xi_2)$ – количество информации о ξ_1 в принятом сигнале ξ_2.

Предположим, что распределение вероятности входного сигнала ξ_1 не может быть произвольным и ограничено определенными требованиями, например, оно должно принадлежать классу W. Величина $C = \sup I(\xi_1, \xi_2)$, где верхняя грань берется по всем возможным распределениям $P_1 \in W$, называется емкостью канала и характеризует максимальное количество информации, которое может быть передано по данному каналу связи (теорема Шеннона).

Предположим далее, что передача сообщений $\xi_0 \to \xi_3$ должна удовлетворять определенным требованиям точности, например, совместное распределение вероятностей $P_{\xi_0 \xi_1}$ передаваемого и принимаемого сообщений ξ_0 и ξ_3 должно принадлежать некоторому классу V. Величина $H = \inf I(\xi_0 \ \xi_3)$, где нижняя грань берется по всем возможным распределениям $P_{\xi_0 \xi_3} \in \mathrm{V}$, характеризует минимальное количество информации, которое должно заключать в себе принимаемое сообщение ξ_3 о ξ_0, чтобы было выполнено условие точности передачи.

Величина H называется энтропией источника сообщений.

Если возможна передача $\xi_0 \to \xi_1 \to \xi_2 \to \xi_3$ с соблюдением требований V и W, то есть существуют соответствующие способы кодирования и декодирования (существуют условные распределения P_1, P_2 и P_3), то $H \le C$.

Для выполнения этого неравенства передача является возможной, т.е. возможна передача последовательно поступающих сообщений $\xi_0^{(1)}, \xi_0^{(2)}, \ldots, \xi_0^{(n)}$.

Предположим, что совокупность X_0 всех возможных сообщений x_0 является дискретной (имеется не более чем счетное число различных сообщений x_0, поступающих с соответствующими вероятностями $P_0(x_0)$, $x_0 \in X_0$) и условие точности передачи v состоит в том, что принимаемое сообщение ξ_3 должно просто совпадать с переданным сообщением $\xi_3 = \xi_0$ с вероятностью 1. Тогда

$$H = -\sum_{x_0} P_0(x_0) \log P_0(x_0) = -M \log P_0(\xi_0).$$

Предположим далее, что имеется лишь конечное число N различных входных сигналов x_1 и нет никаких ограничений на вероятности $P\{\xi_1 = x_1\}$, $x_1 \in X_1$. Кроме того, предположим, что передаваемые сигналы принимаются без искажений, то есть с вероятностью 1 $\xi_2 = \xi_1$. Тогда емкость канала выражается формулой $C = \log_2 N$, т.е. передаваемое количество информации $I(\xi_1, \xi_2)$ будет максимальным в том случае, когда сигналы $x_1 \in X_1$ равновероятны.

Если сообщения $\xi_0^{(1)}, \xi_0^{(2)}, \ldots, \xi_0^{(n)}$ поступают независимо друг от друга, то количество информации, которое несет группа сообщений $\xi_{0m} = (\xi_0^{(1)}, \xi_0^{(2)}, \ldots, \xi_0^{(n)})$ есть

$$H_n = -\sum_{x_{0n}} P_0(x_{0n}) \log P(x_{0n}) = -M \log P(\xi_{0n}) = nH,$$

где $x_{0n} = (x_0^{(1)}, \ldots, x_0^{(n)})$ – группа сообщений, поступающая на кодирование с вероятностью $P(x_{0n}) = P_0(x_0^{(1)}) P_0(x_0^{(2)}) \ldots P_0(x_0^{(n)})$.

Пусть $H<C$, положим также $\delta = (1/2)(C - H)$. Согласно закону больших чисел, примененному к последовательности независимых и одинаково распределенных случайных величин $\log P_0(\xi_0^{(k)}) (k = 1,2,\ldots)$ с математическим ожиданием $M \log P_0(\xi_0^{(k)}) = -H$ для любого $\varepsilon > 0$ найдется такое $n(\varepsilon)$, что при всех $n \geq n(\varepsilon)$

$$P\{-H - \delta \leq (1/n) \log P(\xi_{0n}) \leq H + \delta\} \geq 1 - \varepsilon,$$

где $\log P(\xi_{0n}) = \sum_{k=1}^{n} \log P_0(\xi_0^{(k)})$.

Полученное неравенство говорит о том, что все группы сообщений x_{0n} можно разбить на два класса. К первому классу X^1_{0n} относятся высоковероятные сообщения x_{0n}, для которых $P(x_{0n}) \geq 2^{-n(H+\delta)}$ и количество которых M_n не больше чем $2^{n(H+\delta)}$:

$$M_n \leq 2^{n(H+\delta)}$$

Ко второму классу X^2_{0n} относятся все остальные маловероятные сообщения x_{0n}:

$$P\left\{\xi_{0n} \in X^2_{0n}\right\} \leq \varepsilon .$$

Каждую группу высоковероятных сообщений x_{0n} можно в принципе передать, закодировав ее соответствующей комбинацией сигналов $x_{1n} = (x_1^{(1)},...,x_1^{(n)})$. Число всевозможных комбинаций такого вида есть N_n=2^{nC}, и видно, что M_n<N_n. Имеется N_n различных сигналов x_{1n}, с помощью которых можно закодировать и передать безошибочно все M_n высоковероятных сообщений $x_{0n} \in X^1_{0n}$ Если в дополнение к этому при поступлении любого маловероятного сообщения $x_{0n} \in X^2_{0n}$ передавать некоторый один и тот же сигнал x^0_{1n} (отличный от сигналов, при помощи которых передаются высоковероятные сообщения $x_{0n} \in X^1_{0n}$), то с вероятностью, не меньшей чем $1-\varepsilon$, на выходе канала связи будет приниматься последовательность $\xi_0^{(1)},\xi_0^{(2)},...,\xi_0^{(n)}$:

$$P\{\xi_3^{(1)} = \xi_0^{(1)}, \xi_3^{(2)} = \xi_0^{(2)},..., \xi_3^{(n)} = \xi_0^{(n)}\} \geq 1-\varepsilon .$$

При выполнении неравенства $H < C$ оказывается возможной передача достаточно длинных сообщений $\xi_0^{(1)},\xi_0^{(2)},...,\xi_0^{(n)}$ с той оговоркой, что с вероятностью ε (ε – наперед заданное сколь угодно малое положительное число) может быть допущена ошибка. Имеется целое семейство каналов связи и источников сообщений, зависящих от параметра *n*.

Количество информации $I(\xi_0, \xi_3)$ для абстрактных случайных величин ξ_0 и ξ_3 со значениями в пространствах X_0 и X_3 может быть записано в виде:

$$I(\xi_0, \xi_3) = Mi(\xi_0, \xi_3),$$

где $i(x_0, x_3) = \dfrac{P_{\xi_0\xi_3}(dx_0 dx_3)}{P_0(dx_0)P_3(x_3)}$ – информационная плотность. Последовательность пар (ξ_{0n}, ξ_{3n}) называется информационно устойчивой, если при $n \to \infty$ $I(\xi_0, \xi_3) \to \infty$ и $\dfrac{i(\xi_{0n}, \xi_{3n})}{I(\xi_{0n}, \xi_{3n})} \to 1$ (по вероятности).

Рассмотренная выше последовательность (ξ_{0n}, ξ_{3n}), $\xi_{3n}=\xi_{0n}$ поступающих сообщений $\xi_{0n} = \left(\xi_0^{(1)}, \xi_0^{(2)}, ..., \xi_0^{(n)}\right)$ обладает свойством информационной устойчивости, что в конечном счете и определило возможность передачи сообщений ξ_{0n} с точностью до ε. Этот факт допускает широкое обобщение. Например, если C_n – пропускная способность канала $\xi_{1n} \to \xi_{2n}$, H_n – минимальное количество информации, необходимое для соблюдения требуемой точности передачи $\xi_{0n} \to \xi_{3n}$, причем

$$\lim \frac{H_n}{C_n} < 1$$

(при $n \to \infty$), и существуют информационно устойчивые последовательности пар (ξ_{0n}, ξ_{3n}) и (ξ_{1n}, ξ_{2n}), для которых одновременно

$$\frac{1}{H_n} I(\xi_{0n}, \xi_{3n}) \to 1 \text{ и } \frac{1}{C} I(\xi_{1n}, \xi_{2n}) \to 1,$$

то при весьма широких предположениях для любого наперед заданного $\varepsilon > 0$ существует такое $n(\varepsilon)$, что по всем каналам связи с параметром $n \geq n(\varepsilon)$ возможна передача с точностью до ε.

1) канал связи с изменяющимися состояниями

Как было указано выше, канал характеризуется условными распределениями $З_2$, задающими вероятности тех или иных искажений посылаемого сигнала x_1. Несколько изменим схему канала связи, считая, что имеется некоторое множество Z возможных состояний z канала связи, причем если канал находится в некотором состоянии z и на входе возникает сигнал x_1, то независимо от других предшествующих обстоятельств канал переходит в другое состояние z_1. Этот переход подвержен случайностям и описывается условными распределениями $P(C|x_1,z)$ ($P(C|x_1,z)$ – вероятность того, что новое состояние z_1 будет входить в множество $C \subset Z$). При этом уже считается, что выходной сигнал x_2 однозначно определяется состоянием канала z_1, т.е. существует некоторая функция $\varphi = \varphi(z)$ на пространстве z возможных состояний канала такая, что $x_2 = \varphi(z_1)$. Эта более общая схема позволяет учитывать те изменения, которые в принципе могут возникать в канале по мере его работы.

Рассмотрим стационарный режим работы канала связи. Предположим, что последовательно передаваемые сигналы $\ldots,\zeta_1(-1),\zeta_1(0),\zeta_1(1),\ldots$, соответствующие состояниям канала $\ldots,\xi(-1),\xi(0),\xi(1),\ldots$, и определяемые ими сигналы $\ldots,\xi_2(-1)$, $\xi_2(0)$, $\xi_2(1),\ldots$, на выходе образуют стационарные и стационарно связанные случайные последовательности. Величина $C=\sup I(\xi_1,\xi_2)$, где $I(\xi_1,\xi_2)$, означает скорость передачи информации о стационарной последовательности $\{\xi_1(n)\}$ последовательностью $\{\xi_2(n)\}$ и верхняя грань берется по всем допустимым распределениям вероятностей входной последовательности $\{\xi_1(n)\}$, называется пропускной способностью канала связи.

Предположим, что поступающие на вход канала связи сообщения $\{\xi_0(n)\}$, $n=\ldots,-1,0,1,\ldots$, образуют случайную последовательность. Будем считать правило кодирования заданным, если при всех k, m и $k_1,\ldots,k_m \geq k$ определены условные вероятности

$$P\{\xi_1(k_1)\in B_1,\ldots,\xi_1(k_m)\in B_m \mid \xi_0(-\infty,k)\}$$

Того, что при поступлении последовательности сообщений

$$\xi_0(-\infty,k)=\ldots,\xi_0(k-1),\xi_0(k)$$

на соответствующих местах будут переданы сигналы $\xi_1(k_1),\ldots,\xi_1(k_m)$, входящие в указанные множества $B_1,\ldots,B_\mathrm{m}$. Эти вероятности считаются стационарными в том смысле, что они не меняются при одновременной замене индексов k и $k_1,\ldots,k_m$ на $k+l$ и $k_1+l,\ldots,k_m+l$ при любом целом l. Аналогичными вероятностями

$p\{\xi_3(k_1)\in D_1,\ldots,\xi_3(k_m)\in D_m|\xi_2(-\infty,k)\}$ задается правило декодирования.

Определим величину H формулой $H=\inf I(\xi_0, \xi_3)$, где $I(\xi_0, \xi_3)$ – скорость передачи информации о стационарной последовательности $\{\xi_0(n)\}$ последовательностью $\{\xi_3(n)\}$, $n = ...,-1,0,1,...$ (эти последовательности предполагаются стационарно связанными), и нижняя грань берется по всем допустимым распределениям вероятностей, удовлетворяющим требованиям точности передачи $\{\xi_0(n)\} \to \{\xi_3(n)\}$.

Неравенство $H \le C$ является необходимым условием возможности передачи $\{\xi_0(n)\} \to \{\xi_1(n)\} \to \{\xi_2(n)\} \to \{\xi_3(n)\}$.

Напомним, что каждое сообщение $\xi_0(n)$ представляет собой некоторый элемент x_0 из совокупности X_0. Можно интерпретировать X_0 как некоторый алфавит, состоящий из символов x_0. Предположим, что этот алфавит X_0 является конечным и требование точности передачи состоит в безошибочном воспроизведении передаваемых символов: $P\{\xi_3(k)=\xi_3(k)\}=1$ для любого целого k.

Предположим также, что имеется лишь конечное число входных сигналов x_1 и состояний канала z. Обозначим состояния канала целыми числами $1,2,\dots,N$, и пусть $p(k, x_1,j)$ – соответствующие вероятности перехода из состояния k в состояние j при входном сигнале x_1:

$$p(k, x_1, j) = P\{\zeta(x+1) = j \mid \zeta(n) = k, \xi_1(n+1) = x_1\}.$$

Дополнительно предположим, что любые произведения вида

$$p(k_0, x_1(1), k_1)\, p(k_1, x_1(2), k_2) ... p(k_{n-1}, x_1(n), k_n)$$

являются стохастическими матрицами, задающими эргодические цепи Маркова. Это условие будет выполнено, если, например, каждая из переходных матриц $\{p(k,x_1,j)\}$ имеет положительный коэффициент эргодичности. Тогда при выполнении неравенства $H<C$ и соблюдении условия эргодичности стационарной последовательности $\{\xi_0(n)\}$ сообщений на входе передача возможна с точностью до любого $\varepsilon > 0$, т.е. при соответствующих способах кодирования и декодирования принимаемая последовательность сообщений $\{\xi_3(n)\}$ будет обладать тем свойством, что $p\{\xi_3(k) \neq \xi_0(k)\} < \varepsilon$ для любого целого k.

Пусть $\xi_1 = \{\xi(t),\ t \in T_1\}$ и $\xi_2=\{\xi(t),\ t \in T_2\}$ – два семейства случайных величин, имеющих совместное гауссово распределение вероятностей, и пусть H_1 и H_2 – замкнутые линейные оболочки величин $\xi(t)$, $t \in T_1$, и $\xi(t)$, $t \in T_2$, в гильбертовом пространстве $L^2(\Omega)$. Обозначим буквами P_1 и P_2 операторы проектирования на пространства H_1 и H_2 и положим $P^{(1)} = P_1P_2P_1$, $P^{(2)} = P_2P_1P_2$. Количество информации $I(\xi_1, \xi_2)$ о семействе величин ξ_1, содержащееся в семействе ξ_2, конечно тогда и только тогда, когда один из операторов $P^{(1)}$ или $P^{(2)}$ представляет собой ядерный оператор, т.е. последовательность $\lambda_1, \lambda_2, \ldots$ его собственных значений (все они неотрицательны) удовлетворяет условию $\sum_k \lambda_k < \infty$. При этом

$$I(\xi_1, \xi_2) = -\frac{1}{2}\sum_k \log(1-\lambda_k).$$

В случае, когда ξ_1 и ξ_2 образованы конечным числом гауссовых величин:

$\xi_1=\{\xi(1),\ldots,\xi(m)\}, \xi_2=\{\xi(m+1),\ldots,\xi(m+n)\}$, причем корреляционная матрица B общей совокупности $\xi(1),\ldots,\xi(m+n)$ является невырожденной, количество информации $I(\xi_1, \xi_2)$ может быть выражено следующей формулой:

$$I(\xi_1, \xi_2) = \frac{1}{2}\log\frac{(\det B_1)(\det B_2)}{\det B},$$

где B_1 и B_2 – корреляционные матрицы соответствующих совокупностей ξ_1 и ξ_2.

Гауссовы распределения обладают следующим экстремальным свойством. Для произвольных распределений вероятностей величин

$\xi_1 = \{\xi(1),\ldots, \xi(m)\}$ и $\xi_2 = \{\xi(m+1),\ldots,\xi(m+n)\}$

с соответствующими корреляционными матрицами B_1, B_2 и B количество информации $I(\xi_1, \xi_2)$ удовлетворяет неравенству

$$I(\xi_1, \xi_2) \le \frac{1}{2}\log\frac{(\det B_1)(\det B_2)}{\det B}.$$

Пусть $\xi=(\xi_1,\ldots,\xi_n)$ и $\eta = (\eta_1,\ldots,\eta_n)$ – векторные случайные величины в n-мерном евклидовом пространстве X и $\rho(x,y)$ – некоторая неотрицательная функция, определяющая условие близости величин ξ и η, которое выражается следующим соотношением:

$$M\rho(\xi,\eta) \le \varepsilon .$$

Величину $H=H_\varepsilon$, определенную как H_ε = inf $I(\xi,\eta)$, обычно называют ε-энтропией случайной величины ξ (нижняя грань берется по всем случайным величинам η, удовлетворяющим указанному условию ε-близости случайной величине ξ).

Пусть $\rho(x,y) = \rho(|x-y|)$ и существует производная $\rho'(0)$, $0<\rho'(0)<\infty$. Тогда при $\varepsilon \to 0$ имеет место асимптотическая формула, в которой логарифмы берутся по основанию e:

$$H_\varepsilon = n\log\frac{1}{\varepsilon} + h(\xi) + \log\frac{\Gamma\left(\frac{n}{2}\right)\left[np'(0)\right]^n}{(2\pi)^{n/2}(n-1)!e^n} + O(1),$$

где $\gamma()$ – гамма функция и $h(\xi)$ – дифференциальная энтропия случайной величины ξ:

$$h(\xi) = -\int_X p_\xi(x)\log p_\xi(x)dx,$$

($p_\xi(x)$ – плотность распределения вероятностей, удовлетворяющая весьма широким условиям, которые выполняются, например, если плотность $p(x)$ ограничена и

$h(\xi)>-\infty$). Пусть $\rho(x,y) = \left(\frac{1}{n}\sum_{k=1}^{n}|x_k - y_k|^\alpha\right)^\beta$ ($\alpha,\beta > 0$).

Тогда

$$H_\varepsilon = \frac{n}{\alpha\beta}\log\frac{1}{\varepsilon} + h(\xi) - \log\left\{\left(\frac{\alpha\beta e}{n}\right)^{\frac{n}{\alpha\beta}}\left[\frac{2}{\alpha}\Gamma\left(\frac{1}{\alpha}\right)\right]^n\frac{\Gamma\left(\frac{n}{\alpha\beta}\right)}{\beta\Gamma\left(\frac{n}{\alpha}\right)}\right\} + O(1). \quad (8.1)$$

В частности, при α =2, β =1 имеет место асимптотическая формула

$$H_{\varepsilon} = \frac{n}{2}\log\frac{1}{\varepsilon} + h(\xi) - n\log\sqrt{2\pi e} + O(1)$$

Пусть пара случайных процессов ($\xi_1(t), \xi_2(t)$) образует стационарный в узком смысле процесс, $\xi^{[u,v]}$ – совокупность значений $\xi(t)$, $u \le t \le v$, и пусть $I\{\xi_1, \xi_2^{[t_0,t_2]} \mid \xi_2^{[-\infty,t_0]}\}$ – условное количество информации о процессе $\xi_1 = \xi_1^{[-\infty,\infty]}$, содержащееся в отрезке $\xi_2^{[t_0,t]}$ процесса ξ_2. Среднее количество указанной информации представляет собой линейно растущую функцию от t:

$$MI\{\xi_1, \xi_2^{[t_0,t]} \mid \xi_2^{[-\infty,t_0]}\} = \left(t - t_0\right) I(\xi_1, \xi_2),$$

Фигурирующая здесь величина $I(\xi_1, \xi_2)$ называется средней скоростью передачи информации стационарным процессом ξ_2 о стационарном процессе ξ_1 или просто – скоростью передачи информации [289].

Скорость передачи информации $I(\xi_1, \xi_2)$ обладает рядом свойств, аналогичных свойствам количества информации. Но она имеет и специфические свойства. Так для всякого сингулярного случайного процесса ξ_2, т.е. такого процесса, все значения $\xi_2(t)$ которого являются функциями от совокупности величин $\xi_2^{[-\infty,t_0]}$ (t_0 может быть выбрано любым), имеет место равенство $I(\xi_1, \xi_2)=0$.

Для всякого регулярного случайного процесса ξ_2 равенство $I(\xi_1, \xi_2)=0$ справедливо лишь тогда, когда случайный процесс ξ_1 не зависит от процесса ξ_2 (это говорит о том, что в некоторых случаях $I(\xi_1, \xi_2) \neq I(\xi_1, \xi_2)$).

При дополнительных условиях типа регулярности скорость передачи информации $I(\xi_1, \xi_2)$ совпадает с пределом

$$I(\xi_1, \xi_2) = \lim_{t-t_0 \to \infty} \frac{1}{t-t_0} I\left(\xi_1^{[t_0,t]}, \xi_2^{[t_0,t]}\right),$$

где $I\left(\xi_1^{[t_0,t]}, \xi_2^{[t_0,t]}\right)$ – количество информации об отрезке процесса $\xi_1^{[t_0,t]}$, заключенное в $\xi_2^{[t_0,t]}$. Так будет, например тогда, когда время меняется дискретно, а отдельные величины $\xi_1(t)$ и $\xi_2(t)$ могут принимать лишь конечное число различных значений или когда распределение вероятностей процессов ξ_1 и ξ_2 является гауссовым. В случае непрерывного времени t так будет для гауссовых процессов, когда спектральная плотность $f(\lambda)$ процесса $\xi_2(t)$ удовлетворяет условию $0 < c \leq \lambda^{2n} f(\lambda) \leq c < \infty$

Пусть стационарный процесс $\xi = \xi(t)$ представляет собой последовательность величин, каждая из которых принимает значения из некоторого алфавита x, состоящего из конечного числа символов $x_1, x_2, \ldots, x_n$. Предположим, что вероятность появления на фиксированном месте определенного символа x_i есть p_i, а вероятность появиться за ним символу x_j не зависит от предшествующих x_i значений и есть p_{ij}:

$$P\{\xi(t) = x_i\} = p_i, P\{\xi(t+1) = x_i x_j \mid \xi(t) = x_i, \xi(t-1), \ldots,\} = p_{ij}$$

Другими словами $\xi = \xi(t)$ – стационарная цепь Маркова с переходными вероятностями $\{p_{ij}\}$ и стационарным распределением $\{p_i\}$. Тогда скорость передачи информации стационарным процессом $\xi(t)$ будет $I(\xi, \xi) = -\sum_{ij} p_i p_{ij} \log p_{ij}$.

В частности, если $\xi = \xi(t)$ – последовательность независимых величин (в случае $p_{ij} = p_j$), то

$$I(\xi, \xi) = -\sum_j p_j \log p_j .$$

Пусть $\xi_1 = \xi_1(t)$ и $\xi_2 = \xi_2(t)$ – стационарные гауссовы процессы со спектральными плотностями $f_{11}(\lambda)$, $f_{22}(\lambda)$ и взаимной спектральной плотностью $f_{12}(\lambda)$ причем процесс $\xi_2 = \xi_2(t)$ является регулярным. Тогда

$$I(\xi_1, \xi_2) = -\frac{1}{4\pi} \int \log\left[1 - \frac{|f_{12}(\lambda)|^2}{f_{12}(\lambda) f_{22}(\lambda)}\right] d\lambda,$$

Рассмотрим следующее условие близости гауссовых стационарных процессов $\xi_1(t)$ и $\xi_2(t)$:

$$M|\xi_1(t) - \xi_2(t)|^2 \le \delta^2.$$

Наименьшая скорость передачи информации $H = \inf I(\xi_1, \xi_2)$, совместимая с указанным условием «δ-точности», выражается следующей формулой:

$$H = \frac{1}{4\pi} \int\limits_{f_{11}(\lambda) \ge \theta^2} \log \frac{f_{11}(\lambda)}{\theta^2} d\lambda = -\frac{1}{4\pi} \int \log\left[1 - \frac{|f_{12}(\lambda)|^2}{f_{11}(\lambda) f_{22}(\lambda)}\right] d\lambda,$$

где $f_{22}(\lambda) = \begin{cases} f_{11}(\lambda) - \theta \text{ при } f_{11}(\lambda) \ge \theta^2, \\ 0 \text{ при } f_{11}(\lambda) < \theta^2, \end{cases}$

$f_{12}(\lambda) = f_{22}(\lambda)$, а параметр θ^2 определяется из равенства

$$\int [f_{11}(\lambda) - f_{22}(\lambda)] d\lambda = \delta^2.$$

Эта формула показывает, какого типа спектральная плотность $f_{22}(\lambda)$ должна быть у регулярного стационарного процесса $\xi_2(t)$, который несет минимальную информацию $I(\xi_1, \xi_2) \approx H$ о процессе $\xi_1(t)$. В случае дискретного времени, когда $f_{11}(\lambda) \ge \theta^2$ при всех λ, $-\pi \le \lambda \le \pi$, нижняя грань H скорости передачи достигается для такого процесса $\xi_2(t)$ (со спектральной плотностью $f_{22}(\lambda)$, задаваемой приведенной выше формулой), который связан с процессом $\xi_1(t)$ формулой

$$\xi_2(t) = \xi_1(t) + \zeta(t),$$

где $\zeta(t)$ – стационарный гауссов шум, не зависящий от процесса $\xi_2(t)$; в общем случае формула $f_{22}(\lambda)$ задает предельный вид соответствующей спектральной плотности регулярного процесса $\xi_2(t)$.

В случае, когда спектральная плотность $f_{11}(\lambda)$ приближенно выражается формулой

$$f_{11}(\lambda) \approx \begin{cases} \dfrac{\sigma^2}{2w} \text{при} |\lambda \pm \lambda_0| \le \dfrac{1}{2} w; \\ 0 \text{ при остальных} = \lambda \end{cases}$$

соответствующая минимальная скорость передачи информации H может быть вычислена по приближенной формуле

$$H \approx \frac{w}{2\pi}\log\frac{\sigma^2}{\delta^2},\ \sigma^2 = M[\xi(t)]^2.$$

2) симметричный канал без памяти

Рассмотрим симметричный канал передачи данных без памяти с конечным числом входных сигналов x_1, когда передаваемый сигнал x_1 с вероятностью $1–p$ правильно принимается на выходе канала связи, а с вероятностью p искажается, причем все возможные искажения равновероятны: вероятность того, что на выходе будет сигнал x_2, равна $\frac{p}{N-1}$ для любого $x_2 \neq x_1$, где N – общее число сигналов. Для такого канала связи пропускная способность $c = \sup I(\xi_1, \xi_2)$ достигается в случае, когда на вход поступает последовательность независимых и равномерно распределенных сигналов $\ldots, \xi_1(-1),\ \xi_1(0),\ \xi_1(1), \ldots$; эта пропускная способность выражается формулой

$$C = \log_2 N - (1-p)\log_2(1-p) - p\log_2\frac{p}{N-1}.$$

Рассмотрим канал связи, на входе которого сигналы образуют стационарный процесс $\xi_1 = \xi_1(t)$, $M[\xi_1(t)]^2 < \infty$.

Пусть при прохождении сигнала $\xi_1 = \xi_1(t)$ он подвергается линейному преобразованию $A\varphi$ со спектральной характеристикой $\varphi(\lambda)$ и, кроме того, на него накладывается аддитивный стационарный гауссов шум $\zeta = \zeta(t)$, так что на выходе канала имеется случайный процесс $\xi_2(t)$ вида $\xi_2(t) = a\varphi\ \xi_1(t) + \zeta(t)$.

Предположим также, что ограничение на входной процесс состоит в том, что $M[\xi_1(t)]^2 \le \Delta^2$ (постоянная Δ^2 ограничивает среднюю энергию входного сигнала). Пропускная способность такого канала может быть вычислена по формуле

$$C = \frac{1}{4\pi}\int\limits_{|\varphi(\lambda)|^2\theta^2 \ge f_{\zeta\zeta}(\lambda)} \log\frac{|\varphi(\lambda)|^2\theta^2}{f_{\zeta\zeta}(\lambda)}d\lambda =$$
$$= -\frac{1}{4\pi}\int\log\left[1 - \frac{|\varphi(\lambda)|^2 f(\lambda)}{|\varphi(\lambda)|^2 f(\lambda) + f_{\zeta\zeta}(\lambda)}\right]d\lambda.$$

В последнем выражении интегрирование ведется в пределах $-\pi \le \lambda \le \pi$ для дискретного времени *t* и в пределах $-\infty < \lambda < \infty$ для непрерывного *t*), где $f_{\zeta\zeta}(\lambda)$ – спектральная плотность гауссова процесса ζ (t), функция $f(\lambda)$ имеет вид

$$f(\lambda) = \begin{cases} \theta^2 - f_{\zeta\zeta}(\lambda)|\varphi(\lambda)|^{-2} & при \quad \theta^2 \ge f_{\zeta\zeta}(\lambda)|\varphi(\lambda)|^{-2} \\ 0 \quad противном & случае \end{cases},$$

а параметр θ^2 определяется из равенства $\int f(t)d\lambda = \Delta^2$.

Нужно сказать, что если функция $f(\lambda)$ представляет собой спектральную плотность регулярного стационарного гауссова процесса $\xi_1(t)$, то этот процесс, рассматриваемый как входной сигнал, обеспечивает максимальную скорость передачи информации: $I(\xi_1, \xi_2)=C$. Однако в наиболее интересных случаях, когда время *t* меняется непрерывно, функция $f(\lambda)$ обращается в нуль на тех интервалах частот λ, где уровень шума сравнительно высок (отличные от нуля значения $f(\lambda)$ сосредоточены в основном на тех интервалах частот λ, где уровень шума сравнительно мал), и поэтому не может служить спектральной плотностью регулярного процесса. Более того, если в качестве входного сигнала выбрать процесс $\xi_1(t)$ с спектральной плотностью $f(\lambda)$, то этот сигнал будет сингулярным и соответствующая скорость передачи информации $I(\xi_1, \xi_2)$ будет равна нулю, а не максимально возможному значению *C*, указанному выше.

Тем не менее, приведенные выражения полезны, так как позволяют приблизительно представить вид спектральной плотности $f(\lambda)$ регулярного входного сигнала $\xi_1(t)$, обеспечивающей скорость передачи $I(\xi_1, \xi_2)$, близкую к максимальному значению *C*. С практической точки зрения наиболее интересен случай, когда канал связи имеет ограниченную полосу *w* пропускаемых частот, т.е. когда спектральная характеристика выражается формулой

$$|\varphi(\lambda)|^2 \approx \begin{cases} 1 \quad при \quad |\lambda \pm \lambda_0| \le \frac{1}{2}W, \\ 0 \quad при \quad остальных \quad \lambda \end{cases}$$

а проходящий через канал шум имеет равномерный спектр:

$$f_{\zeta\zeta}(\lambda) \approx \begin{cases} \dfrac{\sigma^2}{2W} & при \quad |\lambda \pm \lambda_0| \le \dfrac{1}{2}W, \\ 0 \quad при \quad остальных \quad \lambda \end{cases}$$

В этом случае пропускная способность может быть вычислена по приближенной формуле

$$C \approx \frac{W}{2\pi} \log\left(1 + \frac{\Delta^2}{\sigma^2}\right).$$

При этом входной сигнал $\xi_1(t)$, обеспечивающий скорость передачи информации $I(\xi_1, \xi_2)$, близкую к максимальной, является гауссовым стационарным процессом со спектральной плотностью $f(\lambda)$ вида

$$f(\lambda) \approx \begin{cases} \dfrac{\Delta^2}{2W} & при \quad |\lambda \pm \lambda_0| \le \dfrac{1}{2}W, \\ 0 \quad при \quad остальных \quad \lambda \end{cases}$$

так что параметры Δ^2 и σ^2 имеют следующий физический смысл:

$\Delta^2 = M|\xi_1(t)|^2$ – энергетический уровень входного сигнала,

$\sigma^2 = M|\zeta(t)|^2$ – энергетический уровень шума.

8.2. Пропускная способность непрерывного канала. Теорема Шеннона

Пропускная способность непрерывного канала в расчете на один отсчет передаваемого сигнала, определяется как

$$\begin{aligned} C_{отсч} &= \max_{\omega(X)} I(X,Y) = \max_{\omega(X)} [h(X) - h(X/Y)] = \\ &= \max_{\omega(X)} [h(X) - h(Y/X)]. \end{aligned} \tag{8.2}$$

Здесь X и Y – случайные величины, отсчеты процессов $X(t)$ и $Y(t)$ на входе и выходе канала. Пропускная способность C определяется как сумма значений $C_{отсч}$, взятая по всем отсчетам за секунду. При этом дифференциальные энтропии в (8.2) должны вычисляться с учетом вероятностных связей между отсчетами.

Вычислим *пропускную способность* непрерывного канала без памяти с аддитивным белым (т.е. имеющим равномерный энергетический спектр и полностью некоррелированные несовпадающие отсчеты) гаусовском шумом, имеющим полосу пропускания F, если средняя мощность сигнала (дисперсия X) не превышает заданной величины P_c. Мощность (дисперсию) шума в полосе F обозначим $P_ш$. Поскольку шум аддитивный, отсчеты входного X и выходного Y сигналов и шума N связаны равенством

$$Y = X + N. \tag{8.3}$$

т.к. N имеет нормальное распределение с нулевым математическим ожиданием, то и условная плотность вероятности при фиксированном x будет также нормальной с математическим ожиданием x и дисперсией $P_ш$. Найдем пропускную способность в расчете на один отсчет (8.4):

$$C_{отсч} = \max_{\omega(X)} \left[h(Y) - h(Y/X) \right]. \tag{8.4}$$

Дифференциальная энтропия гауссовского распределения $h(Y/X)$ в соответствии со своим свойством 2 не зависит от математического ожидания и равна $\log\sqrt{2\pi \cdot e \cdot P_ш}$. Поэтому для определения $C_{отсч}$ следует найти такую плотность распределения $D(x)$, при которой максимизируется $h(Y)$. Из (8.4) учитывая, что X и N-независимые случайные величины, имеем для дисперсий:

$$D(Y) = D(X) + D(N) = P_c + P_ш \tag{8.5}$$

таким образом, дисперсия Y фиксирована, так как P_c и $P_ш$ заданы. В соответствии со свойством 4 дифференциальной энтропии максимальная дифференциальная энтропия при фиксированной дисперсии обеспечивается гауссовским распределением. Известно, что при нормальном распределении X распределение Y будет также нормальным и, следовательно, обеспечивается максимум дифференциальной энтропии (8.6):

$$\max_{\omega(X)} h(Y) = \log\sqrt{2\pi \cdot e \cdot (P_c + P_ш)}, \tag{8.6}$$

откуда

$$C_{отсч} = \log\sqrt{2\cdot\pi\cdot e\cdot(P_с + P_ш)} - \log\sqrt{2\cdot\pi\cdot e\cdot P_ш} = \frac{1}{2}\log\frac{P_с + P_ш}{P_ш}.$$

Переходя к пропускной способности C в расчете на секунду, заметим, что *количество информации*, содержащейся в следующих друг за другом отсчетах, максимально в том случае, когда отсчеты независимы. Этого можно достичь, если процесс $X(t)$ выбрать таким, чтобы его спектральная плотность была равномерной в полосе частот F. В этом случае отсчеты, разделенные интервалами Δt, кратными $1/(2F)$, взаимно некоррелированы, а для гауссовских величин некоррелированность означает независимость. Поэтому *пропускную способность* C (за секунду) можно найти, сложив пропускные способности (8.5) для $2F$ независимых отсчетов:

$$C = 2F\cdot C_{отсч} = F\cdot\log\left(1+\frac{P_с}{P_ш}\right). \quad (8.7)$$

Она реализуется, если $X(t)$ – гауссовский процесс с равномерной спектральной плотностью в полосе частот F (квазибелый шум).

Соотношение (8.7) называют формулой Шеннона. Формула Шеннона указывает на возможность обмена полосы пропускания на мощность сигнала и наоборот. Однако поскольку C зависит от F линейно, а от $P_с/P_ш$ – по логарифмическому закону, компенсировать возможное сокращение полосы пропускания увеличением мощности сигнала, как правило, не выгодно. Более эффективным является обратный обмен мощности сигнала на полосу пропускания. Рассмотрим, как меняется пропускная способность гауссовского канала с изменением полосы пропускания. Для этого выразим мощность шума в канале через его одностороннюю спектральную плотность N_0. Имеем $P_Ш = N_0\cdot F$, поэтому

$$C = F\cdot\log\left(1+\frac{P_с}{N\cdot F}\right) = F\cdot\log e\ln\left(1+\frac{P_с}{N_0\cdot F}\right). \quad (8.8)$$

При увеличении F пропускная способность C сначала быстро возрастает, а затем асимптотически стремиться к пределу

$$C_\infty = \lim_{F\to\infty} C = \left(\frac{P_с}{N_0}\right)\cdot\log e \quad \frac{бит}{с} \quad (8.9)$$

Результат (8.9) можно получить, если учесть, что при $|\varepsilon| << 1$ (т.е. при больших F) $\ln(1+\varepsilon) \approx \varepsilon$. Зависимость C_∞ от F показана на рис. 8.1.

Исходя из (8.9) можно показать, что для передачи заданного количества информации по каналу с шумом отношение энергии сигнала к спектральной плотности шума $h^2 = \dfrac{P_c \cdot T}{N_0}$ должно превысить некоторую пороговую величину. Действительно, если на передачу сообщения затрачено время T, то среднее количество переданной информации $T \cdot I'(X,Y) < T \cdot C_\infty$, т.к. пропускная способность канала при любой полосе F не может превысить предельное значение.

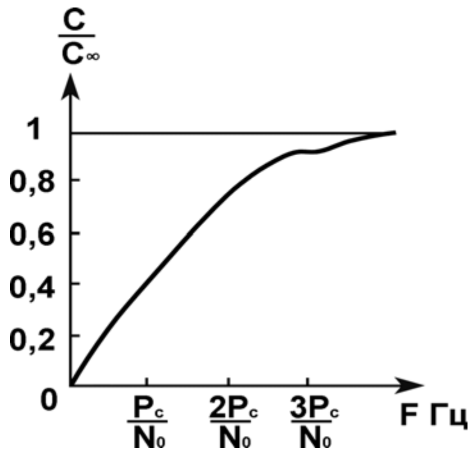


Рис. 8.1. Зависимость C_∞ от F

Таким образом $T \cdot I'(X,Y) < \left(P_c \cdot \dfrac{T}{N_0}\right) \cdot \log e$ и, следовательно, для передачи одного бита (т.е. $T \cdot I'(X,Y) = 1$) информации необходима энергия сигнала

$$P_c \cdot T > N_0 / \log e = N_0 \cdot \ln 2 = 0{,}693 N_0$$

Эти рассуждения устанавливают потенциально взаимосвязь между количеством переносимой сигналом информации и энергией сигнала.

Отметим, что формула Шеннона (8.7) справедлива только для канала с постоянными параметрами и аддитивным гауссовским белым или квазибелым шумом. Если аддитивный шум не гауссовский и его спектр неравномерен в полосе пропускания канала, то его пропускная способность больше, чем вычисленная по формуле (8.7). Мультипликативные помехи (замирание сигнала), обычно снижают пропускную способность по сравнению с результатом (8.7).

Рассмотрим теперь вопрос согласования источника непрерывных сообщений с непрерывным каналом. Передача непрерывных сообщений по *каналу* без помех не представляет интереса, так как в этом теоретическом случае проблема связи вообще не возникает. Одним импульсом, амплитуда которого на приемной стороне воспринимается с неограниченной точностью, может быть передано бесконечно большое количество информации, однако этот результат не может быть использован в практике, так как этот импульс нельзя точно измерить.

Для канала с шумом с пропускной способностью С, на вход которого подключен источник с производительностью $H'_{\varepsilon}(X)$ Шеннон доказал следующую теорему. Если при заданном критерии эквивалентности сообщений источника ε_0^2 его *е-производительность* меньше пропускной способности канала $H'_{\varepsilon}(X) < C$, то существует способ кодирования и декодирования, при котором неточность воспроизведения сколь угодно близка к ε_0^2 (прямая теорема). При $H'_{\varepsilon}(X) > C$ такого способа не существует (обратная теорема).

Доказательство теоремы осуществляется аналогично доказательству основной теоремы кодирования для канала с шумом.

Термин «кодирование» здесь понимается в широком смысле, так как он определен во введении.

Не доказывая теорему, поясним возможность осуществления указанного в ней способа передачи. Если сообщения должны воспроизводиться с определенной верностью, то из бесконечного множества непрерывных сообщений длительностью *Т* передавать необходимо только конечное подмножество воспроизводящих сообщений. Процесс кодирования при этом заключается в отождествлении полученного от источника сообщения с ближайшим воспроизводящим и сопоставлением ему конкретного сигнала из множества разрешенных сигналов, специально подобранных для передачи, с учетом действующей в канале помехи. При декодировании полученный сигнал отождествляется с ближайшим разрешенным и ставится в соответствие воспроизводящему сообщению. Ошибки не произойдет, если принятый сигнал попадет в некоторою собственную область соответствующего разрешенного сигнала, размеры которой зависят от средней мощности помехи. При определенном уровне средней мощности передаваемых сигналов можно создать ограниченное число разрешенных сигналов с не перекрывающимися собственными областями. Оно (это число) и определяет предельную скорость передачи с обеспечением заданного уровня верности.

Поскольку обычно допускается возможность появления любого значения помехи, вероятность воспроизведения другого разрешенного сигнала остается конечной. Однако при доказательстве теоремы показано, что она стремится к нулю при неограниченном увеличении длительности передаваемых сигналов.

При этом из теоремы Шеннона следует, что при выполнении условия $H'_{\varepsilon}(X) < C$, можно преобразовать сообщение в сигнал так, чтобы отношение сигнал-шум на выходе приемника (декодера) было больше значения ρ_0, обеспечивающего эквивалентность переданного и принятого сообщений, хотя в канале (т.е. на входе приемника) отношение сигнал-шум может быть во много раз меньше ρ_0.

Однако до сих пор оптимальное кодирование непрерывных сообщений (без преобразования в дискретные) в непрерывном канале не находит приемлемой реализации. Более предпочтительным в настоящее время представляется преобразование непрерывных сообщений в дискретные с последующим использованием эффективного и помехоустойчивого кодирования.

Контрольные вопросы

1. Представить математическую вероятностную модель канала связи и раскрыть ее содержание.
2. Как можно представить кодированное сообщение, передаваемое по каналу связи?
3. Что такое емкость канала связи и чем она характеризуется?
4. Дать определение энтропии источника сообщений и как она определяется?
5. Как определяется количество информации для сообщений $\xi_0^{(1)}, \xi_0^{(2)}, ..., \xi_0^{(n)}$, поступающих в канал связи независимо друг от друга?
6. Описать условия, при которых возможна передача достаточно длинных сообщений $\xi_0^{(1)}, \xi_0^{(2)}, ..., \xi_0^{(n)}$.

7. Как можно представить количество информации I(ξ_0, ξ_3) для абстрактных случайных величин ξ_0 и ξ_3 со значениями в пространствах X_0 и X_3?

8. Что представляет собой канал связи с изменяющимися состояниями?

9. Описать стационарный режим работы канала связи.

10. Раскрыть содержание ε-энтропии случайной величины ξ.

11. Что такое скорость передачи информации I(ξ_1, ξ_2) и каковы ее свойства?

12. Что такое стационарный гауссов шум ζ(t)?

13. Раскрыть содержание симметричного канала без памяти.

14. Как определяется пропускная способность симметричного канала без памяти

15. Как определяется пропускная способность канала связи, имеющего ограниченную полосу w пропускаемых частот?

16. Раскрыть содержание выражения для расчета пропускной способности непрерывного канала в расчете на один отсчет передаваемого сигнала.

17. Как вычисляется пропускная способность непрерывного канала без памяти с аддитивным белым (т.е. имеющим равномерный энергетический спектр с полностью некоррелированными несовпадающими отсчетами) гаусовский шум?

18. Представить формулу Шеннона, описывающей возможность обмена полосы пропускания на мощность сигнала и наоборот.

19. Представить математическое выражение зависимости пропускной способности гауссовского канала от изменения полосы пропускания канала связи.

20. Описать потенциальную взаимосвязь между количеством переносимой сигналом информации и энергией сигнала.

21. Представить математическое выражение для описания канала связи с постоянными параметрами и аддитивным гауссовским белым или квазибелым шумом.

22. Обосновать возможность или невозможность вычисления пропускной способности канала связи с аддитивными не гауссовским шумом и неравномерным спектром в полосе пропускания канала использованием формул Шеннона $C = 2F \cdot C_{отсч} = F \cdot \log\left(1 + \frac{P_с}{P_ш}\right)$.

23. Как может быть осуществлено согласование источника непрерывных сообщений с непрерывным каналом?

ГЛАВА 9. ПОДХОДЫ К ФОРМАЛИЗАЦИИ ПОНЯТИЯ ЦЕННОСТИ ИНФОРМАЦИИ

Может ли существовать единое определение ценности информации? По-видимому, нет. Существует множество совершенно не сходных между собой ситуаций, в которых использование информации приносит (или может принести) выгоду. Этим объясняется, что уже сейчас существует немало формализаций понятия ценности информации, а появление новых продолжается.

Предположим, что понятие количества информации еще не формализовано. При самом поверхностном взгляде ясно, что между понятиями ценности и количества информации (оба понятия сейчас рассматриваются на интуитивном уровне) много общего и что трудно провести между ними четкую границу. Действительно, если ценность информации (для определенности – ценность каждого из нескольких возможных сообщений) так или иначе установлена, то возникает вопрос, нельзя ли просто отождествить количество информации с ее ценностью, измерять количество информации, скажем, не в битах, а в рублях. Количество информации (вспомним о формализации Шеннона) выражается уменьшением неопределенности (энтропии), что можно рассматривать как выигрыш особого (важного именно для задач связи) вида. Таким образом, понятия ценности информации и ее количества переходят друг в друга.

9.1. Апостериорный подход

Поясним сказанное выше о взаимном переходе друг в друга понятий ценности и количества информации. Рассмотрим некоторую целесообразно действующую систему. Вместо того чтобы количественно оценивать информационные сообщения, поступающие в систему, можно оценивать действия (их выгоду или невыгоду), осуществляемые системой на основании (как предполагается) поступившей информации, не заботясь о том, что представляет собой и как, измеряется поступившая информация. Естественно считать, что действия системы (и связанный с ними выигрыш) косвенно характеризуют и информацию, на основании которой эти действия совершаются. Иными словами, рассматривается только «реализованная» системой информация. Реализация же представляет собой акт выбора системой одного из допустимых действий, вариантов поведения. Назовем такой подход к оценке количества (или ценности) информации апостериорным.

Апостериорный подход не нов: вопрос о том, знал или не знал некто определенные обстоятельства, часто с большой достоверностью может быть решен на основании действий субъекта, их успеха или неуспеха.

Отметим интересную особенность апостериорного подхода. Он позволяет – наряду с полезной информацией – рассматривать (и количественно оценивать) и «дезинформацию», поступающую в систему. Поступление «дезинформации» проявляется в таких действиях, которые ухудшают положение системы, ведут к росту проигрыша. Шенноновская теория информации имеет дело только с неотрицательными величинами – энтропией и количеством информации, а потому и не может описать важный случай поступления дезинформации. Модификации шенноновской теории, позволяющие сделать это, основаны на применении апостериорного подхода.

Апостериорный подход, является лишь общим принципом – слишком общим, чтобы применяться непосредственно. Однако он дает возможность построить классификацию мер количества и ценности информации и прояснить сходство и различие между ними достаточно четко и с единой точки зрения.

Можно выделить три направления изучения количественных свойств информации, три группы теорий.

1. *«Чистые теории информации»*. Выигрыш или проигрыш измеряется степенью уменьшения или увеличения неопределенности – специальной величины, характеризующей, грубо говоря, степень незнания (или неуверенности в знании) состояния среды или системы. Понятие неопределенности, в свою очередь, допускает различные уточнения.

2. *«Теории абсолютной ценности информации»*. Рассматривается только «материальный» выигрыш или проигрыш (т.е. выраженный, например, в рублях, тоннах сэкономленного горючего). Количественные и структурные аспекты информации игнорируются. Это направление близко к теории статистических решений, в которой основным является понятие средних потерь, или «риска». Степень риска характеризует качество принимаемых решений и – косвенно – информацию, на основании которой принимаются решения.

Здесь принято к нему относить известные разделы теории исследования операций (смотри библиографию), в которых принимаются определенные предположения об информированности объектов (например: «центр полностью информирован о целях подсистем», «центр не имеет никакой информации о локальных ограничениях для подсистем, но полностью информирован о моделях подсистем и их параметрах»), если такие предположения раз и навсегда фиксируются в данном разделе. В этом случае собственно информация, очевидно, фактически не входит в рассмотрение, а употребление этого термина для описания поведения систем и подсистем – вопрос удобства и наглядности.

3. *«Теории ценности количества информации»*. Апостериорный принцип применяется одновременно и к материальному выигрышу и к выигрышу в смысле уменьшения неопределенности. Сравнение этих двух видов выигрыша (значение одного из которых, как правило, фиксируется, а другого – максимизируется) позволяет установить максимальную материальную пользу, которую способно принести данное фиксированное количество информации (т. е. фиксированное уменьшение неопределенности), и минимальное количество информации, необходимое (при условии наилучшего использования этого количества) для обеспечения уровня материального выигрыша не ниже заданного. Мера количества информации (или мера неопределенности) в этом случае вовсе не обязана быть сходной с шенноновской; важно лишь наличие в теории этого направления не менее двух видов штрафа, не менее двух способов измерения информации.

Далее подробно рассмотрим наиболее важный случай, когда мера неопределенности – Шенноновская. Здесь же заметим, что с обосновываемой точки зрения сюда попадают те разделы теории исследования операций, в которых «степень информированности» становится существенно переменным параметром (пусть даже и принимающим всего два значения: «полная информированность» и «абсолютная неинформированность»).

В данном разделе рассмотрим лишь некоторые модификации теории информации, связанные с различными уточнениями понятия неопределенности.

Теории, принадлежащие второму направлению, применимы в тех случаях, когда можно пренебречь ограничениями на пропускную способность каналов, емкость запоминающих устройств, затраты, связанные с передачей и хранением информации, время и сложность ее обработки. Например, в тех случаях, когда ожидается поступление редких, неважных сообщений, выигрыш от применения которых неизмеримо превосходит эти затраты, обычно игнорируются (хотя бы и нервом приближении) вопросы экономного использования средств передачи, хранения (а часто – и обработки) информации. Так, «полная информированность центра» в теории исследования операций может означать на практике получение центром как одного бита информации, так и огромных таблиц, в которых существенны все символы. В этой теории важен факт информированности или неинформированности подсистем – факт, непосредственно проявляющийся в законах функционирования подсистем, – а не количество информации. Теории второго направления, таким образом, игнорируют ограничения, связанные с реальным процессом передачи и обработки информации.

Теории третьего направления имеют достаточно ясно очерченную область применения: изучение таких моделей, в которых информационные ограничения существенны. Роль этих ограничений определяется на практике в основном следующими двумя (не исключающими друг друга) причинами:

– затраты на передачу, хранение и обработку информации становятся сравнимыми с возможным выигрышем от ее использования;

– существуют неустранимые (для данной системы) ограничения (любой природы – от физических законов до прозаического отсутствия лишнего блока памяти) на емкость устройств, скорость передачи и сложность обработки – ограничения, заставляющие искать наиболее выгодные способы использования данного фиксированного (или ограниченного сверху) количества информации.

Ясно, что именно в этих случаях четкое различие между ценностью информации и количеством информации становится необходимым. Если первое понятие связано главным образом с внешним эффектом (выигрышем или штрафом), то второе – с внутренними затратами на передачу, хранение, воспроизведение, обработку информации.

Как отмечено выше, количество информации можно рассматривать, как выигрыш особого вида – выигрыш в уменьшении неопределенности. Это означает, что произвол в выборе меры количества информации переносится на выбор меры неопределенности. Последнее понятие обычно формализуют в виде безразмерной величины, допускающей интерпретацию, как степени богатства, разнообразия, неожиданности состояний среды или системы, являющейся источником информации.

В отношении материального выигрыша, входящего, наряду с количеством информации, в число основных понятий третьей группы теорий заметим следующее. Одной функции выигрыша далеко не всегда достаточно для удовлетворительного отражения цели функционирования системы. Как правило, требуется оптимизация по многим критериям, а вместо простой экстремальности приходится часто рассматривать те или иные виды компромисса. При анализе систем высокой сложности вообще трудно говорить о критериях, о цели. Например, оптимальное решение может характеризоваться некоторой системой требований, аксиоматически описывающих такие содержательные понятия, как «равноправие», «справедливость». Вполне очевидно, что возможность удовлетворить такой системе требований не может не зависеть от информационных ограничений, действующих в системе. Однако данное направление лишь начинает разрабатываться.

9.2. Формализация понятий неопределенности и количества информации, отличные от подхода Шеннона

В соответствии с проведенной выше классификацией рассмотрим сначала «чистую теорию информации», но в вариантах, отличных от шенноновского. В основном речь будет идти о формализации понятия «неопределенность». Именно оно служит обычно основой для определения количества информации.

Интересной и очень простой формализацией понятия количества информации является определение меры целесообразности управления, предложенное А. А. Харкевичем. Сам А. А. Харкевич называет введенную им меру «ценностью информации». Определим ее как «меру целесообразности управления», так как в данной книге понятие «ценность информации» имеет более широкое содержание. Формализм работы применим в тех случаях, когда единственной задачей, стоящей перед системой, является достижение определенной цели.

Интересно, что в модели А. А. Харкевича апостериорный подход применен «в полную силу». Физическая природа сигналов, логическая структура сообщений, их длина полностью игнорируются. Считается, что на основании некоторой информации, поступившей к системе, – информации, природа которой безразлична, – система принимает решение, изменяющее вероятность достижения цели. Естественно требовать, чтобы увеличению вероятности достижения цели отвечал случай положительного значения вводимой меры, уменьшению вероятности – отрицательного значения, а сохранению прежнего значения вероятности нулевого значений меры. Этим требованиям удовлетворяет мера целесообразности управления (одновременно являющаяся и мерой ценности информации, на основании которой система принимает решение), определяемая как

$$\log_2(p_1 / p_0), \tag{9.1}$$

где p_1 – вероятность достижения цели после выполнения принятого (на основании поступившей информации) решения; p_0 – вероятность достижения цели до принятия решения (или, что здесь то же самое, до поступления информации).

Мера (9.1) не только проста и прозрачна, но и, по-видимому, весьма подходит для использования там, где выигрыш или проигрыш не может быть описан в виде приращения или потерь одних лишь материальных ресурсов. Особенно отчетливо невозможность такого описания видна тогда, когда все средства, которыми располагает информационная система, направлены на избежание катастрофического проигрыша.

Рассмотрим модель, которую можно считать крайним упрощением ситуации, в которой движущийся объект имеет ограниченный временной ресурс (например, время полета самолета ограничено запасом топлива, а, следовательно, цель – посадка самолета, пусть даже и не в аэропорту назначения, – должна быть достигнута за время, не превышающее заданного).

Пусть объект за один шаг может переместиться из точки k в точку $k + 1$ или и k-1 (k целое); исходное положение – точка $k = 0$; цель есть точка $m > 0$; эта точка должна быть достигнута не более чем за n ($n \geq m$) шагов. При отсутствии информации переходы из k в $k+1$ и $k–1$ равновероятны.

Целесообразность управления (и ценность информации, на основе которой оно выбрано) легко подсчитывается. Заметим, что значение целесообразности здесь зависит не только от текущей координаты, но и от номера шага (т. е. от количества оставшегося временного ресурса; в более сложных случаях – от всей истории движения объекта). Если, например, $n = 3$, а $m = 1$, то целесообразность перемещения из точки 0 в точку –1, на первом шаге равна

$$\log_2 \frac{1/4}{(1/2) + (1/2)^2} \approx -1{,}32\, бит\,,$$

что соответствует, как и следовало ожидать, случаю дезинформации.

Сформулируем теперь два неявно использованных в приведенных примерах принципа.

1. Вероятности p_0 и p_1 в (9.1) подсчитывают, исходя из предположения, что дальнейшее поведение объекта будет совершенно случайным (иначе говоря, предполагается худший случай – сообщение, получаемое объектом на каждом шаге, является последним).

2. На каждом шаге объекту сообщается информация о необходимом поведении только на этом шаге (а не на несколько шагов вперед).

На практике оба предположения всегда ослабляются (так, с одной стороны, диспетчер дает пилоту указания, рассчитанные на возобновление связи через некоторое время; с другой стороны, он может дать указания на несколько шагов вперед). Изучение модели в более общей форме требует существенного усложнения аппарата (например, привлечения методов динамического программирования).

Сравним меру Харкевича с количеством информации по Шеннону на основе апостериорного подхода. Это можно сделать, если сопоставить обычной задаче передачи информации формальную цель: угадывание на приемном конце сообщений, передаваемых по каналу.

Будем считать, что объект всегда точно выполняет полученную инструкцию (ошибки, проявляющиеся в нецелесообразном поведении объекта, отнесем на счет передающего конца линии связи). Тогда траектория объекта есть не что иное, как своеобразная запись последовательности полученных сообщений. Ясно, что на каждом шаге уменьшается энтропия ансамбля всех возможных траекторий объекта (или, что то же, энтропия ансамбля ожидаемых допустимых сообщений). Но уменьшение неопределенности поведения объекта (совпадающей в данном случае с шенноновской энтропией) вовсе не означает, что объект ведет себя благоприятным для достижения цели образом. Целесообразность, таким образом, может быть как положительной, так и отрицательной. А неопределенность поведения может только уменьшаться.

Таким образом, возможность описания в модели случая «дезинформации» связана с *наличием цели, отличной от цели угадывания поведения объекта*.

Если же цель состоит именно в угадывании, то мера целесообразности переходит в шенноновскую меру количества информации. Пусть целью (в смысле Харкевича) системы связи является угадывание сообщений длины 3 в алфавите {0, 1}. Априори все сообщения равновероятны (вероятность каждого равна 1/8). Вероятность достижения цели до прихода первого символа равна 1/8; после прихода первого символа она равна 1/4. Целесообразность по Харкевичу полученного сообщения (т.е. первого символа), или, что то же, принятого на этом основании «решения» (о том, что первый символ есть 0 или 1) равна $\log_2(p_1/p_0) = \log_2[(1/4)/(1/8)] = 1$бит), что совпадает с количеством информации (по Шеннону), содержащейся в первом символе.

Одновременное рассмотрение мер Шеннона и Харкевича приводит к постановке задачи в духе теории ценности количества информации, причем мера Харкевича, очевидно, играет роль «материального» выигрыша. Какую наибольшую пользу в деле увеличения целесообразности может играть данное фиксированное количество информации по Шеннону? Интерпретация этой задачи естественна: связь с объектом поддерживается через канал с ограниченной пропускной способностью.

2. Интересное обобщение меры Шеннона естественно вытекает из одной простой задачи теории распознавания образов. Это обобщение осуществляется на иной основе, чем у А. А. Харкевича, в то время как неопределенность в его работе связана с вероятностью достижения цели, неопределенность оценивается средним числом попыток (проб) до достижения цели (предполагается, что она рано или поздно достигается с вероятностью единицы). Подобная мера неопределенности подходит для случаев, в которых система действует «методом проб и ошибок» и таким образом приобретает сведения, которых не имела раньше. Удобной мерой оказывается логарифм среднего числа проб до момента правильного решения задачи: в этом случае получаются наиболее простые соотношения между пропускной способностью канала связи (по которому система может получать некоторую информацию, заменяющую – с точки зрения системы – информацию, извлекаемую из эксперимента методом проб и ошибок) и максимально возможным сокращением неопределенности. При этом шенноновские меры количества информации и неопределенности становятся частными случаями более общих выражений.

Рассматривая известную модель, в которой дано $M = \{m_1,...,m_r\}$ – конечное множество, разбитое на n непересекающихся подмножеств (классов) $A_1,...,A_n$ и для M задано распределение вероятностей

$p(m_i)(i=1,...,r)$, в соответствии с этим распределением наугад выбирается элемент $m_l \in M$, решение соответствующей задачи означает:

найти номер j из множества $\{1,..., n\}$, такой, что выбранный элемент принадлежит классу A_j.

При этом предполагается, что: а) номер j выбирается наугад в соответствии с распределением $q=\{q_i\}(i=1,...,n)$; б) после каждой попытки осуществляется проверка правильности решения (действительно ли $m_l \in A_j$?).

Если подмножество угадано правильно, то задача решена; если же нет, то делается новая попытка. Можно допустить, что распределение изменяется от попытки к попытке в зависимости от предшествующих результатов (например, естественно исключать номера, оказавшиеся ошибочными).

Будем говорить, что задан решающий алгоритм α, если задано начальное распределение q и закон изменения q от пробы к пробе в зависимости от результатов предыдущих проб. Для данных m_l и решающего алгоритма число проб $K_\alpha(m_l)$ есть случайная величина. *Неопределенностью задачи для пары* (α, m_l) назовем логарифм ожидаемого числа проб до решения задачи:

$$N_\alpha(m_l) = \log \overline{K}_\alpha(m_l).$$

Усредняя по всем элементам, получаем величину

$$N_\alpha = \sum_{l=1}^{r} p(m_l) N_\alpha(m_l) = \sum_{l=1}^{r} p(m_l) \log \overline{K}_\alpha(m_l),$$

которую и назовем *неопределенностью задачи для данного решающего алгоритма* α.

Энтропия становится частным случаем неопределенности задачи: она оказывается равной неопределенности задачи для решающего алгоритма (из класса алгоритмов, не меняющих q в процессе решения), «знающего» и наилучшим образом «использующего» распределение вероятностей ответов задачи.

Действительно, легко видеть, что для алгоритмов (указанного класса) наилучшим является (в смысле уменьшения неопределенности) алгоритм β с распределением $q = \{q_i\}$: $q_j = p_j \left(j = 1,\dots,n \right)$, где $p_j = \sum_{m_l \in A_j} p(m_l)$. Неопределенность задачи для этого алгоритма β

$$N_\beta = -\sum_{i=1}^{n} p_i \log p_i = H\left(p_{i,\dots,} p_n\right) \tag{9.2}$$

т. е. равна энтропии распределения $\{p_j\}$ (или, что здесь то же $\{q_j\}$). Для любого алгоритма (указанного класса) т. е. для произвольного распределения q неопределенность задачи

$$N_\alpha\left(p/q\right) = -\sum_{i=1}^{n} p_i \log q_i \,,$$

что, как известно, не меньше, чем величина (9.2).

Докажем это равенство. Пусть $m_l \in A_i$. При этом условии $\overline{K}_\alpha^l(m_i) = \sum_{k=1}^{\infty} kP$ { класс A_i будет угадан ровно за k попыток} $= \dfrac{1}{q_i}$.

Поэтому

$$N_\alpha\left(p/q\right) = -\sum_{i=1}^{n} p_i \log q_i \overline{K}_\alpha^i\left(m_l\right) = -\sum_{i=1}^{n} p_i \log q_i$$

Таким образом, среди алгоритмов указанного класса («необучающихся») наилучшим является такой, у которого распределение $\{q_i\}$совпадает с распределением $\{p_i\}$. Выбор иного распределения q можно трактовать как предположение системы («гипотезу наблюдателя»), что истинное распределение $\{p_j\}$ есть $\{q_j\}$.

Рассмотрим в качестве примера распределение p: $p_j = 2^{-j}\left(j = 1,2.. \right)$ и оптимальный необучающихся алгоритм $\beta\left(q_j = p_j\right)$. Неопределенность

$$N_\beta = N_\beta\left(p/p\right) = -\sum_{j=1}^{n} p_j \log p_j = \sum j \cdot 2^{-i} = 2\,.$$

Интересно отметить, что неопределенность задачи оказывается равной (для оптимального алгоритма) конечной величине, хотя среднее число попыток до решения задачи бесконечно:

$$\overline{K}_{\beta}^{i} = \frac{1}{p_i};\ \sum_{i=1}^{\infty} p_i \overline{K}_{\beta}^{i} = \infty,$$

(здесь $\overline{K}_{\beta}^{i}$ – среднее число проб до успеха, если $m_l \in A_i$). То же верно, очевидно, всегда, когда число классов бесконечно.

На этом примере можно продемонстрировать, какое влияние на неопределенность окажет замена необучающегося алгоритма обучающимся. Пусть распределение меняется от попытки к попытке по следующему закону: если оказалось, что выбранный номер *j* не является номером класса *Ai* такого, что $m_l \in A_i$, то перед следующей попыткой полагается $g_j = 0$, а остальные вероятности умножаются на $(1 - g)$, где q' – старая вероятность номера *j*, какой она была до попытки.

Пусть исходное распределение (перед первой попыткой) есть $q_i = p_j$ (для всех *j*). Нетрудно видеть, что для такого алгоритма (обозначим его γ) среднее число попыток до успеха при условии, что $m_l \in A_i$, в случае распределения $\{p_j\} = \{2^{-j}\}$ равно $\log_2(1/p_i)$ и неопределенность задачи

$$N_{\gamma} = \sum_{j=1}^{\infty} p_j \log_2 \log_2 \frac{1}{p_j} = \sum_{j=1}^{\infty} 2 - j \log_2 j = 0{,}7326 \text{бит}.$$

Отметим, что обучение здесь эффективно только в случае высокой неравномерности распределения p_j (как в данном примере). В случае равномерного распределения $p_j = 1/n$ (мы опять рассматриваем случай конечного *М*) число шагов (среднее) до успеха сокращается лишь в два раза при переходе от необучающегося алгоритма к обучающемуся. Далее рассматриваются только необучающиеся алгоритмы.

Теперь предположим, что под влиянием информации, поступающей по каналу связи, распределение q^1 было заменено на распределение q^2. Соответствующее изменение неопределенности учеными принято принимать за меру пришедшей по каналу информации, так называемой «полезной информации»:

$$\mathrm{I} = \mathrm{N}_{\alpha}\left(p/q^2\right) - \mathrm{N}_{\alpha}\left(p/q^1\right).$$

Если в качестве нулевого уровня неопределенности задачи взять уровень при равномерном распределении $q_i^0 = 1/n$, то запас полезной информации в распределении *q* можно охарактеризовать величиной

$$I_q = -\sum_{i=1}^{n} p_i \log \frac{1}{n} + \sum_{i=1}^{n} p_i \log q_i = \log n - N_\alpha(p/q).$$

Если никакой информации по каналу не поступает, то, как следует из сказанного, неопределенность задачи – даже для оптимального решающего алгоритма – не может стать ниже величины $H(p_1, p_2, \dots p_n)$. Но если после выбора объекта (напомним, что этот выбор осуществляется в соответствии с распределением *p*) по каналу будет передаваться информация (возможно, с искажениями) о том, какому из классов *A*; принадлежит выбранный объект m_l, то неопределенность задачи может стать меньше указанного уровня. Например, в идеальном случае, когда номер класса становится полностью известным системе, она может выбрать решающий алгоритм с распределением *q*, сосредоточенным на номере этого класса; тогда неопределенность станет равной нулю, так как угадывание осуществляется за одну попытку.

Формально вся модель описывается тройкой $\langle p, P, Q \rangle$, где $p = \{p_i\}$ – априорное распределение на множестве {*1, ..., n*} номеров классов *Aj*; *P* –матрица $\|P_{ik}\| \ (i = 1, \dots, n; k = 1, \dots, m)$, описывающая канал без памяти с *n* входными и *m* выходными символами; $Q = \{q^1, \dots, q^m\}$ – множество допустимых распределений, на основе которых система строит решающий алгоритм $\left(q^k = \{q_1^k, \dots, q_n^k\}; k = 1, \dots, m\right)$.

Сначала производится (средой) выбор элемента m_l; и, следовательно, номера *j* класса *Aj* (с распределением *p*); затем сообщение о номере *j* передается по каналу системе и последняя выбирает на основании сообщения одно из распределений q^k, строит соответствующий (необучающийся) алгоритм и пытается с его помощью угадать номер класса, как это описано выше. Канал, получив на входе символ *i* (номер класса), с вероятностью P_{ik} выдаст на выходе символ *k* (номер «рекомендуемого» распределения q^k), и система будет применять в процессе угадывания распределение $q^k \left(\sum_k P_{ik} = 1 \right)$ для всех *i*.

Неопределенность задачи при условии, что $m_l \in A_i$, равна

$$-\sum_{k=1}^{m} P_{ik} \log q_i^k \tag{9.3}$$

так как с вероятностью P_{ik} придет сигнал k, и, следовательно, будет сделано в среднем $\left(q_i^k\right)^{-1}$ попыток до успеха. Усредняя (9.3), получаем полную неопределенность задачи

$$N_{p,P,Q} = -\sum_{i=1}^{n} p_i \sum_{k=1}^{m} P_{ik} \log q_i^k$$

Сформулируем результат, показывающий связь между неопределенностью и шенноновским количеством информации, передаваемой по каналу:

1. Имеет место

$$N_{p,P,Q} \geq H(p) - I(p,P), \tag{9.4}$$

где $H(p) = -\sum_{i=1}^{n} p_i \log p_i$ — энтропия распределения p;

I (p, P) – шенноновское количество информации между случайными величинами I (вход) и i (выход), передаваемой по каналу:

$$I(p,P) = -\sum_{k=1}^{m} \overline{P_k} \log \overline{P_k} + \sum_{i=1}^{n} P_i \sum_{k=1}^{m} P_{ik} \log P_{ik},$$

где $\overline{P_k} = \sum_{i=1}^{n} p_i P_{ik}$.

Так как I (p, P) не превосходит пропускной способности канала (отнесенной, естественно, к одному акту выбора $m_l \in M$), то неопределенность задачи не может быть уменьшена (в результате введения канала) на значение, большее его пропускной способности.

2. Минимум неопределенности задачи достигается, а именно равен правой части (9.4), если p, P, Q таковы, что

$$q_i^k = p_i P_{ik} / \overline{P_k} \left(i = 1,\ldots,n; k = 1,..,m\right).$$

В теории распознавания образов могут применяться и другие меры неопределенности (и соответственно количества информации), более или менее сходные с рассмотренной выше. Например, вместо среднего числа попыток до успеха можно рассматривать, наоборот, логарифм минимального числа «показов» (т. е. длины обучающей последовательности), необходимого для того, чтобы среднее число ошибок на «экзамене» не превосходило определенного уровня. Неопределенность задачи здесь тем больше, чем дольше приходится обучать систему находить правильные решения. Это вполне согласуется с таким возможным интуитивным смыслом понятия неопределенности: неопределенность множества ситуации тем выше, чем больше усилий необходимо затратить, чтобы научиться правильно ориентироваться в ситуациях, выбирать верные решения.

Итак, мера трудности решения задачи может выступать как мера неопределенности, а, следовательно, уменьшение неопределенности (т. е. количество информации) – оцениваться уменьшением трудности решения задачи.

Отметим, что для абсолютно детерминированных (алгоритмических) процедур решения задачи совершенно естественной мерой трудности является мера сложности вычислений (например, число шагов алгоритма, объем используемой памяти). Хотя здесь исходные данные полностью определяют результат, фактическое получение результата требует определенных затрат времени и других ресурсов. В некотором смысле можно считать, что результат становится «все более определенным» по мере развертывания вычислительного процесса.

3. Рассмотрим теперь одну общую меру – меру накопления информации в адаптивных системах управления – так называемую естественную функцию неопределенности.

Поиски формализации, которая адекватно описывала бы наиболее общие свойства понятия неопределенности с точки зрения задач теории адаптивных систем управления, привели к построению в известной модели, проясняющей вопрос о соотношении мер количества информации и ценности информации для задач управления. По принятой классификации теорию можно отнести как к первой, так и ко второй группе теорий. Но сразу заметим, что модель не может дать ничего для теории ценности количества информации, так как в модели рассматривается лишь одна величина, которую можно интерпретировать либо как «материальный» выигрыш, либо как выигрыш в уменьшении неопределенности. В этом отношении теория похожа на модель А. А. Харкевича. Существенное усложнение состоит в следующем.

Выбор управления в адаптивной системе должен производиться с учетом его влияния не только на текущее состояние системы (и на «текущий выигрыш»), но и на темп накопления информации, названный в научной литературе принципом дуальности управления А. А. Фельдбаума. Оптимальная стратегия управления всегда дуальна – за исключением тех случаев, когда темп накопления информации не зависит от выбора управления. Однако на практике приходится часто пользоваться субоптимальными стратегиями: выбор оптимальной стратегии слишком сложен. Здесь вопрос о дуальности приобретает самостоятельное значение, так как, «подправляя» субоптимальную стратегию в сторону увеличения темпа уменьшения неопределенности (или, что то же, темпа накопления информации), часто удается ускорить процесс адаптации и, как следствие, уменьшить средние потери.

В связи с этим возникает вопрос о введении некоторой функции неопределенности, с помощью которой можно было бы описать процесс накопления информации в системе, – такой функции, которая была бы тесно связана с целью управления (т.е. с минимизацией среднего штрафа). Основное положение состоит в том, что, подобно тому как положение тела, его скорость и т.д. могут быть измерены только по отношению к фиксированной системе координат, так и информация, заключенная в произвольном сообщении, может быть измерена только по отношению к некоторой фиксированной управляемой системе, причем при переходе к другой системе наши выводы могут изменяться не только количественно, но и качественно (как в механике при переходе от инерциальной к неинерциальной системе координат).

Цель функционирования системы – минимизировать функцию потерь выбором управляющих воздействий $u_1,...,u_n$ ($\underline{u_t}$ – управляющее воздействие в момент t): $R\ (\vartheta, u_1,...,u_n) \to \min$, где параметр и характеризует устройство объекта управления и имеет следующую интерпретацию: он включает в себя все то, что существенно для цели управления, но неизвестно нам. Однако распределение вероятностей параметра ϑ может быть известным. Поэтому предлагается рассматривать распределение $m_l \in A_j$ на пространстве Θ всех возможных значений параметра ϑ. Вводимая ниже функция неопределенности φ является функцией от распределения ν на Θ:

$$\varphi : M(\Theta) \to R,$$

где $M(\Theta)$ – множество всех вероятностных мер на Θ; R – множество действительных чисел.

В моменты $t = 1,...,n$ на объект управления подается воздействие u_t из множества U_t всех допустимых в момент t воздействий. В результате получаем (после каждого такта, кроме последнего) информацию об объекте – сообщение $y_t \in Y_t$ (где Y_t – множество всех допустимых в момент t сообщений о состоянии объекта управления), причем это сообщение случайным образом зависит от воздействий $u_1,...,u_t$, подававшихся в моменты 1, 2,..., t. Для описания подобной зависимости, очевидно, следует считать, что управляющие воздействия и значение параметра ϑ определяют не конкретное значение сообщения y_t, а распределение на Y_t. Это описывается заданием отображений

$$\mu_t : \Theta \times U_1 \times U_2 \times \ldots \times U_t \to M(Y_t),$$

где $M(Y_t)$ – множество всех вероятностных мер на пространстве $Y_t (t = 1,2,...,n-1)$. Функция штрафа есть отображение

$$R : \Theta \times U_1 \times ... \times U_n \to R.$$

Задачу ее минимизации можно понимать по-разному из-за того, что значение параметра ϑ не известно точно. Чаще всего можно считать известным распределение ν на θ («байесовский случай») и минимизация понимается в смысле минимизации среднего значения функции штрафа R при выбранной стратегии управления $\delta = \{\delta_t\}$:

$$\delta_t : \theta \times U_1 \times Y_1 \times ... \times U_{t-1} \times Y_{t-1} \to U_t$$

($t=1,2,..,n$). Если стратегия δ выбрана, то среднее значение штрафа определено однозначно, ибо каждая стратегия δ порождает меру на пространстве $\theta \times U_1 \times ... \times U_n$ и среднее значение функции R есть среднее по этой мере. Следовательно, задача управления сводится к минимизации указанного среднего по этой мере путем выбора оптимальной стратегии управления δ.

Пятерка $<\theta, \{U_t\}_t =_{1,...,n,}, \{Y_t\}_t =_{1,...,n-1}, \{\mu_t\}_t =_{1,...,n-1}, R>$, называется n-шаговой управляемой системой. Можно перенести рассмотрение на случай бесконечношаговой системы с сохранением смысла обозначений.

Аппарат n-шаговых управляемых систем предложен для описания функционирования адаптивных систем управления. В то же время данная конструкция описывается учеными на языке управляемых марковских процессов с неполной информацией. И по их мнению, подобные системы являются не новым объектом изучения, а скорее удобным языком описания целенаправленной деятельности.

На практике часто встречается случай, когда заранее не известно, какая функция потерь R будет описывать качество функционирования системы. Однако предложенный формализм позволяет описать этот случай, если положить $\theta = \theta_1 \times \theta_2$ (считая, что все μ_t зависят только от $\vartheta_1 \in \theta_1$), а множество возможных функций потерь $\{R_{\vartheta_2}\}(\vartheta_2 \in \theta_2)$ задать в виде одной функции с параметром $\vartheta = (\vartheta_1, \vartheta_2)$:

Итак, управляемая n-шаговая система задана. Она и играет роль «системы координат», относительно которой будет измеряться неопределенность (порожденная неточностью знания значения параметра $\vartheta \in \theta$).

Желание сохранить соответствие вводимого понятия с интуитивным смыслом термина «неопределенность» заставляет наложить на функцию неопределенности $\varphi : M(\theta) \to R$: следующие ограничения:

1) $\forall \nu \in M(\theta)[\varphi(\nu) \geq 0]$ – неопределенность не может быть отрицательной ни для какого распределения ν параметра ϑ;

2) $\varphi(\nu) = 0$, если мера ν сосредоточена в одной точке $\vartheta \in \theta$ (неопределенность равна нулю, если значение ϑ известно точно);

3) для любого действительного α из отрезка.[0,1] должно быть выполнено $\varphi(\alpha\nu_1 + (1-\alpha)\nu_2) \geq \alpha\varphi(\nu_1) + (1-\alpha)\varphi(\nu_2)$ (функция неопределенности должна быть вогнутой) .

Первые два свойства не вызывают возражений, а третье обосновывается тем, что в случае его невыполнения для каких-нибудь двух распределений ν_1 и ν_2 удается построить эксперимент, такой, что после его проведения неопределенность становится больше, чем до эксперимента, а это вряд ли осмысленно.

Свойства 1–3 никак не связаны с функцией штрафа R. Ясно, что по отношению к последней функция неопределенности должна обладать следующим свойством: чем меньше значение функции неопределенности, тем меньше (при прочих равных условиях) должны быть средние потери.

Какой штраф связан с неточностью знания параметра ϑ?

Пусть $W(\nu)$ – нижняя грань средних значений функции штрафа R (нижняя грань берется по всем стратегиям управления) при условии, что известно только распределение ν параметра ϑ (но точное значение ϑ не известно). Другими словами, $W(\nu)$ – наименьшие потери в этих условиях.

Предположим, что каждый раз, когда производится случайный выбор значения параметра ϑ (в соответствии с распределением ν), это значение становится точно известным.

В этих условиях наименьший штраф (для каждого отдельного акта выбора ϑ и последующей процедуры управления) будет равен $q(\vartheta) = \inf \; R(\vartheta, u_1, ..., u_n)$ и остается усреднить эту величину по мере ν, чтобы вычислить наименьшие средние потери в условиях, когда значение) всегда точно известно (хотя и случайно выбрано): $Q(\nu_1) = \int\limits_{\theta} q(\vartheta) d\nu(\vartheta)$.

Разность $D(\nu) = W(\nu) - Q(\nu)$ (всегда неотрицательна из-за вогнутости операции взятия нижней грани) и представляет собой штраф, обусловленный незнанием точно значения ϑ.

Теперь четвертое условие, накладываемое на функцию неопределенности ϑ, можно сформулировать так:

4) $\varphi(\nu_1) < \varphi(\nu_2)$ тогда и только тогда, когда $D(\nu_1) < D(\nu_2)$.

Заметим, что функция $D : M(\theta) \to R\,O$ сама удовлетворяет свойствам 1– 4. Ее называют «естественной функцией неопределенности».

Функция D соответствует нашим представлениям о неопределенности. Но этого аргумента, конечно, недостаточно, чтобы быть уверенным в общности ее определения. Но исключено, что можно построить иную функцию неопределенности, удовлетворяющую свойствам 1 – 4. Поэтому важное значение имеет полученный учеными результат: ни каких других функций, удовлетворяющих свойствам 1 – 4 не существует. Более точно: какова бы ни была функция φ: $M(\theta) \to R$ удовлетворяющая свойствам 1 – 3, найдете: n-шаговая управляемая система, для которой (φ будет естественной функцией неопределенности (т.е. будет совпадать с функцией D для этой системы).

Так как шенноновская энтропия удовлетворяет свойствам 1 – 3, то она оказывается естественной функцией неопределенности некоторой вполне определенной управляемой системы (а именно, системы, целью которой является минимизация логарифма среднего числа попыток до успешного угадывания номера класса A_j, в который попал элемент m_l при распределении $p = \{p_1, ..., p_{n'}\}$ вероятностей попадания в классы. При этом, очевидно Θ есть $\langle 1,2,...,n' \rangle$; $U_t = \Theta$ для всех t; $Y_t = \{0,1\}$; $\mu(\vartheta, u_1, ..., u_t)$ – распределение, при котором вероятность единицы равна единице («сосредоточенное на 1»), если среди чисел $i = 1,2,...,t$ найдется хотя бы одно такое, что $u_i = \vartheta$, и распределение, «сосредоточенное на 0», если не так $(t = 1, ..., n-1)$; $R(\vartheta, u_1, ..., u_n) = \log s$, где s – наименьшее из чисел, таких, что $u_s = \vartheta$; роль ν играет распределение $p = \{p_1, ..., p_{n'}\}$).

Данная модель убедительно демонстрирует релятивность и взаимный переход друг в друга понятий «ценности» и «количества» информации. Однако существует мнение ученых, что не всегда можно согласиться с тем, что, что разделение количественных характеристик на «количество» и «ценность» нецелесообразно потому, что рассматривается одна функция (которая выступает в роли либо ценности, либо количества информации, но, разумеется, не одновременно в обеих ролях), хотя ясно, что наиболее содержательно одновременное рассмотрение двух (как минимум) функций, одна из которых связана с задачами обработки информации, а другая – с материальным выигрышем от использования информации.

Интерес к одновременному рассмотрению двух указанных функций обусловлен прежде всего тем, что затраты на обработку информации (включая передачу, хранение, вычисления и т.п.) в настоящее время становится сравнимыми с выгодами от использования информации. Поэтому необходимо «раздваивать» эти две стороны понятия информации. Такое «раздвоение» осуществляется в теориях третьей группы.

Все представленные подходы объединяет одна общая черта – стохастическое описание среды, пространства допустимых реакций и процесса выработки выходной реакции по входной информации.

Отсюда вытекает недостаток этих подходов – невозможность применения теории ценности информации к отдельным, рассматриваемым вне зависимости от ансамбля, состояниям системы и среды. Кроме того – необходимость выбирать конкретные значения вероятностей, что при недостатке надежных статистических данных приводит к ненадежности количественных результатов. Если учесть, что весьма сложная оптимизация, определяемая принципом экстремальности (а без него теория ценности информации некорректна), применяются в условиях весьма малой достоверности значений вероятностей, то станет ясно, что лишь в исключительно простейших случаях удается получить точную зависимость среднего штрафа от количества информации.

Контрольные вопросы

1. Перечислить причины, обусловливающие трудности разработки единого определения ценности информации.
2. В чем сущность апостериорного подхода в определении ценности и количества информации?
3. В чем сущность формализации понятий неопределенности и количества информации, отличающихся от подхода Шеннона?
4. Раскрыть содержание количественной меры целесообразности информации управления.
5. Провести сравнительную оценку мер количества информации по Харкевичу и Шеннону.
6. Какие меры неопределенности при оценке количества информации используются в теории распознавания образов?
7. В чем сущность общей меры накопления информации в адаптивных системах управления?

ГЛАВА 10. ОПРЕДЕЛЕНИЕ ГРАНИЦ ПРИМЕНИМОСТИ КЛАССИЧЕСКОЙ ТЕОРИИ ИНФОРМАЦИИ

Мысль о нелинейной динамике, как о некотором плодотворном методологическом субстрате подтверждается возникновением на ее основе одного из «побегов» – динамической теории информации, в которой ставятся и решаются вопросы о механизмах генерации и рецепции информации, оценке ее ценности и эволюции ценности, новизне, различии целей, математических моделях развития мозаичных систем. В этой теории стадии информационного процесса как бы нанизаны на единый «синергетический стержень».

Заметим, что методология выявления связи информации и самоорганизация до сих пор не проработана в методологических исследованиях. Причинами этого являются множественность разнообразных, непохожих друг на друга трактовок понятия «информация», порой не выходящих за пределы метафор; неопределенность интерпретации на уровне термодинамических аналогий.

До сих пор методологами не вполне осознано различие роли информации в кибернетических и самоорганизующихся системах: во-первых рассматривается процесс динамического равновесия между системой и ее окружением, что позволяет объяснить лишь сохранение достигнутого сконструированного человеком порядка (гомеостазис); во-вторых достигается порядок за счет внутренних детерминант сильно удаленной от состояния равновесия системы, для которой радикальное значение имеет влияние флуктуаций, «организующее» когерентное поведение элементов системы, многовариантность путей эволюции.

Концепция и понятийный базис традиционной теории информации создавался в условиях смешения языков формализма различных узкоспециальных отраслей знания. Это привело к ряду парадоксальных ситуаций, очевидных с позиций постнеклассических представлений, но вполне согласованных с традицией видения мира, жестко детерминированного, безальтернативного, описываемого линейными (или квазилинейными) закономерностями. В этом мире для сложного многостадийного процесса требовалось (и было достаточно) единственное определение. Ускользающая при этом сущность информационного процесса является результатом попытки сделать характеристику этапа процесса определением всего феномена.

Плюральность мира и многовариантность развития в постнеклассической науке находят свое отражение в полифундаментальности описания. С этих позиций в данной работе реализована обобщенная схема феномена информации, как целостного информационного процесса, содержащего отдельные стадии, которые невозможно описать на единственном языке. Это возникновение информации, ее рецепция, кодирование, запоминание, передача, декодирование, целенаправленное действие, воспроизведение. Этот методологический подход упорядочивает спектр определений феномена, сложившийся в конкретно-научных областях, каждое из которых раскрывает его значимые особенности, но не объемлет его во всей целостности. Таким образом удается снять противоречия частнонаучных определений.

Некоторые выводы могут показаться представителям гуманитарных наук тривиальными, ибо они уже используются на интуитивном уровне или априори. Вместе с тем тот факт, что естественные науки описали важные свойства динамики систем на лаконичном языке математики, выясняя отношения между элементами систем, представляется значимым и для гуманитарных наук. Исследование сложных систем, развитие которых сопровождается сменой доминирующих элементов (например, политических систем, научных школ, ареалов культуры), оптимально при обращении к неформализованным выводам динамической теории информации.

Представления ученых о слагаемых реального процесса эволюции логической информации позволили им установить рациональные границы для экстраполяций выводов сложившихся известных культурологических школ, в рамках которых был представлен не только широкий спектр трактовок сущности культуры, но и результаты исследования функционирования структурных элементов системы, а также было выявлено, что структуры, создаваемые культурой, не могут считаться естественными, так как генетически они не предопределены. Их нельзя считать искусственными, ибо они не являются запланированными продуктами интеллектуального творчества. Это делает сложной связь между развитием культуры и расширением знаний. Отсюда и противоречия между «двумя культурами», между естественным и гуманитарным знанием. Истоки противоречий в статическом взгляде на природу, характерном для классического образца рациональности, который не мог уживаться с фактом самопроизвольного создания структур культурой.

Обращение к постнеклассической картине мира позволяет увидеть множество неравновесных сложных открытых систем, самоорганизация которых демонстрирует роль хаоса как созидательного начала и конструктивного нелинейного многовариантного механизма эволюции. Выход из хаоса всегда связан с генерацией информации благодаря тому, что самоорганизующаяся система совершает выбор одного из представленных природой путей развития. Причем выбор этот случаен. Таким образом, эволюция любой самоорганизующейся системы – это, прежде всего, повышение ценности ее информации. Важность этих событий в эволюционном развитии очевидна. Тем не менее, вопросы об участии информации в эволюции самоорганизующихся систем при всей их актуальности ставились лишь в отдельных работах естественнонаучного профиля (Г. Николиса, Дж. Николиса, И.Р. Пригожина, Г. Хакена, Д.С. Чернавского).

В работе рассмотрены случаи, когда роль аттракторов процессов самоорганизации играют стационарные структуры. При этом за пределами рассмотрения остается не менее важное направление синергетических исследований научной школы А.А. Самарского и С.П. Курдюмова, существенно нестационарных, усложняющихся и деградирующих в режимах с обострением структур. Структуры при этом приобретают характер процессов, ибо связаны со становлением. Модели этих процессов методологически значительно сложнее, ибо содержат внутри себя возможности перехода на режим противоположного характера.

Эйфория пятидесятых-семидесятых годов в связи с представлением об информации как о некотором всеобщем свойстве материи, связанном с уровнем ее организации, т. е. свойстве, противоположном энтропии (негэнтропии), сменилась разочарованием и пессимистическим отношением к эвристичности информационного подхода.

С чем связана эта метаморфоза? Вероятно, не с умалением значения информации в жизни человека и окружающего его мира: мы постоянно убеждаемся в значимости информации в политической, экономической, социальной ситуациях; в трудах естествоиспытателей и философов в связи с обостренным интересом к событиям в открытых системах мы читаем, что это – системы, пронизываемые потоками вещества, энергии и информации; наконец, расшифровка в середине XX столетия генетического информационного кода всего живого сделала бесспорной определяющую роль информации в феномене жизни. Так что актуальность механизмов изучения информационных процессов не оставляет сомнений.

Учеными и специалистами в этой области были обнаружены существенные фактические и методологические неточности, мешающие развитию общей теории информации. Так, оказалось весьма важным различать макро- и микроинформацию. Последнюю легко сопоставить с энтропией, но невозможно запомнить. Это нарушает ключевое свойство всякой информации – ее фиксируемость. Методологическим следствием этого различия является нарушение связи между переданным системе количеством макроинформации, определяемом по формуле Шеннона [289], и ее энтропией.

Применение информационного подхода к экологическим, научным и социокультурным системам невозможно без методологического ключа, позволяющего определить правомерность информационно-энтропийных соотношений. В качестве такого ключа предлагается использовать термодинамический подход к оценке характера той или иной величины (мера вещества, функция состояния, количественная мера процесса). Естественно разъединяется термин «информация» с термином «количество информации», использованным в классической теории информации. Становится ясным, что эта количественная теория не может быть распространена за пределы теории связи, что она не имеет никакого отношения ни к экономике, ни к искусству, ни к науке.

Требуется работа методолога по собиранию всего сделанного в направлении создания теории информации, в которой будет уделяться внимание не только количественной, но и содержательной, смысловой, а также аксиологической стороне вопроса.

Задачей теории информации является выявление самых общих свойств любой информации независимо ни от ее семантики, ни от ее природы. Другой, скорее методологической, чем научной, задачей является различие между зачастую неадекватно используемыми характеристиками информации – качеством, ценностью. Согласно традиционной теории ценность информации зависит от той цели, к которой стремится получающий информацию объект. Считалось, что цель ставится извне. Это придавало информации антропный характер, а главное, исключало возможность самопроизвольного изменения ценности информации.

10.1. Случайный выбор, как источник генерации информации

В настоящее время содержание термина «информация» рассматривается в значительном многообразии и разнообразии известных формулировок и не принято его однозначное общее определение. Широкое распространение получило содержание информация, означающее сумму сведений, которые получает человек об окружающем мире, об интересующих его явлениях, о других людях, о самом себе. Жизнь личности, общества основаны на сообщении и получении информации.

В середине XX века возникла новая наука – кибернетика, изучавшая процессы управления и связи в искусственных устройствах и живых организмах. Речь идет, в том числе и об известных работах Винера [37,38]. Благодаря кибернетике, одним из основных понятий которой является информация, стало ясным огромное значение информационных процессов при функционировании всего живого: растений, простейших организмов, животных. Казалось, что между живой и неживой природой имеется существенное «кибернетическое» различие. Неживые системы самопроизвольно переходят к состоянию максимальной неупорядоченности, отвечающему максимальной энтропии (второе начало термодинамики). Живые системы, обладающие заданной функциональной организацией, напротив, переходят к состояниям биологической упорядоченности. Эту особенность живых систем справедливо объяснили на основе сочетания открытости и сильной неравновесности, создающих предпосылки для протекания информационных процессов. Оказалось, что любое биологическое упорядочение направляется информацией. Именно поэтому при развитии живого энтропия утрачивает характер жесткой альтернативы: в то время как одни системы вырождаются, другие достигают высокого уровня организации, их энтропия самопроизвольно понижается. Живые системы являются самоорганизующимися.

В семидесятых годах учеными было четко показано, что самоорганизация не является прерогативой живых систем. Природа изобилует примерами неживых открытых систем, способных в неравновесных условиях использовать энтропию как источник организации, переходить «от хаоса к порядку», т. е. в этом отношении вести себя подобно живым системам. В связи с этим оказался возможен объединяющий, а вовсе не взаимоисключающий подход к явлениям физики, химии, геологии, с одной стороны, и биологии – с другой. В рамках этого подхода, оказавшегося в настоящий момент в центре новейшей естественнонаучной (постнеклассической) картины мира, встает вопрос о роли информации в процессах самоорганизации (вообще всех, а не только живых) систем. Трудно предположить, что явления самоорганизации в живой природе направляются потоками информации, а в неживой, – каким-то иным способом.

Становится очевидным, что на фоне развития теории самоорганизации (синергетики) понимание информации как суммы сведений весьма ограничено. Одно из наиболее содержательных определений, позволяющих понять роль информации в самоорганизующихся системах, а также подойти к оценке количества информации и единице ее измерения дано Генри Кастлером: информация есть случайный запоминаемый выбор варианта из многих возможных и равноправных.

Строго говоря, эта часто цитируемая формулировка относится не к определению содержания понятия «информация», а способу ее создания – особенностям *генерации информации.* Согласно определению Кастлера выбор варианта должен быть:

1) случайным;

2) совершаемым из многих возможных и равноправных (но не обязательно равновероятных) вариантов;

3) запоминаемым.

Рассмотрим последовательно эти выборы.

1. Выбор должен быть случайным, если речь идет о возникновении новой информации. Если выбор подсказан на основе предшествующей информации, то речь может идти о ее восприятии, т. е. о рецепции.

Поскольку выбор экспериментальной базы классической науки исключает «задачи с развилками», т.е. в которых решение является детерминированным, то становится ясным, что генерация информации даже в столь простых и наглядных вариантах выносится «за скобки», как этой науки, так и классического детерминизма. В дальнейшем этот тезис будет развит вплоть до утверждения о том, что без введения дополнительных аксиом в виде концептуальных положений теории самоорганизации невозможно раскрыть смысл феномена информации.

Нетрудно видеть, что случайный выбор должна совершать самоорганизующаяся физико-химическая система, когда оказывается достигнутым порог ее устойчивости. При некотором значении варьируемого параметра состояния системы нарушается устойчивость термодинамической ветви, система оказывается в критической ситуации. Это отвечает точке бифуркации, после перехода через которую система может оказаться в одном из двух стационарных состояний. На рис. 10.1. приведена типичная бифуркационная кривая. По оси абсцисс – параметр состояния, по оси ординат – стационарные значения некоторой переменной *x*.

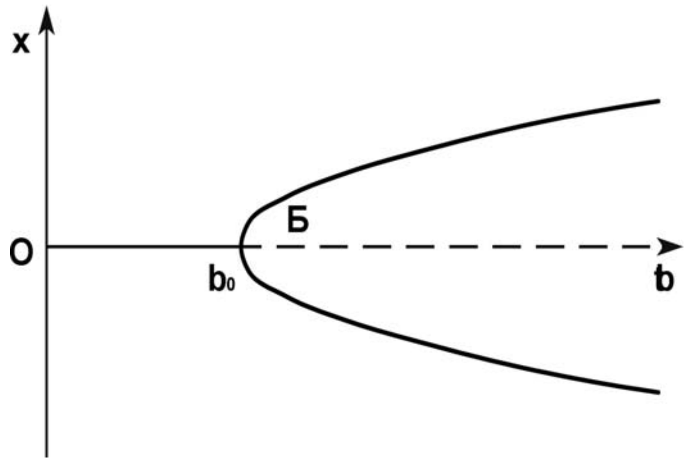


Рис. 10.1. Симметричная бифуркационная диаграмма

ОБ – термодинамическая ветвь, становящаяся неустойчивой правее точки Б. При $b > b_0$ существуют два стационарных состояния.

Существование этих состояний предоставляет системе выбор: она может отдать предпочтение одной из возможностей. Каким образом система совершает выбор? В нем присутствует элемент случайности: нелинейное макроскопическое уравнение, описывающее временную эволюцию системы, предсказывая многовариантность траекторий эволюции, не в состоянии предсказать, по какой из возможных (разрешенных законами природы) траекторий пойдет дальнейшее развитие. Не помогает и обращение к микроскопическому описанию. Перед нами случайные явления, аналогичные исходу бросания игральной кости. В процессе случайного выбора, который совершает система, генерируется информация.

Случайность выбора создает платформу для статистической оценки количества информации и выбора единицы ее измерения.

2. Рассмотрим выбор вариантов, принадлежащих одному множеству. Это и соответствует термину в определении Кастлера «равноправный». Мы уже отметили, что равноправность не означает равновероятность, но, тем не менее, вероятности выбора вариантов не очень далеки друг от друга.

3. Рассмотрим требование запоминаемости (рецепции) выбора. Если речь идет о человеке, сделавшем тот или иной выбор, то при фактической сложности процесса запоминания кажется более или менее ясным: человек обладает памятью, данной ему природой, он способен запоминать.

А если речь идет о «бифуркационном» выборе? Что значит в этом случае «запомнить»? На физическом языке это значит привести систему в определенное устойчивое состояние. При этом возникает упорядоченность (структура), олицетворяющая нарушение прежней пространственной симметрии и установление новой асимметричной и в то же время более упорядоченной формы координации частиц, из которых состоит система. В термодинамически устойчивой изолированной системе закон максимума энтропии создает ограничения в виде требования пространственной инвариантности распределения вещества. Никакая информация не может запомниться равновесной системой, ибо она имеет лишь одно единственное состояние. Запоминать может только система, находящаяся в далеко неравновесных условиях, способная формировать диссипативные структуры, описанные в серии статей и монографий одного из создателей синергетики (учения о самоорганизации на основе термодинамики неравновесных структур), а также в сотнях статей ученых-синергетиков.

Наличие порогов и четкая локализация в пространстве и времени являются очень эффективным путем осуществления регуляторных функций, которые обеспечивают запоминание. Создающиеся в результате выбора структуры – это результат информационного процесса. Запоминать могут и биологические молекулы, способные находиться в нескольких состояниях.

Таким образом, кастлеровское требование к информации «запомнить» реализуется в полном объеме только в рамках синергетики. Интуитивно ясное из обыденного опыта требование «запомнить» превращается во вполне научный термин, указывающий на определенные структурно-динамические изменения внутри системы, способной к самоорганизации.

Требование запоминания выбора реализуется в виде общего свойства любой информации, названного «фиксируемостью», т. е. с «записью» информации, не имеющей материальной природы, на материальном носителе. Это вовсе не означает, что информация присутствует всюду, что она как некий статический компонент (подобно гипотетическому эфиру) заполняет весь материальный мир. Представление об информации как о субстанции – это такая же фикция, как гипотеза теплорода. Представление об информации должно быть связано с понятием процесса. В самом деле, выбор – это процесс; рецепция или прием, запоминаемой информации – это процесс изменения состояния системы; передача информации по каналам связи, ее кодирование, декодирование, трансформация тоже, несомненно, являются процессами.

Только в специальных искусственных или естественных устройствах-ячейках может храниться законсервированная память о протекших информационных процессах; тогда можно говорить о «накопленной» информации. Ясно, что «накопить» процесс невозможно, но ведь с успехом можно накопить последствия процесса. Так, нельзя накопить совершенную механическую работу, которая тоже является процессом, но результат работы может быть накоплен, например в виде увеличения внутренней энергии адиабатно сжимаемого газа.

Во всех реальных запоминающих устройствах время запоминания ограничено, ибо ограниченное время сохраняется то состояние, в которое перешел рецептор под влиянием переданной ему информации. Важно, чтобы интервал времени сохранения памяти был не меньше, чем длительность того процесса, в котором данная информация может понадобиться.

Таким образом, благодаря синергетике появляется возможность возвести идею о значимости информационных процессов в окружающем нас мире в ранг способа видения этого мира. Вместе с тем необходимо отметить, что процесс генерации информации не имеет аналогов в традиционной науке. Это – событие, синтезирующее «случай» и «необходимость», поскольку случайный выбор в момент генерации возможен только из числа состояний, разрешенных природой.

Информация бывает условная и безусловная. Условная информация это прежде всего код, соответствие между условными символами и реальными предметами. Условность информации при выборе любого кода людьми (например, телеграфного кода, азбуки Морзе) очевидна. Условной является и информация, содержащаяся в алфавите и словарном запасе языка. Заметим, что в последнее время наше представление о языке необычайно расширилось, что легко объясняется: большой интерес к проблемам языка связан с важностью изучения процессов мышления. В связи с этим развиваются такие области, как математическая лингвистика, разделяющаяся на статистическую и структурную. Перед математиками постоянно стоит задача построения искусственных языков для взаимодействия человека с вычислительной машиной. Для специалистов-кибернетиков изучение языка – это одна из основных задач, поскольку структура управления – это структура языка системы.

Как полагают ряд ученых, мы оставаясь на позициях кибернетики, рассматриваем язык как некоторый организм, который возникнув под влиянием определенных, может быть, не всегда понятых нами, в том числе и в деталях, причин, который продолжает самостоятельно эволюционно развиваться, оказывая часто решающее влияние на иерархически вышестоящие системы, такие как мышление человека.

Современная наука считает, что возникновение речи т. е. замена звуковых сигналов словами, положило начало человеческой истории. Связано ли появление языка с генерацией условной информации? Для положительного ответа на этот вопрос необходимо выяснить роль случайности при выборе того или иного способа общения с помощью языка. О том, что случайность здесь играет существенную роль, свидетельствует факт множественности языков (специалисты насчитывают около 3000 различных разговорных языков, на которых говорят народы Земли). Но даже если бы к настоящему моменту все языки слились в некий единый международный язык, мы не могли бы исключить влияния случая на его формирование, ибо это слияние могло бы быть результатом эволюции языковых систем, а не насаждением единого языка извне.

Специалисты считают, что логическая кодовая условная информация есть результат общественной деятельности, ибо кодовая информация может быть полезной только в том случае, если ею владеют несколько человек. С помощью языка в свое время возник новый вид информации – так называемая логическая информация. Однако кодовой является и условная генетическая информация, генерация которой произошла задолго до появления существ, способных к общественной деятельности. Вместе с тем удается свести генетику к формальному описанию явлений в терминах языка, причем языка весьма жесткого и закрытого, так как в его словаре не происходит изменений, ибо любые изменения словаря приводят к летальному исходу для носителя информации. Это язык с застывшим словарем. Замечательно, что, в отличие от обычных языков, язык генетического кода является единым, его структура действительна как для человека, так и для растений.

Безусловной является информация:

– о реально происходящих событиях, т. е. смысловая, не возникающая случайно, а рецептирующаяся из окружающей действительности;

– генерирующаяся в процессах самоорганизации.

По мнению ученых, каждое сообщение содержит как условную, так и безусловную информацию, и разделить их иногда непросто. Так, например, математический формализм, унифицированный на нижних уровнях иерархической лестницы, кажется безусловной информацией. Однако следует помнить, что унификация математического аппарата произошла не сразу. Генерировалась условная информация рядом ученых-математиков. Унификация произошла в результате эволюции, приведшей к отбору наиболее приемлемого варианта.

В принципе математика – аксиоматизированная область знаний, что делает ее единой наукой, имеющей свою особенную логическую структуру. Идеал языка такой науки – это система правил оперирования со знаками. Чтобы задать «исчисление», необходимо составить алфавит первичных элементов-знаков, задать начальные слова исчисления, построить правила получения новых слов из начальных слов.

Таким образом, математическая мысль и система кодов неразделимы. Символы имеют для математика принципиальное значение: они – предмет изучения математики. Так, методологические установки Гильберта по Клини состоят в том, что символы сами по себе являются окончательными предметами и не должны использоваться для обозначения чего-либо, отличного от них самих, математик смотрит на них, а не через них, и не на то, что за ними; таким образом, они являются предметами без интерпретации или значения.

Математическое знание содержится в кратких высказываниях – математических структурах. Возникает вопрос: а содержат ли (и если да, то какую) информацию доказательства теорем? Очевидно, да, содержат безусловную информацию, поскольку процесс доказательства не является строго формализованным. Эта информация является дочерней по отношению к базовой и рецептируется благодаря определенным генерируемым математическим идеям, что иллюстрируется, например, следующим высказыванием великого математика Карла Гаусса по поводу одного из его открытий: «Наконец, два дня назад я добился успеха, но не благодаря моим величайшим усилиям, а благодаря Богу. Как при вспышке молнии проблема внезапно оказалась решенной. Не могу сказать сам, какова природа путеводной нити, которая соединила то, что я уже знал, с тем, что принесло мне успех». Очевидно, речь идет о случайном выборе на уровне бессознательного, т. е. о генерации информации.

Мы затронули вопрос о «дочерней» информации, вопрос, связанный с ее качеством. В следующем разделе рассмотрим «качество» наряду с такими характеристиками информации, как ее «количество» и «ценность».

10.2. Современные взгляды на характеристики информации

Качество информации. В работах ученых и специалистов, приведенных в библиографии, термин «качество» употребляется в смысле «ценность», что требует уточнения в используемой терминологии искомой области. Определение содержания понятия качества информации связано с существованием иерархической структуры информации. Эта характеристика иллюстрируется продолжением выбора пути при разветвлениях. Это разветвление имеется только на уже выбранном ранее направлении, т. е. этот новый выбор совершается на основе прежнего, причем прежний не предрешает последующего. Поэтому, несмотря на связь выборов, необходимо различать их уровни, т. е. уровни генерируемой информации.

Примером последовательных выборов системой пути своего развития является бифуркационная кривая (рис. 10.2).

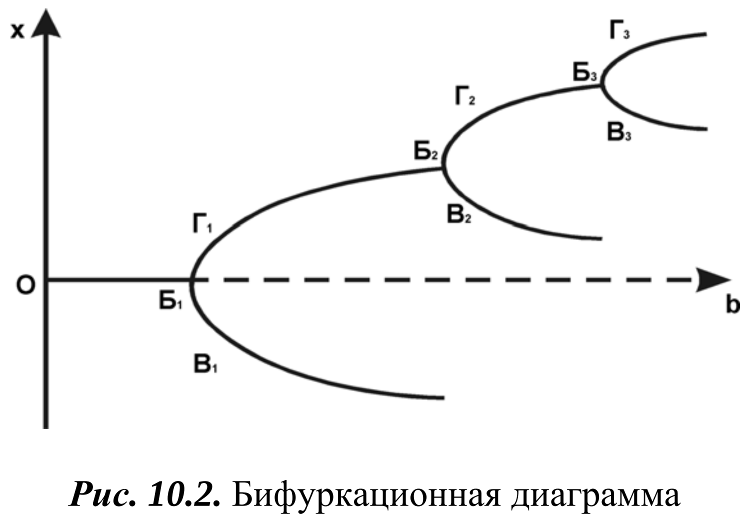


Рис. 10.2. Бифуркационная диаграмма

По оси абсцисс – параметр состояния системы, по оси ординат – значения переменной *x*, описываемой нелинейным уравнением – моделью самоорганизующейся системы

В точке $Б_1$ система совершает выбор между ветвями $Б_1В_1$ и $Б_1Г_1$ если выбор сделан в пользу $Б_1В_1$, то в точке $Б_2$ система снова оказывается перед выбором: ветвь $Б_2В_2$ или ветвь $Б_2Г_2$. Очевидно, только предыдущий выбор ветви $Б_1В_1$ делает актуальным этот второй, затем третий и т. д. Выборы, как и в предыдущем примере. Ни один из последующих выборов не предрешается предыдущим. Отсюда – необходимость различать уровни генерируемой информации, каждый из которых отвечает определенному *качеству*.

Эта разнокачественность обеспечивает иерархическую структуру. Информации на уровнях не только качественно различаются, образуя ряд подмножеств, но и связаны друг с другом, ведь для информации высокого уровня совершенно необходима более «древняя» информация нижних уровней. Существует специальный термин «тезаурус», который означает информацию более нижнего уровня, которая необходима для генерации или рецепции информации на верхнем уровне, качественно отличном от нижнего. Выбор варианта делается не между вариантами разных уровней, а всегда только на одном уровне. Это заложено в определении того, что есть генерация информации: выбор одного варианта, среди нескольких возможных и равноправных. «Равноправные» варианты – это варианты одного *качества,* т. е. одного иерархического уровня.

1) количество информации

Как было отмечено раньше, формула, позволяющая определить количество информации, была получена Шенноном. Возможность обобщения этой формулы на все виды языков, передачу и использование любых видов сигналов, которые служат для передачи информации, наиболее полно проанализирована В.И. Корогодиным. Целесообразно изложить его точку зрения потому, что многократные попытки приложить теорию Шеннона к самым различным областям человеческой деятельности встречаются с явными трудностями. Остановимся на двух из них.

Во-первых, в процессе упомянутых попыток определенные элементы сообщений совершенно неправомерно отождествлялись с информацией, как таковой. Эти элементы связаны лишь закономерностями передачи по каналам связи различных сигналов, не связанных с семантикой сообщения. Если сохранить все буквы осмысленного сообщения, переставив их, то осмысленность исчезнет, а количество информации, определяемое формулой Шеннона, останется прежним.

Получается бессмысленная информация, т. е. вообще не информация. Получившееся в результате перестановок букв бессодержательное нечто не может называться информацией. В работе Корогодина предлагается назвать частотную характеристику элементов сообщения, определяемую правой частью формулы $I(X) = -\sum_{i=1}^{N} p_i \log_2 p_i$, *информационной тарой*, которая может содержать полезную информацию, а может быть пустой. Очевидно, что емкость информационной тары имеет вполне определенное числовое значение, не зависящее от степени заполнения тары.

Во-вторых, остается неясным вопрос о том, как сопоставить количество информации данного сообщения при переходе к другому языку. Другими словами, не известно, как в общем случае изменяется количество информации при переводе с одной системы записи на другую. Если *H* – емкость информационной тары и если текст сообщения абсолютно компактен, то количество информации сообщения *(I)* равно емкости информационной тары. И только в этом случае можно пользоваться для определения *I* формулой $I(X) = -\sum_{i=1}^{N} p_i \log_2 p_i$. В общем случае $I < H$.

Стремление некоторых авторов использовать эту формулу для определения количества информации, связанной с любыми событиями основано на допущении о том, что информация, будучи общим свойством материи, содержится во всех объектах и событиях природы. Эта неверная посылка и абсурдные следствия из нее свидетельствуют, как известно из литературы о том, что вероятностное определение количества информации, перенесенное из математической теории связи в обобщенную теорию информации, утрачивает свою эвристичность.

2) ценность информации и ее эффективность

Математическая теория информации полностью игнорирует содержание информации. Поэтому вопрос о ее ценности не ставится. Рассчитывая пропускную способность канала связи, бессмысленно принимать во внимание содержание телеграмм.

Вопрос о ценности возникает, прежде всего, в биологии. Биологическая эволюция необратима и направлена. Исходный материал для эволюции – случайные мутации генов – не имеют заданной направленности, тем не менее, работает мощный направляющий фактор – естественный отбор, основанный на повышении *ценности* информации, трансформированной в итоге мутации. Таким образом, для биологии существенна не столько количественная, сколько ценностная характеристика информации. Информация может быть более или менее ценной в зависимости, от преследуемой цели, происхождение которой до недавнего времени в теории информации не обсуждалось. Ценной информацией считается та, которая помогает достижению цели.

Следует обратить внимание на следующее различие оценок «количество» и «ценность». В отличие от шенноновского определения количества информации, передаваемой по каналам связи, ценность проявляется в результатах рецепции. Она непосредственно связана с рецепцией. Ю.А. Шрейдеру принадлежит следующий наглядный пример: «Имеется том 2 «Курса высшей математики» В.И. Смирнова. Эта книга содержит богатую информацию. Какова ее ценность? В ответ приходится спросить – для кого? Для школьника информация этой книги нулевая, так как он не обладает достаточной подготовкой, достаточным уровнем рецепции и не в состоянии эту информацию воспринять. Для профессора математики ценность тоже нулевая, так как он все это хорошо знает. Максимальной ценностью эта книга обладает для студентов того курса, которым книга предназначена, поскольку речь идет об очень хорошем учебнике. Зависимость ценности от уровня подготовки, от предшествующего запаса информации – тезауруса… – проходит через максимум» (рис. 10.3).

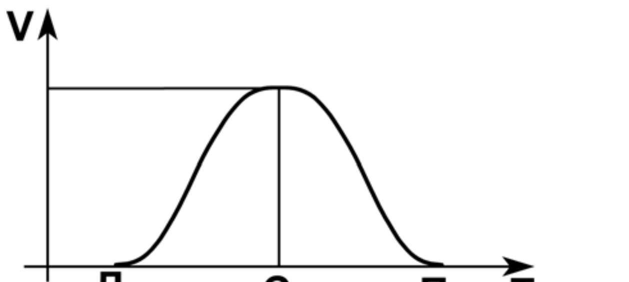


Рис. 10.3. Зависимость ценности информации от тезауруса:
Д – дошкольник, С – студент, П – профессор

Известны несколько способов количественного определения ценности. Все они основаны на представлении о цели, достижению которой способствует полученная рецептором информация. Чем в большей мере информация помогает достижению цели, тем более ценной она считается.

1. Если цель наверняка достижима, и притом несколькими путями, то возможно определение ценности (V) по уменьшению материальных, или временных затрат благодаря использованию информации. Так, например, сочетание хороших предметного и алфавитного каталогов библиотеки, наличие библиографических справочников сокращают время на составление списка литературы по конкретному интересующему читателя вопросу.

2. Если достижение цели не обязательно, но вероятно, то используется один из следующих критериев:

а) мерой ценности, предложенной М.М. Бонгартом и А.А. Харкевичем, является величина

$$V = \log_2 \frac{P}{p},$$

где p – вероятность достижении цели до получения информации, а P – после; учитывая, что p и P могут изменяться от 0 до 1, заключим, что пределы изменения V – от $-\infty$ до $+\infty$;

б) мерой ценности, предложенной В.И. Корогодиным, является величина

$$V = \frac{P - p}{1 - p};$$

при этом V изменяется от 0 до 1.

Очевидно, величину V для некоторой информации невозможно задать одним единственным числом. Определенное значение ценности можно получить только лишь для известной пары источник-рецептор (например, учебник – студент).

Ценность информации, получаемой рецептором, зависит от ее количества. Так, если целью изучения учебника является овладение методом решения определенного цикла задач, то прочтение двух-трех параграфов в лучшем случае дает возможность решать лишь малую долю задач. С ростом количества информации увеличивается величина V.

В области малых значений I скорость увеличения V мала в силу изложенных выше соображений, но и в области больших I темп роста уменьшается, поскольку, начиная с некоторых значений V, дальнейший рост этой величины уже не влияет на успех решения задач из цикла: любая задача может быть решена. Это значит, что кривая $V=f(I)$ для данной пары источник – рецептор имеет вид кривой с насыщением (рис.10.4).

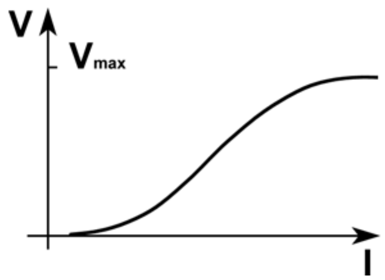


Рис. 10.4. Зависимость ценности информации от ее количества

По ходу кривой изменяется приращение ценности, приходящейся на единичный интервал I. Это позволяет ввести еще одну характеристику – эффективность информации ε.

В работах Корогодина, за эффективность принято отношение $\varepsilon = V / I$, что нельзя признать удачным. Предлагается, для ε следующую величину:

$$\varepsilon = \frac{dV}{dI}.$$

Другими словами, эффективность информации равна первой производной функции $V = V(I)$ для каждого заданного значения I.

Зависимость $\varepsilon = f(I)$ представляет собой кривую с максимумом, причем только с одним (рис. 10.5). Площадь под кривой $\varepsilon = f(I)$ и осью абсцисс называют «информационным полем».

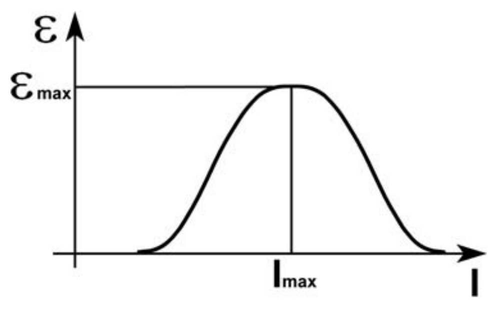


Рис. 10.5. Зависимость эффективности информации (ε) от ее количества (I):
I_{max} – оптимальное количество информации,
ε_{max} – максимальное значение эффективности

10.3. Синергетический подход к классификации информации

В методологической и научной литературе неоднократно обсуждалась возможность сопоставлять количество информации (*I*) и величину термодинамической энтропии (*S*) системы. Анализ известных работ, указанных в библиографии книги, позволил выявить необходимость методологического уточнения корректности вышеупомянутого сопоставления в различных ситуациях.

Людвиг Больцман был первым ученым, обратившим внимание на связь величин *I* и *S*.

Говоря о необратимом увеличении энтропии идеального газа при его изотермическом расширении в пустоту (за 50 лет до формулы Шеннона), он заметил, что этот процесс сопровождается потерей информации о местонахождении (в общем случае о состояниях) молекул. Эта идея получила дальнейшее развитие после работ по математической теории связи, выполненных сотрудниками фирмы Be1l Te1ephon. Оказалось возможным использовать формулу $I(X) = -\sum_{i=1}^{N} p_i \log_2 p_i$ для подсчета количества информации о состояниях молекул, если принять, что P_i есть вероятность нахождения молекулы в ее *i*-м состоянии. Если учесть, что по Больцману энтропия, приходящаяся на одну молекулу, есть

$$S = -k\sum_{i=1}^{M} P_i \ln P_i, \tag{10.1}$$

где *k* – постоянная Больцмана, *M* – число состоянии молекул, то эквивалентность между увеличением термодинамической энтропии, происходящем при совершении некоторого процесса, и потерей информации о молекулах кажется очевидной.

Изменение энтропии однозначно связано с количеством информации:

$$I_{МИКР} = \frac{1{,}44}{k}(S_0 - S_1), \tag{10.2}$$

или

$$\frac{k}{1{,}44} I_{МИКР} = S_0 - S_1,$$

где $k = 1{,}38*10^{-23}$ *Дж/К*; $1{,}44 = \log_2 e$, а $S_0 > S_1$.

Мы применили обозначение $I_{МИКР}$ ибо (10.2) получено на основе представлений о микросостояниях. Это количество *микроинформации*. Один бит микроинформации может быть условно выражен в энтропийных единицах умножением на (1,38/1,44) · 10^{-23} Дж/К. Тогда согласно формуле (10.2)

$$I_{МИКР} = S_0 - S_1, \tag{10.3}$$

где $I_{МИКР}$ выражено в энтропийных единицах.

Смысл связи I и S может быть продемонстрирован на следующем примере. Допустим, замораживается сосуд с водой. При этом убывает энтропия воды и возрастает информация о месторасположении ее молекул. Последнее связано с тем, что переход от жидкого состояния к кристаллическому сопровождается фиксацией молекул воды в узлах кристаллической решетки, что отвечает известному упорядочению, т. е. уменьшению энтропии. В этом случае речь идет о микроинформации. И.Р. Пригожин, определяя меру информации при продвижении на один шаг вдоль марковской цепи, пишет о самой непосредственной связи этой величины с энтропией. Однако в рассматриваемом в случае речь идет также о микроинформации.

Микроинформация имеет мало общего с макроинформацией, во-первых, потому что ее создание не связано со случайным выбором; во-вторых, система молекул тела (при заданных условиях) имеет только одно устойчивое состояние (аттрактор) – это равновесное состояние в то время, как система, рецептирующая кастлеровскую информацию должна иметь минимум два стационарных состояния; в-третьих (и это очень важно), эта информация не может быть заполнена. Так, информацией можно считать набор координат и скоростей молекул газа в данный момент времени. Запоминать это бессмысленно, поскольку через короткое время система о нем «забудет». Что же касается количества макроинформации, то в приведенном примере оно равно нулю, даже если все координаты и скорости молекул известны.

Д.С. Чернавский отмечает, что Микроинформация не обязательно связана с микрочастицами. Любая незапоминаемая информация – это Микроинформация.

В отличие от микроинформации, информация о микросостояниях «оплачивается» энтропией в неэквивалентной мере: возрастание энтропии во много раз превышает количество полученной информации. Это хорошо видно на простом примере. Бросанию монеты должно отвечать изменение энтропии, примерно равное 10^{-23} Дж/К, но выделение энтропии при работе мышц (сила бросания), теплоты при ударе об пол монеты несоизмеримо больше. Заметим, что по I это изменение вообще нельзя определить, ибо оно зависит от массы монеты, ее упругости, степени шероховатости пола и т. д.

Рассматривая вопрос о математическом виде связи энтропии и информации, Л. Бриллюэн ограничился случаем микроинформации, получив формулу (8.3). Именно поэтому этот феномен называют «информация в смысле Бриллюэна», в отличие от рассмотренной выше макроинформации по Кастлеру.

Очевидной ошибкой Бриллюэна была неправомерная экстраполяция (10.3) на случай макроинформации, т. е. запись в виде

$$I = S_0 - S_1. \qquad (10.4)$$

Отвечая положительно на вопрос об однозначности связи между информацией и энтропией, Бриллюэн комментирует это следующим образом: дополнительные знания, т. е. информация, уменьшают энтропию рецепиента, значит, формула (10.3) имеет универсальный характер. Пример Бриллюэна свидетельствует о том, что не верный, но и ошибочный научный вывод может служить движению вперед, ибо он охватил более широкий круг явлений, нежели тот, к которому первоначально относилась задача. Следствия неверной формулы (10.4) вылились в методологическую некорректность, что привело к сильнейшей путанице понятий, связанной с пресловутым негэнтропийным принципом.

Далее рассмотрим предпосылки генерации информации, ее характеристики и свойства. Все эти эпитеты, а также представления об эволюции информационных систем, имеют смысл по отношению к макроинформации. Поэтому в дальнейшем будет использоваться термин «информация», в понимании под ним только макроинформации.

10.4. Методологический анализ рецепции информации

Подробное рассмотрение наиболее распространенных противоречий и ошибок в методологии изучения явления информации имеет под собой попытку вскрыть корни заблуждений, так как это является и главной причиной отставания в развитии общей теории информации, для которой уже созданы необходимые предпосылки. Подмена одного не очень ясного термина «информация» другим, еще более непонятным, таким, как «негэнтропия» или «отражение», создает только видимость прогресса. Что же касается печально знаменитого негэнтропийного принципа (т. е. утверждения о том, что количество информации служит определенным отрицательным вкладом в энтропию), то, как нетрудно показать, он оказался скомпрометирован своими же собственными следствиями. Но если не учитывать этих следствий (не думать о них) и не знать, в чем именно кроется то заблуждение, которое привело к неверному результату, то начинает казаться, что не так уж порочен и сам результат. А если учесть, что независимо друг от друга авторами неточностей были не кто-нибудь, а Винер и Бриллюэн, авторитеты которых были заслуженно очень велики, то нетрудно предвидеть живучесть заблуждения. В [162] показана некоторая некорректность выводов Бриллюэна на случай запоминаемой информации. Получено, что информация дает отрицательный вклад в энтропию. Отсюда – легко прижившийся термин «негэнтропия» (негэнтропия равна количеству микроинформации со знаком минус и не имеет отношения к запоминаемой информации). Однако дело не в названии, а в тех «захватывающих возможностях», которые открывает негэнтропийный принцип. Получается, что нельзя изменить энтропию тела не создав или не уничтожив некоторое количество информации.

Одной из основных предпосылок негэнтропийного принципа является утверждение зависимости термодинамической энтропии от знания. Экспериментальное изучение какой-либо системы есть извлечение из нее информации, что согласно (10.3) должно сопровождаться возрастанием энтропии изучаемого объекта и ее уменьшением в голове экспериментатора, получившего порцию негэнтропии. Таким образом, с помощью информации удается регулировать энтропию исследователей. Поскольку энтропия присуща всем термодинамическим системам, то и негэнтропия (информация), как дополняющая ее характеристика, должна быть приписана всей материи. Отсюда следуют такие некорректные, но до сих пор еще довольно популярные методологические положения.

1. Информация содержится в каждом материальном объекте, она вездесуща и, следовательно, является одним из свойств материи.

2. Существуют две характеристики степени порядка материальных объектов: неупорядоченность (энтропия) и упорядоченность (негэнтропия, равная информации).

Первое утверждение противоречит представлению об информации как о процессе, ибо процесс, да еще связанный со случайным выбором, может протекать, а может и отсутствовать. Генерация информации – это событие, которое не является детерминированным и уж конечно не представляет собой характеристику любого материального объекта.

Второе следствие является методологическим абсурдом, что иллюстрирует принципиальное неблагополучие с породившим его негэнтропийным принципом. Получается, что одна и та же сущность – степень упорядоченности – потребовала для своего описания двух характеристик: энтропии (степень беспорядка) и негэнтропии (степень порядка).

Количественной мерой степени упорядоченности служит энтропия. Разные системы упорядочены в большей или меньшей степени. Если в «большей» – энтропия относительно мала, если в «меньшей» – относительно велика. Поэтому энтропия является одновременно характеристикой упорядоченности и неупорядоченности. Ведение негэнтропии, как величины, *обратной по знаку энтропии*, лишено физического смысла.

Заметим, что начало этой эквилибристики терминами и знаками положено Н. Винером: «Как количество информации в системе есть мера организованности системы, – говорит он, – точно так же энтропия системы есть мера дезорганизованности системы, одно равно другому, взятому с обратным знаком». Это определение содержит противоречие внутри самого себя. По Винеру количество информации I равно энтропии S со знаком минус

$$I = -S. \tag{10.5}$$

Но, как уже отмечалось выше, S – функция состояния, существенно положительная ($S>0$), следовательно, отрицательным является количество информации, что противоречит, как здравому смыслу, так и формуле Шеннона, из которой со всей определенностью следует $I > 0$. О недоразумении со знаками и размерностями в выражениях количества информации писал У. Эшби, но его замечание осталось незамеченным.

Формальное сходство между статистически определяемыми по формулам Шеннона и Больцмана величинами I и S, да еще вместе со случайной ошибкой в знаке, послужило поводом к далеко идущим обобщениям и к множеству некорректных расчетов количества информации в различных объектах.

По соотношению (10.5) информация является функцией состояния, что резко противоречит высказываниям самого Винера: «Процесс получения и использования информации является процессом нашего приспособления к случайностям внешней среды и нашей жизнедеятельности в этой среде. Потребности и сложность современной жизни предъявляют гораздо большие, чем когда-либо раньше, требования к этому процессу информации, и наша пресса, наши музеи, научные лаборатории... должны удовлетворить потребности этого процесса». В этой характеристике понятие информации связано с представлением о процессе, а вовсе не с функцией состояния – негэнтропией.

Описанная парадоксальная ситуация лишний раз свидетельствует о том, что даже крупные ученые не сговариваясь, могут допускать аналогичные ошибки, корень которых уходит в непроработанность методологии того или иного вопроса.

Негэнтропийный принцип уже давно подвергнут заслуженной критике, однако в семидесятых годах он был в известной степени реставрирован. Дело в том, что с развитием термодинамики неравновесных систем появилась возможность поставить вопрос о том, что возникновение информации (запомненного выбора из набора возможных состояний) сопровождается образованием структуры, а значит, локальным понижением энтропии системы. Следовательно, снова возникает (теперь уже в связи с макроинформацией) вопрос о связи I и S. Какова эта связь? Точки зрения различных авторов на этот вопрос расходятся. Рассмотрим те из них, которые являются полярными.

Так, Е.А. Седов, сторонник информационно-энтропийного подхода к описанию процессов самоорганизации, утверждал: «Методы теории информации, разработанные К. Шенноном для чисто прикладных задач техники связи, оказываются универсальным средством анализа процессов самоорганизации как простейших физических тел, так и сложнейших интеллектуальных и социальных систем». И далее: «Мера, найденная Шенноном, оказалась единой универсальной мерой упорядоченности для всех существующих в мире систем».

Полностью отрицает связь информации с энтропией В.И. Корогодин, который, детально проанализировав информационные процессы в живых системах, приходит к безапелляционному выводу о некорректности использования формулы Шеннона в случаях, выходящих за пределы задач теории и практики связи. Отсюда и вывод о невозможности сопоставления информации и энтропии. «Зачем отождествлять информацию с энтропией и выражать ее количество в термодинамических единицах эрг*град$^{-}$? Получается как в том анекдоте, когда семинарист, окрестив поросенка «карасем», спокойно слопал его в Великий пост... Не лучше ли за всеми этими феноменами сохранить присущие наименования и организацию продолжать именовать организацией, энтропию – энтропией, а информацию – информацией, выявив сущность этого понятия, а, не производя подмену одних терминов другими?».

Последний упрек можно отнести к Г. Хакену, который в своей монографии проводит анализ изменения информации Шеннона в самоорганизующейся физико-химической системе. Как известно, формула Шеннона позволяет измерить количество информации, передаваемое в единицу времени при использовании заданного набора сигналов. Поэтому перенесение метода Шеннона на решение информационных проблем самоорганизации в самых разных системах требует специальных разъяснении. Хакен исследует изменение информации Шеннона при неравновесном фазовом переходе в состояние с большей степенью упорядоченности. При таком неравновесном фазовом переходе система приобретает способность хранить информацию о том выборе (или отборе), который привел ее к переходу «от хаоса к порядку». Это дает основание автору трактовать энтропию как информацию, что в рамках развиваемого им подхода означает возможность термин «информация» предпочесть термину «энтропия». Что это дает?

По-видимому, не так уж много. В самом деле, вблизи порога, т.е. в области точки бифуркации, из-за критических флуктуаций информация сильно возрастает, поэтому автор возвращается к термину «энтропия», более уместному, по его мнению, в этой области. Однако при использовании энтропии Шеннона в качестве меры упорядоченности возникает трудность: в часто обсуждаемом Хакеном переходе в лазере через порог лазерного излучения энтропия, посчитанная по Шеннону, оказывается больше, чем в исходном «равновесном» состоянии. В этой ситуации трактовка энтропии Шеннона, как информации, снова оказывается предпочтительней, так, как состояние генерации оптической среды не может быть более хаотичным, чем равновесное. Таким образом, анализ включения информационных процессов в самоорганизацию подменяется подбором удобных для каждого конкретного случая терминов. Все это снижает методологическую ценность информационно-энтропийного подхода к самоорганизующимся системам.

Рассмотрим наиболее прогрессивный взгляд на вопрос о связи информации и энтропии. При генерации информации в самоорганизующейся системе энтропия уменьшается. Однако это не дает основания говорить о связи информации и энтропии по следующей причине. I и S это величины, разные в термодинамическом отношении. Информация – это величина, которая характеризует процесс, ибо связана с выбором, запоминанием, кодированием, передачей и т. д. Подобные величины, характеризующие меру протекания процесса, известны в термодинамике. Это, например, тепло и работа – величины, характеризующие процессы. Они отличаются от функций состояния таких, как внутренняя энергия, энтропия, свободная энергия тем, что обладают полными дифференциалами. В связи с этим первое начало термодинамики можно прочитать как утверждение о том, что приращение внутренней энергии (dU) равно сумме количества подведенного тепла (Q) и совершенной внутренними силами работе (A'):

$$dU = \delta Q + \delta A'. \qquad (10.6)$$

Речь не идет о соотношении U и Q или U и A'. Эти величины несопоставимы, сравнить можно только dU с δQ и $\delta A'$.

К сожалению, часто приходится встречаться с непродуманностью методологических подходов к термодинамическим величинам. Так, в процитированной монографии Г. Хакен пишет: «Первое начало утверждает, что в замкнутой системе энергия сохраняется, причем энергия может принимать различные формы, такие, как внутренняя энергия, совершенная работа или тепло». Это неверная формулировка, ибо работа и тепло – не формы энергии, а способы ее изменения.

Можно накопить энергию, но невозможно накопить работу. Речь может идти не о переходе энергии в работу или тепло, а об *изменении* энергии системы, за счет совершения работы внешними силами или подвода к ней тепла. Это существенно, ибо в принципе исключает тождество Q и U. Только изменение величины U может быть сопоставлено с количеством подведенного тепла.

Точно так же обстоит дело с соотношением количества переданной информации и функцией состояния S. Энтропия системы изменяется при рецепции ею информации. В работах Винера и Шеннона процессам рецепции практически не уделялось внимания. От рецептора требовалось лишь умение отличать один кодовый символ от другого.

Самые общие соображения о процессе рецепции позволяют утверждать, что рецепция – процесс необратимый, так как информация не может самопроизвольно возвращаться вспять. Рецепция – процесс неравновесный, ибо потоки информации между источником и рецептором неуравновешенны. После того, как были сформулированы основные положения синергетики, стало ясно, что поскольку рецепция информации означает возникновение определенной упорядоченности в воспринимающей системе, то это процесс, далекий от равновесия. Другими словами, рецепторная система есть система диссипативная, переходящая под влиянием информационного потока в состояние с уменьшенной энтропией.

Таким образом, рецепция информации есть необратимый, неравновесный процесс перехода системы из менее устойчивого состояния в более устойчивое. Этот процесс сопровождается уменьшением энтропии рецепторной системы.

Все сказанное позволяет заключить, что:

во-первых, количество информации – это величина, характеризующая процесс ее рецепции, не являющаяся функцией состояния;

во-вторых, количество информации, изменяя энтропию рецептирующей системы, связано не с функцией состояния системы, не с энтропией, а с величиной ее изменения.

Если убыль энтропии обозначить ΔS, то между количеством информации и убылью энтропии существует связь

$$I \sim -\Delta S.$$

Важно понимать, что ΔS вовсе не энтропия системы, а ее убыль.

Переход от знака пропорциональности к знаку равенства возможен при введении коэффициента пропорциональности, который, в отличие от случая микроинформации, *не является постоянным, а* зависит от конкретной ситуации, и даже для одной и той же системы, получающей разное количество информации, может зависеть от этой величины. Другими словами, связь между I и S не только нелинейная, но и неопределенная. В некоторых случаях этот коэффициент очень велик, т. е. последствия получения информации несоизмеримо велики по сравнению с ее количеством.

Так, весь поток транспорта меняет направление при получении водителями одного бита информации – замены красного сигнала светофора на зеленый. Это – так называемая триггерная ситуация, постоянно реализующаяся в биологических системах.

Таким образом, нет оснований говорить, об определенной количественной связи между информацией системы и изменением ее энтропии. Кажущаяся реставрация синергетикой негэнтропийного принципа бросает тень на это научное направление, вызывая настороженное отношение к ее возможностям играть роль общенаучной концепции.

Контрольные вопросы

1. Каковы причины недостаточной проработанности методологии выявления связи информации и самоорганизации?
2. В чем состоит методологическое различие роли информации в кибернетических и самоорганизующихся системах?
3. Каковы особенности создания концепции и понятийного базиса традиционной теории информации?
4. Какова роль плюральности мира и многовариантности развития в постнеклассической науке исследований феномена информации?
5. В чем сущность применения информационного подхода к экологическим, научным и социокультурным системам?
6. В чем различия в содержании терминов «информация» и «количество информации», используемых в классической теории информации?
7. В чем состоит задача теории информации в информационном обществе?
8. Какова роль работ Винера в развитии теории информации?
9. В чем сущность концепции оценки количества информации и единиц ее измерения Генри Кастлера?

10. Какова роль синергетики в появлении идеи о значимости информационных процессов об окружающем нас мире как о способе видения этого мира?

11. В чем сущность условной и безусловной информации?

12. Что такое бифуркация и как она соотносится с качеством информации?

13. В чем сущность постановки вопроса оценки ценности информации и ее эффективности?

14. Какова сущность известных способов количественного определения ценности информации?

15. Какова зависимость ценности информации от тезауруса?

16. Какие существуют меры ценности информации?

17. Как зависит ценность информации от ее количества?

18. Представить зависимость эффективности информации (ε) от ее количества (I)

19. В чем сущность синергетического подхода к классификации информации?

20. Что такое микроинформация и как она оценивается?

21. В чем сущность методологического анализа рецепции информации?

22. Что такое негэнтропия?

23. В чем сущность генерации информации?

ГЛАВА 11. ВЗГЛЯДЫ НА ИНФОРМАЦИЮ В КОНТЕКСТЕ ПОСТНЕКЛАССИЧЕСКОЙ НАУКИ

Трудности однозначного определения понятия «информация» объясняются попытками описать типично неклассический феномен в терминах классической науки. С классической точки зрения должно существовать единственное объективное описание явления, которое не зависит ни от способа, ни от времени наблюдения. Впервые с парадоксами фундаментального классического описания столкнулись физики в вопросах о локализации частицы, а также о волновой и одновременно корпускулярной их природе. Оказалось невозможным разрешить парадокс силами самой физики. Потребовался философский анализ, завершившийся генерацией идеи огромной значимости – принципом дополнительности, выдвинутым Нильсом Бором, принципом, выходящим далеко за пределы физики, стимулирующим переоснащение методологического инструментария в связи с многомерностью. Может оказаться, что содержание процесса, системы не исчерпывается каким-либо одним теоретическим языком, посредством которого можно было бы выразить параметры, способные принимать вполне определенные значения. Различные языки описания и точки зрения на то или иное явление могут оказаться дополнительными. Они связаны с одной и той же реальностью, но не сводятся к одному единственному описанию.

Полифундаментальность присуща тем областям предметного знания, в которых:

а) наиболее остро взаимодействуют философские основания теории с ее собственными исходными принципами;

б) системы понятий (информация, язык, память, код, связь и т.д.) используются в других, не обязательно терминологических смыслах;

в) при парадигмальных сдвигах обнаруживаются различные способы организации эмпирического материала, относящегося к одному и тому же феномену, что выражается в существовании различных независимых понятийных подсистем, рассмотренных Николис Г., Пригожиным И.

Трудно найти полифундаментальности более яркий пример, чем определение понятия «информация» в современный период становления постнеклассической картины мира. В работе этих авторов полифундаментальность определяется, как новое явление в развитии понятийной формы мышления, допускающее равную гносеологическую ценность альтернативных теорий. Заметим, что применительно к обсуждаемой познавательной ситуации конкурирующие альтернативности уступают место дополнительности.

Значение принципа дополнительности усилилось после утверждения парадигмы постнеклассической науки, описывающей саморазвитие, самоорганизацию материальных структур. В философии стало развиваться направление, связанное с преобразованиями в логике мышления, переосмысление классических представлений Кастлера Г. о сущности замены дискретной логики мышления на континуальную, предметом которой является полифундаментальный многомерный мир. Эти изменения касаются и содержания категорий развития, его направленности. С учетом этих изменений и представлений о развитии в настоящем разделе рассматривается информационный процесс, отдельные акты которого порождают систему определений, не противоречивых, а дополняющих друг друга, что лишает информацию ореола таинственности и божественности в согласии с мнением И. Р. Пригожина и И. Стенгерс: «Неустранимая множественность точек зрения на одну и ту же реальность означает невозможность существования божественной точки зрения, с которой открывается «вид» на всю реальность». Это относится и к точке зрения на саму эту множественность, что объясняет возникновение многомерной методологии, направленной на возможность отдельные измерения («срезы», «ипостаси») целого видеть равноправными. Если многоликость мира в этой методологии видится метафорически, то перед исследователями нелинейной динамики систем она предстает вполне конкретно в виде многовариантности (С. П. Курдюмов), многомодальности (Г. Р. Иваницкий), многослойности (Ю. М. Лотман).

Принципиально важно при изучении информационного процесса понимать равноправность всех его стадий.

Значительную роль в жизни общества играет информация, кодирующая специальные механизмы (технологии), общественные и государственные институты, принимающая форму алгоритма или программы. Именно так информация, сама по себе не являющаяся материальной, преображает материальный мир. Для понимания функционирования информации-программы очень важно знать ее свойства. Комплекс свойств информации, связи между ними, их иерархия выявлены известными авторами на примере биологических систем. Вместе с тем эти свойства изоморфны для всех видов информации, что еще раз подчеркивает общность самого феномена. Представляется, что выявление общности свойств, практика их использования для выяснения механизмов информационных процессов будет играть существенную роль в построении обобщенной теории информации.

Конец раздела посвящен интерпретации предпосылок информации – асимметрии и отбору, играющим особую роль в современном мире. Одним из удивительных примеров грандиозной асимметрии этого мира является нарушение симметрии между веществом и антивеществом, как предпосылка синергетической информации, обеспечивающей процесс дифференциации частиц, из которых возникли звезды, планеты и все живые существа. Ранняя история Вселенной отмечена переходом от состояния теплового равновесия к неравновесному состоянию, провоцируемому действием гравитации. В сильно неравновесном мире могут иметь место крупномасштабные переходы с нарушением симметрии между концентрациями частиц и античастиц. Благодаря этому, современное вещество представляется, как результат некоторого первичного неравновесия – асимметрии.

Если обратиться к структурам привычных масштабов, то можно убедиться, что информация (ее функционирование и воспроизводство) всегда связана с отбором некоторых предпочтительных асимметричных форм. В живом мире асимметрия имеет вездесущий характер. Так, например, в природе возможны две конфигурации аминокислот, несовмещаемые путем вращения. Аминокислоты в белках абсолютно отобраны в пользу только одной из конфигураций. Другим примером является асимметрия стартовых сигналов, позволяющих читать генетическую информацию в виде троек нуклеотидов в заданном направлении от точки старта. Эта асимметрия фундаментальна по своему значению.

В современной литературе, приведенной в библиографии, придается большое значение асимметрии правого и левого полушарий мозга человека в образовании семиотических структур, в синергетике интуиции, а также в процессе творчества.

11.1. Многозначность определения понятия информации

Роль информационных взаимодействий в природных процессах стала вырисовываться три десятилетия тому назад в связи с развитием в ходе новой научно-технической революции кибернетики – науки, которая занимается изучением систем любой природы, способных воспринимать, перерабатывать информацию и использовать ее для управления и регулирования.

Перевод на язык известной теории информации процесса дарвиновской эволюции видов (Кастлер Г. и другие ученые) был не только шагом вперед по пути развития исторической биологии, он продемонстрировал перспективность общенаучного повода, использования общенаучных понятий и самых общих представлений теории информации в естественных науках. Стало ясно, что в основе биотических процессов, как и в основе биологической эволюции, лежит принцип обратной связи, что характерно для кибернетических систем.

Следует признать, что некоторые попытки поставить понятие «информация» на службу науке о неживой природе были весьма неудачными. Так, Г. Ф. Хильми пишет в своей книге: «Прежде всего мы заметим, что во многих случаях после наступления одного определенного явления *A* всегда или в существенной части случаев наступает другое явление *B*; ... в подобных случаях осуществление явления *A* с большей или с меньшей уверенностью позволяет предвидеть наступление явления B, и поэтому естественно говорить, что A есть носитель информации о B. Такую связь мы будем символически записывать посредством формулы $A = infB$». И далее: «Приведем некоторые примеры подобных связей. Горному обвалу предшествует образование крупных трещин в скалах; весеннему разливу рек предшествует таяние снега и т. д. ... Подобных примеров можно привести сколько угодно».

Академик Н. Н. Моисеев по аналогичному поводу замечает, что в естествознании существует очень полезная традиция: каждый новый термин вводится лишь тогда, когда без него нельзя обойтись.

Можно ли обойтись без понятия «информация»? По мнению академика А. И. Берга – невозможно: «Ни вещества, ни энергии, не связанных с информационными процессами, не существует. Не энергия, а информация выйдет, наверное, в XX веке на первое место в мире научных и практически действенных понятий».

В каких же случаях плодотворность информационного подхода к изучаемым явлениям не вызывает сомнений?

1) Во всех живых системах – от растений, микроорганизмов до человека и человеческих сообществ. Во многих случаях, когда мы имеем дело с обратной связью. Без информационных процессов понятие обратной связи зачастую вообще лишено смысла.

2) В теории связи, задачей которой является повышение качества передачи информации; информация в теории связи является статистической.

3) В самоорганизующихся системах при переходе системы из одного состояния с энтропией S_1 в другое состояние с S_2 ($S_2 < S_1$) происходит генерация информации. Г. Хакен предложил назвать этот вид информации синергетической информацией. Это – случай эволюции диссипативных систем, когда существенную роль играют не силовые, а кооперативные (корреляционные) связи.

4) Если речь идет об увеличении разнообразия элементов системы и повышения ее сложности, в том числе о накоплении знаний.

5) При изучении систем, развитие которых определяется эффективностью коммуникативных связей, использующих естественные и мягкие языки, создающие обилие знаковых ситуаций – социокультурных систем.

Современное понимание того, что такое информация и какую роль она играет в естественных и искусственно созданных системах, складывается постепенно. Оно представляет собой совокупность знаний, полученных разными науками: философией, физикой, биологией, общей теорией систем, теорией связи, математикой. Вместе с тем, как пишет И.В. Мелик-Гайказян, «В свете достижений информатики, как в прикладном, так и теоретическом аспектах может показаться парадоксальным тот факт, что до сих пор наукой еще не предложено единое общепринятое определение информации, отражающее давно сложившееся бытовое представление об информации, как о сообщении новых сведений, так и вновь открытый (пока еще, в основном, интуитивно) антиэнтропийный характер самопроизвольно возникающих и протекающих в природе информационных процессов».

По мнению Н. Н. Моисеева, «...строгого и достаточно универсального определения информации не только нет, но оно и вряд ли возможно». Естественно возникает вопрос: с чем это связано? Почему, несмотря на большое число работ по информатике, теории информации, кибернетике, философии определение общенаучного понятия «информация» до сих пор действительно не сформулировано. Известна многоплановость понятия «энтропия», используемого в самых разных аспектах. Очевидно, что это относится и к понятию «информация»: ее определение, даваемое тем или иным ученым, зависит от сферы его деятельности или научных интересов.

И.В. Мелик-Гайказян утверждает: «Как это ни парадоксально звучит, но для развития теории информации в ее современном виде вообще не требуется определения информации как таковой. Необходимым и достаточным для построения теории является понятие количества информации». С этим можно согласиться, если речь идет об использовании обсуждаемого понятия в теории связи. Но более сложная задача – обсудить возможность информационного подхода к развитию различных систем от механических и термоконвективных до знаковых – не может быть решена без анализа вопроса о сущностях информации, а использование некорректных определений общенаучного понятия «информация» неизбежно приведет к ошибкам. Весьма характерно, что развитие понятия «информация», развитие синергетики, аналогия между информационными процессами в биологии, физике и химии приводят к пересмотру позиций автора. Он также дает следующее определение: «Информация в обыденном смысле это сведения, известия, в научно-технических приложениях – то, что несет на себе сигнал». Определение относится к любой передаче информации, которая всегда осуществляется с помощью сигналов. Однако отдельные сигналы или буквы, передаваемые по каналам связи, сами по себе не несут информации, для передачи которой устроена система связи. Информацию содержат лишь осмысленные наполненные содержанием сочетания букв.

Вместе с тем сигналы играют в информационных системах важную роль. Если энергетические и вещественные потоки, образно говоря, питают систему, то потоки информации, переносимые сигналами, организуют все ее функционирование, управляют ею. Однако приведенное здесь определение, как и множество других, не исчерпывают богатства содержания понятия «информация».

В науке уже имела место аналогичная ситуация, когда в результате исследований сил различной природы возникла потребность объединить множество взаимодействий тел единым понятием «энергия». В связи с этим представляется весьма конструктивным подход Ю. П. Петрова, попытавшегося проанализировать специфические черты процессов которые считаются информационными, черты, существенно отличные от энергетических процессов, описываемых в привычных терминах классической физики, для интерпретации которых вполне достаточно известных законов (законы Ньютона, законы сохранения).

Справедливо отмечено, что в природе имеют место взаимодействия между системами, между системой и внешней средой, между элементами системы, результаты которых не зависят от получаемой системой энергии, или от количества переданного ею тепла, импульса, момента импульса. В технике это радиоприемник, который на входе принимает радиосигнал, а на выходе выдает звуковой сигнал практически независимо от энергии радиосигнала. Это отнюдь не противоречит закону сохранения энергии, ибо приемник питается внутренними источниками энергии, а радиосигнал лишь управляет этой энергией, распределяя ее по разным каналам.

Все информационные процессы подобны тем, что происходят в радиоприемнике: их результат нельзя объяснить только с помощью обычных классических характеристик, требуется новая, которой и является информация, как мера особого класса взаимодействий материальных систем.

Особенности информационных взаимодействий:

– их результат не находится в однозначной зависимости от энергии взаимодействия;

– взаимодействие несимметрично, так как системы воздействующая и воспринимающая неинвариантны (приемник и передатчик нельзя поменять местами);

– воспринимающая информацию система является открытой, неравновесной системой, она по-разному может изменять свое состояние, т. е. результаты воздействия сигналов могут быть различными.

Дальнейшее прояснение содержательной стороны понятия информации лежит в сфере научных идей, ибо информация является общенаучным понятием, хотя в его определении безусловно имеются методологические и философские проблемы. В рамках настоящего раздела ставится задача выяснения причин тех трудностей методологического порядка, которые стоят на пути выработки единого, достаточно общего подхода к определению феномена информации.

Анализ библиографии позволяет говорить, что в настоящее время существуют лишь отдельные фрагменты «информационной» теории, в том числе: работы создателей кибернетики по развитию концепции регулирования; работы по математической теории связи; создание начал динамической теории информации на основе идей синергетики; работы по изучению свойств, характеристик и специфики живых информационных систем.

Противоречивость высказываний создателей кибернетики по вопросу о сущности феномена информации свидетельствует о своеобразном «информационном парадоксе»: кибернетика широко использовала информацию, как средство управления сложными системами, но оказалась неспособной дать адекватное определение явлению информации.

Это показывает, что несмотря на заинтересованность в конкретном определении информации, несмотря на то, что вопросам этого определения занимались ведущие кибернетики, стремящиеся к постановке больших теоретических проблем, им не удалось раскрыть сущность понятия «информация». К этому выводу приводит анализ обзорных работ, посвященных философии кибернетики, теории информации. В них поднимаются интересные и актуальные вопросы:

– о связи информации и психологии;
– роли информации в развитии общества;
– философской проблеме надежности кибернетических систем;
– социальных проблемах информатики;
– философских проблемах кибернетики.

Вместе с тем основной; важный для построения обобщенной теории информации вопрос о том, что представляет собой феномен информации, остается открытым. В рамках теории связи и кибернетики порождена разноголосица в определениях, трактовках и интерпретациях, что бросает тень сомнения на реальность самого феномена и целесообразность создания его теории. По-видимому, речь должна идти не о субъективных сложностях, с которыми сталкивались отдельные ученые, а о некоторой принципиальной особенности познавательной ситуации. По-видимому, здесь мы встречаемся с тем парадоксальным случаем, когда экспликация термина оказывается невозможной в силу малоизученности круга соответствующих ему явлений. Эта ситуация будет, очевидно, преодолена лишь после того, как данные, полученные наукой, достигнут некоторой «критической массы», будут обнаружены неизвестные в настоящее время закономерности и прояснятся системы связей между отдельными аспектами общей проблемы. Пока же, по-видимому, допустимо прибегать к такому определению информации, которое операционно удобно для данного типа исследований.

Нельзя не согласиться с тем, что трудности определения природы информации и не могли быть преодолены без известных теперь нам инноваций и без представлений об информации, как о процессе, начальные стадии которого связаны с эволюцией синергетической системы. В досинергетический период науки мысль исследователей об информации не могла разорвать «заколдованный круг», охватывающий множество разнообразных определений одного и того же явления. В каком только образе не представала информация: информация-выбор; информация-программа; сигнал; сведения; знания; негэнтропия; отражение в сознании, процессы отражения. Неудивительно, что подобная неоднозначность вызывает сомнения в реальности информации.

Поиска информации как субстанции или как функции состояния совершенно бесперспективны. Они, так или иначе, навязывают информации сугубо термодинамический статус, но при этом игнорируется, что информация – это процесс.

Считать информацию фикцией, все равно, что считать фикцией количество переданного телу тепла только на том основании, что гипотеза о теплороде оказалась мертворожденной. Никто не видел тепло ни как субстанцию, ни как свойство (функцию состояния). Тем не менее, теплопередача – реальный способ изменения внутренней энергии среды.

Поскольку материалистическая теория познания установила, что всеобщим свойством материи является отражение, информация стала казаться таким же всеобщим свойством материи, как, например, движение.

Каково же соотношение между информацией и отражением? А.Д. Урсул: «Информация есть отраженное разнообразие». Это определение сводится к тому, что отражаться может не только разнообразие, но и однообразие, информация же связана с разнообразием. Но отражение по определению, по своему смыслу также связано с разнообразием, поэтому в рамках определения Урсула невозможно отделить информацию от отражения. Позже автор дает иное определение: «Информация не тождественна отображению, а есть лишь его инвариантная часть, поддающаяся опредмечиванию, объективированию, передаче». Автор поясняет, что отражение далеко не всегда можно перекодировать, перенести с одного материального объекта на другой, информация же перекодируется, передаются, создавая образы, в которых инвариантом является она сама.

В подобных определениях смешиваются понятия «отражение» и «разнообразие», «генерация» и «рецепция». В самом деле, «разнообразие» входит в определение генерации информации (случайный выбор из «разнообразных», нетождественных элементов), а «отражение» – это элемент рецепции информации. Термин «отражение» не охватывает, таким образом, всего феномена. Как возникает информация? Какие цели при этом достигаются? Как оценить новизну, ценность информации, стимулируемую ею целенаправленную деятельность? Эти вопросы остаются без ответа.

Итак, можно заключить, что информация как отражение это не более чем оптическая метафора; информация как субстанция или как функция состояния – это фикция, ее нет в природе. Информация реальна, объективна, как *процесс*. Информация – это дитя многомерного мира самоорганизующихся систем и необратимых процессов, мира, открывшегося людям сравнительно недавно, еще не вполне и далеко не всеми осознанного. Информация не раскрыла своей сущности в рамках кибернетики, поскольку традиционная кибернетика не рассматривала самоорганизующихся систем в русле синергетики, концептуальный аппарат которой является исходной точкой новых представлений в науке.

Информация – сложный процесс, состоящий из элементарных актов. Элементарность эта относительна, так как не все механизмы соответствующих процессов еще изучены. При переходе от одной стадии процесса к другой информация всякий раз предстает в новом обличье. Увидеть это, тем более идентифицировать отдельные стадии с теми или иными определениями, не входя в область представлений постнеклассической науки, невозможно, ибо традиционной науке чуждо описание создания нового, ранее не существовавшего, чуждо процессуальное описание становления. Таким образом, становится очевидным, что корни трудностей определения феномена информации заключаются в попытках рассмотреть постнеклассический объект – феномен информации – с позиций и методами традиционной науки. История науки знает немало примеров подобных неудач. Всякий раз выход был в отказе от моделей, неадекватных реальности, в переходе к новой парадигме.

11.2. Информационные процессы в нелинейном мире

В настоящее время признано, что в период традиционной науки значительная часть конкретного мира вокруг нас ускользала из ячеек научной сети. Это в полной мере относится к понятию «информация», строгое, четкое и однозначное определение которого неизменно ускользало от исследователей. В предыдущем подразделе было высказано предположение о том, что неудачи в этом направлении связаны с попытками подойти к постнеклассическому объекту с типично классическими мерками. Само стремление дать, во что бы то ни стало однозначное определение, и разочарования по поводу невозможности этого следует отнести к стратегии классической науки с ее фундаменталистским образом, со стремлением интерпретации множества явлений, с помощью единой формулы.

В постнеклассический период развития науки теория самоорганизации открытых систем различной природы выявила несовершенство модели фундаментализма, которое стало ощутимо при соприкосновении с гетерогенными, полиморфными системами, состоящими из разнородных, порой несопоставимых элементов. Общее, что объединяет эволюции таких систем (в большом числе случаев), есть информация с инвариантными законами ее функционирования, характеристиками и свойствами. Для общей теории информации в рамках синергетики важно, что постнеклассическая наука вместо поиска причинных отношений стала апеллировать к таким целостным факторам, как симметрия, упорядоченность и др.; описание объектов, основывавшееся прежде на модальности долженствования, стало осуществляться теперь в модальности возможности, когда выбор одного из разрешенных состояний не предписывается необходимостью.

Для того чтобы убедиться в адекватности феномена информации миру процессов, изучаемых постнеклассической наукой, перечислим ее основные концептуальные положения, опираясь при этом, как на конкретно-научные результаты (Николис, Пригожин, Хакен, Курдюмов), так и на философские обобщения (Пригожин-Стенгерс, Курдюмов).

Наука прошлого столетия оперировала обратимыми (классическая механика) и необратимыми (равновесная термодинамика) процессами. Наряду с этим в природе реализуются необратимые процессы, приводящие к созданию пространственно-временных (диссипативных) структур. Такая необратимость играет существенную конструктивную роль, из нее следует различие между прошлым и будущим («стрела времени»), обусловленное не особенностями нашего описания, а самой природой вещей.

Первое концептуальное положение постнеклассической науки имеет прямое отношение к процессу генерации информации, являющемуся типично необратимым, поскольку генерация это выбор, и, коль скоро он сделан, происходит событие, разделяющее прошлое и будущее. Можно заключить, что если бы не различие между прошлым и будущим, игнорируемое традиционной наукой, информация не могла бы возникать. Это еще раз подчеркивает: место феномена информации только в постнеклассической картине мира.

Неравновесные состояния скрытых систем весьма конструктивны благодаря их способности к самоорганизации, (т. е. к пороговому локальному самоупорядочению) с образованием диссипативных структур.

Такое поведение – необходимый элемент процесса рецепции информации, т. е. второе концептуальное положение, как и первое, имеет отношение к информации, так как рецепция есть процесс неравновесный. Кроме того, поскольку рецепция информации означает возникновение определенной упорядоченности в воспринимающей системе, это не только неравновесная, но далекая от равновесия система. Рецепторная система, следовательно, есть система диссипативная.

Важнейшим состоянием синергетической системы является хаос, или, точнее, хаотическая динамика. Как утверждается учеными, связанная с разупорядоченностью неустойчивость движения позволяет системе непрерывно определять собственное пространство состояний, создавая тем самым информацию.

Нет необходимости комментировать связь информации с третьим концептуальным положением постнеклассической науки, о ней говорится в самом этом положении. Уместно привести пример, указывающий на эту связь. Показано, что предельно неустойчивой хаотической системой является мозг здорового бодрствующего человека. Генерация логической информации человеком имеет благодаря этому неограниченные возможности.

Известно, что учеными выявлена неформальная связь информации с идеями теории самоорганизации, представляющей собой основу постнеклассической науки. Представление мира, интерпретируемого этой наукой, как мира не только траекторий, но и процессов. Неопределенность мира, не являющегося ни автоматом, ни хаосом, не исключает значимой деятельности индивида в нем. Возможность значимой деятельности здесь вытекает из феномена логической информации; для определения ее сущности рассмотрим методологические особенности мира постнеклассической науки.

Одной из основных таких особенностей является то, что «новый» мир не поддается описанию одной единственной истиной. Поэтому методологической платформой для определения сущности информации является многомерность этого феномена.

Задача дискриминации различных определений не должна возникать, поскольку разные модели имеют право на существование. По-видимому, вместо дискриминации необходимо установление связи между альтернативными описаниями, относящимися к одному и тому же многомерному феномену. Почва для такого синтеза видится в раскрытии природы сложности и полифундаментальности понятия «информация». Основа для решения этой задачи – рассмотрение единого информационного процесса, разные стадии которого как раз и порождают множество определений, ибо на каждой отдельной стадии информация проявляет свои специфические черты. К сожалению, идея об отсутствии инвариантной «базовой истины» в отношении информации оказалась чуждой кибернетике. Ниже целесообразно рассмотреть черты, объединяющие и разъединяющие кибернетику и синергетику. Эти общенаучные направления развились на одной и той же основе – теории систем, оба оперируют такими понятиями, как «система», «структура», «организация». Оба направления связаны с изучением сложных динамических систем, различных по своей природе, оба абстрагируются от вещественного содержания систем, стремясь сформулировать общие законы их организации.

В сороковых годах в рамках кибернетики были предприняты попытки создания теории самоорганизующихся технических систем без выяснения конкретных механизмов самоорганизации. Правильные догадки о такой возможности не были развиты. К синергетике впервые пришли отнюдь не кибернетики, а физико-химики и физики, изучавшие нелинейные колебания, колебательные химические реакции и стимулированное излучение лазера. Следует признать огромное практическое значение кибернетики и принципиальную невозможность создать в ее рамках единую теоретическую концепцию.

В отличие от этого появление синергетики воспринимается как рождение нового объединяющего направления в науке, знаменующего новое мировоззрение. В рамках синергетики удается раскрыть сущность феномена информации на основе перехода современного познания от парадигм бытия к парадигме становления. Рассмотрим принципиальные различия кибернетического и синергетического подходов.

1. Кибернетика изучает законы управления и связи, оперируя законами управления и информационных взаимодействий (с помощью обратных связей) на нескольких уровнях одновременно.

Синергетика имеет дело с системами неуправляемыми, их развитие можно только направлять, при этом небольшое энергетическое воздействие-укол в нужном месте пространства-времени оказывается достаточным, чтобы система перестроилась и возник новый тип структур.

2. Основным фактором кибернетических систем служит информация и в силу этого кибернетика отвлекается от энергетических и вещественных субстратов. Однако кибернетические системы действуют в рамках традиционной науки (классической и квантовой механики) и в их поведении нет ничего, что не могло бы быть предсказано, с помощью этих законов. Они вполне детерминированы.

Поведение синергетических систем не противоречит законам традиционной науки, но не прогнозируется ими.

3. Кибернетика использует информацию, как нечто данное. Выяснение причин и механизмов ее возникновения не входит в ее задачи.

С позиций синергетики для генерации информации принципиально важны хаотические состояния систем, выход из которых сопровождается созданием новой информации.

4. Синергетика позволяет проследить эволюцию информационной системы, как эволюцию ценности и новизны созданной информации.

Кибернетика, как и математическая теория связи (традиционная теория информации), не рассматривает понятия ценности и новизны в информации.

В изучении многомерного мира одним из ведущих принципов является принцип дополнительности, согласно которому существует связь между альтернативными описаниями одного и того же сложного объекта. Каждое из определений не охватывает полноты сущностных характеристик объекта как целого, но пополняет другие. Таким образом, если различные определения одного и того же явления не содержат ошибок, они позволяют соединить его альтернативные модели для охвата всей многомерной реальности. Подобный антифундаментализм безусловно характерен для информации. Он-то и является причиной беспокойства тех исследователей, которые считают, что логический фундаментализм имеет неограниченную область применения, а потому существует жесткое определение термина «информация».

Такая лаконичная и жесткая дефиниция действительно невозможна: большинство определений верны, но все они являются частными, ибо отражают либо лишь определенный этап процесса (например, «передача сведений»), либо его цель «приобретение новых знаний»).

Выше была определена ценность информации, как мера повышения вероятности достижения цели, что предполагает некоторое целенаправленное действие. В.И. Корогодиным в связи с этим вводится понятие оператора, как устройства (механизма), обеспечивающего достижение цели. Отсюда – следующее определение: «Совокупность приемов, правил и сведений, необходимых для построения оператора, будем называть информацией». Итак, информация приобретает вид плана, алгоритма, программы, что согласуется с алгоритмическим определением А.Н. Колмогоровым информации. Это определение также является частным, ибо оставляет открытым вопрос: откуда берется программа?

Рассмотрим это на примере. Допустим, перед конструктором стоит цель существенно улучшить в заданном направлении параметры некоторой машины. Действительно, последним этапом работы будет проектирование технологий для создания усовершенствованного варианта машины. Это и есть организация оператора по некоторой программе. Но необходимым и важным элементом работы является генерация существенно новой и ценной идеи усовершенствования. Отсутствие такой идеи сделает бессмысленной всю дальнейшую деятельность. Понадобится еще целая серия информационных актов: рецепция и отбор самим автором, кодирование (составление чертежей), рецепция и отбор оптимального варианта техническим советом, рецепция информации технологами и ее кодирование и уж как завершающий этап – создание действующей модели. В данном случае речь идет о разумной генерации новой информации и заранее сформулированной цели. При этом «узкая» цель построения конкретной машины может расходиться с целями и общими тенденциями научно-технического прогресса (например, некий конструктор совершенствует турбовинтовой самолет, а общей тенденцией является создание реактивных аппаратов).

Известно, что феномен информации это многостадийный процесс. В главе 8 было показано, что две необходимых стадии – генерация и рецепция информации – существенно неравновесные и необратимые процессы. Это делает информацию объектом постнеклассической науки и раскрывает причину прежних неудачных попыток дать универсальное определение сложного феномена, рассматривая всякий раз лишь какой-либо один из множества его аспектов. Такой подход позволяет, не дискриминируя, примирить друг с другом различные частные определения и на их основе построить представление о едином информационном процессе, состоящем из отдельных; связанных друг с другом стадий. Порядок стадий и их число в конкретных случаях различны. Промежуточным этапом может быть запоминание информации с помощью специальных устройств, обладающих памятью. При этом передача информации в «память» возможна на разных этапах процесса. Поэтому в блок-схеме информационного процесса, которая предлагается (рис. 11.1), отражены возможные стадии, последовательность которых не является универсальной.

Так, например, в некоторых случаях необходимо двойное кодирование: после генерации в сознании «творящего» и после рецепции уже закодированной в его сознании информации для ее передачи в «память» или для дальнейших целенаправленных действий. В результате этого стадия рецепции не обязательно предваряется генерацией новой информации.

Следует оговориться относительно термина «информация». Под ним будем понимаем процесс, но, тем не менее, употреблять его с глаголами «хранить», «запасать», «обладать». Во всех этих случаях необходимо понимать изменения, которые происходят с системой в результате протекания информационного процесса.

Каждой стадии информационного процесса (рис. 11.1) возможно привести в соответствие то или иное определение, взятое из литературных источников.

Приведем в соответствие отдельным актам, представленным блоками 1-8, определения понятия «информация».

Блок 1. «Генерация информации есть случайный запоминаемый выбор» (определение Кастлера).

Блок 2. Рецепция информации – необратимый, неравновесный процесс. Рецепторная система – диссипативная система, способная

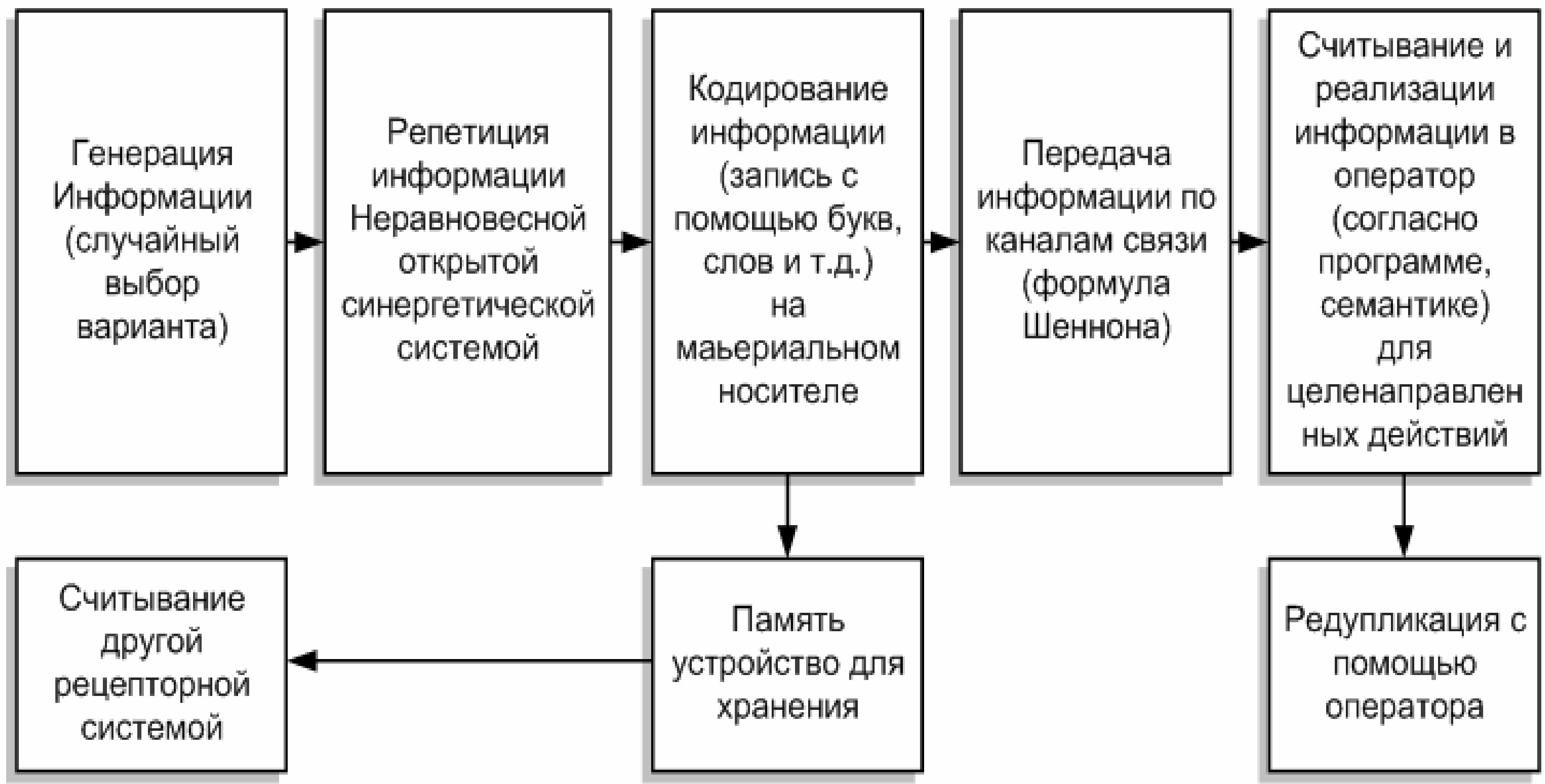


Рис. 11.1. Элементарные акты информационного процесса

самопроизвольно повышать степень своей упорядоченности. На стадии рецепции созданная информация воспринимается неравновесной материальной средой, способной создавать асимметричные структуры, отбираемые с помощью соответствующих правил (законы сохранения, термодинамики), естественного отбора, отбора разумными элементами (людьми или созданными людьми устройствами).

Как считают ученые, понятие об информации на стадии ее рецепции ассоциируется с образованием структур, созданием порядка из беспорядка, что отвечает соответствующему содержанию информации: «…информация означает порядок, коммуникация есть осознание порядка из беспорядка, или, по крайней мере, увеличение степени той упорядоченности, которая существовала до получения сообщения».

Блок 3. Кодирование – подготовка сообщения для передачи по каналу связи к блоку управления и переработки или в блок памяти (блок 7). В соответствии с этим дается известное в литературе определение: «Информация – это сведения, содержащиеся в данном сообщении и рассматриваемые как объект передачи, хранения и переработки».

Блок 4. Для задачи передачи информации по каналу связи важно не понятие об информации, а математическая величина – количество информации, – вычисленная Шенноном. До сих пор принято считать шенноновское определение исчерпывающим.

Блоки 5, 6 отвечают реализации некоторой программы. В связи с этим, специалисты считают, что наряду с кастлеровским определением информации может рассматриваться и другое ее содержание, где информация, представляет план строения клетки и, следовательно, всего организма.

Обращаясь к известным определениям содержания информации Блюменау, Харкевича, Шилейко и других авторов, в части того, что можно использовать определение информации, содержание которого отражает реальную действительность и которое удобно для данного типа исследований, то становится понятным, что феномен информации в ее разнообразии и многообразии содержания может рассматриваться под разными углами зрения. Так, для биолога – это программа синтеза белков, закодированная в генах, для геодезиста – план местности, для телефониста – система отобранных согласно коду сигналов, для поэта композитора – это образы, рождающиеся из хаоса на интуитивном уровне.

Каким бы сложным нам ни казался информационный процесс, необходимо помнить, что своим происхождением он обязан материальной среде; генерируемая информация не может быть запомнена вне связи с материей, а результат процесса никогда не находится в противоречии с законами природы, хотя и не прогнозируется ими.

Учитывая все сказанное, и в отличие от ФЗ «Об информации, информационных технологиях и защите информации», где «информация – сведения (сообщения, данные), независимо от формы их представления», приходим к следующему определению.

Феномен информации есть многостадийный, необратимый процесс становления структуры в открытой неравновесной системе, начинающейся со случайного запоминаемого выбора, который эта система делает, переходя от хаоса к порядку, и завершающийся целенаправленным действием согласно алгоритму или программе, отвечающим семантике выбора.

На основе изучения общих свойств информации разных иерархических уровней, можно попытаться понять природу взаимоотношений между информацией и объектами материального мира и убедиться, что информация становится активной, лишь фиксируясь на материальных носителях.

Контрольные вопросы

1. Чем объясняются трудности однозначного определения понятия «информация» для неклассического подхода?
2. В чем сущность полифундаментальности взглядов на информацию?
3. В чем сущность ассиметрии взглядов на содержание информации?
4. Привести примеры концепций и взглядов на многозначность определения понятия информации.
5. Привести примеры и случаи плодотворности информационного подхода к изучению феномена информации?
6. Каково соотношение понятий «энергия», «материя», «информация»?
7. Каковы особенности информационных взаимодействий материальных объектов?
8. Перечислить актуальные проблемы информации в философии кибернетики, теории информации.
9. Какие существуют взгляды ученых на соотношение между информацией и отражением? В чем сущность информационных процессов в нелинейном мире?
10. В чем принципиальные различия кибернетического и синергетического подходов во взглядах на феномен информации?
11. Перечислить элементарные акты информационного процесса.

ГЛАВА 12. КВАНТОВАЯ ТЕОРИЯ ИНФОРМАЦИИ (ВМЕСТО ЗАКЛЮЧЕНИЯ)

12.1. Источники формирования квантовой информатики: основные понятия, достижения, проблемы

Квантовые компьютеры, квантовые вычисления, квантовая связь, квантовая криптография – все это вызывает в настоящее время повышенный интерес. Идеи квантовой информатики, квантовых вычислений являются продуктами новейшего времени и не успели войти в практику вузовского и школьного образования, в учебники и учебные пособия. Даже классические учебники по квантовой механике не содержат важнейших понятий квантовой физики, лежащих в основе квантовой информатики (понятий запутанных состояний, например). Ниже, на основе использования материалов, представленных в библиографии будет сделана попытка рассмотреть основные понятиями и методы, относящиеся к области квантовых вычислений и квантовой информации.

Квантовая механика – это математическая платформа, или совокупность правил, предназначенная для построения физических теорий.

Например, существует физическая теория, известная, как квантовая электродинамика, описывающая взаимодействие атомов со светом. Квантовая электродинамика построена на основе квантовой механики, но содержит специфические правила, не определяемые квантовой механикой. Связь квантовой механики с конкретными физическими теориями, например, с квантовой электродинамикой, в чем-то похожа на связь операционной системы компьютера с конкретной прикладной программой – операционная система задает некоторые базовые параметры и режимы работы, но не определяет, каким образом прикладные программы будут выполнять свои специфические задачи.

Принципы квантовой механики просты, но даже специалисты находят их противоречащими интуиции. Истоки квантовых вычислений и квантовой информации можно усмотреть в постоянном желании физиков лучше понять квантовую механику. Одной из задач области квантовых вычислений и квантовой информации является разработка инструментов, которые развивали бы наше интуитивное понимание квантовой механики и делали ее предсказания более доступными для человеческого разума.

Например, в начале 80-х гг. ученых стало интересовать, можно ли использовать квантовые эффекты для передачи сигнала со скоростью, превышающей скорость света, что безоговорочно запрещено эйнштейновской теорией относительности. Решение этой проблемы свелось к выяснению того, можно ли копировать неизвестное квантовое состояние. Если бы копирование оказалось возможным, то при помощи квантовых эффектов можно было бы передавать сигнал со скоростью, превышающей скорость света. Однако копирование, столь легко выполнимое для классической информации, в общем случае оказывается невозможным в квантовой механике. Эта *теорема о невозможности копирования* (no-cloning theorem), сформулированная в начале 80-х гг., является одним из самых первых результатов в области квантовых вычислений и квантовой информации. С тех пор к ней было сделано много уточнений, и теперь есть концептуальные инструменты, позволяющие понимать, насколько точно сможет работать устройство (всегда несовершенное) квантового копирования. Эти инструменты, в свою очередь, были применены для понимания аспектов квантовой механики.

Другое историческое направление, внесшее вклад в развитие квантовых вычислений и квантовой информации, зародилось в 70-х гг. в связи с интересом к *получению полного контроля над одиночными квантовыми системами*. В применениях квантовой механики до 70-х гг. XX века обычно осуществлялся общий контроль над объемным образцом, содержащим невообразимое количество квантовомеханических систем, ни одна из которых не была доступна напрямую. Например, квантовая механика замечательно объясняет сверхпроводимость. Но поскольку сверхпроводник представляет собой огромный (по сравнению с атомными масштабами) образец проводящего металла, то можно распознать лишь немногие аспекты его квантовомеханической природы. При этом отдельные квантовые системы, составляющие сверхпроводник, остаются недоступными. Такие устройства, как ускорители частиц, позволяют получать неограниченный доступ к отдельным квантовым системам, но по-прежнему не для полного контроля над элементарными системами.

Начиная с 70-х гг. было разработано много методов управления одиночными квантовыми системами. В качестве примера можно привести методы удержания одиночного атома в «атомной ловушке», обеспечивающие его изоляцию от всего остального мира и позволяющие с невероятной точностью исследовать различные аспекты его поведения. При помощи сканирующего туннельного микроскопа удается перемещать отдельные атомы, составляя из них заданные массивы. Были продемонстрированы электронные устройства, работа которых основана на переносе единичных электронов.

Специалисты и ученые в этой области считают, что работы над проблемой получения полного контроля над одиночными квантовыми системами позволяют исследовать нетронутые области познания в надежде открыть новые, неожиданные явления. Квантовые вычисления и квантовая информация естественным образом вписываются в эту программу. Они ставят ряд практических задач разных уровней сложности для людей, ищущих способы лучшего манипулирования одиночными квантовыми системами, стимулируют развитие новых экспериментальных методик и показывают наиболее интересные направления, в которых нужно ставить эксперименты. И наоборот возможность управления одиночными квантовыми системами играет существенную роль в практической плоскости использования квантовой механики применительно к квантовым вычислениям и квантовой информации.

Несмотря на большой интерес к рассматриваемой области, усилия по построению систем обработки квантовой информации принесли на сегодняшний день скромные результаты. Современная техника для квантовых вычислений представлена маленькими квантовыми компьютерами, способными выполнять десятки операций над несколькими квантовыми битами (кубитами). Были продемонстрированы экспериментальные прототипы устройств для реализации квантовой криптографии – способа секретной связи на больших расстояниях – и даже на таком уровне, когда они могут быть полезны в некоторых реальных приложениях. Однако разработка технологий для реализации крупномасштабной обработки квантовой информации остается серьезной задачей для физиков и инженеров будущего [169, 276 и др.].

Важным достижением XX века является появление информатики (computer science) с ее современным началом, положенным великим математиком Аланом Тьюрингом в 1936 г. в известной работе по описанию абстрактного понятия – программируемого компьютера через модель вычислений, машину Тьюринга. Его универсальная машина может использоваться для моделирования любой другой машины. Она полностью обеспечивает решение задачи алгоритмическими средствами. Т.е., если алгоритм может быть выполнен на любом физическом устройстве, например, на современной ЭВМ, то существует эквивалентный алгоритм для универсальной машины Тьюринга, который решает ту же самую задачу, что и алгоритм, выполняемый на персональном компьютере.

Это утверждение, называемое тезисом Чёрча – Тьюринга, устанавливает эквивалентность между физическим понятием класса алгоритмов, выполнение которых возможно на некотором физическом устройстве, и строгим математическим понятием универсальной машины Тьюринга. Широкое признание этого тезиса положило начало развитию обширной теории информатики.

Вскоре после появления работы Тьюринга были построены первые компьютеры на электронных компонентах. Джон фон Нейман разработал простую теоретическую модель, объясняющую, как на практике собрать компьютер, обладающий всеми свойствами универсальной машины Тьюринга. Тем не менее, настоящая разработка аппаратного обеспечения началась только в 1947 г., когда Джон Бардин, Уолтер Браттейн и Уилл Шокли создали транзистор. С этого момента мощь компьютера стала увеличиваться. В 1965 г. Гордон Мур даже сформулировал закон этого роста, известный как *закон Мура*, согласно которому производительность компьютеров, обеспечиваемая при одной и той же цене, будет удваиваться примерно каждые два года.

При всей справедливости закона Мура, сегодня большинство специалистов считают, что этот рост прекратится где-то в конце первых двух десятилетий двадцать первого века. Традиционные подходы к разработке компьютерной технологии начинают упираться в фундаментальные трудности, связанные с размерами. По мере того, как электронные устройства становятся все меньше и меньше, в их функционирование постепенно вмешиваются квантовые эффекты.

Одним из возможных решений проблемы, связанной с прекращением действия закона Мура, является *переход к другой вычислительной парадигме*. Одна из таких парадигм представляется квантовой теорией вычислений, основанной на идее использования для выполнения вычислений квантовой механики, а не классической физики.

Так, несмотря на возможность применения классического компьютера для моделирования квантового компьютера, эффективное осуществление такого моделирования невозможно. Таким образом, квантовые компьютеры существенно превосходят по скорости классические компьютеры. Это преимущество в скорости настолько значительно, что по мнению многих исследователей никакой мыслимый прогресс в классических вычислениях не поможет преодолеть разрыв в производительности между классическим и квантовым компьютерами.

Важным вопросом является вопрос «эффективного» или «неэффективного» моделирования квантового компьютера. Многие ключевые понятия, необходимые для ответа на этот вопрос, фактически появились еще до того, как возникла идея квантового компьютера. В частности, понятия эффективного и неэффективного алгоритма обрели математическую точность в теории сложности вычислений. При этом, эффективным является алгоритм, время выполнения которого полиноминально зависит от объема решаемой задачи. Для выполнение неэффективного алгоритма, напротив, требуется сверхполиномиальное (обычно экспоненциальное) время.

В конце 60-х и начале 70-х гг. было замечено, что машина Тьюринга обладает, как минимум такой же эффективностью, как и любая другая модель вычислений, в том смысле, что задача, которая может быть эффективно решена в рамках некоторой модели вычислений, может быть эффективно решена и на машине Тьюринга путем использования машины Тьюринга для моделирования другой модели вычислений. Это наблюдение было сформулировано в виде усиленной версии тезиса Чёрча–Тьюринга: *любой алгоритмический процесс может быть эффективно смоделирован на машине Тьюринга*.

Усиление этой версии тезиса Чёрча-Тьюринга заключено в слове «эффективно». Если тезис Чёрча-Тьюринга верен, то из него следует, что независимо от типа машины, используемой для выполнения алгоритмов, эта машина может быть эффективно смоделирована при помощи стандартной машины Тьюринга. Это важное дополнение, поскольку оно подразумевает, что для анализа возможности эффективного выполнения данной вычислительной задачи мы можем ограничиться анализом машины Тьюринга.

Однако уже после Тьюринга различные группы исследователей обнаружили, что некоторые типы аналоговых компьютеров могут эффективно решать задачи, не имеющие, по всей видимости, эффективного решения на машине Тьюринга. На первый взгляд такие аналоговые компьютеры нарушают сильную форму тезиса Чёрча-Тьюринга. К сожалению, если сделать реалистичные предположения о наличии шума в аналоговых компьютерах, то они окажутся неэффективными во всех известных реализациях и не смогут решать задачи, не имеющие эффективного решения на машине Тьюринга.

Это заключение, состоящее в том, что при оценке эффективности модели вычислений необходимо учитывать влияние реального шума, стало одним из первых крупных вызовов, брошенных квантовым вычислениям и квантовой информации. Ответом на него стала разработка теории кодов, исправляющих квантовые ошибки, и устойчивых к ошибкам квантовых вычислений. Таким образом, в отличие от аналоговых вычислений квантовые вычисления в принципе допускают наличие конечного уровня шума, сохраняя свои вычислительные достоинства.

Первое серьезное возражение против тезиса Чёрча–Тьюринга появилось в середине 70-х гг. XX века, когда Роберт Соловей и Волкер Штрассен показали, что проверить, является ли целое число простым или составным, можно с помощью вероятностного алгоритма.

В тесте Соловея–Штрассена случайность использовалась как существенная часть алгоритма. Алгоритм не давал достоверного ответа на вопрос, является ли данное целое число простым или составным, определив это лишь с некоторой вероятностью. Повторяя тест Соловея–Штрассена несколько раз, можно определить это почти наверняка. Нужно особо отметить, что во время появления теста Соловая–Штрассена не было известно какого-либо эффективного детерминированного алгоритма для проверки целых чисел на простоту. Получалось, что компьютеры, имеющие доступ к генератору случайных чисел, могли эффективно выполнять вычислительные задачи, для которых не было эффективного решения на традиционной детерминированной машине Тьюринга. Это открытие послужило толчком к поиску других вероятностных алгоритмов, который полностью оправдал себя, приведя к созданию успешно развивающейся области исследований.

Вероятностные алгоритмы поставили под сомнение тезис Чёрча-Тьюринга, показав, что существуют эффективно решаемые задачи, которые, тем не менее, не могут быть эффективно решены на детерминированной машине Тьюринга.

Возникшее затруднение было легко устранено простой модификацией тезиса: *любой алгоритмический процесс может быть эффективно смоделирован на вероятностной машине Тьюринга*.

Эта модификация обусловила появление проблемы разработки модели вычислений, позволяющей эффективно решать задачи, не имеющие эффективного решения в рамках модели вычислений Тьюринга, т.е. модели вычислений, которая бы эффективно моделировала любую другую модель вычислений.

Решая эту проблему Дэвид Дойч в 1985 г. решил выяснить, можно ли использовать законы физики для вывода еще более сильной версии тезиса Чёрча–Тьюринга. Вместо принятия специальных гипотез Дойч стал искать физическую теорию для обоснования тезиса Чёрча–Тьюринга, которое было бы столь же надежным, как и статус самой этой теории. В частности, Дойч попытался описать вычислительное устройство, которое было бы способно эффективно моделировать произвольную физическую систему. Поскольку законы физики, в конечном счете, являются квантовомеханическими, Дойч пришел к рассмотрению вычислительных устройств, основанных на принципах квантовой механики. От этих устройств – квантовых аналогов машин, описанных Тьюрингом – ведет свое начало концепция современного квантового компьютера.

Сегодня не ясно, достаточно ли универсального квантового компьютера Дойча для эффективного моделирования произвольной физической системы. Доказательство или опровержение этой гипотезы представляет собой одну из проблем в области квантовых вычислений и квантовой информации.

Модель квантового компьютера Дойча позволила оспорить форму тезиса Чёрча–Тьюринга. Дойчем была поставлена задача о возможности квантового компьютера эффективно решать вычислительные задачи, не имеющие эффективного решения на классическом компьютере, даже если это вероятностная машина Тьюринга. Решение этой задачи было основано на разработке простого примера, показывающего, что квантовые компьютеры действительно могут превосходить по вычислительной эффективности классические компьютеры.

Развитием работ Дойча стали исследования Питера Шора. В 1994 году он продемонстрировал, решение двух задач, поиска простых сомножителей целого числа и вычисления дискретного логарифма на квантовом компьютере. Результаты Шора показывали, превосходство квантовых компьютеров по производительности перед машинами Тьюринга, включая их вероятностный вариант.

Следующим доказательством эффективности квантовых компьютеров является решение задачи поиска в некотором неструктурированном поисковом пространстве, осуществленное в 1995 году Ловом

Гровером. Алгоритм Гровера не давал такого эффектного ускорения, как алгоритмы Шора, но ввиду широкого применения методологий, основанных на поиске, он вызвал значительный интерес.

В это же время ученые разрабатывают идею Ричарда Фейнмана, высказанную им в 1982 г. Фейнман указал, что моделирование квантовомеханических систем на классических компьютерах сопряжено с существенными трудностями. Он предположил, что построение компьютеров на основе принципов квантовой механики позволило бы этих трудностей избежать. Использование квантовых компьютеров для эффективного моделирования систем, не имеющих какой-либо известной эффективной модели на классическом компьютере, позволяет утверждать, что в будущем главным их применением станет моделирование квантовомеханических систем, слишком сложных для моделирования на классическом компьютере.

С другой стороны, разработка алгоритмов для квантовых компьютеров трудна в силу следующих проблем:

– если прибегнуть к интуиции при разработке алгоритмов, то алгоритмические идеи, к которым придем, будут классическими. Для создания хороших квантовых алгоритмов необходимо «отключить» классическую интуицию, используя для достижения желаемого результата чисто квантовые эффекты;

– недостаточно разработать алгоритм, который просто является квантомеханическим, но он должен быть лучше, чем любой из существующих классических алгоритмов, в силу того, что найденный алгоритм, использующий чисто квантовые аспекты квантовой механики, не будет представлять большого интереса из-за существования классических алгоритмов со сравнимой производительностью.

С учетом рассмотренного содержания проблем возникает необходимость:

– обобщения знаний по вопросу оценки производительности квантовых компьютеров по сравнению с классическими;

– выявления причин повышения эффективности квантовых компьютеров;

– определения классов задач, предпочтительных для решения на квантовом компьютере и соотнесения этих классов с классами задач, решаемых на классических ЭВМ;

Не менее актуальной является и проблема использования квантовых вычислений и квантовой информации в теории связи (communication), в основу которой, в 1948 году, положены работы теории информации и связи Клода Шеннона, в том числе и математическое определение понятия информации, и источника информации, вопросы обмена информацией по каналу связи, определение ресурсов, необходимых для передачи информации по каналу связи, помехозащищенной передаче информации по каналу связи.

Решение этих задач представлены Шенноном в двух доказанных им фундаментальных теоремах теории информации.

Теорема 1. О кодировании для канала без шума. Здесь определено количество физических ресурсов, необходимых для хранения выходных данных источника информации.

Теорема 2. О кодировании для канала с шумом. В ней определено количество информации, которое можно надежно передать по каналу связи в присутствии шума и показано, что для достижения надежной передачи в присутствии шума можно использовать коды, исправляющие ошибки.

Теорема Шеннона о кодировании для канала с шумом устанавливает верхний предел защиты информации, обеспечиваемой кодами, исправляющими ошибки. Но теорема не дает явного вида кодов, при помощи которых можно было бы достичь этого предела на практике. С момента опубликования работ Шеннона и до настоящего времени исследователи разрабатывают все новые и лучшие классы кодов, исправляющих ошибки, пытаясь приблизиться к пределу, установленному теоремой Шеннона. Существует сложная теория кодов, исправляющих ошибки, которая предлагает пользователю, желающему разработать хороший код, множество вариантов выбора. Такие коды широко применяются; они используются, например, в проигрывателях компакт-дисков, компьютерных модемах и спутниковых системах связи.

Так же развивалась и квантовая теория информации.

В 1995 г. Бен Шумахер доказал аналог теоремы Шеннона о кодировании в отсутствие шума, по ходу дела определив «квантовый бит», или «кубит», как реальный физический ресурс. Однако до сих пор неизвестно никакого аналога теоремы Шеннона о кодировании для канала с шумом применительно к квантовой информации. Несмотря на это, по аналогии с классическими эквивалентами была разработана теория исправления квантовых ошибок, которая позволяет квантовым компьютерам эффективно проводить вычисления в присутствии шума, а также осуществлять надежную связь по квантовым каналам с шумом.

Классические идеи исправления ошибок оказались очень важными для разработки и понимания кодов, исправляющих квантовые ошибки. В 1996 г. независимо работавшие Роберт Калдербанк с Питером Шором и Эндрю Стин открыли важный класс квантовых кодов, называемых сейчас CSS-кодами, по первым буквам их фамилий. Впоследствии эти коды были отнесены к категории симплектических (стабилизирующих) кодов, независимо разработанных Роберто Калдербанком, Эриком Рейнсом, Питером Шором и Нейлом Слоуном, а также Даниэлем Готтесманом. Эти открытия, опирающиеся на основные идеи классической теории линейного кодирования, в значительной степени способствовали быстрому пониманию кодов, исправляющих квантовые ошибки, к их применению в области квантовых вычислений и квантовой информации.

Теория кодов, исправляющих квантовые ошибки, была разработана с целью защиты квантовых состояний от шума. В 1992 г. Чарльз Беннет и Стивен Уиснер объяснили, как передавать два классических бита информации путем передачи от отправителя к получателю только одного квантового бита. Это было названо сверхплотным кодированием.

Еще больший интерес представляют результаты в области распределенных квантовых вычислений. Учеными было доказано, что квантовые компьютеры могут потребовать экспоненциально меньшего количества передач для решения определенных задач по сравнению с классическими сетевыми компьютерами. Но эти задачи пока не представляют особого интереса в реальных условиях, и имеют некоторые нежелательные технические ограничения. Важным вопросом, в области квантовых вычислений и квантовой информации, в будущем, является поиск практически важных задач, для которых распределенные квантовые вычисления имеют значительное преимущество над распределенными классическими вычислениями [169, 276].

Теория информации начинается обычно с изучения свойств одиночного канала связи. На практике представляют интерес сетевые структуры, состоящие из нескольких каналов. Свойства таких сетей, обеспечивающих передачу информации, изучаются в сетевой теории информации.

Сетевая квантовая теория информации и исследование методов и способов передачи информации по сетям квантовых каналов являются актуальными научными направлениями. В современной литературе представлены некоторые предварительные результаты по исследованиям этих направлений, но единой сетевой теории информации для квантовых каналов пока не создано.

В качестве примера, подтверждающего важность этого направления, является известный пример по передаче информации по квантовому каналу с шумом от ее отправителя к получателю. Если этот канал имеет нулевую пропускную способность для квантовой информации, то по нему нельзя надежно передавать никакую информацию. При допуске, что рассматриваются две копии канала, работающие синхронно, с интуитивной точки зрения очевидно (что, как известно уже доказано наукой), что для квантовой информации такой канал также имеет нулевую пропускную способность. Но если меняется направление одного из каналов на обратное, то оказывается, что иногда можно получить ненулевую пропускную способность для передачи информации от источника к потребителю.

Если в классическом случае имеется два канала с сильным шумом и нулевой пропускной способностью, работающих параллельно, то объединенный канал также имеет нулевую пропускную способность. Если изменить направление одного из каналов на обратное, то по-прежнему получается нулевая пропускная способность. В квантомеханическом случае обращение одного из каналов с нулевой пропускной способностью может позволить передавать информацию.

Противоречащие интуиции свойства наподобие только, что описанного иллюстрируют феноменологическую природу квантовой информации. Понимание возможностей передачи информации по сетям квантовых каналов представляет собой проблему в области квантовых вычислений и квантовой информации.

Наиболее практические результаты демонстрируются на известных примерах применения квантовой теории информации в криптографии. Известно, что криптография решает проблему осуществлении связи или вычислений с участием двух и более сторон, которые не могут доверять друг другу. Решается известная криптографическая проблема, связанная с передачей конфиденциальных сообщений, т.е. с так называемым, известным засекречиванием сообщений. Примером такого засекречивания может быть передача продавцу номера своей кредитной карты в обмен на товары, с условием защиты этого номера от перехвата третьей стороной с использованием соответствующих криптографических методов и способов симметричного (криптосистем с секретным ключом – private key) и нессиметричного шифрования (криптосистемам с открытым ключом – public key).

Работа симметричной криптосистемы основана на том, что две стороны используют для связи один и тот же секретный ключ, известный только им. Такие симметричные криптосистемы имеют недостатки, и один из них – сложность распределения ключа, исключающая возможность его перехвата третьими лицами.

Известно одно из открытий в области квантовых вычислений и квантовой информации, характеризующихся тем, что квантовая механика позволяет исключить нарушение конфиденциальности при распределении ключей. Здесь идет речь о процедуре известной в квантовой криптографии, которая заключается в том, что используется квантомеханический принцип, согласно которому наблюдение в общем случае возмущает наблюдаемую систему. Если злоумышленник попытается вести подслушивание во время передачи ключа между двумя лицами, то его присутствие будет проявляться в виде возмущения канала связи, используемого для согласования ключа. Это позволит пользователям отбросить биты ключа, принятые во время подслушивания, и начать все заново.

Принципы квантовой криптографии были впервые предложены Стивеном Уиснером в конце 60-х гг. В 1984 г. Чарльз Беннет и Джилльз Брассар, опираясь на более раннюю работу Уиснера, предложили квантомеханический протокол распределения ключей, исключающий любую возможность их компрометации. Это позволило специалистам разработать множество квантовых криптографических протоколов и их экспериментальных прототипов, которые, как известно, почти достигли такого состояния, когда они могут быть полезны в реальных приложениях ограниченного масштаба.

Вторым важным типом криптосистем являются криптосистемы с открытым ключом. Такие криптосистемы не опираются на передачу секретного ключа. В этом случае публикуется «открытый ключ», доступный всем желающим. Этот открытый ключ используется для шифрования сообщения, но третья сторона не может использовать открытый ключ получателя для расшифровки сообщения. Чтобы расшифровать сообщение, получатель использует секретный ключ, который формируется по определенному закону на основе известного, открытого ключа отправителя. При этом секретный ключ известен только получателю, что обеспечивает ему определенную степень уверенности, в том, что к сообщению не имеет доступа третье лицо. Для получения содержания этого секретного ключа необходим значительный вычислительный ресурс, чтобы получить секретный шифр по открытому ключу. Такие криптосистемы решают проблему распределения ключей, исключая необходимость передачи секретного ключа перед установлением связи.

Криптография с открытым ключом получила широкое распространение с середины 70-х гг., что отражено в известных работах Уитфилда Диффи и Мартина Хеллмана, а также Ральфа Меркле. Рональд Райвест, Ади Шамир и Леонард Эдельман разработали криптосистему RSA, которая является наиболее распространенной, сочетающей в себе безопасность и практичность.

Безопасность криптосистем с открытым ключом основана на том, что обращение стадии шифрования только при наличии открытого ключа в любом случае должно быть затруднительным. Так, задача обращения стадии шифрования RSA тесно связана с задачей факторизации. Предположение о безопасности RSA во многом обусловлено тем, что задачу факторизации трудно решить на классическом компьютере. Однако быстрый алгоритм факторизации на квантовом компьютере, разработанный Шором, мог бы использоваться для взлома RSA. Что касается других криптосистем с открытым ключом, то они также могли бы быть взломаны, если бы был известен быстрый классический алгоритм решения задачи о вычислении дискретного логарифма, подобный шоровскому квантовому алгоритму вычисления дискретного логарифма. Это практическое применение квантовых компьютеров – взлом криптографических кодов, в значительной степени стимулировало интерес к квантовым вычислениям и квантовой информации.

В последнее время произошел стремительный переход квантовой механики непосредственно в сферу практического применения. Это касается и перспективного для практических приложений квантового ресурса – «*запутанные состояния*» (entangled states).

Запутанность есть особая квантовая форма корреляций составных систем, не имеющая классического аналога, которая возникает в системе, состоящей из двух и более взаимодействующих подсистем (или взаимодействовавших ранее, а затем разделенных), и представляет собой суперпозицию альтернативных (взаимоисключающих с классической точки зрения) состояний, которая не может быть реализована в классической физике.

Для таких систем флуктуации отдельных частей взаимосвязаны, но не посредством обычных взаимодействий путем обмена энергией (*классических корреляций*), ограниченных, например, скоростью света, а посредством *нелокальных квантовых корреляций*, когда изменение одной части системы в тот же самый момент времени сказывается на остальных ее частях (даже разделенных в пространстве, в пределе и на бесконечно больших расстояниях).

Математически это выражается в том, что вектор состояния системы, как единого целого, не может быть представлен в виде произведения (тензорного) векторов состояния своих подсистем. В этом случае невозможно разделить систему на локальные объекты и сказать, что вот это – один объект, а вот это – другой. Всегда есть некоторая часть системы, которая принадлежит обоим объектам в равной степени. Подсистемы переплетены, запутаны между собой подобно сиамским близнецам, и составляют единое целое, пусть даже в какой-то незначительной своей части. Описание таких систем в рамках «локальной объективной теории», которая предполагает наличие независимых объектов, становится невозможным. Точнее, классическую физику можно рассматривать, как некоторое приближение при описании физической реальности, когда квантовые корреляции незначительны по сравнению с теми классическими корреляциями, на которых мы останавливаем свое внимание, т.е. на тех физических характеристиках системы, которые характеризуют локальный объект.

Например, если взять сиамских близнецов, классическая физика будет способна описать характеристики каждого из близнецов по отдельности и такое описание будет в чем-то достаточно разумным. Но при таком описании невозможно будет учесть самого главного, что такие близнецы неразрывно связаны друг с другом, пусть даже самым незначительным участком своего тела, и не могут, например, перемещаться независимо друг от друга. Хотя, согласно классическому описанию, ничто не запрещает им находиться в разных комнатах. Согласитесь, ценность такого описания сразу резко падает. В отличие от этого, квантовая механика может описать объект и как единое целое, и как отдельные локальные его части. Классическое описание становится, при этом, частным случаем квантомеханического описания, когда преднамеренно исключаются отдельные свойства всей системы, как единого целого. При таком подходе уже можно понять, с какой целью и для чего используется классический подход, не забывая о границах его применимости, и что такое описание дает исчерпывающую информацию об объекте.

Может сложиться обманчивое впечатление, что поскольку квантовые корреляции в макроскопическом мире незначительны, ими можно полностью пренебречь. Классическая физика так и поступает. Но при этом не учитывается одно существенное обстоятельство – свойства этих корреляций столь необычны, удивительны и всеобъемлющи, что легко могут «перевесить» самые сильные классические корреляции. Пренебрегая квантовыми корреляциями, классическая физика в результате резко ограничивает свои возможности при описании физической реальности, сводя ее практически к бесконечно малой части всей *совокупной квантовой реальности*. Отсюда неспособность классической физики описать огромное количество «сверхъестественных» явлений. Постепенно приходит понимание этого обстоятельства, и на первый план, в качестве самого пристального объекта внимания ученых, выходят квантовые корреляции запутанных состояний системы.

Одно из основных достижений квантовой теории последних лет состоит в том, что был сделан переход к количественному описанию квантовой запутанности. Были введены различные меры запутанности, появилась возможность теоретически рассчитывать эти величины и сопоставлять полученные значения с результатами физических экспериментов. Необходимость в количественном описании квантовой запутанности была вызвана тем, что в последнее время запутанные состояния стали важным практическим ресурсом для многих новых прикладных дисциплин: квантового компьютера, квантовой криптографии, квантовой телепортации, физики квантовых вычислений и др.

Запутанные состояния привлекли самое пристальное внимание ученых. Постепенно в качестве самостоятельных разделов квантовой теории выделились ее прикладные направления: теория декогеренции, теория запутанных состояний, квантовая теория информации.

Современных исследователей явление квантовой запутанности привлекает тем, что позволяет целенаправленно задействовать для практических нужд нелокальные квантовые ресурсы системы. Нелокальность является специфической особенностью квантовой запутанности, и ее проявления кажутся «магическими» с точки зрения классической физики.

Контроль и управление степенью квантовой запутанности в системе невозможно реализовать чисто-классическими методами. Речь идет о создании принципиально новых технических устройств, и запутанные состояния являются многообещающим практическим ресурсом с качественно новыми и неожиданными сферами применения.

Запросы практики и потребности общества в новых перспективных технологиях, обусловили науку перейти от привычной полуклассической копенгагенской интерпретации квантовой механики, подразумевающей обязательное наличие классического наблюдателя (измерительного прибора) к чисто квантовому подходу, в котором уже не осталось место классическому «пережитку». В результате, квантовый подход к описанию окружающей реальности стал самодостаточной согласованной теорией, построенной из единых общих принципов, логично включающей в себя классическую физику, как частный случай квантового описания.

Еще одно достижение квантовой теории состоит в том, что была теоретически доказана и экспериментально подтверждена возможность «манипуляции» квантовой запутанностью. Меру запутанности системы можно изменять, как усиливая ее (очищение, дистилляция запутанности), так и уменьшая (разбавление запутанности, декогеренция окружением).

Декогеренция – это физический процесс, который сопровождается уменьшением квантовой запутанности (потерей когерентности квантовых суперпозиций) в результате взаимодействия системы с окружением.

Таким образом, во многом благодаря практическим нуждам, вошли в сферу внимания научного сообщества и стали объектом тщательного (как теоретического, так и экспериментального) исследования важнейшие фундаментальные физические процессы в окружающей нас реальности, которые раньше наука не рассматривала. Пришло понимание, что мера квантовой запутанности системы, ее динамика и физические процессы, ведущие к усилению или уменьшению квантовой запутанности, – это основополагающие характеристики системы. А фундаментальность новых (для науки) физических процессов обусловлена тем, что они являются неотъемлемым свойством любого элемента реальности.

За прошедшие сто лет квантовая теория достаточно сильно окрепла для того, чтобы приступить к описанию физической реальности в глобальных космологических масштабах.

Теория декогеренции сегодня в состоянии предложить космологическую модель возникновения, проявления, «видимости» окружающей нас классической реальности из единого квантового источника.

Запутанность – это уникальный квантомеханический ресурс, который играет ключевую роль во многих наиболее интересных применениях квантовых вычислений и квантовой информации. Запутанность считается фундаментальным ресурсом Природы, сравнимым по важности с энергией, информацией, энтропией или любым другим фундаментальным ресурсом. В последние годы предпринимаются огромные усилия, направленные на лучшее понимание ее свойств. Хотя законченной теории запутанности пока нет, к настоящему времени удалось достичь некоторого прогресса в понимании этого странного понятия квантовой механики.

Многие исследователи надеются, что дальнейшее изучение свойств запутанности даст сведения, которые будут способствовать разработке ее новых применений в области квантовых вычислений и квантовой информации.

Квантовые вычисления и квантовая информация позволяют думать о вычислениях физически, и этот подход открывает много новых возможностей в области связи и обработки информации. Специалисты по информатике и теории информации получили новую плодотворную парадигму для исследований. Любая физическая теория, а не только квантовая механика, может служить базисом для теории обработки информации и теории связи. В результате этих исследований могут быть созданы устройства обработки информации, намного превосходящие по своим возможностям современные вычислительные и коммуникационные системы, что будет иметь свои положительные и отрицательные последствия для всего общества.

Конечно, квантовые вычисления и квантовая информация ставят перед физиками задачи, но при этом не совсем понятно, что эта область предлагает физике в долгосрочной перспективе. Физика традиционно является дисциплиной, где основное внимание сосредоточено на понимании «элементарных» объектов и простых систем, однако многие интересные аспекты природы проявляются лишь с ростом размеров и сложности. Такие явления отчасти исследуются в химии и инженерных науках, но всякий раз довольно специфическим образом квантовые вычисления и квантовая информация предоставляют новые инструменты, позволяющие перебрасывать мост от простого к относительно сложному: в сфере вычислений и алгоритмов есть систематические средства для построения и изучения таких систем. Применение идей из этих областей уже начинает приводить к выработке новых взглядов на физику. Можно надеяться, что в последующие годы этот подход будет успешно применяться во всех ее разделах.

В настоящее время весьма актуальными являются проблемы:

– оценки эффективности квантового компьютера по сравнению с классическим, экспериментальной реализации обработки квантовой информации;

– томографии квантовых состояний и томографии квантовых процессов на основе квантовых вычислений и квантовой информации.

Томография квантовых состояний представляет собой метод определения квантового состояния системы, состоящий в повторном приготовлении одного и того же состояния и измерении его разными способами. Это позволяет преодолеть «скрытую» природу квантового состояния (состояние нельзя определить путем прямого измерения) и составить его полное описание. Томография квантовых процессов, будучи тесно связана с томографией квантовых состояний, претендует на большее, а именно, на полное описание динамики квантовой системы. Например, с ее помощью можно узнать характеристики некоторого квантового элемента и квантового канала связи или определить типы и амплитуды различных шумовых процессов в системе.

Кроме применений в области квантовых вычислений и квантовой информации, томография квантовых процессов может стать важным диагностическим инструментом, по описанию и совершенствованию базовых операций в любых областях науки и техники, где существенны квантовые эффекты.

Большой интерес представляют также различные базовые коммуникационные операции. Так в квантовой криптографии это практическое приложение по передаче небольших объемов информации (ключа) с большой секретностью.

Применение квантовой телепортации может быть использовано для передачи квантовых состояний между удаленными узлами сети в присутствии шума. Идея состоит в том, чтобы сосредоточить усилия на передаче состояния Белла или ЭПР-пар (Эйнштейна, Подольского, Розена) между узлами, намеревающимися связываться друг с другом. ЭПР-пары могут искажаться в процессе передачи, но специальные протоколы очищения запутанности могут затем «очищать» эти пары, позволяя использовать их для телепортации квантовых состояний из одного места в другие. Фактически, протоколы на основе очищения запутанности и телепортации превосходят более традиционные технологии исправления квантовых ошибок в том, что касается обеспечения свободной от шумов передачи кубитов.

Важным применением устройств обработки квантовой информации является моделирование квантовой системы, содержащей всего несколько десятков «кубитов» (или их эквивалента в терминах какой-либо другой базовой системы), для которого не хватит ресурсов даже самых больших суперкомпьютеров.

Так для выполнения вычисления системы с 50 кубитами, только для описания ее состояния, необходимо $2^{50} \approx 10^{15}$ комплексных амплитуд. Если амплитуды хранятся со 128-битовой точностью, то для каждой амплитуды требуется 256 бит или 32 байта. В сумме это составляет примерно 32 тысячи терабайтов информации, что значительно превышает возможности существующих компьютеров и примерно соответствует объему памяти, которого можно ожидать у суперкомпьютеров во втором десятилетии двадцать первого века, если закон Мура останется в силе. Для 90 кубитов, при том же уровне точности, требуется $32\text{x}10^{27}$ байтов, для хранения которых, даже если представлять биты одиночными атомами, нужны килограммы вещества.

Таким образом, возникает вопрос о полезности квантового моделирования. Очевидно, что традиционные методы по-прежнему будут использоваться для определения элементарных свойств материалов, например, энергии связи и основных спектроскопических характеристик. Однако после того, как основные свойства будут хорошо понятны, квантовое моделирование должно найти широкое применение в качестве «лаборатории» для конструирования новых молекул и проверки их свойств.

Для тестирования всего разнообразия возможных структур молекул в традиционной лаборатории требуется и значительное многообразие и разнообразие оборудования и материалов. На квантовом компьютере все это оборудование можно смоделировать программно, что будет намного дешевле и быстрее. Безусловно, что окончательное конструирование и тестирование необходимо выполнять с использованием реальных физических систем, однако квантовые компьютеры могут обеспечить изучение гораздо большего числа потенциальных структур и тем самым добиваться лучшего конечного результата.

Известно, что «прямые» расчеты для разработки новых молекул пробовали проводить и на классических компьютерах, однако эти попытки имели ограниченный успех из-за огромных требований к вычислительным ресурсам, необходимым для моделирования квантовой механики на классическом компьютере. Следует ожидать, что квантовые компьютеры справятся с этим гораздо лучше.

Помимо квантового моделирования и квантовой криптографии, факторизации и вычисления дискретного логарифма, представляет интерес квантовый поиск, который может широко использоваться из-за большого практического значения поисковой эвристики.

Актуальной проблемой является практическая реализация этих достижений в реальных физических системах. Так известно, что для малых масштабов (уровня нескольких кубитов) уже предложено несколько устройств обработки квантовой информации. Среди них – реализации, основанные на оптических технологиях, т. е. на использовании электромагнитного излучения.

Для элементарных манипуляций с фотонами подходят простые устройства типа зеркал и светоделительных пластинок. Здесь основная трудность состоит в выдаче одиночных фотонов по требованию. Вместо этого экспериментаторы решили использовать схемы, где одиночные фотоны генерируются время от времени случайным образом. При помощи таких оптических технологий были реализованы квантовая криптография, сверхплотное кодирование и квантовая телепортация.

Главное достоинство оптических технологий в том, что фотоны проявляют себя как наиболее стабильные носители квантомеханической информации, а главный недостаток – фотоны не взаимодействуют друг с другом непосредственно. Взаимодействие необходимо осуществлять через какое-то промежуточное звено, например, атом, а это вносит дополнительные шумы и усложняет эксперимент. Таким образом, реальное взаимодействие двух фотонов есть двухступенчатый процесс: первый фотон взаимодействует с атомом, который, в свою очередь, взаимодействует со вторым фотоном.

Известна альтернативная схема, которая основана на методах захвата различных типов атомов: существуют ионные ловушки, где в ограниченном пространстве заключено небольшое количество заряженных атомов, и ловушки нейтральных атомов для захвата незаряженных атомов. В схемах обработки квантовой информации, основанных на атомных ловушках, для хранения кубитов используются атомы. Электромагнитное излучение в этих схемах тоже используется, но совсем не так, как при «оптическом» подходе. Фотоны применяются здесь для манипулирования информацией, хранящейся в атомах, а не как самостоятельные элементы хранения информации.

Однокубитовые квантовые элементы можно реализовать, воздействуя импульсами электромагнитного излучения на отдельные атомы. Соседние атомы могут взаимодействовать друг с другом посредством, например, дипольных сил, обеспечивающих работу многокубитовых квантовых элементов. Более того, взаимодействие соседних атомов можно модифицировать, воздействуя на них подходящими импульсами электромагнитного излучения. Это позволяет выбирать элементы, которые реализуются в системе. Наконец, для осуществления квантовых измерений в таких системах подходит хорошо известный метод квантовых переходов, который позволяет исключительно точно проводить измерения в вычислительном базисе, используемом для квантовых вычислений.

Другой класс схем обработки квантовой информации основан на ядерном магнитном резонансе (ЯМР). В таких схемах квантовая информация хранится в ядерном спине атомов, входящих в молекулу, а манипулирование этой информацией осуществляется при помощи электромагнитного излучения. Использование таких схем сопряжено со специфическими трудностями, поскольку в ЯМР невозможно обращаться напрямую к отдельным ядрам. Вместо этого используется гигантское число (как правило, порядка 10^{15}) практически одинаковых молекул, находящихся в растворе.

Электромагнитные импульсы воздействуют на образец, заставляя каждую молекулу реагировать примерно одним и тем же образом. Каждую молекулу следует рассматривать как независимый компьютер, а образец в целом – как совокупность огромного числа компьютеров, работающих параллельно (в классическом смысле). Обработка квантовой информации при помощи ЯМР сопряжена с тремя специфическими трудностями, которые сильно отличают ее от других схем обработки квантовой информации.

Во-первых, молекулы обычно приготавливают путем приведения их в равновесное состояние при комнатной температуре, которая настолько высока по сравнению с типичной энергией переворота спина, что спины приобретают почти полностью случайную ориентацию.

Из-за этого начальное состояние становится значительно более «шумным», чем было бы желательно для обработки квантовой информации. Преодоление этого шума представляет собой интересную задачу.

Во-вторых, – класс измерений, которые могут выполняться при исследовании ЯМР, не содержит большинства общих измерений, которые хотелось бы выполнять при обработке квантовой информации. Тем не менее, измерений этого класса достаточно для многих задач по обработке квантовой информации.

В-третьих, поскольку при использовании ЯМР к молекулам нельзя обращаться по отдельности, может возникнуть вопрос: как же манипулировать отдельными кубитами? К счастью, разные ядра в молекуле могут иметь разные свойства, что позволяет обращаться к ним по отдельности, или, по крайней мере, с таким разрешением, которого достаточно для выполнения операций, необходимых при квантовых вычислениях.

В существующих предложениях можно найти многие из элементов, требуемых для осуществления крупномасштабной обработки квантовой информации: в ионной ловушке можно прекрасно приготавливать состояния и проводить квантовые измерения над небольшим числом кубитов; с помощью ЯМР можно реализовать великолепную динамику в малых молекулах; технология производства твердотельных систем позволяет отлично масштабировать конструкции. Объединение всех этих элементов в одну систему стало бы большим шагом на пути к гипотетическому квантовому компьютеру. К сожалению, все эти системы очень различаются, и от больших квантовых компьютеров нас отделяют многие и многие годы.

Однако, как считают ученые и специалисты, наличие всех этих свойств у существующих (пусть и различных) систем служит хорошим предзнаменованием, указывающим на возможность появления в далекой перспективе процессоров для крупномасштабной обработки квантовой информации. Более того, специалисты утверждают о целесообразности развития гибридных конструкций, сочетающих в себе лучшие черты двух или более существующих технологий.

Например, сейчас ведется большая работа по захвату атомов в электромагнитных резонаторах. Это позволяет гибко манипулировать атомом внутри резонатора при помощи оптических методов, а также открывает возможность управления одиночными атомами в реальном масштабе времени с использованием обратной связи такими способами, которые недоступны в традиционных атомных ловушках.

В известной литературе отмечается, что обработку квантовой информации ни в коем случае нельзя считать просто еще одной технологией обработки информации. Например, есть соблазн отмахнуться от квантовых вычислений, посчитав их очередной технологической модой в эволюции компьютера, которая со временем пройдет, так было с другими модными идеями, скажем, памятью на цилиндрических магнитных доменах, широко рекламировавшейся в начале 80-х гг. XX в. как следующее большое достижение в технологии запоминающих устройств. Специалисты считают, что это будет ошибкой, поскольку квантовые вычисления представляют собой абстрактную парадигму обработки информации, которая может иметь множество различных технических реализаций. Можно сравнивать два разных предложения по квантовым вычислениям в отношении их технологических достоинств, как сравнивают «хорошее предложение с «плохим», но даже очень посредственное предложение по квантовому компьютеру в качественном отношении радикально отличается от самого замечательного проекта классического компьютера.

12.2. Квантовая информация – основной объект квантовой информатики

В области квантовых вычислений и квантовой информации термин *«квантовая информация»* имеет два разных значения:

– применяется в качестве общего названия для всех видов деятельности, связанных с обработкой информации на основе квантовой механики и охватывает квантовые вычисления, квантовую телепортацию, теорему о невозможности копирования;

– является более специализированым и относится к изучению элементарных задач по обработке квантовой информации, например, не охватывая построение квантовых алгоритмов, поскольку детали конкретных квантовых алгоритмов выходят за рамки «элементарных».

Во избежание путаницы целесообразно использовать термин *«квантовая теория информации»* для этой более специализированной области параллельно с широко распространенным термином «(классическая) теория информации» для описания соответствующей известной классической области.

Термин «квантовая теория информации» имеет свой недостаток – отражает только теорию. Но экспериментальная демонстрация элементарных процессов, изучаемых в квантовой теории информации, представляет также большой интерес.

Целью рассмотрения вопроса о квантовой информации является введение в основные идеи квантовой теории, информации. Будучи ограниченной элементарными задачами по обработке квантовой информации, эта теория может выглядеть в виде начинающего набора из множества, как будто, не связанных друг с другом предметов, подпадающих под рубрику «квантовая теория информации». Это объясняется тем, что данная дисциплина все еще находится в состоянии разработки. При этом выделяют несколько фундаментальных целей, придающих единство квантовой теории информации [169, 276 и др.]:

1) определение элементарных классов статических ресурсов в квантовой механике. Примером служит кубит. Другой пример – бит; классическая физика представляет собой частный случай квантовой физики, поэтому элементарные статические ресурсы, используемые в классической теории информации, должны иметь большое значение в квантовой теории информации. Еще одним примером элементарного класса статических ресурсов является состояние Белла, разделенное между двумя удаленными друг от друга сторонами.

2) определение элементарных классов динамических процессов в квантовой механике. Пример – запоминание, т. е. способность сохранять квантовое состояние на протяжении некоторого периода времени. Менее тривиальными процессами являются передача квантовой информации между двумя сторонами; копирование (или попытка копирования) квантового состояния, а также защита обрабатываемой квантовой информации от влияния шума.

3) определение затрат ресурсов на реализацию элементарных динамических процессов. Например, определение минимальных ресурсов для надежной передачи квантовой информации между двумя сторонами при использовании канала связи с шумом.

Подобные цели и задачи ставятся и в классической теории информации; однако квантовая теория информации шире классической, поскольку она включает в себя все статические и динамические элементы классической теории, а также дополнительные статические и динамические элементы.

Рассмотрим пример, который покажется очень знакомым специалистам по классической теории информации: проблемы передачи классической информации по квантовому каналу.

12.3. Классическая информация в квантовых каналах

Фундаментальными результатами классической теории информации являются: теорема о кодировании для канала без шума и теорема о кодировании для канала с шумом, доказанные Шенноном. *Теорема о кодировании для канала без шума* устанавливает, сколько битов требуется для хранения информации, выдаваемой источником информации, а теорема о кодировании для канала с шумом устанавливает, какое количество информации можно надежно передать по каналу связи в присутствии помех.

Определение содержания понятия источника информации является фундаментальной проблемой классической и квантовой теории информации, к которой мы будем неоднократно возвращаться.

Так, специалистами предлагается следующее предварительное определение.

Источник классической информации описывается набором вероятностей p_j, $j = 1, 2, . . ,d$. Каждое обращение к источнику приводит к выдаче «буквы», выбираемой случайным образом с вероятностью p_j независимо от предыдущих обращений к источнику.

Например, если источник представляет собой английский текст, то числа могут соответствовать буквам алфавита и знакам препинания, а вероятности p_j – давать относительные частоты, с которыми различные буквы встречаются в обычном английском тексте. Хотя на самом деле в английском языке буквы не встречаются независимо, для наших целей это будет достаточно хорошим приближением.

Обычный английский текст в значительной степени избыточен, и этой избыточностью можно воспользоваться для сжатия текста. Например, буква «*e*» встречается в обычном английском тексте гораздо чаще буквы «*z*». Следовательно, в хорошем алгоритме сжатия английского текста для представления буквы «*e*» будет использоваться меньше битов информации, чем для представления буквы «*z*». Теорема Шеннона о кодировании для канала без шума определяет, насколько хорошо можно заставить работать такой алгоритм сжатия.

Точнее говоря, эта теорема утверждает следующее: классический источник, описываемый вероятностями p_j, может быть сжат так, что результат каждого обращения к источнику можно представить в среднем при помощи $H(p_j)$ битов информации, где $H(p_j) \equiv -\sum_j p_j \log(p_j)$ есть функция распределения вероятностей источника, называемая энтропией Шеннона. Более того, теорема о кодировании для канала без шума устанавливает, что попытка представить источник при помощи меньшего числа битов приведет к большой вероятности ошибки при развертывании текста.

Теорема Шеннона о кодировании для канала без шума служит хорошим примером одновременного достижения всех перечисленных выше целей, стоящих перед теорией информации. В современной науке определены два статических ресурса:

– бит и источник информации. Определен двухэтапный динамический процесс;

– сжатие данных от источника информации и последующее их развертывание для восстановления информации;

– найден количественный критерий для определения ресурсов, потребляемых оптимальным алгоритмом сжатия данных.

Второй значительный результат Шеннона – теорема о кодировании информации для канала шумом – устанавливает, какое количество информации может быть надежно передано по каналу в присутствии помех.

Так, при передаче информации в пространстве и времени, выдаваемой некоторым источником, по каналу с шумом, необходимо обеспечить хранение информации в присутствии шума. Задача состоит в том, чтобы закодировать выдаваемую информацию при помощи кодов, исправляющих ошибки так, чтобы любой шум, внесенный каналом, можно было устранить на конце этого канала. В кодах, исправляющих ошибки, это достигается за счет введения в посылаемую по каналу информацию избыточности, достаточной для того, чтобы даже после искажения некоторой части информации можно было восстановить исходное сообщение.

Так, если по каналу с шумом передаются одиночные биты, а шум в канале таков, что для достижения надежной передачи каждый бит, выдаваемый источником, необходимо перед пересылкой по каналу кодировать двумя битами. Известно, что такой канал имеет пропускную способность в половину бита, поскольку каждое обращение к каналу можно использовать для надежной доставки примерно половины бита информации. *Шенноновская теорема о кодировании для канала с шумом дает общую процедуру для вычисления пропускной способности произвольного канала с шумом.*

В теореме Шеннона о кодировании для канала с шумом также достигаются все три стоящие перед теорией информации цели, о которых говорилось выше. Используются два типа статических ресурсов:

– источник информации и биты, пересылаемые по каналу;

– три динамических процесса. Основной процесс – это шум в канале. Чтобы его уменьшить, выполняются дополняющие друг друга процедуры кодирования и декодирования состояния, применяя код, исправляющий ошибки;

– для заданной модели шума теорема Шеннона устанавливает, какую избыточность должна внести оптимальная схема исправления ошибок, чтобы достичь надежной передачи информации.

В обеих теоремах о кодировании Шеннон ограничился хранением выходных данных источника информации в классических системах – битах и им подобных. В квантовой теории информации встает естественный вопрос – что произойдет, если изменить среду хранения так, чтобы классическая информация передавалась при помощи квантовых состояний.

Например, отправитель может захотеть сжать некоторую классическую информацию, выдаваемую источником, и передать сжатую информацию получателю, который затем развернет ее. Если в качестве среды хранения сжатой информации используется квантовое состояние, то теорема Шеннона о кодировании для канала без шума неприменима для определения оптимального алгоритма сжатия и развертывания. При этом, возникает вопрос о возможности на основе использования кубитов получения лучшего сжатия, чем в классическом случае. Известно, что, на самом деле кубит не обеспечивает никакого существенного выигрыша в объеме данных, требуемых для передачи информации по каналу без шума.

Следующим шагом является исследование проблемы передачи классической информации по квантовому каналу с шумом. В идеале нужен результат, который позволял бы определять пропускную способность такого канала применительно к передаче информации. Вычисление пропускной способности представляет собой серьезную проблему по нескольким причинам.

Квантовая механика, используя непрерывное пространство, дает огромное разнообразие моделей шума, и совсем не очевидно, как приспособить классические методы исправления ошибок для борьбы с этим шумом. Возникает проблема о получении выигрыша, при условии если кодировать классическую информацию при помощи запутанных состояний и передавать их затем друг за другом по каналу с шумом. Также возникает вопрос об эффективности декодирования при помощи запутанных измерений.

В известной литературе приводится доказательство теоремы ХШВ (Холево–Шумахера–Вестморленда), которая устанавливает нижний предел пропускной способности такого канала. Вообще, принято считать, что теорема ХШВ дает точное значение пропускной способности, хотя полное доказательство этого до сих пор неизвестно. Под вопросом остается только возможность использования кодирования при помощи запутанных состояний для увеличения пропускной способности сверх нижнего предела, установленного теоремой ХШВ. Все известные на сегодня факты свидетельствуют, что это не поможет увеличить пропускную способность, но определение истинности или ложности такого предположения пока остается интересной открытой проблемой квантовой теории информации.

12.4. Квантовая информация в квантовых каналах

Конечно, классическая информация – это не единственный статический ресурс, доступный в квантовой механике. Квантовые состояния сами являются естественным статическим ресурсом, даже более естественным, чем классическая информация. Рассмотрим различные квантовые аналоги теорем Шеннона о кодировании применительно к сжатию и развертыванию квантовых состояний.

Квантовое понятие источника информации по аналогии с классическим, определяется разными способами. Для определенности принят предварительный вариант: квантовый источник описывается набором вероятностей p_j и соответствующих квантовых состояний $|\psi_j\rangle$.

Каждое обращение к источнику дает состояние $|\psi_j\rangle$ с вероятностью p_j, причем разные обращения к источнику не зависят друг от друга. Можно ли сжать выходные данные такого квантомеханического источника? Известен случай источника кубитов, который переводит состояние $|0\rangle$ с вероятностью p и состояние $|1\rangle$ с вероятностью $1-p$.

Как видно, по содержанию, он ничем не отличается от классического источника, выдающего одиночный бит 0 с вероятностью p или *1* с вероятностью $1 - p$, поэтому с помощью подобных методов можно сжать источник так, что для хранения сжатой информации потребуется только $H\ (p,\ 1 - p)$ кубитов, где $H(\cdot)$ – снова функция энтропии Шеннона.

Если источник выдает состояние $|0\rangle$ с вероятностью p и состояние $(|0\rangle + |1\rangle)/\sqrt{2}$ с вероятностью $1 - p$, то стандартные методы сжатия классических данных больше не применимы, поскольку в общем случае мы не можем различать состояния $|0\rangle$ и $(|0\rangle + |1\rangle)/\sqrt{2}$. Можно ли в этом случае по-прежнему выполнять операцию сжатия какого-либо типа?

Как известно, сжатие некоторого типа возможно даже в этом случае. Оно может перестать быть безошибочным в том смысле, что квантовые состояния на выходе источника могут слегка искажаться процедурой сжатия-развертывания. Тем не менее, требуется, чтобы это искажение становилось очень малым и, в конце концов, пренебрежимо малым при переходе к сжатию больших блоков выходных данных источника. Чтобы качественно охарактеризовать искажения, введем для алгоритма сжатия меру совпадения (fidelity), которая определяет среднее искажение, вносимое этим алгоритмом.

Сущность сжатия квантовых данных состоит в том, что сжатые данные должны восстанавливаться с очень большой точностью. В этом случае применяется известная мера совпадения, как аналог вероятности корректного выполнения развертывания. В пределе больших длин блоков она должна стремиться к 1, что означает отсутствие ошибок.

Теорема Шумахера о кодировании для канала без шума устанавливает, какое количество ресурсов требуется для сжатия квантовых данных при условии, что источник можно восстановить с точностью, близкой к 1. В случае источника, выдающего ортогональные квантовые состояния $|\psi_j\rangle$ с вероятностями p_j, теорема Шумахера сводится к утверждению о том, что возможно сжатие не более, чем до классического предела $H(p_j)$.

Однако, в общем случае неортогональных состояний, выдаваемых источником, теорема Шумахера устанавливает до какого предела их можно сжать. Здесь используют не шенноновскую энтропию $H(p_j)$, а новую энтропийную величину – энтропию фон Неймана. Энтропия фон Неймана совпадает с энтропией Шеннона тогда и только тогда, когда состояния $|\psi_j\rangle$ ортогональны. В противном случае энтропия фон Неймана для источника p_j, $|\psi_j\rangle$ в общем случае строго меньше, чем энтропия Шеннона $H(p_j)$.

Так, например, в случае источника, который выдает состояние $|0\rangle$ с вероятностью p и $(|0\rangle+|1\rangle)/\sqrt{2}$ с вероятностью $1-p$, можно надежно произвести сжатие с использованием меньшего, чем $H(p,1-p)$, числа кубитов в расчете на одно обращение к источнику.

Основная причина такого уменьшения требуемых ресурсов заключается в следующем. Предполагается, что источник, выдающий состояния $|0\rangle$ с вероятностью p и $(|0\rangle+|1\rangle)/\sqrt{2}$ с вероятностью $1-p$, используется большое число раз, чем n. Тогда по закону больших чисел источник с большой вероятностью выдаст примерно np копий $|0\rangle$ и $n(1-p)$ копий $(|0\rangle+|1\rangle)/\sqrt{2}$, что позволяет, выдаваемое источником состояние аппроксимировать суперпозицией членов.

Фактически теорема Шумахера о кодировании для канала без шума дает несколько лучшую оценку. В результате это фактически приводит к увеличению избыточности. Конечно, это особенность конкретного алгоритма и в общем решении сжатие данных достигается за счет гораздо более рационального использования избыточности.

Теорема Шумахера о кодировании для канала без шума является аналогом теоремы Шеннона о кодировании для канала без шума применительно к сжатию и развертыванию квантовых состояний. Можно ли найти аналог теоремы Шеннона о кодировании для канала с шумом? Благодаря использованию теории кодов, исправляющих квантовые ошибки, в этом важном вопросе достигнут большой прогресс, но полного аналога до сих пор не найдено.

12.5. Квантовая различимость

Рассмотренные динамические процессы – сжатие, развертывание, шум, кодирование и декодирование с использованием кодов, исправляющих ошибки, присущи как в классической, так и квантовой теории информации. Однако введение новых типов информации, таких как квантовые состояния, расширяет класс динамических процессов за рамки тех, что рассматриваются в классической теории информации. Примером является проблема различимости квантовых состояний.

Известно, что в классическом случае есть возможность различать неодинаковые элементы информации, по крайней мере, в принципе. Конечно, на практике смазанная буква «а» на странице может быть трудноотличима от буквы «о», но в принципе возможно достоверно различать два различных варианта.

В квантомеханическом случае, напротив, не всегда можно различить произвольные состояния. Например, не существует такого процесса, допускаемого квантовой механикой, который бы позволил надежно различать состояния $|0\rangle$ и $\left(|0\rangle+|1\rangle\right)/\sqrt{2}$. Строгое доказательство этого факта требует инструментов, которых пока нет, но на примерах очень легко убедиться, что это различение невозможно [169, 276].

Эта неразличимость неортогональных квантовых состояний лежит в основе квантовых вычислений и квантовой информации. Она составляет суть утверждения о том, что квантовое состояние содержит скрытую информацию, недоступную для измерения, и тем самым играет ключевую роль в квантовых алгоритмах и квантовой криптографии. Одной из центральных проблем квантовой теории информации является разработка мер для количественного определения степени различимости неортогональных квантовых состояний.

12.6. Создание и преобразование запутанности – важный динамический процесс квантовой теории информации

Запутанность представляет собой еще один элементарный статический ресурс квантовой механики. Его свойства отличаются от свойств ресурсов, знакомых, главным образом, по классической теории информации. Рассмотрим две известные в литературе теоретико-информационные проблемы, связанные с запутанностью.

Создание запутанности – это простой динамический процесс, изучаемый в квантовой теории информации, в рамках которой решаются задачи:

Определения количества кубит, которыми должны обменяться две стороны, чтобы создать заданное запутанное состояние, разделенное между ними, при условии, что перед этим они не разделяли никакой запутанности: определение содержания динамического процесса, представляющего преобразование запутанности из одной формы в другую.

Рассматривается пример состояния Белла, которое разделено между отправителем и получателем информации, которое они хотят преобразовать в запутанное состояние какого-то другого типа. При этом возникают задачи:

– определение потребного ресурса для выполнения задачи преобразования в запутанное состояние;

– решение поставленной задачи без установления связи между отправителем и получателем;

– определение количества квантовых передач при наличии квантовой связи.

Решение этих задач о создании и преобразовании запутанности приводит к формированию самостоятельной области исследований в области квантовых вычислений.

Например, распределенное квантовое вычисление можно рассматривать просто, как метод создания запутанности между двумя и более сторонами; тогда нижние пределы количества передач, необходимых для выполнения такого распределенного квантового вычисления, определяются нижними пределами количества передач, которое требуется для создания подходящих запутанных состояний.

Квантовая теория информации представляет собой наиболее абстрактную часть области квантовых вычислений и квантовой информации, а в некотором смысле и наиболее фундаментальную. Развитие квантовой теории информации и, в конечном счете всей сферы квантовых вычислений и квантовой информации, ставит следующие проблемные вопросы:

– что делает возможным обработку информации?

– что разделяет квантовый и классический миры?

– какие ресурсы, недоступные в классическом мире, могут использоваться в квантовых вычислениях?

Варианты решения этих задач на сегодня, как известно, еще недостаточно полны и требуют дальнейших исследований четкости представления, возможностей и ограничений процессов обработки квантовой информации.

Контрольные вопросы

1. Каковы источники формирования квантовой информации?
2. Дать определение квантовой механики и раскрыть ее роль в формировании квантовой информатики.
3. Какова роль машины Тьюринга в формировании квантовой информатики?
4. В чем сущность закона Мура, характеризующего развитие НИТ?
5. Каковы перспективы использования для выполнения вычислений квантовой механики?
6. В чем актуальность проблемы использования квантовых вычислений и квантовой информации в теории связи и роль в ней теорем Шеннона?
7. В чем сущность теории кодов, исправляющих квантовые ошибки?
8. В чем сущность проблемы реализации распределенных квантовых вычислений?
9. В чем сущность сетевой квантовой теории информации и исследования методов и способов передачи информации по сетям квантовых каналов?
10. Что такое декогеренция и запутанность?
11. Перечислить актуальные проблемы развития квантовой информатики?
12. В чем сложности обработки квантовой информации при помощи ЯМР?
13. Раскрыть содержание квантовой информации как базового понятия квантовой теории информации
14. Каковы фундаментальные цели квантовой теории информации?
15. Как описывается классическая информация в квантовых каналах?
16. Как осуществляется передача квантовой информации в квантовых каналах?
17. В чем состоит сущность сжатия информации?
18. Дать понятие квантовой различимости в квантовой теории информации.
19. В чем сущность создания и преобразования запутанности, как важного динамического процесса квантовой теории информации?

ЛИТЕРАТУРА

1. Афанасьев, В.Г. Социальная информация / В.Г. Афанасьев. – М.: Наука, 1994.

2. Афанасьев, В.Г. Социальная информация и управление обществом / В.Г. Афанасьев. – М.: Политиздат, 1975.

3. Артамонов, Г.Т. Информатика: теория и практика / Г.Т. Артамонов // НТИ. – Сер.1 –1999.– № 6.– С. 36–43.

4. Артамонов, Г.Т. Информатика: теория и практика / Г.Т. Артамонов // НТИ. – Сер.1 –1997.– № 8; 1998.– №№: 1, 4, 6, 12.

5. Афанасьев, В.Г. Научно-техническая революция, управление, образование / В.Г. Афанасьев. – М.: Прогресс, 1972.

6. Авраам, Г.Д. (США) Перспективы создания национальной информационной системы США / Г.Д. Авраам // НТИ. – № 9. – 1993. – С.22-27.

7. Афанасьева, Т.А. Информационное обеспечение органов управления в свете концепции информационного менеджмента / Т.А. Афанасьева // Зарубежная радиоэлектроника. – №4. – 1995. – С.45-53.

8. Альянах, И.Н. Моделирование вычислительных систем / И.Н. Альянах. – Л.: Машиностроение. Ленингр. отд-ние, 1988. – 223 с.

9. Баранов, В.А. Пакет символьной математики MathCad / В.А. Баранов, И.Ю. Баранов. – Орел: ВИПС, 1998. – 128 с.

10. Бауэр, Ф. Информатика. Задачи и решения / Ф. Бауэр, Г. Гооз. – М.: Мир, 1976.

11. Берг, А.И. Управление, информация, интеллект / А.И. Берг, Б.В. Бирюков, Н.Н. Воробьев и др. – М.:Мысль, 1976.

12. Берлекэмп, Э. Алгебраическая теория кодирования / Э. Берлекэмп. – М.: Мир, 1971.

13. Беспроводная цифровая связь. – М.: ЭКО-ТРЕНДЗ, 2001. – 285с.

14. Биллингслей, П. Эргодическая теория и информация / П. Биллингслей. – М.: Мир, 1969.

15. Биркгоф, Г. Современная прикладная алгебра / Г. Биркгоф, Т. Барти. – М.: 1976.

16. Биркгофф, Г. Математика и психология / Г. Биркгофф. – М.: Сов. радио, 1977.

17. Бокаревич, Ю.Б. СУБД Access для Windows 95 / Ю.Б. Бокаревич, Н.В. Пушкина. – СПб.: VHB

18. Блейхаут, Р. Теория и практика кодов, контролирующих ошибки / Р. Блейхаут. – М.: Мир, 1986. – 576 с.

19. Блох, Э.Л. О методе декодирования для кодов Боуза-Чоудхури, исправляющих тройные ошибки / Э.Л. Блох // Известия АН СССР. Техническая кибернетика. – 1964. – № 3 (9).

20. Богатырь, Б.Н. Концепция системной интеграции информационных технологий в высшей школе / Б.Н. Богатырь и др. – М.: РосНИИСИ, 1993. – 72 с.

21. Богумирский, Б.С. MS DOS 6.2/6.22 / Б.С. Богумирский. – СПб.: «Питер», 1995.

22. Богумирский, Б.С. Руководство пользователя ПЭВМ: В 2-х ч. Ч.1 / Б.С. Богумирский. – СПб: Ассоциация OILKO, 1992. – 357 с.: ил.

23. Бойс, Дж. Осваиваем Windows 95: пер. с англ. / Дж. Бойс. – М.: БИНОМ. – 400 с.

24. Бониц, М. Информация – знание – информатика / М. Бониц // Международ. форум по информ. и докум. – 1990. – Т.15, № 2.– С. 3–6.

25. Бриллюэн, Л. Наука и теория информации: пер. с англ. / Л. Бриллюэн. – М.: Физматгиз, 1960.

26. Бриллюэн, Л. Научная неопределенность и информация: пер. с англ. / Л. Бриллюэн. – М.: Мир, 1966.

27. Бусленко, Н.П. Моделирование сложных систем / Н.П. Бусленко. – М.: Наука, 1978. – 400 с.

28. Баева, Н.Н. Многоканальные системы передачи / Н.Н. Баева, В.Н. Гордиенко и др. – М.: Радио и связь, 1996.

29. Баркун, М.А. Цифровые системы синхронной коммутации / М.А. Баркун, О.Р. Ходасевич. –М.: ЭКО-ТРЕНДЗ, 2001. – 188 с.

30. Брюханов, Ю.А. Теория дискретных и цифровых сигналов и цепей / Ю.А. Брюханов, А.И. Кренов. – Ярославль: 1991. – 114 с.

31. Василькова, В.В. Порядок и хаос в развитии социальных систем: (Синергетика и теория социальной самоорганизации) / В.В. Василькова. – СПб.: Лань, 1999.

32. Васильев, Ф.П.: Информационные технологии управления в органах внутренних дел / Ф.П. Васильев и др.: под ред. профессора Минаева В.А. – М.: Академия управления МВД России, 1997. – 704 с.

33. Воеводин, В.В. Математические модели и методы в параллельных процессах / В.В. Воеводин. – М.: Наука, 1986. – 296 с.

34. Гуркин, В.Ф. Развитие подвижной связи в России / В.Ф. Гуркин, И.В. Николаев. – М.: Радио и связь, 2000. – 160 с.

35. Вентцель, Е.С. Теория вероятностей / Е.С. Вентцель. – М.: Высшая школа, 1998.

36. Вентцель, А.Д. Курс теории случайных процессов / Е.С. Вентцель. – М.: Наука, 1996.

37. Винер, Н. Кибернетика или Управление и связь в животном и машине / Н. Винер. – М.: Советское радио, 1968.

38. Винер, Н. Кибернетика и общество / Н. Винер. – М.: Изд-во иностр. литературы, 1958. – 200 с.

39. Волькенштейн, М.В. Энтропия и информация / М.В. Волькенштейн. – М.: Наука, 1980.

40. Васильев, В.И, Теория передачи и преобразования информации / В.И. Васильев. – Б/м: 1995. – 96 с. с илл.

41. Введение // Социальные проблемы информатики: Сб. статей. – Л.:Лгик, 1974. – С.3 – 5.

42. Влэдуц, С.Г. Алгеброгеометрические коды. Основные понятия / С.Г. Влэдуц, Д.Ю. Ногин, М.А. Цфасман. – М.: МЦНМО, 2003. – 504 с.

43. Водяхо, А.И. Высокопроизводительные системы обработки данных / А.И. Водяхо, Н.Н. Горнец, Д.В. Пузанков. – М.: Высшая школа, 1997. – 304 с.

44. Возенкрафт, Дж. Последовательное декодирование / Дж. Возенкрафт, Б. Рейфен. – М.: ИЛ, 1963.

45. Волфовиц, Дж. Теоремы кодирования в теории информации / Дж. Волфовиц. – М.: Мир, 1967.

46. Гаврилов, О.А. Информатизация правовой системы России. Теоретические и практические проблемы / О.А. Гаврилов. – М.: Изд-во «Юридическая книга», при участии изд-ва «ЧеРо», 1998. – 144 с.

47. Герасименко, В.А. Основы информационной грамоты / В.А. Герасименко. – М.: Энергоатомиздат, 1996. – 320 с.

48. Герасименко, В.А. Концепция современной информатики / В.А. Герасименко // Зарубежная радиоэлектроника. – 1994. – № 4. – С. 77-91

49. Герасименко, В.А. Основы информатики. Ч. 1. Введение в информатику / В.А. Герасименко // МГИАИ. – М., 1996. Деп. В ВИНИТИ 16.07.91, № 3718-В91, 1991. – 134 с.

50. Герасименко, В.А. Основы информатики. Ч. 2. Мировозренческие основы информатики / В.А. Герасименко // МГИАИ. – М., 1996. Деп. В ВИНИТИ 16.07.91, № 3719-В91

51. Гилула, М.М. Модели данных и модели информации в информационных системах / М.М. Гилула // Системные

исследования. Методологические проблемы. Ежегодник 1992-1994 / Гл. ред. Д.М. Гвишиани. – М.: Эдиториал УРСС, 1996. – С.212-236.

52. Гиляревский, Р.С. Что такое информатика? / Р.С. Гиляревский // НТИ. Сер.1. – 1989. – № 11. – С.18–21.

53. Гиляревский, Р.С. Роль интеллектуальных информационных систем в развитии информатики / Р.С. Гиляревский //НТИ. Серия 2. – 1987. – № 9. – С. 5–9.

54. Горский, Ю.М. Информационные аспекты управления и моделирования / Ю.М. Горский. – М.: Наука, 1978.

55. Готт, В.С. Социальная роль информатики / В.С. Готт, Э.П. Семенюк, А.Д. Урсул. – М.: Знание, 1977. – 64 с.

56. Галлагер, Р. Теория информации и надежная связь / Галлагер. – М.: Сов. Радио, 1974.

57. Гиляревский, Р.С. Научные коммуникации и проблема информационной потребности / Р.С. Гиляревский, В.А. Маркусова, А.И. Черный // НТИ. – Сер. 1. Орг. и методика информ. работы. – 1993. – № 9. – С.1-7.

58. Гоппа, В.Д. На непроводимых кодах достигается пропускная способность ДСК / В.Д. Гоппа // Проблемы передачи информации. – 1974. – Вып. 1. – С. 111-112.

59. Доронин, С.И. Роль и значение квантовой теории в свете ее последних достижений / С.И. Доронин [Электронный ресурс] – http://quantmagic.narod.ru/volumes/VOL112004/p1101.html .

60. Дмитриев, В. И. Прикладная теория информации: учебник для вузов по специальности «Автоматизированные системы обработки информации и управления» / В.И. Дмитриев. – М.: Высшая школа, 1989.

61. Духин, А.А. Теория информации: учебное пособие / А.А. Духин. – М.: Гелиос АРВ, 2007.

62. Дюк, В.А. Компьютерная психодиагностика / В.А. Дюк. – СПб.: Изд. «Братство», 1994. – 364 с.

63. Денисов, М.Ю. Цифровые системы передачи / М.Ю. Денисов. – Орел: ВИПС, 1996.

64. Диалоговые системы моделирования / В.В. Пирогов, Л.П. Богомолов, С.Ф. Гайстеров и др. – Рига: Зинатне, 1977. – 176 с.

65. Еременко, В.Т. Методологические, технологические и социокультурные аспекты информатики: монография / В.Т. Еременко, С.Ю. Лачинов, О.В. Третьяков. – Орел: Изд-во ОРАГС, 2007. – 188 с.

66. Еременко, В.Т., Теория информации и информационных процессов: монография / В.Т. Еременко, Н.А. Орешин, Н.Г. Подчерняев, О.В. Третьяков. – Орел: Орловский юридический институт МВД России, 2000. – 187с.

67. Ершов, А.П. Информатизация: от компьютерной грамотности учащихся к информационной культуре общества /А.П. Ершов // Коммунист. – 1988. – № 2. – С. 92-92.

68. Ершов, А.П. Отношение методологии и технологии программирования / А.П. Ершов // Технология программирования. Тез. Докл. II Всес. конф. – информационные материалы. – Киев: ИК АН УССР. – 1986. – С. 10-12.

69. Жигарев, А.Н. Основы компьютерной грамоты / А.Н. Жигарев и др. – Л.: Машиностроение, 1988. – С.5-10, 52-61.

70. Закорюкин, В.Б. Теория и средства передачи информации: Учебное пособие / В.Б. Закорюкин // Московский государственный институт радиотехники, электроники и автоматики. Технический университет. – М., 1999. – 95 с.

71. Железнов, И.Г. Сложные технические системы (оценка характеристик) / И.Г. Железнов. – М.: Высш. шк., 1984. – 119 с.

72. Зингиренко, А.М. Системы многоканальной связи / А.М. Зингиренко, Н.Н. Баева, М.С Тверецкий. – М.: Связь, 1980.

73. Зюко, А.Г. Теория передачи сигналов / А.Г. Зюко, Л.М. Финк и др. – М.: Связь, 1980. – 288 с.

74. Зуюс, Ю.К. Региональная информационная политика и вопросы ценообразования / Ю.К. Зуюс // НТИ. –Сер. 1. – Орг. и методика информ. работы. – № 11. – 1989. – С. 45-46.

75. Зюко, А. Г. Теория передачи сигналов: учебник для вузов. 2-е изд. перераб. и доп. / А.Г. Зюко, Д.Д. Кловский, М.В. Назаров, и др. – М.: Радио и связь, 1986. 304 с.

76. Информатика и информационная безопасность. учебное пособие/ Под ред. Минаева В.А., Фисуна А.П.– Хабаровск, Дальневосточный ЮИ, 2002.– 528 с.

77. Информатика и информационные технологии в юридической деятельности: учебное пособие / под ред. В.А. Минаева, А.П. Фисуна, А.Н. Шаковца. – Хабаровск: Дальневосточный юридический институт МВД РФ, 2006.– 424 с.

78. Информатика. В 2-х т. Изд. 2-е – расширенное и доп. Т. 1. Концептуальные основы: учебник / под общ. науч. ред. В.А. Минаева, А.П. Фисуна, С.В. Скрыля, С.В. Дворянкина, М.М. Никитина, Н.С. Хохлова. – М.: Маросейка, 2008.– 464 с.

79. Информатика: В 2-х т. Изд. 2-е – расширенное и доп. Т. 2. Средства и системы обработки данных: учебник / под общ. науч. ред. В.А. Минаева, А.П. Фисуна, С.В. Скрыля, С.В. Дворянкина, М.М. Никитина, Н.С. Хохлова. – М.: Маросейка, 2008. – 544 с.

80. Информатика: учебник / под ред. проф. Н.В. Макаровой. – М.: Финансы и статистика, 1997. – 768 с.

81. Информатика: практикум по технологии работы на компьютере /Под ред. Н.В. Макаровой. – М.: Финансы и статистика, 1997. – 384 с.

82. Информатика. Концептуальные основы: учебник / под общ. ред. С.В. Скрыля. – Орел: Изд-во «Орлик», 2007. – 372 с.

83. Информатика и информационная безопасность. учеб. пособие / под ред. Минаева В.А., Фисуна А.П.– Хабаровск: Дальневосточный ЮИ, 2002.– 528 с.

84. Информатика. В 2-х т. Изд. 2-е – расширенное и доп.. Т. 1. Концептуальные основы: учебник / под общ. науч. ред. В.А. Минаева, А.П. Фисуна, С.В. Скрыля, С.В. Дворянкина, М.М. Никитина, Н.С. Хохлова. – М.: Маросейка, 2008.– 464 с.

85. Информатика: В 2-х т. Изд. 2-е – расширенное и доп.. Т. 2. Средства и системы обработки данных: учебник / под общ. науч. ред. В.А. Минаева, А.П. Фисуна, С.В. Скрыля, С.В. Дворянкина, М.М. Никитина, Н.С. Хохлова. – М.: Маросейка, 2008.– 544 с.

86. Ипатов, В. П. Основы теории связи: учеб. пособие / В.П. Ипатов. – СПб.: Санкт-Петербургский государственный электротехнический университет, 1999. – 79 с., с илл.

87. Историческая информатика / под ред. Л..И. Бородкина , И.М. Гарской. – М.: Мосгорархив, 1996.

88. Кавалеров, Г.И. Введение в информационную теорию измерений / Г.И. Кавалеро, С.М. Мандельштам. – М.: Энергия, 1974.

89. Кадомцев, Б.Б. Динамика и информация. 2-е изд. / Б.Б. Кадомцев. – М.: Редакция журнала «Успехи физических наук», 1999.

90. Казанцев, Э.Ф. Технологии исследования биосистем / Э.Ф. Казанцев. – М.: Машиностроение, 1999. – 177 с.

91. Кальоти, Дж. От восприятия к мысли: пер. с нем. / Дж. Кальоти. – М.: Мир, 1998.

92. Камша, В.П. О парадигме компьютерной лингвистики / В.П. Камша, Л.С. Камша // НТИ. – Сер. 2. Информ. процессы и системы. – 1993. – № 8. – С. 1-8.

93. Камша, В.П. Роль качественных аспектов информации

в лингвоинформировании / В.П. Камша // НТИ. – Сер. 2. Информ. процессы и системы. – 1995. –№ 8. – С. 8 –21.

94. Каныгин, Ю.М. Социально-экономические проблемы создания и использования искусственного интеллекта / Ю.М. Каныгин, Г.И. Калитич. – Киев: УкрНИИНТИ, 1989 – 36 с.

95. Карташевский, В.Г. Сети подвижной связи / В.Г. Карташевский, С.Н. Семенов. – М.: ЭКО-ТРЕНДЗ, 2001. – 299 с.

96. Кричевский, Р.Л. Если Вы – руководитель. Элементы психологии менеджмента в повседневной работе / Р.Л. Кричевский. – М.: Дело, 1993. – 352 с.

97. Киндлер, Е. Языки моделирования: пер. с чеш. / Е. Киндлер. – М.: Энергоатомиздат, 1985. – 288 с.

98. Клещев, Н.Т. Телекоммуникации / Н.Т. Клещев. – М.: Радио и связь, 1999. – 500 с.

99. Каныгин, Ю.М. Социально-экономические проблемы информатизации / Ю.М. Каныгин, Г.И. Калитич. – Киев: УкрНИИНТИ, 1990, – 48 с.

100. Каныгин, Ю.М. Основы теоретической информатики / Ю.М. Каныгин, Г.И. Калитич. – Киев: Наукова думка, 1990.

101. Каныгин Ю.М., Яковенко Ю.І. З позицій коллективного розуму. Новий повгляд на проблему штучного интеллекта / Ю.М. Каныгин, Ю.І. Яковенко // Висник Академии наук Украинской РСР.– 1989. – № 9.- С. 88-91.

102. Каранчук, В.П. Основы применения ЭВМ / В.П. Каранчук. – М.: Радио и связь, 1988. – С.5-10, 29- 40.

103. Кастлер, Г. Возникновение биологической организации / Г. Кастлер. – М.: Мир, 1967.

104. Кассами, Т. Верхняя граница для k/n аффинно инвариантных кодов с фиксированным d/n. Кибернетический сборник. Новая серия / Т. Касами. – М., 1971. – Вып. 8. – С. 5-11.

105. Кастлер, Г. Возникновение биологической организации / Г. Кастлер. – М.: Мир, 1967.

106. Кириллов, В.И. Логика: Учебник для юридических факультетов и институтов / В.И. Кириллов, А.А. Старченко. – М.: Юристъ. 1995.

107. Кларк, Дж. Кодирование с исправлением ошибок в системах цифровой связи: пер. с англ / Дж. Кларк, Дж. Кейн. – М.: Радио и связь. 1987. –392 с.

108. Колесник, В.Д. Циклические коды Рида-Маллера и их де-

кодирование / В.Д. Колесник, Е.Т. Мирончиков // Проблемы передачи информации. – 1968. – Вып. 4. – С. 20-25.

109. Колесник, В.Д. Введение в теорию информации (кодирование источников): учебное пособие / В.Д. Колесник, Г.Ш Полтырев. – Л.: Ленинградский университет, 1980. – 163 с. с илл.

110. Колесник, В.Д. Курс теории информации / В.Д. Колесник, Г.Ш. Полтырев. – М.: Наука, 1982.

111. Колесников, А. Excel 7.0 для Windows / А. Колесников. – К.:BHV, 1996. – 480 с.

112. Колин, К.К. Информационное общество и проблема образования / К.К. Колин // Информационное общество. – № 2-3. – С.18-19.

113. Колин, К.К. Фундаментальные основы информатики: Социальная информатика: учеб. пособие / К.К. Колин. – М.: Академический проект; Екатеринбург: Деловая книга, 2000. – 350 с.

114. Колин, К.К. Социальная информатика – научная база постиндустриального общества / К.К. Колин. – М., 1993.

115. Колмогоров, А.Н. Три подхода к понятию количества информации / А.Н. Колмогоров. – Проблемы передачи информации. – 1965, – Т.1. – Вып.1. – С.3 –11.

116. Колмогоров, Л.Я. К логическим основам теории информации и теории вероятностей / Л.Я. Колмогоров // Проблемы передачи информации. – 1969. – Т. 5. – № 3.

117. Колмогоров, А.И. Теория информации и теория алгоритмов / А.И. Колмогоров. – М.: 1987.

118. Копылов, В.А. Информационное право: учеб. пособие / В.А. Копылов.– М.: Юристъ, 1997. – 472 с.

119. Коржик, В.И. Взаимосвязь свойств двоичного группового кода и его нуль-пространства / В.И. Коржик. – Проблемы передачи информации. – 1966. – Вып. 4. – С. 91-95.

120. Коржик, В.И. Оценка d_{mH} Для циклических кодов / В.И. Коржик. – Проблемы передачи информации. – 1966. – Вып. – 2. – С. 78.

121. Корогодин, В.И. Определение понятия информации и возможности его использования в биологии / В.И. Корогодин // Биофизика. – 1983. – Т. 28, вып. 1.– С. 171-177.

122. Корогодин, В.И. Информация и феномен жизни / В.И. Корогодин. – Пущино: АН СССР, 1991. – 200 с.

123. Котляров, В.П. Гипертекстовая среда как инструментарий проектирования программного проекта / В.П. Котляров, М.В. Токарев // Пользовательский интерфейс. – № 3. – 1993 – С. 39-53

124. Кочегуров, В.А. Введение в прикладную теорию информа-

ции: учеб. пособие. Министерство общего и профессионального образования Российской Федерации. Томский политехнический университет. – Томск, 1999. – 95 с.

125. Краткий словарь по социологии / под общ. ред. Д.М. Гвишиани, Н.И.Лапина. – М.: Политиздат, 1988.

126. Кретов, В.С. Некоторые аспекты создания интеллектуальных информационных систем в политологии / В.С. Кретов, И.Е. Власов, И.В. Фролов // НТИ. – Сер.2 Информационные процессы и системы. – 1994. – № 11. – С.9-15.

127. Кричевский, Р.Е. О числе ошибок, исправляемых кодом Рида-Маллера / Р.Е. Кричевский. – ДАН – СССР. – 1970, Т. 191. – Вып. 3. – С. 544-547.

128. Крылов, В.В. Расследование преступлений в сфере информации / В.В. Крылов. – М.: Изд. «Городец», 1998. – 264 с.

129. Крылов, В.В. Информационные компьютерные преступления / В.В. Крылов.– М.: Издательская группа ИНФРА.М–НОРМА, 1997. – 285 с.

130. Куликовский, Л.Ф. Теоретические основы информационных процессов / Л.Ф. Куликовский, В.В. Мотов. – М.: Высш. шк, 1987. –248 с.

131. Кузнецов, А.В. Теория кодов, исправляющих дефекты и ошибки, и её приложения. Автореф. диссертации д.т.н. 05.25.01. Институт проблем передачи информации / А.В. Кузнецов. – М., 1988. – 26 с.

132. Кук, Д. Компьютерная математика / Д. Кук, Г. Бейз. – М., 1990.

133. Курдюмов, В.А. Креативно-когнитивная функция языка и лингво-технические приемы достижения эффекта убеждения / В.А. Курдюмов // НТИ. – Сер.2. Информ. процессы и системы. – 1997. – № 8. – С.31-36.

134. Курушин, В.Д. Компьютерные преступления информационная безопасность / В.Д. Курушин, В.А. Минаев. – М.: Новый Юрист, 1998.– 256 с.

135. Левин, М.Ш. О третьей грамотности / М.Ш. Левин // НТИ. – Сер.2 Информационные процессы и системы.– 1995.– № 6.– С.20–30.

136. Левин, М.Ш. О третьей грамотности / М.Ш. Левин // НТИ. – Сер.2. Информ. процессы и системы. 1995. – № 6. – С. 20-30.

137. Лобанов, А.С. Семиотика: вчера, сегодня, завтра / А.С. Лобанов // НТИ. Сер. 2. Информ. процессы и системы. – 1995. – № 7. – С. 9-18.

138. Логика: учеб. пособие для общеобразоват. учеб. заведений,

шк. и классов с углубл. изуч. логики, лицеев, гимназий/ А.Д. Гетманова, А.Л. Никифоров, М.И. Панов и др. – М.: Дрофа, 1995.

139. Ломакин, М.И. Военно-социальная информация / М.И. Ломакин, А.А. Чертополох, А.В. Костин и др.. – М.: Военный университет, 1997. – 195 с.

140. Лебедев, А.Н. Основы теории моделирования: конспект лекций / А.Н. Лебедев. – Пенза: ППИ, 1977. – 81 с.

141. Мак-Вильяме, Ф.Дж. Теория кодов, исправляющих ошибки / Ф.Дж. Мак-Вильяме, И.Дж.А. Слоэн. – М., Связь, 1979.

142. Мамиконов, А.Г. Принятие решений и информация / А.Г. Мамиконов. – М.: Наука, 1983. – 184 с.

143. Марков, А.А. Введение в теорию кодирования / А.А. Марков. – М.: Наука, 1982.

144. Массарский, Л.В. Имитационный комплекс взаимодействия АСУ и производственной модели объекта управления / Л.В. Массарский, Л.Л Шуб. – Калинин: Центрпрограммсистем, 1980. – 36 с.

145. Математический энциклопедический словарь. – М.: Научное изд-тво «Большая Российская энциклопедия», 1995.

146. Мелик–Гайказян, И.В. Информационные процессы и реальность / И.В. Мелик–Гайказян. – М.: Наука, Физматлит, 1998. – 192 с.

147. Мельников, В.В. Защита информации в компьютерных системах / В.В. Мельников. – М.: «Финансы и статистика», «Электроинформа», 1997. – 364 с.

148. Минаев, В.А. Информатика и информационные технологии в юридической деятельности: учеб. пособие / В.А. Минаев и др. под ред. В.А. Минаева, А.П. Фисуна, А.Н. Шаковца. – Хабаровск: Дальневосточный юридический институт МВД РФ, 2006.– 424 с.

149. Минаев, В.А. Основы информационной безопасности / В.А. Минаев, С.В. Скрыль, А.П. Фисун. – Воронеж: ВИ МВД РФ, 2001. – 452 с.

150. Матвеева, М.В. Об исправлении тройных ошибок в кодах Боуза-Чоудхури над полем GF(3) / М.В. Матвеева // Проблемы передачи информации. – 1968. – Вып. 1. – С. 20-27.

151. Микулина, О.А. Основные понятия статистической теории информации: учеб. пособие / О.А. Микулина. Министерство образования Российской Федерации. Московский государственный инженерно-технический институт. – М., 2000. – 92 с.

152. Минаев, В.Н. Концептуальный подход подготовки

специалистов в области информационной безопасности / В.Н. Минаев, А.П. Фисун, А.Н. Касилов // Материалы Международной конференции «Информатизация правоохранительных систем» (2 – 3 июля 1996 г.). Тезисы докладов. Часть 1. – М.: МАИ, Академия МВД России, 1996. – С. 135 – 137

153. Мириманова, М.С. Информативно-когнитивные процессы и их роль в информатизации / М.С. Мириманова // НТИ. Сер.1.–1989. – С.62-64.

154. Михайлов, А.И. Информатика – новые названия теории научной информации / А.И. Михайлов, А.И. Черный, Р.С. Гиляревский // НТИ. – 1966. – № 12. – С. 1-3.

155. Михайлов, А.И. Основы информатики / А.И. Михайлов, А.И. Черный, Р.С. Гиляревский. – М.: Наука, 1968. – 756 с.

156. Михайлов, А.И. Научные коммуникации и информатика / А.И. Михайлов, А.И. Черный, Р.С. Гиляревский. – М.: Наука, 1976. – 435 с.

157. Михайлов, А.И. Основы построения телекоммуникационных систем и сетей общего пользования / А.И. Михайлов. – Орел, ВИПС, 1998.

158. Морозов В.К. Основы теории информационных сетей / В.К. Морозов, А.В. Долганов. – М.: Высш. шк., 1987. – 271 с.

159. Могилев, А.В. и др. Информатика: учеб. пособие для вузов / А.В. Могилев, Н.И. Пак, Е.К. Хоннер; Под ред. Е.К. Хоннера. – М.: Изд. центр «Академия», 2000. – 816 с.

160. Москвин, В.Д. Словарь основных терминов и определений / В.Д. Москвин и др. //Справочное пособие. Основные положения развития Взаимоувязанной сети связи Российской Федерации не перспективу до 2005 года . – Руководящий документ. – М.: ГКЭС России, 1996. – 27 с.

161. Мелик–Гайказян, И.В. Информация и самоорганизация: Методологический анализ / И.В. Мелик–Гайказян. – Томск: Изд-во ТПУ, 1995.

162. Мелик–Гайказян, И. В. Информационные процессы и реальность / И.В. Мелик–Гайказян. – М.: Наука, Физматлит, 1997.

163. Моисеев, Н.Н. Человек и ноосфера / Н.Н. Моисеев. – М.: Молодая гвардия, 1990.

164. Моль А. Социодинамика культуры / А. Моль. – М.: Прогресс, 1973.

165. Мышеев, А.В. Основы прикладной теории кодирования ин-

формации / А.В. Мышеев. – Обнинск: Обнинский институт атомной энергетики. 1998. – 78 с.

166. Никитин, Г.И. Помехоустойчивые циклические коды: учеб. пособие / Г.И. Никитин, С.С. Поддубный. – СПб., 1998. – 71 с.

167. Начала информационной теории управления / Петров Б.Н. и др. // Итоги науки. Техническая кибернетика. – 1966 –1975 гг. – № 1-6.

168. Нильсен, М. Квантовые вычисления и квантовая информация / М. Нильсен, И. Чанг; пер. с англ. – М.: Мир, 2006. – 824 с.

169. Николис, Г. Самоорганизация в неравновесных системах / Г. Николис, И. Пригожин; пер. с англ. – М.: Мир, 1979.

170. Николис, Г. Познание сложного. Введение / Г. Николис, И. Пригожин; пер. с англ. – М.: Прогресс, 1990.

171. Николаев, В.И. Системотехника: методы и приложения / В.И. Николаев, В.М. Брук. – Л.: Машиностроение, 1985. – 199 с.

172. Ожегов, С.И. Словарь русского языка / С.И. Ожегов; под ред. д.ф.н., проф. Шведовой Н.Ю. – 14 изд., стереотипное. – М.: «Русский язык», 1982. – 816 с.

173. Орнстейн, Д. Эргодическая теория, случайность и динамические системы / Д. Орнстейн. – М.: Мир, 1978.

174. Основы государственного управления в сфере информатизации в Российской Федерации. – М.: Юристъ, 1997.– 334 с.

175. Основы общей теории систем. Часть 1. А.А. Попов, И.М. Телушкин, С.Н. Бушуев и др. – М.: ВАС, 1992.

176. Основные проблемы информатики и библиотечно-библиографическая работа / Под ред. А.В. Соколова. – Л.: ЛГИК, 1976. – 319 с.

177. Основы моделирования сложных систем / Л.И. Дыхненко, В.Ф. Кабаненко, И.В. Кузьмин и др. – Киев: Вища шк., 1981. – 359 с.

178. О создании квазинатурной модели комплекса технических средств АСУ / В.А. Бункин, В.Ю. Ралль, И.Н. Альянах и др. // Пробл. системотехники; Под ред. В.И. Николаева. – Л., 1980. – С. 165 – 167.

179. Основы теории вычислительных систем / С.А. Майоров, Г.И. Новиков, Т.И. Алиев и др.; Под ред. С.А. Майорова. – М.: Высш. шк., 1978. – 408 с.

180. Парамонов, Ю.В. Введение в теорию и методы защиты информации / Ю.В. Парамонов. – М.: Информ-связьиздат, 1999. – 56 с.

181. Пранявичюс, Г. Модели и методы исследования вычислительных систем / Г. Пранявичюс. – Вильнюс: Мокслас, 1982. – 228 с.

182. Прохоров И.В. Телекоммуникационные сети / И.В.

Прохоров, А.И. Толстой. – М.: МИФИ,1996.
183. Першиков, В.И. Толковый словарь по информатике / В.И. Першиков, В.М. Савинков. – 2-е изд., доп. – М.: Финансы и статистика, 1995. – 554 с.
184. Петров, Ю.П. О различных формах и видах информации: Информационные проблемы изучения биосферы / Ю.П. Петров. – М.: Наука, 1988.
185. Петров, В.В. Информационная теория синтеза оптимальных систем контроля и управления. (Непрерывные системы) / В.В. Петров, А.С. Усков. – М.: Энергия, 1975.
186. Плоткин, М. Двоичные коды с заданным минимальным расстоянием / М. Плоткин // Кибернетический сборник. – Вып. 7. – М.: ИЛ, 1963.
187. Побельский, В.В. Язык СИ++: учеб. пособие / В.В. Побельский. – М.: Финансы и статистика, 1999. – 560 с.
188. Пойма, Д. Как решать задачу / Д. Пойма. – М.: Учпедгиз, 1959. – С. 143, 202-203.
189. Пойма, Д. Математическое открытие. Решение задач: основные понятия, изучение / Д. Пойма. – М.: Наука. 1970.
190. Попов, Э.В. Алгоритмические основы интеллектуальных роботов и искусственного интеллекта / Э.В. Попов, Г.Р. Фирдман. – М.: Наука, 1976.
191. Поспелов, Г.С. Душа и сердце новейшей информационной технологии / Г.С. Поспелов. – М.: Знание, 1988. – Вып. 21. – С. 8-32.
192. Поспелов, Г.С. Искусственный интеллект – основа новой информационной технологии / Г.С. Поспелов. – М.: Наука, 1988; М.: Радио и связь, 1990. – 528 с.
193. Потапов, В.Н. Теория информации. Кодирование дискретных вероятностных источников: учеб. пособие / В.Н. Потапов. – Новосибирск: Нов. гос. ун-т., 1999.
194. Правовая информатика и кибернетика: учебник / Под ред. Н.С. Полевого.– М.: Юрид. лит., 1993.– 528 с.
195. Пригожин, И. Порядок из хаоса. Новый диалог человека с природой: пер. с англ. / И. Пригожин, И. Стенгерс. – М.: Прогресс, 1986.
196. Пригожин, И. Время, хаос, квант: пер. с англ / И. Пригожин, И. Стенгерс. – М.: Прогресс, 1986.
197. Программирование. Э.З. Любимский, В.В. Мартынюк, В.П. Трифонов. – М.: Наука. Главная редакция физико-математической

литературы, 1980.

198. Пушкин, В.Г. Проблемы надежности / В.Г. Пушкин. – М.: Наука, 1971.

199. Преснухин Л.Н., Нестеров П.В. Цифровые вычислительные машины: учеб. пособие / Л.Н. Преснухин, П.В. Нестеров. – М.: Высшая школа,1981. – 511 с.

200. Проблемы передачи информации: [журнал]. – 1999. – Т.35.

201. Пятибратов, А.П. Вычислительные системы, сети и телекоммуникации: учебник / А.П. Пятибратов и др. – М.: Финансы и статистика, 1998. – 400 с.

202. Рабинер, А. Теория и применение цифровой обработки сигналов: пер. с англ. / А. Рабинер, Б. Гоулд. – М.: Мир, 1978. – 848 с.

203. Райков, А.Н. Интеллектуальные информационные технологии в аналитических исследованиях социально-политических объектов / А.Н. Райков // НТИ. – Сер.2. Информационные процессы и системы.– 1994.– № 11.– С.1–8.

204. Ракитов, А.И. Информатизация общества и стратегия ускорения / А.И. Ракитов // Правда. – 1987. – 23 янв.– С. 2-3.

205. Ракитов, А.И. Информатизация советского общества – реальность и перспективы / А.И. Ракитов // НТИ. – Сер.1.– 1989. – № 11 – С.8–15.

206. Рассолов, М.М. Информационное право: учеб. пособие / М.М. Рассолов. – М.: Юристъ, 1999. – 400 с.

207. Ракитов, А.И. Информатизация советского общества – реальность и перспективы / А.И. Ракитов // НТИ. – Сер. 1. Орг. методика информ. работы. – 1989. – № 11. – С. 8-18.

208. Реляционно-функциональная концепция информации и ее приложения// НТИ. – Сер. 2. Информ. процессы и системы. – 1997. – № 8. – С. 8 -17.

209. Рид, И.С. Класс кодов с исправлением нескольких ошибок и схема кодирования: Кибернетический сб. / И.С. Рид. – М., 1960. – Вып. 1.

210. Рябко, Б.Я. Сжатие данных с помощью «мнимого скользящего окна» /Б.Я. Рябко // Проблемы передачи информации. – 1996.– Т. 32. – № 2. – С. 22-30,

211. Седов, Е.А. Информационно-энтропийные свойства социальных систем / Е.А. Седов // Общественные науки и современность. – 1993. – № 5. – С. 92-102.

212. Смирнов, Г.А. Логические аспекты целостности / Г.А. Смирнов // Системные исследования. Методологические проблемы.

Ежегодник 1995-1996 / Гл. ред. Д.М. Гвишиани. – М.: Эдиториал УРСС, 1996. – С.108-127.

213. Семаков, Н.В. Эквидистантные q-е коды с максимальнымп расстоянием и разрешимые уравновешенные неполные блок-схемы. / Н.В. Семаков, В.А. Зиновьев // Проблемы передачи информации. – 1968. – Вып. 2. – С. 3-10.

214. Семаков, Н.В. Совершенные и квазисовершенные равновесные коды / Н.В. Семаков, В.Л. Зиновьев // Проблемы передачи информации. – 1969. – Вып. 2. – С. 14-18.

215. Семенюк, Э.П. Информатизация общества и развитие методологических проблем информатики / Э.П. Семенюк // НТИ. – Сер.2. Информационные процессы и системы. – 1990.– № 12.– С.2–9.

216. Семенюк, Э.П. Информатика: как ее понимать? / Э.П. Семенюк // НТИ. – Сер.2. Информационные процессы и системы. – 1984.– № 7.– С.1–8.

217. Семенюк, Э.П. Информатика: достижения, перспективы, возможности / Э.П. Семенюк. – М.: Наука, 1988.– 176 с.

218. Семенюк, Э. П. Информатизация общества и развитие методологических проблем информатики / Э.П. Семенюк // Информационные процессы и системы. – 1990. – № 12. – С. 2-9.

219. Семенюк, Э.П. Информационный подход к познанию действительности / Э.П. Семенюк. – Киев: Наукова думка, 1988. – 240.

220. Словарь иностранных слов. – 15-е изд., испр. – М.: Рус. Яз., 1988.

221. Словарь по кибернетике/ Под ред. В.С. Михалевича. – 2-е изд. – К.: Гл. ред. УСЭ им. М.П. Бажана. 1989. –751 с.

222. Слоэн, Н.Дж.А. Обзор конструктивной теории кодирования и таблица двоичных кодов с наибольшими известными скоростями. Кибернетический сборник. Новая серия / Н.Дж.А. Слоэн. – М., 1973. – Вып. – 10. – С. 5-32.

223. Смирнов, А.С. Основы теории кодирования. Линейные групповые и циклические коды: учеб. пособие /А.С. Смирнов. – СПб.: СГТУ, 1998. – 147 с.

224. Соболев, С. Да, это вполне серьезно! / Возможное и невозможное в кибернетике / С. Соболев. – М.: АН СССР, 1963. – С. 82-88.

225. Советов, Б.Я. Построение сетей интегрального обслуживания / Б.Я. Советов, С.А. Яковлев. – Л.: Машиностроение. Ленингр. отд-ние, 1990. – 332 с.

226. Советов, Б.Я. Моделирование систем / Б.Я. Советов, С.А. Яковлев. – М.: Высш. шк., 1985. – 271 с.

227. Соколов, А.В. Объекты и предметы библиотековедения, библиографоведения и информатики (метатеоретический анализ) / А.В. Соколов // Связь библиотечно–научных дисциплин с информатикой: Сб. научных трудов / ЛГИК им Н.К. Крупской. – 1982. – Т.68.– С. 10–46.

228. Соколов, А.В. Социальная информатика / А.В. Соколов, А.И. Манкевич // Социальные проблемы информатики: Сб. статей.– Л.: ЛГИК, 1974.– С.3-5.

229. Соколов, А.В. Взаимосвязь информатики и библиотечно-библиографических дисциплин / А.В. Соколов, А.И. Манкевич, Т.Н. Колтыпина // Научные и технические библиотеки СССР. – 1974. – Вып. 4(126). – С. 28-36.

230. Соколов, А.В. Информатика в перспективе (к вопросу о классификации видов информации и системе наук коммуникационного цикла) / А.В. Соколов, А.И. Манкевич // НТИ. – Сер. 2. – 1971. № 10. – С. 5-9.

231. Соловьев, Г.Н. Операционные системы ЭВМ: учеб. пособие / Г.Н. Соловьев, В.Д. Никитин. – М.: Высшая школа, 1989. – 255 с.; СПб; 1997. – 400 с.

232. Социальная психология: Краткий курс / Под ред. Г.П. Предвечного и Ю.А. Шерковина. – М.: Политиздат, 1975.

233. Стипахно, И.И. Сложные двойные цепи Маркова: Управление и передача информации. Автореферат диссертации на соискание ученой степени к.ф.-м.н. 01.01.05. – АН УССР. – Институт математики. – Киев, 1989. – 9 с.

234. Системный анализ и структуры управления. Под редакцией В.Г. Шорина. – М.: Знание, 1975. - 304 с.

235. Соболь, И.М. Численные методы Монте-Карло / И.М. Соболь. – М.: Наука, 1973. – 311 с.

236. Темников, Ф.Е. Теоретические основы информационной техники/ Ф.Е. Темников В.А. Афонин, В.И. Двитриев В.И. - М.: Энергия, 1979. - 512 с.

237. Тараканов, К.В. Информатика / Под ред, доктора технических наук, профессора К.В. Тараканова. - М.: Книга, 1986. – 304 с.

238. Таха, Х.А. Введение в исследование операций. В 2-х кн. Кн.2. Пер. с англ. / Х.А. Таха. – М.: Мир, 1985.

239. Теория информации и информационных процессов: монография /Под ред. д.т.н. В.Т. Еременко, д.т.н. А.П. Фисуна. – Орел: Изд. ГОУ ВПО «Орловский государственный университет», 2008.– 478 с.

240. Токура, И., Теория кодирования / И Токура, Т. Касами, И. Ивадари, Я. Инагаки. – М.: Мир, 1978.

241. Толковый словарь по вычислительным системам / Под ред. В. Иллингуорта и др.: Пер. с англ. А.К. Белецкого и др.; Под ред. Е.К. Масловского. – М.: Машиностроение, 1991.- 560 с.

242. Урсул, А.Д. Эволюция. Космос. Человек (Общие законы развития и концепция антропокосмизма) / А.Д. Урсул, Т.А. Урсул. – Кишинев: Штиница, 1986.

243. Урсул, А.Д. На пути к устойчивому развитию цивилизации: информационные факторы / А.Д. Урсул, Т.А. Урсул. // Информационное общество. – 1997. – № 2-3. – С. 20-27.

244. Урсул, А.Д. Природа информации. Философский очерк / А.Д. Урсул. – М.: Политиздат, 1968.

245. Урсул, А.Д. Информация. Методологические аспекты / А.Д. Урсул. – М.: Наука, 1971.

246. Урсул, А.Д. Отражение информация / А.Д. Урсул. – М.: Мысль, 1973.

247. Урсул, А.Д. Проблема информации в современной науке. Философские очерки / А.Д. Урсул. – М.: Наука, 1975.

248. Урсул, А.Д. Развитие информатики и информатизация общества: вопросы методологии / А.Д. Урсул // НТИ. – Сер.1. – 1989. –№ 1. – С.2–9.

249. Урсул, А.Д. Информатизация: системно-деятельностный подход / А.Д. Урсул. // НТИ. – Сер.2.– 1989. –№ 11. – С.2–8.

250. Урсул, А.Д. Информатизация общества: Введение в социальную информатику / А.Д. Урсул. – М., 1990.

251. Урсул, А.Д. Социальная информатика: две концепции развития / А.Д. Урсул. // НТИ. – Сер. 1. – 1990. № 1.– С. 2–7.

252. Филимонов, А.Ф. О разработке в США системы мер по защите национальной информационной инфраструктуры / А.Ф. Филимонов // Информационное общество. – № 1. – 1997.

253. Фаддеев, Д.К. К понятию энтропии конечной вероятностной схемы / Д.К. Фаддеев. – М.: УМН. т. XI, вып. 1, 1956.

254. Файнстейн, А. Основы теории информации / А. Файнстейн. – М.: Иностранная литература , 1960.

255. Фано, Р. Передача информации. Статистическая теория связи / Р. Фано. – М.: Мир, 1965.

256. Фёдоров, С.В. Основы теории информации: учеб. пособие / С.В. Фёдоров. – Владимир: Владимирский государственный университет, 1998. – 75 с.

257. Флейшман, Б.С. Основы системологии / Б.С. Флейшман. – М.: Радио и связь, 1982. – 368 с.

258. Феллер, В. Введение в теорию вероятностей и ее приложения. В двух томах. Т.1: Пер. с англ / В. Феллер. – М.: Мир, 1984.

259. Фисун, А.П. Правовые основы обеспечения защиты информации / Под ред. А.П. Фисуна. – Орел: ВИПС, 1997. – 131 с.

260. Фисун, А.П. Метод разработки содержания теоретических основ компьютерной графики: сб. научных статей / А.П. Фисун, Ф.М. Дорохов, А.Н. Касилов. – // Информационные технологии в деятельности органов внутренних дел. – Орел: Ор.ЮИ, 1999. – С. 22–33.

261. Фисун, А.П. О разработке программы исследования проблем информационных систем на основе построения их концептуальной классификационной модели / А.П. Фисун, А.Н. Касилов и др. // Материалы Международной научно-практической конференции (29 мая –2 июня 1995 г.) «Языки мозга и тела человека: проблемы и практическое использование в деятельности органов внутренних дел». – Орел: МАИ, Ор.ВШ МВД России, 1996. – С. 243–248.

262. Фисун, А.П. Анализ вариантов и направлений развития существующих государственных образовательных стандартов высшего профессионального образования в области защиты информации и информационной безопасности / А.П. Фисун, А.Н. Касилов // Сборник научных работ «Информационные технологии

263. в деятельности органов внутренних дел». – Орел: Ор.ЮИ МВД России, 1998. – С. 51–56.

264. Фисун, А.П. О государственных образовательных стандартах высшего профессионального образования в области информационной безопасности / А.П. Фисун, А.Н. Касилов // Материалы Международной конференции «Информатизация правоохранительных систем» (30 июня – 1 июля 1998 г.). Тезисы докладов. Часть 1. – М.: МАИ, Академия МВД России, 1998. – С. 147 – 149.

265. Фисун, А.П. Информатика. Часть 1. Информация и информационные системы как объект обеспечения

информационной безопасности: Курс лекций (Рукопись) / А.П. Фисун – Орел: ВИПС, Кафедра информатики и вычислительной техники, 1998. – 274с.

266. Фисун, А.П. Информатика и информационная безопасность: учеб. пособие / А.П. Фисун, А.Н. Касилов, А.Г. Мешков. – Орел: ОГУ, 1999.- 282 с.

267. Фисун, А.П. Теоретические основы информатики и информационная безопасность / А.П. Фисун, В.А. Минаев, В.Н. Саблин; Под ред. д-ров техн. наук, профессоров В. А. Минаева, В. Н. Саблина. – М.: Радио и связь, 2000. – 468 с.

268. Хакен, Г. Информация и самоорганизация / Г. Хакен. – М.: Мир, 1993.

269. Хастингс, Н. Справочник по статистическим распределениям / Н Хастингс, Дж. Пикок. – М.: Статистика, 1980. – 95 с.

270. Харкевич, А.А. О ценности информации / А.А. Харкевич // Проблемы кибернетики. – Вып.4. – М.: Физматгиз, 1960.

271. Хаффман, Д.А. Метод построения кодов с минимальной избыточностью: Кибернетический сборник. Вып. 3 / Д.А. Хаффман. – М.: ИЛ, 1961.

272. Хоор, Ч.Э. Непротиворечивые взаимодополняющие теории семантики языков программирования / Ч.Э. Хоор, П.Е. Лауэр. – М.: Мир, 1980. – С. 196 – 221.

273. Хоффман, Л.Дж. Современные методы защиты информации / Л.Дж Хоффман; пер. с англ., под ред. Герасименко В.А. – М.: Сов. радио, 1980. – 263 с.

274. Хинчин, А.Я. Об основных теоремах теории информации / А.Я. Хинчин. – М.: УМН, т. XI, вып. 1, 1956.

275. Хинчин, А. Я. Понятие энтропии в теории вероятностей / А.Я. Хинчин. – М.: УМН, т. VIII, вып. 3, 1953.

276. Холево, А.С. Введение в квантовую теорию информации / А.С. Холево – М.: МЦНМО, 2002.

277. Хэмминг, И.Р. Коды с обнаружением и исправлением ошибок / И.Р. Хэмминг. – М., 1956. – С. 7-22 Ц, А].

278. Цифровая обработка сигналов и ее применения. Digital signal processing and its applications: 1-я Международная конференция, 30 июня - 3 июля 1998 г. – М., 1998. – 21 с.

279. Цвиркун, А.Д. Имитационное моделирование в задачах синтеза структуры сложных систем (оптимизационно-имитационный

подход) / А.Д. Цвиркун, В.К. Акинфиев, В.А. Филиппов. – М.: Наука, 1985. – 174 с.

280. Цимбал, В.П. Теория информации и кодирование / В.П. Цимбал. – Киев: ВШ, 1982. – 304 с.

281. Цимбал, В.П. Теория информации и кодирование / В.П. Цимбал. – Киев: ВШ, 1982. – 304 с.

282. Цапенко, М.П. Измерительные информационные системы. Принципы построения / М.П. Цапенко. – М.: Энергия, 1974. – 320 с.

283. Чернавский, Д.С. Синергетика и информация (динамическая теория информации). Изд. 2-е, испр. и доп. / Д.С. Чернавский – М.: Едиториал УРСС, 2004. – 288 с.

284. Черри, К. Человек и информация: пер. с англ / К. Черри. – М.: Связь, 1972.

285. Чисар, И. Теория информации / И. Чисар, Я Кернер. – М.: Мир, 1985.

286. Шавелько, И.К. Основы теории кодирования и передачи информации / И.К. Шавелько, В.И. Мощиль. – М.: МИИГАиК, 1999. – 47 с.

287. Шаров, А.А. Биосемиотика: функционально–эволюционный подход к анализу и смыслу информации / А.А. Шаров // НТИ. – Сер. 2. – 1990. – № 12. – С. 10-20.

288. Шварцман, В.О. Теория передачи дискретной информации: учеб. для вузов связи / В.О. Шварцман, Г.А. Емельянов. – М.: Связь, 1979. – 424 с.

289. Шеннон, К. Работы по теории информации и кибернетике / К. Шеннон. – М.: Изд. иностр. лит., 1963.

290. Шепель, В.М. Настольная книга бизнесмена и менеджера: Управленческая гуманитарология / В.М. Шепель. – М.: Финансы и статистика, 1992.

291. Шемакин, Ю.И. Семантика информационной технологии / Ю.И. Шемакин // НТИ. – Сер.2. Информ. процессы и системы. – №11. – 1995. – С. 5-10.

292. Шерковин, Ю.А. Психологические проблемы массовых информационных процессов / Ю.А. Шерковин. – М.: 1973.

293. Шеннон, Р. Имитационное моделирование систем – искусство и наука: Пер. с англ. / Р. Шеннон. – М.: Мир, 1978. – 418 с.

294. Шилейко, А.В. Введение в информационную теорию систем / Под. ред. А.В. Шилейко. – М.: Радио и связь, 1985. – 280 с.

295. Шилейко, А.В. Энтропия и информация / А.В. Шилейко // НТИ. – Сер. 2. – 1993. – № 7. – С. 1-11.

296. Шилейко, А.В. Введение в информационную теорию систем / Под ред. В.А. Шилейко. – М.: Радио и связь, 1985.

297. Шнепс, М.А. Системы распределения информации. Методы расчета: Справ. Пособие / М.А. Шнепс. – М.: Связь, 1979. – 334 с.

298. Шрейдер, Ю.А. О феномене информационного продукта / Ю.А. Шрейдер // НТИ. – Сер.1. Орг. и методика информ. работы. №11. – С.21-24.

299. Шрейдер, Ю.А. Двойной облик современной информатики / Ю.А. Шрейдер // Природа. – 1988.– № 5.– С.64–71.

300. Шрейдер, Ю.А. Социальные аспекты информатики / Ю.А Шрейдер // НТИ. – Сер.2. – 1989. – №1. – С.2-9.

301. Шоломов, Л.А. Основы теории дискретных логических и вычислительных устройств / Л.А. Шоломов. – М.: Наука, 1980.

302. Шувалов, В.П. Передача дискретных сообщений: учеб. для вузов / Под ред. В.П. Шувалова. – М.: Радио и связь, 1990. – 464 с.

303. Шумаков, П.В. Дельфи 4. Руководство разработчика баз данных / П.В. Шумаков, В.В. Фаронов. – М.: НОЛИДЖ, 1999. – 560 с.

304. Энциклопедический словарь. – М.: Большая Советская энциклопедия, 1955.

305. Эшби, У.Р. Принципы самоорганизации / У.Р. Эшби // Принципы самоорганизации. – М.: Мир, 1966.

306. Эшби У.Р. Введение в кибернетику: пер. с англ. / У.Р. Эшби. – М.: ИЛ, 1959.

307. Яковлев, В.А. Защита информации на основе кодового зашумления / В.А. Яковлев. Под ред. В.И. Коржака. – СПб.: ВАС, 1993. – 245 с.

308. Якубайтис, Э.А. Информационные сети и системы. Справочная книга / Э.А. Якубайтис. – М.: Финансы и статистика. 1996. – 368 с.

309. Эдуард Кофлер– Entscheidungen bei teilweise bekannter Verteilung der Zustände, Zeitschrift für OR, Vol. 18/3, 1974

310. Эдуард Кофлер- Extensive Spiele bei unvollständiger Information, in Information in der Wirtschaft, Gesellschaft für Wirtschafts- und Sozialwissenschaften, Band 126, Берлин 1982

311. Эдуард Кофлер- Equilibrium Points, Stability and Regulation in Fuzzy Optimization Systems under Linear Partial Stochastic Information (LPI), Proceedings of the International Congress of Cybernetics and Systems, AFCET, Париж 1984, pp. 233-240

312. Эдуард Кофлер- Decision Makingunder Linear Partial Information[3]. Proceedings of the European Congress EUFIT, Ахен, 1994, p. 891-896.
313. Эдуард Кофлер- Linear Partial Information with Applications. Proceedings of ISFL 1997 (International Symposium on Fuzzy Logic), Цюрих, 1997, p.235-239.
314. Яглом, А.М. Вероятность и информация / А.М. Яглом, И.М. Яглом. – М.: Мир,1973.
315. Ahlswede R., Korner J. On the connections between the entropies of input and output distributions of discrete memoryless channels.//Proceedings of the Fifth Conference on Probability Theory, Brasov 1974, Editura Academiei Rep. Soc. Romania, Bucuresti 1977, pp. 13-23.
316. SI. Ahmed N., Bares R.M., Rao K.R. Multidimensional B1FORE Transform. Electronics Letters. 6. 1970. 237-238.
317. Ahmed N„ Rao K.R. Discrete Fourier and Hadamad transforms. Eletronics Letters. 6. 1970. 221-224.
318. Ahmed N.$_t$ Rao K.R. Complex BIFORE transform. Eletronics Letters. 6. 1970. 256-258, 387.
319. Ahmed N., Rao K.R. Additional properties of complex BIFORE transform. IEEE Trans. Audio Electroacoustics. 19. 1971. 252-253.
320. Algebraic and combinatorial theory: ACCT-VI: Sixth Intern Workshop. Sept. 6-12. 1998, Pskov, Russia: Proceedings - (M., 1998]. - VII. 261 с: илл. Библ. В конце докл. - Указ. 98-39161.
321. Arimoto S. An algorithm for computing the Capacity of Arbitrary Discrete Memoryless Channels, IEEE-IT 1972, 18, 14-20.
322. Billingsley P. Ergodic Theory and Information. – New York: Wiley, 1965.
323. Claude E. Shannon, Warren Weaver. The Mathematical Theory of Communication. Univ of Illinois Press, 1963.
324. Csiszar I. Simple proofs of some theorems on noiseless channels. 1C 1969, 14, 285-298.
325. Feistel H. Cryptography and computer privacy. Scientific American. 228. № 5, 1973, 15-2.
326. Feistel H., Nortz W.A., Smith J.L. Some cryptographic techniques for machine-to-machine data communications. Proc. IEEE. 63. 1975. 1545-1554.
327. Gej/e P.R. An open tetter to communication engineers. Proc. IEEE. 55. 1967, 2173 [I, 14].
328. Geffe P.R. How to protect data with ciphers that are really hard to break. Electronics. 46. 1973, 99-101.

329. Gilbert E.N., Mac-Williams F.J., Shane N.I.A. Codes which detect deception. Bell Syst. Tech. J. 53. 1974, 405-424 ll].

330. Hamming R. V. Error detecting and error correcting codes. BSTJ 1950, 29, 147-160.

331. Hardy G.H., Littlewood I.E., Polya G. Inequalities. Cambridge Univ. Press, 1934.

332. Harmuth H.F. Applications of Walsh functions in communications. IEEE Spectrum. 6. November, 1969, 81-91.

333. Harmuth H.F. Transmission of Information by Orthogonal Functions. Spinger. New York, 1970.

334. Hartley R.V.L. Transmission of information. BSTJ 1928, 7, 535.

335. Huffman DA. A Method for the Construction of Minimum Redudancy Codes. - Proc. IRE, 1952, т. 40, с. 1098-1101.

336. IEEE Transactions on Electromagnetic Compatibility, Special Issue on Applications of Walsh functions. 13. August 1971.

337. Le Garf A. Dictionnaire de l’informatique. – Paris:Presses Universitaires de France? 1982.

338. Leech J., Sloane N.J.A. Sphere packings and error-correcting-codes. Canad. J. Math. 23. 1971 718-745.

339. Li Ming. Introduction to Kolmogorov complexity and its applications. NY etc: Springer Cop. 1993. XX. 546 с. С илл.

340. Machlup F., Mansfield U. Cultural diversity in studies of information // The study of information: Interdisciplinary message.–New York: Wiley, 1883. –P/ 6–7, 18–23.

341. Marton K. A coding theorem for the discrete memory less broadcast channel. IEEE-IT. 1979.25,306-311.

342. Maxwell's Demon: Entropy, Information, Computing, H. S. Leff and A. F. Rex, Editors, Princeton University Press, Princeton, NJ (1990).

343. Pratt W.K., Kane J., Andrews H.C. Hadamard transform image coding. Proc. IEEE. 57. 1969 57-68.

344. Robert M. Gray. Entropy and information theory/ Nyetc. Springer. Cop. 1990 - XXIII, 24 с

345. R. Landauer, Information is Physical Proc. Workshop on Physics and Computation PhysComp'92 (IEEE Comp. Sci.Press, Los Alamitos, 1993, pp. 1-4.

346. Schroeder Manfred Robert. Number theory in science and communication. With applications in cryptography, physics, digital information

computing, a self similarity. Berlin etc. Springer, 1990. XIX. 374 c. с илл.

347. Saracevic T. An essay of the past and future of information science education// Inform. Processing & Management. – 1979. – Vol. 15.– P.1–15.

348. Sixth Joint Swedish-Russian international Workshop on information theory. August 22-27. 1993. Moile, Sweden Proceeding Lund. Studentlitt, 1993.464 c

349. Slamecka V., Pearson C. Information science // Encyclopedia of computer science end engineering.– 2nd ed. – Neww York: Van Nostrand, 1982.– P. 725–726.

350. Thomas M. Cover, Joy A. Thomas. Elements of information theory New York: Wiley, 1991.

351. Usher M.J. Information theory for information technologes.

352. Zorkoczy P. Information Technology: An Introduction.– White Plains (N.Y.): Knowledge Industry Publikations, 1983, IX.– 140 p.

Учебное издание

Еременко Владимир Тарасович
Минаев Владимир Александрович
Фисун Александр Павлович
Константинов Игорь Сергеевич
Коськин Александр Васильевич
Зернов Владимир Алексеевич
Белевская Юлия Александровна
Дворянкин Сергей Владимирович

ТЕОРИЯ ИНФОРМАЦИИ

Учебник

Редактор И.А. Хлюпина
Технический редактор Д.В. Агарков

Орловский государственный технический университет
Лицензия ИД №00670 от 05.01.2000 г.

Подписано к печати 31.08.2010 г. Формат 60x84 1/16.
Печать офсетная. Усл. печ. л. 27,7. Тираж 1000 экз.
Заказ №______

Отпечатано с готового оригинал-макета
на полиграфической базе ОрелГТУ,
302030, г. Орел, ул. Московская, 65.