МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РЕСПУБЛИКИ КАЗАХСТАН

Казахский агротехнический университет им.С.Сейфуллина

МУКУШЕВ Б. А., ТУРДИНА А.Б.

ОСНОВЫ ТЕОРИИ ИНФОРМАЦИИ

Утверждено Академическим советом университета в качестве учебного пособия

Нур-Султан 2022

УДК 004 (075.8)
ББК 32.973 я73
М90

Мукушев Б.А., Турдина А.Б. Основы теории информации: учебное пособие. – Нур-Султан: КАТУ им.С.Сейфуллина, 2022. - 112 с.

Рецензенты:
Мурзабекова Г.Е. - доцент кафедры ИКТ Казахского агротехнического университета им.С.Сейфуллина, кандидат физико-математических наук.
Сулеймен Р.Н. – доцент кафедры «Технической физики» НАО ЕНУ им. Л.Н.Гумилева, PhD(физика)

ISBN 978-601-257-383-1

Учебное пособие содержит необходимые теоретические материалы, практические задания, контрольные вопросы и задания для самостоятельной работы студентов. Пособие состоит из трех глав, где приведены основные вопросы теории информации и представлены задания для самостоятельной работы.

Данное издание предназначено для студентов, магистрантов и преподавателей вуза, а также лиц, занимающихся самообразованием.

УДК 004 (075.8)
ББК 32.973 я73
М90

ISBN 978-601-257-383-1

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Теорией информации называется наука, изучающая количественные закономерности, связанные с получением, передачей, обработкой и хранением информации. Возникнув в 40-х годах нашего века из практических задач теории связи, теория информации в настоящее время становится необходимым математическим аппаратом при изучении всевозможных процессов управления.

Как известно, основа теории информации была заложена благодаря стремительному развитию кибернетики, именно в кибернетике произошло углубление и обогащение понятия информации, было определено место информации в системах управления в живых организмах, в общественных и технических системах.

Темы учебного пособия расположены в последовательности, которая устанавливается исходя из логики курса. Каждое занятие носит комбинированный характер: на уроке выдается теоретический материал, который подкрепляется практическими занятиями. Для эффективной работы по темам рекомендуется организация самостоятельной работы студентов.

I. ИНФОРМАЦИЯ, СВОЙСТВА ИНФОРМАЦИИ И ЕЕ ИЗМЕРЕНИЕ

Понятия информации, данных, знаний. Виды информации, ее свойства и формы представления. Единицы измерения информации. Содержательный и алфавитный подходы к измерению количества ин формации. Формулы Хартли и Шеннона, закон аддитивности информации.

1.1 Основные понятия теории информации

В настоящее время признана фундаментальность понятия информации. Сегодня существует гипотеза о том, что информационные процессы в живой и неживой природе, в технике и обществе имеют общую основу, подчиняются единым законам и закономерностям. Осознание информации в качестве особого вида объективной реальности, присущей всему сущему, вызвало огромный интерес всего научного мира. Изучение главных свойств информации и особенностей их проявления в различных информационных средах является одной из важнейших проблем фундаментальной науки на ближайшие десятилетия. Ныне укрепляется убеждение в том, что практически все существующие в природе взаимосвязи носят информационный характер. Осмысливание феномена информации современной наукой и осознание ее главенствующей роли стало причиной появления нового фундаментального метода научного познания - *информационного подхода*, обладающего междисциплинарным качеством, призванного формировать информационную картину мира, новое научное мировоззрение и новую информационную культуру человека и общества.

В понятии информации различают два аспекта. Во-первых, информация представляет собой меру организации системы. В теории информации допускается такое название, как структурная информация, выражающая внутреннее достояние системы. Для социальной системы применяется понятие *информационной энтропии*, определяющейся по среднему значению информации. Во-вторых, от структурной информации следует отличать

относительную информацию, которая тесно связана с отражением одного процесса через другой процесс.

Теория информации рассматривается как существенная часть кибернетики.

Кибернетика - это наука об общих законах получения, хранения, передачи и переработки информации. Ее основной предмет исследования - это так называемые кибернетические системы, рассматриваемые абстрактно, вне зависимости от их материальной природы. Примеры кибернетических систем: автоматические регуляторы в технике, мозг человека или животных, биологическая популяция, социум. Часто кибернетику связывают с методами искусственного интеллекта, т.к. она разрабатывает общие принципы создания систем управления и систем для автоматизации умственного труда. Основными разделами (они фактически абсолютно самостоятельны и независимы) современной кибернетики считаются: теория информации, теория алгоритмов, теория автоматов, исследование операций, теория оптимального управления и теория распознавания образов.

Основоположниками кибернетики (датой ее рождения считается 1948 год, год соответствующей публикации) считаются американские ученые Норберт Винер (Wiener) и Клод Шеннон (Shannon, он же основоположник теории информации). Винер ввел основную категорию кибернетики – управление, показал существенные отличия этой категории от других, например, энергии, описал несколько задач, типичных для кибернетики, и привлек всеобщее внимание к особой роли вычислительных машин. Выделение категории управления позволило Винеру воспользоваться понятием информации, положив в основу кибернетики изучение законов передачи и преобразования информации.

Принцип управления лежит в основе организации и действия любых управляемых систем: автоматических устройств, живых организмов и т. п. Подобно тому, как введение понятия энергии позволило рассматривать все явления природы с единой точки зрения и отбросило целый ряд ложных теорий, так и введение понятия информации позволяет подойти с единой точки зрения к изучению самых различных процессов взаимодействия в природе.

Основные разделы теории информации – кодирование источника (сжимающее кодирование) и канальное (помехоустойчивое) кодирование. Теория информации тесно связана с криптографией и другими смежными дисциплинами.

Получение, обработка, передача и хранение различного рода информации - непременное условие работы любой управляющей системы. В этом процессе всегда происходит обмен информацией между различными звеньями системы. Любая информация для того, чтобы быть переданной, должна быть соответственным образом «закодирована», т.е. переведена на язык специальных символов или сигналов. Сигналами, передающими информацию, могут быть электрические импульсы, световые или звуковые колебания, механические перемещения и т.д.

Одной из задач теории информации является отыскание наиболее экономных методов кодирования, позволяющих передать заданную информацию с помощью минимального количества символов. Эта задача решается как при отсутствии, так и при наличии искажений (помех) в канале связи.

Другая типичная задача теории информации ставится следующим образом: имеется источник информации (передатчик), непрерывно вырабатывающий информацию, и канал связи, по которому эта информация передается в другую инстанцию (приемник). Какова должна быть пропускная способность канала связи для того, чтобы канал «справлялся» со своей задачей, т.е. передавал всю поступающую информацию без задержек и искажений?

Ряд задач теории информации относится к определению объема запоминающих устройств, предназначенных для хранения информации, к способам ввода информации в эти запоминающие устройства и вывода ее для непосредственного использования.

Чтобы решать подобные задачи, нужно, прежде всего, научиться измерять количественно объем передаваемой или хранимой информации, пропускную способность каналов связи и их чувствительность к помехам (искажениям).

Таким образом, теория информации представляет собой раздел прикладной математики, посвященный измерению информации, ее потока, «размеров» канала связи и т. п., особенно применительно к радио, телеграфии, телевидению и к другим средствам связи. Кроме

того, теория информации изучает методы построения кодов, обладающих полезными свойствами.

Как и любая математическая теория, теория информации оперирует с математическими моделями, а не с реальными физическими объектами (источниками и каналами связи).

Понятия информации, данных, знаний. Термин «информация» происходит от латинского «infomation», что означает «разъяснение, осведомление, изложение». В широком смысле информация – это общенаучное понятие, включающее в себя обмен сведениями между людьми, обмен сигналами между живой и неживой природой, людьми и устройствами. Понятие информации является одним из фундаментальных в современной науке вообще и базовым для информатики. Однако, если задаться целью формально определить понятие «информация», то сделать это чрезвычайно сложно, т.к. информация является первичным и неопределяемым в рамках науки понятием.

Понятие «информация» достаточно широко используется в обычной жизни современного человека, поэтому каждый имеет интуитивное представление, что это такое. Но когда наука начинает применять общеизвестные понятия, она уточняет их, приспосабливая к своим целям, ограничивает использование термина строгими рамками его применения в конкретной научной области.

Деятельность людей связана с переработкой и использованием материалов, энергии и информации. Соответственно развивались научные и технические дисциплины, отражающие вопросы материаловедения, энергетики и информатики. Значение информации в жизни общества стремительно растет, меняются методы работы с информацией, расширяются сферы применения новых информационных технологий. Сложность явления информации, его многоплановость, широта сферы применения и быстрое развитие отражается в постоянном появлении новых толкований понятий информатики и информации.

Само понятие информации в научное употребление ввел именно основатель кибернетики Н. Винер. И хотя он не дал определения информации, но, тем не менее, писал про нее в работе «Кибернетика, или Управление и связь в животном и машине»: «Информация есть информация, а не материя и не энергия».

Тем самым он, с одной стороны, указывает на феноменологическую сущность информации, противопоставляя ее понятиям материи и энергии, которые обычно характеризуют все аспекты нашего бытия, а с другой стороны, ставит на один уровень с этими понятиями по степени фундаментальности. Существуют и другие определения информации, отражающие основную суть данного понятия. Информацию понимают как нарушение симметрии в системе любой природы. Она также связана с вероятностным поведением этого явления (нарушения). Под информацией понимаются неоднородности распределения вещества и энергии в пространстве и времени в единстве их семантических, синтаксических и прагматических характеристик. (Семантика изучает знаковые системы как средство выражения смысла; синтактика изучает внутреннюю структуру знаковых систем безотносительно к выполняемым ими функциям, а прагматика — отношение знаковых систем к тем, кто их использует).

Проблема определения информации - одна из глубоких проблем философии, которая подходит к ней с позиции трех концепций: атрибутивной, функциональной и антропоцентрической. С атрибутивной точки зрения информация является свойством всего сущего, мерой упорядоченности, структурированности любой материальной системы. С позиции функциональной концепции информация появилась с возникновением жизни, так как связана с функционированием сложных самоорганизующихся систем, к которым относятся живые организмы и человеческое общество. Согласно антропоцентрической концепции информация существует лишь в человеческом сознании, которая функционирует только в условиях активизации мыслительной деятельности личности.

Рассмотрим множество определений и взглядов на понятие «информация» с различных точек зрения. Так, например, наиболее общее философское определение звучит следующим образом: «Информация есть отражение реального мира. Информация - отраженное разнообразие, то есть нарушение однообразия. Информация является одним из основных универсальных свойств материи» [1]. В узком, практическом толковании определение понятия «информация» представляется так: «Информация есть все

сведения, являющееся объектом хранения, передачи и преобразования» [2].

У подавляющего большинства авторов свое понимание информации, иногда в чем-то пересекающееся, но нередко совсем несовпадающее. Все разнообразие взглядов на информацию более или менее четко укладывается в две ведущие модели, одна из которых трактует информацию как неотъемлемое свойство материи, ее атрибут («атрибутивная концепция»), а другая – как неотъемлемый элемент самоуправляемых (технических, биологических, социальных) систем, как функцию этих систем («функционально-кибернетическая концепция»).

На современном этапе информация превратилась в глобальный, в принципе неистощимый ресурс человечества, вступившего в новую эпоху развития цивилизации. Информация является основной движущей силой всех эволюционных и революционных процессов в обществе. Она - стержень развития и взаимодействия всех его социальных структур. Игнорирование или недооценка роли информации в обществе как важнейшего социального фактора приводит к серьезным политическим ошибкам, о чем красноречиво свидетельствует многовековая история человечества. И, наоборот, умелое использование социальной информации неоднократно позволяло успешно разрешать, казалось бы, непреодолимые противоречия и преодолевать социальные кризисы.

В социальной информатике «Информация (Information) — содержание сообщения или сигнала; сведения, рассматриваемые в процессе их передачи или восприятия, позволяющие расширить знания об интересующем объекте» [3].

«Информация – первоначально — сведения, передаваемые одними людьми другим людям устным, письменным или каким-нибудь другим способом» [4]. В самом общем смысле информация есть обозначение некоторой формы связей или зависимостей объектов, явлений, мыслительных процессов. Информация есть понятие, абстракция, относящееся к определенному классу закономерностей материального мира и его отражения в человеческом сознании. В зависимости от области, в которой ведется исследование, и от класса задач, для которых вводится понятие информации, исследователи подбирают для него различные определения.

Автор теории информации К. Шеннон определил понятие информации как коммуникацию, связь, в процессе которой устраняется неопределенность.

При таком понимании информация – это результат выбора из набора возможных альтернатив. Однако математическая теория информации не охватывает все богатство содержания информации, поскольку она не учитывает содержательную сторону сообщения.

Дальнейшее развитие математического подхода к понятию «информация» отмечается в работах логиков (Р. Карнап, И. Бар-Хиллел) и математиков (А.Н. Колмогоров). В этих теориях понятие информации не связано ни с формой, ни с содержанием сообщений, передаваемых по каналу связи. Понятие «информация» в данном случае определяется как абстрактная величина, не существующая в физической реальности, подобно тому, как не существует мнимое число или не имеющая линейных размеров точка.

С кибернетической точки зрения информация (информационные процессы) есть во всех самоуправляемых системах (технических, биологических, социальных). При этом одна часть кибернетиков определяет информацию как содержание сигнала, сообщения, полученного кибернетической системой из внешнего мира. Здесь сигнал отождествляется с информацией, они рассматриваются как синонимы. Другая часть кибернетиков трактуют информацию как меру сложности структур, меру организации. Вот как определяет понятие «информация» американский ученый Б. Винер, сформулировавший основные направления кибернетики, автор трудов по математическому анализу, теории вероятностей, электрическим сетям и вычислительной техники: «информация - это обозначение содержания, полученного из внешнего мира».

Информация выступает в качестве меры разнообразия. Чем выше упорядоченность (организованность) системы, объекта, тем больше в ней содержится «связанной» информации. Отсюда делается вывод, что информация – фундаментальная естественнонаучная категория, находящаяся рядом с такими категориями как «вещество» и «энергия», что она является неотъемлемым свойством материи и потому существовала и будет существовать вечно.

С 50-60-х годов терминология теории информации стала применяться и в физиологии (Д. Адам). Была обнаружена близкая аналогия между управлением и связью в живом организме и в информационно-технических устройствах. В результате введения понятия «сенсорная информация» (т.е. оптические, акустические, вкусовые, тепловые и прочие сигналы, поступающие к организму извне или вырабатываемые внутри его, которые преобразуются в импульсы электрической или химической природы, передающиеся по нейронным цепям в центральную нервную систему и от нее – к соответствующим эффекторам) появились новые возможности для описания и объяснения физиологических процессов раздражимости, чувствительности, восприятия окружающей среды органами чувств и функционирования нервной системы.

В рамках генетики было сформулировано понятие генетической информации – как программа (код) биосинтеза белков, материально представленных полимерными цепочками ДНК. Генетическая информация заключена преимущественно в хромосомах, где она зашифрована в определенной последовательности нуклеидов в молекулах ДНК. Реализуется эта информация в ходе развития особи.

С правовой точки зрения информация определяется как некоторая совокупность различных сообщений о событиях, происходящих в правовой системе общества, ее подсистемах и элементах и во внешней по отношению к данным правовым информационным образования среде, об изменениях характеристик информационных образований и внешней среды, или как меру организации социально-экономических, политических, правовых, пространственных и временных факторов объекта. Она устраняет в правовых информационных образованиях, явлениях и процессах неопределенность и обычно связана с новыми, ранее неизвестными нам явлениями и фактами.

Таким образом, систематизируя вышеизложенное, можно сделать вывод, что для инженеров, биологов, генетиков, психологов понятие «информации» отождествляется с теми сигналами, импульсами, кодами, которые наблюдаются в технических и биологических системах. Радиотехники, телемеханики, программисты понимают под информацией рабочее тело, которое можно обрабатывать, транспортировать, так

же как электричество в электротехнике или жидкость в гидравлике. Это рабочее тело состоит из упорядоченных дискретных или непрерывных сигналов, с которыми и имеет дело информационная техника. Если попытаться объединить предложенные подходы, то получится следующее:

Информация - это:

- данные, определенным образом организованные, имеющие смысл, значение и ценность для своего потребителя и необходимая для принятия им решений, а также для реализации других функций и действий;
- совокупность знаний о фактических данных и зависимостях между ними, являющихся одним из видов ресурсов, используемых человеком в трудовой деятельности и быту;
- сведения о лицах, предметах, фактах, событиях, явлениях и процессах независимо от формы представления;
- сведения, неизвестные до их получения;
- значение, приписанное данным;
- средство и форма передачи знаний и опыта, сокращающая неопределенность и случайность и неосведомленность;
- обобщенный термин, относящийся к любым сигналам, звукам, знакам и т.д., которые могут передаваться, приниматься, записываться и/или храниться.

Приведенные выше определения понятия «информация» показывают, что понятия «знание», «информация», «данные» часто отождествляются. Однако, эти понятия необходимо различать.

Подходы к трактовке понятия «информация» уже были рассмотрены выше. Теперь остановимся на рассмотрении таких понятий как «данные» и «знания».

Вот как определяет понятие *«данные»* С.В. Симонович: «Мы живем в материальном мире. Все, что нас окружает и с чем мы сталкиваемся относится либо к физическим телам, либо к физическим полям. Все объекты находятся в состоянии непрерывного движения и изменения, которое сопровождается обменом энергией и ее переходом из одной формы в другую. Все виды энергообмена сопровождаются появлением сигналов. При взаимодействии сигналов с физическими телами в последних возникают определенные изменения свойств

– это явление называется регистрацией сигналов. Такие изменения можно наблюдать, измерять или фиксировать иными способами – при этом возникают и регистрируются новые сигналы, т.е. образуются данные» [5].

Известны также следующие трактовки понятия «данные».

Данные это:

- факты, цифры, и другие сведения о реальных и абстрактных лицах, предметах, объектах, явлениях и событиях, соответствующих определенной предметной области, представленные в цифровом, сим- вольном, графическом, звуковом и любом другом формате;

- информация, представленная в виде, пригодном для ее передачи и обработки автоматическими средствами, при возможном участии автоматизированными средствами с человеком;

- фактический материал, представленный в виде информации, чисел, символов или букв, используемый для описания личностей, объектов, ситуаций или других понятий с целью последующего анализа, обсуждения или принятия соответствующих решений.

Из всего многообразия подходов к определению понятия «данные» на наш взгляд справедливо то, которое говорит о том, что данные несут в себе информацию о событиях, произошедших в материальном мире, поскольку они являются регистрацией сигналов, возникших в результате этих событий. Однако данные не тождественны информации [5]. Станут ли данные информацией, зависит от того, известен ли метод преобразования данных в известные понятия. То есть, чтобы извлечь из данных информацию необходимо подобрать соответствующий форме данных адекватный метод получения информации. Данные, составляющие информацию, имеют свойства, однозначно определяющие адекватный метод получения этой информации. Причем необходимо учитывать тот факт, что информация не является статичным объектом – она динамически меняется и существует только в момент взаимодействия данных и методов. Все прочее время она пребывает в состоянии данных. Информация существует только в момент протекания информационного процесса. Все остальное время она содержится в виде данных. Одни и те же данные могут в момент потребления представлять разную

информацию в зависимости от степени адекватности взаимодействующих с ними методов.

По своей природе данные являются объективными, так как это результат регистрации объективно существующих сигналах, вызванных изменениями в материальных телах или полях. Методы являются субъективными. В основе искусственных методов лежат алгоритмы (упорядоченные последовательности команд), составленные и подготовленные людьми (субъектами). В основе естественных методов лежат биологические свойства субъектов информационного процесса. Таким образом, информация возникает и существует в момент диалектического взаимодействия объективных данных и субъективных методов.

Переходя к рассмотрению подходов к определению понятия «*знания*»можно выделить следующие трактовки.

Знания – это:

- вид информации, отражающей знания, опыт и восприятие человека - специалиста (эксперта) в определенной предметной области;

- множество всех текущих ситуаций в объектах данного типа и способы перехода от одного описания объекта к другому;

- осознание и толкование определенной информации, с учетом путей наилучшего ее использования для достижения конкретных целей, характеристиками знаний являются: внутренняя интерпретируемость, структурируемость, связанность и активность. «Знания есть факты плюс убеждения плюс правила».

Основываясь на приведенных выше трактовках рассматриваемых понятий, можно констатировать тот факт, что знание - это информация, но не всякая информация – знание. Информация выступает как знания, отчужденные от его носителей и обобществленные для всеобщего пользования. Другими словами, информация – это превращенная форма знаний, обеспечивающая их распространение и социальное функционирование. Получая информацию, пользователь превращает ее путем интеллектуального усвоения в свои личностные знания. Здесь мы имеем дело с так называемыми информационно-когнитивными процессами, связанными с представлением личностных знаний в виде информации и воссозданием этих знаний на основе информации.

В превращении информации в знание участвует целый ряд закономерностей, регулирующих деятельность мозга, и различных психических процессов, а также разнообразных правил, включающих знание системы общественных связей, – культурный контекст определенной эпохи. Благодаря этому знание становится достоянием общества, а не только отдельных индивидов. Между информацией и знаниями имеется разрыв. Человек должен творчески перерабатывать информацию, чтобы получить новые знания.

Таким образом, учитывая вышеизложенное, можно сделать вывод, что фиксируемые воспринимаемые факты окружающего мира представляют собой *данные.* При использовании данных в процессе решения конкретных задач появляется *информация.* Результаты решения задач, истинная, проверенная информация (сведения), обобщенная в виде законов, теорий, совокупностей взглядов и представлений представляет собой *знания.*

В рамках теории информации существуют 3 наиболее распространенные концепции информации, каждая из которых по-своему объясняет ее сущность.

Первая концепция (концепция К. Шеннона), отражая количественно- информационный подход, определяет информацию как меру неопределенности (энтропию) события. Количество информации в том или ином случае зависит от вероятности его получения: чем более вероятным является сообщение, тем меньше информации содержится в нем. Этот подход, хоть и не учитывает смысловую сторону информации, оказался весьма полезным в технике связи и вычислительной технике и послужил основой для измерения информации и оптимального кодирования сообщений. Кроме того, он представляется удобным для иллюстрации такого важного свойства информации, как новизна, неожиданность сообщений.

При таком понимании *информация – это снятая неопределенность, или результат выбора из набора возможных альтернатив.*

Вторая концепция *рассматривает информацию как свойство материи.* Ее появление связано с развитием кибернетики и основано на утверждении, что информацию содержат любые сообщения, воспринимаемые человеком или приборами. Наиболее

ярко и образно эта концепция информации выражена академиком В.М. Глушковым. Он писал, что «информацию несут не только испещренные буквами листы книги или человеческая речь, но и солнечный свет, складки горного хребта, шум водопада, шелест травы».

То есть, *информация как свойство материи создает представление о ее природе и структуре, упорядоченности и разнообразии. Она не может существовать вне материи, а значит, она существовала и будет существовать вечно, ее можно накапливать, хранить и перерабатывать.*

Третья концепция основана на логико-семантическом подходе, при котором информация трактуется как знание, причем не любое знание, а та его часть, которая используется для ориентировки, для активного действия, для управления и самоуправления. Иными словами, информация – это действующая, полезная часть знаний. Представитель этой концепции В. Г. Афанасьев, развивая логико-семантический подход, дает определение социальной информации*: «Информация, циркулирующая в обществе, используемая в управлении социальными процессами, является социальной информацией. Она представляет собой знания, сообщения, сведения о социальной форме движения материи и о всех других формах в той мере, в какой она используется обществом...»*

Рассмотренные подходы в определенной мере дополняют друг друга, освещают различные стороны сущности понятия информации и облегчают тем самым систематизацию ее основных свойств. Обобщив данные подходы, можно дать следующее определение информации:

Информация — это сведения, снимающие неопределенность об окружающем мире, которые являются объектом хранения, преобразования, передачи и использования.

Сведения — это знания, выраженные в сигналах, сообщениях, известиях, уведомлениях и т.д.

Информация передается в виде *сообщений,* определяющих форму и представление передаваемой информации. Примерами сообщений являются музыкальное произведение; телепередача; команды регулировщика на перекрестке; текст, распечатанный на принтере; данные, полученные в результате работы составленной

вами программы и т.д. При этом предполагается, что имеются «источник информации» и «получатель информации».

Сообщение от источника к получателю передается посредством какой-нибудь среды, являющейся в таком случае «каналом связи» (рис.1).

Рис. 1 - Универсальная схема передачи информации

Так, при передаче речевого сообщения в качестве такого канала связи можно рассматривать воздух, в котором распространяются звуковые волны, а в случае передачи письменного сообщения (например, текста, распечатанного на принтере) канал связи можно считать лист бумаги, на котором напечатан текст.

Виды информации. При классификации информации необходимо учитывать тот факт, что одна и та же информация может быть отнесена к различным классификационным группировкам, в зависимости от ее предметной ориентации.

Классификация информации

1. По форме представления:

- символьная, основанная на использовании символов – букв, цифр, знаков и т.д., является наиболее простой, практически применяется для передачи несложных сигналов о различных событиях (например, дорожные знаки, различный свет светофора и т.п.);
- текстовая, использующая так же символы (буквы, цифры, матема- тические знаки), но информация заложена не только в них, но и вих сочетании;
- графическая (фотографии, чертежи, схемы, рисунки);
- звуковая;
- видео.

2. По способу восприятия:

- визуальная;
- аудиальная;
- тактильная;

- обонятельная;
- вкусовая.

3. По общественному значению:
- массовая;
- специальная;
- личная.

Формы представления информации. Рассмотрим универсальную схему передачи информации (см. рис.1). Чтобы сообщение было передано от источника к получателю, необходима некоторая материальная субстанция – носитель информации. Сообщение, передаваемое с помощью носителя, назовем сигналом. В общем случае сигнал – это изменяющийся во времени физический процесс. Такой процесс может содержать различные характеристики (например, при передаче электрических сигналов могут изменяться напряжение и сила тока). Та из характеристик, которая используется для представления сообщений, называется параметром сигнала.

В случае, когда параметр сигнала принимает последовательное во времени конечное число значений (при этом все они могут быть пронумерованы), сигнал называется *дискретным,* а сообщение, передаваемое с помощью таких сигналов – дискретным сообщением. Информация, передаваемая источником, в этом случае также называется *дискретной.* Если же источник вырабатывает непрерывное сообщение (соответственно параметр сигнала – непрерывная функция во времени), соответствующая информация называется *непрерывной (аналоговой).* Пример дискретного сообщения - процесс чтения книги, информация в которой представлена текстом, т.е. дискретной последовательностью отдельных значков (букв). Примером непрерывного сообщения служит человеческая речь, передаваемая модулированной звуковой волной; параметром сигнала в этом случае является давление, создаваемое этой волной в точке нахождения приемника - человеческого уха.

Непрерывное сообщение может быть представлено непрерывной функцией, заданной на некотором отрезке [a, b]. Непрерывное сообщение можно преобразовать в дискретное (такая процедура называется **дискретизацией).** Для этого из бесконечного множества значений этой функции (параметра сигнала) выбирается

их определенное число, которое приближенно может характеризовать остальные значения. Таким образом, любое сообщение может быть представлено как дискретное, иначе говоря, последовательностью знаков некоторого алфавита.

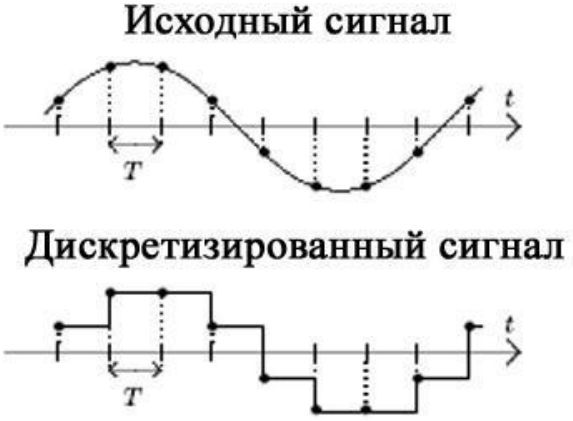


Рис. 2 - Преобразование аналогового сигнала в дискретный

Возможность дискретизации непрерывного сигнала с любой желаемой точностью (для возрастания точности достаточно уменьшить шаг) принципиально важна с точки зрения информатики. Компьютер – цифровая машина, т.е. внутреннее представление информации в нем дискретно. Дискретизация входной информации (см. рис.2.) позволяет сделать ее пригодной для компьютернойобработки.

Существуют и другие вычислительные машины - аналоговые ЭВМ. Они используются обычно для решения задач специального характера и широкой публике практически не известны. Эти ЭВМ в принципе не нуждаются в дискретизации входной информации, так как ее внутреннее представление у них непрерывно. В этом случае все наоборот - если внешняя информация дискретна, то ее «перед употреблением» необходимо преобразовать в аналоговую.

Свойства информации. «Информация является динамическим объектом, образующимся в момент взаимодействия объективных данных и субъективных методов» [6]. Как и всякий объект, информация обладает свойствами. На свойства информации влияют как свойства данных, так и свойства методов, взаимодействующих с данными в ходе информационного процесса. По окончании свойства процесса свойства информации переносятся на свойства новых данных, то есть свойства методов

могут переходить на свойства данных.

Знания, информация – обладают свойствами далеко не обычными. Например, известно высказывание Б.Шоу: «если у вас и у меня имеется по одному яблоку, и мы ими обменялись, то у каждого из нас осталось по одному яблоку; если у вас и у меня имеется по одной идее, и мы ими обменялись, то у каждого из нас будет по две идеи». Однако этим особенности свойств информации не ограничиваются. Информация специфична и с точки зрения старения (т.е. на информацию действует не само время, а появление новой информации, отрицающей или уточняющей данную), и с точки зрения различных вариантов относительно материального носителя или знаковой формы, и с точки зрения воздействия и так далее.

Можно привести немало разнообразных свойств информации. Каждая научная дисциплина рассматривает те свойства, которые ей наиболее важны. Систематизация существующих подходов к выделению свойств информации, позволяет говорить о том, что информации присущи следующие свойства.

1. Атрибутивные свойства – это те свойства, без которых информация не существует. К данной категории свойств относятся:

- *неотрывность информации от физического носителя и языковая природа информации* (одно из важнейших направлений информатики как науки является изучение особенностей различных носителей и языков информации, разработка новых, более совершенных и современных. Необходимо отметить, что хотя информация и неотрывна от физического носителя и имеет языковую природу, она не связана жестко ни с конкретным языком, ни с конкретным носителем);

- *дискретность* (содержащиеся в информации сведения, знания - дискретны, т.е. характеризуют отдельные фактические данные, закономерности и свойства изучаемых объектов, которые распространяются в виде различных сообщений, состоящих из линии, составного цвета, буквы, цифры, символа, знака);

- *непрерывность* (информация имеет свойство сливаться с уже за- фиксированной и накопленной ранее, тем самым, способствуя по- ступательному развитию и накоплению).

2. Прагматические свойства – это те свойства, которые характеризуют степень полезности информации для пользователя,

потребителя и практики. Проявляются в процессе использования информации. К данной категории свойств относятся:

- *смысл и новизна* (это свойство характеризует перемещение информации в социальных коммуникациях, и выделяет ту ее часть, кото- рая нова для потребителя);

- *полезность* (уменьшение неопределенности сведений об объекте, дезинформация расценивается как отрицательные значения полез- ной информации);

- *ценность* (ценность информации различна для различных потреби- телей и пользователей);

- *кумулятивность* (характеризует накопление и хранение информации);

- *полнота* (характеризует качество информации и определяет достаточность данных для принятия решений или для создания новых данных на основе имеющихся; чем полнее данные, тем шире диапа- зон методов, которые можно использовать, тем проще подобрать метод, вносящий минимум погрешностей в ход информационного процесса);

- *достоверность* (данные возникают в момент регистрации сигналов, но не все сигналы являются полезными – всегда присутствует какой-то уровень посторонних сигналов, в результате чего полезные данные сопровождаются определенным уровнем информационного шума; если полезный сигнал зарегистрирован более четко, чем посторонние сигналы, достоверность информации может быть более высокой; при увеличении уровня шумов достоверность информации снижается; в этом случае для передачи того же количества информации требуется использовать либо больше данных, либо более сложные методы);

- *адекватность* (это степень соответствия 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);

- *доступность* (мера возможности получить ту или иную информацию; на степень доступности информации влияют одновременно как доступность данных, так и доступность адекватных методов для их интерпретации; отсутствие доступа к

данным или отсутствие адекватных методов обработки данных приводят к одинаковому результату: информация оказывается недоступной; отсутствие адекватных методов для работы с данными во многих случаях приводит к применению неадекватных методов, в результате чего образуется неполная, неадекватная или недостоверная информация);

- *актуальность* (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);

- *объективность и субъективность* (понятие объективности информации является относительным; это понятно, если учесть, что методы являются субъективными; 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).

3. Динамические свойства – это те свойства, которые характеризуют изменение информации во времени:

- *рост информации* (движение информации в информационных коммуникациях и постоянное ее распространение и рост определяют свойство многократного распространения или повторяемости. Хотя информация и зависима от конкретного языка и конкретного носителя, она не связана жестко ни с конкретным

языком, ни с конкретным носителем. Благодаря этому информация может быть полученаи использована несколькими потребителями.);

- *старение* (информация подвержена влиянию времени).

1.2 Методы и единицы измерения информации

В связи с прогрессом технических средств массовых и других коммуникаций, и в особенности с ростом объема передаваемых сообщений (информаций) появилась необходимость их *измерения* для улучшения условий передачи. Поскольку информация характеризуется некоторыми параметрами, раскрывающими различные ее качества, следует измерять по мере возможности все эти параметры (количество информации, смысл и ценность информации, среднее значение информации и др.)

В процедуре измерения информации центральное место занимает вычисление количества сообщений, которое определяется математически как величина логарифмически обратно пропорциональная степени вероятности того события, о котором идет речь в сообщении. Чем чаще происходит одно и то же событие, тем меньше информации об этом событии. Информационным будет редкое, маловероятное событие.

Изучение научного понятия информации с точки зрения современной науки раскрыло новый аспект материального единства мира, позволило подойти с единой точки зрения ко многим, ранее казавшимся различным процессам: передаче сообщений по техническим каналам связи, функционированию нервной системы и вычислительной машины, разнообразным процессам управления, передаче наследственности потомкам через гены и др.

В неживой природе единственным универсальным научным подходом к изучению явлений широкого класса является понятие *«энергия»*. С единой энергетической точки зрения стало возможным обсуждать и сопоставлять явления электрические и механические, биологические, тепловые, явления, ранее совместно не изучавшиеся. А «обмен энергией» между различными природными объектами является основным условием существования и развития материи. В настоящее время на такую же роль инструмента научного познания претендует понятие

«информация». На основе информационного анализа удалось изучить с единой точки зрения природные, социальные и экономические явления.

Например, при исследовании влияния различных сигналов из внешней среды на животного важными являются их адекватные реакции на эти сигналы, то есть поведения. Не понимая, как животное пользуется получаемой им информацией, нельзя глубоко понять основы его поведения. Также невозможно понять суть работы управляющей системы любой природы, если не знать всех информационных процессов, в ней протекающих. Информация является источником принятия решений. Также через информацию (подаваемые сигналы) реализуется управление системой любой природы.

Измерение информации выступает главной проблемой методологии научного познания, также учебного.

Понятие об измеримости информации связано с тем, что получение информации, ее увеличение одновременно означает уменьшение незнания или информационной неопределенности.

Например, известно, что А. живет в городе С. Сообщение о том, что он живет по улице Иртышской уменьшило неопределенность. Получив такую одноразовую информацию, мы стали знать больше, но информационная неопределенность осталась, хотя уменьшилась. Логарифмически обратную величину информационной неопределенности принято называть вероятностью.

Различные подходы к определению понятия «количество информации». Единицы измерения информации и соотношение между ними. Определить понятие «количество информации» довольно сложно. В решении этой проблемы существуют два основных подхода. Исторически они возникли практически одновременно. В конце 40-х годов XX века один из основоположников кибернетики Клод Шеннон развил «вероятностный подход», а работы по созданию ЭВМ привели к «объемному» подходу.

Шеннон предложил в конце 40-х годов единицу измерения информации - бит. Каждому сигналу в теории приписывалась априорная вероятность его появления. Чем меньше вероятность появления того или иного сигнала, тем больше информации он

несет для потребителя (т.е. чем неожиданнее новость, тем больше ее информативность).

Информация равна нулю, когда возможно только одно событие. С ростом числа событий она увеличивается и достигает максимального значения, когда события равновероятны.

Информация размером в один бит содержится в ответе на вопрос, требующий ответа «да» или «нет».

Сообщение, уменьшающее неопределенность знаний человека в два раза, несет для него 1 бит информации.

Для хранения информации в компьютере используются специальные устройства памяти. Дискретную информацию хранить гораздо проще непрерывной, т.к. она описывается последовательностью чисел. Если представить каждое число в двоичной системе счисления, то дискретная информация предстанет в виде последовательностей нулей и единиц. Присутствие или отсутствие какого-либо признака в некотором устройстве может описывать некоторую цифру в какой-нибудь из этих последовательностей. Например, позиция на дискете описывает место цифры, а полярность намагниченности – ее значение. Для записи дискретной информации можно использовать ряд переключателей, различные виды магнитных и лазерных дисков, электронные триггеры и т.п. Одна позиция для двоичной цифры в описании дискретной информации тоже называется битом (bit, binary digit – «двоичная цифра»).

Непрерывную информацию тоже измеряют в битах.

Бит – это очень маленькая единица, поэтому часто используется величина в 8 раз большая – байт (byte, B). Как и для прочих стандартных единиц измерения для бита и байта существуют производные от них единицы, образуемые при помощи приставок кило (K), мега (M), гига (G или Г), тера (T), пета (P или П) и других. Но для битов и байтов они означают не степени 10, а степени двойки:

Кило – 2^{10} = 1024; Мега – 2^{20}; Гига – 2^{30}; Тера – 2^{40}; Пета – 2^{50}.

Пример 1. Определить сколько бит в 1 килобайте?

Решение: 1 Килобайт = 1*1024=1024 байт*8= 8192 бита

Пример 2. Определить сколько мегабайт составляют 8192 бита?

Решение: 8192 бита = 8192:8 = 1024 байта: 1024 = 1 Килобайт :1024 = 0, 0009765625 Мегабайт

Задания:

1. Объем информационного сообщения 1 572 864 байт выразить в кило-байтах и мегабайтах.
2. Определить количество битов в двух килобайтах, используя для чиселтолько степени 2.

Методы измерения информации. В 1865 г. немецкий физик Рудольф Клаузиус ввел в статистическую физику понятие энтропии или меры уравновешенности системы.

В статистической физике *энтропия* трактуется как мера вероятности пребывания системы в данном состоянии. Чем больше беспорядка, тем больше энтропия. Любая система постепенно переходит к своему более вероятному состоянию.

Энтропия - понятие, пришедшее из физики и распространившееся на все сферы жизнедеятельности. Означает усреднение, распыление, оскудение, хаотичность, беспорядочность.

Энтропия (от греч. en, tropē – поворот, превращение) — в теории информации: величина, характеризующая степень неопределенности системы.

В 1921 г. основатель большей части математической статистики, англичанин Рональд Фишер впервые ввел термин «информация» в математику, но полученные им формулы носят очень специальный характер.

В 1948 г. Клод Шеннон в своих работах по теории связи выписывает формулы для вычисления количества информация и энтропии. Термин энтропия используется Шенноном по совету патриарха компьютерной эры фон Неймана, отметившего, что полученные Шенноном для теории связи формулы для ее расчета совпали с соответствующими формулами статистической физики, а также то, что «точно никто не знает» что же такое энтропия.

Если отвлечься от конкретного смыслового содержания информации и рассматривать сообщения информации как последовательности знаков, сигналов, то их можно представлять битами, а измерять в байтах, килобайтах, мегабайтах, гигабайтах, терабайтах и петабайтах.

Выше было отмечено, что информация может пониматься и интерпретироваться по-разному. Вследствие этого имеются различные подходы к определению методов измерения информации, меры количества информации. Раздел информатики

(теории информации) изучающий методы измерения информации называется *информметрией.*

Количество информации – числовая величина, адекватно характеризующая актуализируемую информацию по разнообразию, сложности, структурированности, определённости, выбору (вероятности) состояний отображаемой системы.

Содержательный подход (вероятностный). Количество информации, заключенное в сообщении, определяется объемом знаний, который несет это сообщение получающему его человеку. Сообщение содержит информацию для человека, если заключенные в нем сведения являются для этого человека *новыми и понятными* и, следовательно, пополняют его знания.

При содержательном подходе возможна качественная оценка информации: *полезная, безразличная, важная, вредная*... Одну и ту же информацию разные люди могут оценить по-разному.

Если рассматривается система, которая может принимать одно из n возможных состояний, то актуальна задача оценки такого выбора, исхода. Такой оценкой может стать мера информации (или события). Мера – это некоторая непрерывная действительная неотрицательная функция, определённая на множестве событий и являющаяся аддитивной т.е. мера конечного объединения событий (множеств) равна сумме мер каждого события.

Формула Хартли. Пусть имеется N состояний системы S или N опытов с различными, равновозможными последовательными состояниями системы. Если каждое состояние системы закодировать, например, двоичными кодами определённой длины d, то эту длину необходимо выбрать так, чтобы число всех различных комбинаций было бы не меньше, чем N. Наименьшее число, при котором это возможно или мера разнообразия множества состояний системы задаётся формулой Р. Хартли:

$$H = k \log_a N,$$

где k – коэффициент пропорциональности (масштабирования, в зависимости от выбранной единицы измерения меры), а - основание системы меры.

Если измерение ведётся:

- в экспоненциальной системе, то k=1, H=lnN (нат);
- в двоичной системе, то k=1/ln2, $H=\log_2 N$ (бит);
- в десятичной системе, то k=1/ln10, H=lgN (дит).

Пусть в некотором сообщении содержатся сведения о том, что произошло одно из N равновероятных событий (равновероятность означает, что ни одно событие не имеет преимуществ перед другими). Тогда количество информации, заключенное в этом сообщении, – x бит и число N связаны формулой:

$$2^x = N.$$

Данная формула является показательным уравнением относительно неизвестной x. Из математики известно, что решение такого уравнения имеет вид:

$x = \log_2 N$ (логарифм от N по основанию 2).

Пример1. В барабане для розыгрыша лотереи находится 32 шара. Сколько информации содержит сообщение о первом выпавшем номере (например, выпал номер 15)?

Решение: Поскольку вытаскивание любого из 32 шаров равновероятно, то количество информации об одном выпавшем номере находится из уравнения: $2^x = 32$. Но $32 = 2^5$. Следовательно, x = 5 бит. Очевидно, ответ не зависит от того, какой именно выпал номер.

Закон аддитивности информации. Количество информации $H(x_1, x_2)$, необходимое для установления пары (x_1, x_2), равно сумме количеств информации $H(x_1)$ и $H(x_2)$, необходимых для независимого установления элементов x_1 , x_2.

$$H(x_1, x_2) = H(x_1) + H(x_2)$$

Пример 2: Используя закон аддитивности и формулу Хартли, подсчитать, какое количество информации несет достоверный прогноз погоды.

Решение: Предположим, что прогноз погоды на следующий день заключается в предсказании дневной температуры (обычно делается выбор из 16 возможных для данного сезона значений) и одного из 4-х значений облачности (солнечно, переменная облачность, пасмурно, дождь).

Тогда, $H(x_1,x_2) = H(x_1) + H(x_2) = \log_2 16 + \log_2 4 = = 4 + 2 = 6$ бит.

Пример 3. Шарик находится в одном из восьми ящиков. Информационная неопределенность равна восьми, а вероятность нахождения шарика в одном из восьми ящиков равна 1/8.

Пример 4. Книга лежит на одной из двух полок – верхней или нижней.

Сообщение, уменьшающее неопределенность ровно вдвое, содержит единицу информации – бит. Сообщение о том , что книга лежит на верней полке, несет один бит информации.

Пример 5. Книга лежит на одной из трех полок – верхней, средней или нижней. Сообщение о том, что книга лежит на средней полке, несет в себе информацию больше, чем бит.

Научный подход к оценке информации был предложен в 1928 году американским инженером Р.Хартли. Он ввел в теорию информации формулу, названную впоследствии его именем – *формулу Хартли:*

$$I = \log_2 N,$$

где N-количество *равновероятных* событий, I-количество бит в сообщении.

Пример 6. Информация о том, что книга лежит на одной из трех полок, содержит

$I=\log_2 3=1{,}585$ бит информации.

Формулу Хартли (1) можно написать по другому. Поскольку каждое из N возможных событий имеет одинаковую вероятность: P=1/N, то N=1/P, значит

$$I = \log_2 N = \log_2 (1/P) = -\log_2 P$$

Задачи:

1. «Вы выходите на следующей остановке?» - спросили человека в автобусе. «Нет», - ответил он. Сколько информации содержит ответ?

2. Какой объем информации содержит сообщение, уменьшающее не определенность знаний в 4 раза?

3. Вы подошли к светофору, когда горел желтый свет. После этого загорелся зеленый. Какое количество информации вы при этом получили?

4. Вы подошли к светофору, когда горел красный свет. После этого загорелся желтый свет. Сколько информации вы при этом получили?

5. Группа школьников пришла в бассейн, в котором 4 дорожки для плавания. Тренер сообщил, что группа будет плавать на дорожке номер 3. Сколько информации получили школьники из этого сообщения?

6. В корзине лежат 8 шаров. Все шары разного цвета. Сколько информации несет сообщение о том, что из корзины достали красный шар?

7. Была получена телеграмма: «Встречайте, вагон 7». Известно, что в составе поезда 16 вагонов. Какое количество информации было получено?

8. В школьной библиотеке 16 стеллажей с книгами. На каждом стеллаже 8 полок. Библиотекарь сообщил Пете, что нужная ему книга находится на пятом стеллаже на третьей сверху полке. Какое количество информации библиотекарь передал Пете?

9. При угадывании целого числа в диапазоне от 1 до N было получено 7 бит информации. Чему равно N?

10. При угадывании целого числа в некотором диапазоне было получено6 бит информации. Сколько чисел содержит этот диапазон?

11. Сообщение о том, что ваш друг живет на 10 этаже, несет 4 бита ин формации. Сколько этажей в доме?

12. Сообщение о том, что Петя живет во втором подъезде, несет 3 бита информации. Сколько подъездов в доме?

13. В коробке лежат 7 разноцветных карандашей. Какое количество информации содержит сообщение, что из коробки достали красный карандаш?

14. Какое количество информации несет сообщение: «Встреча назначена на сентябрь».

15. Какое количество информации несет сообщение о том, что встреча назначена на 15 число?

16. Какое количество информации несет сообщение о том, что встреча назначена на 23 октября в 15.00?

17. При угадывании целого числа в некотором диапазоне было получено8 бит информации. Сколько чисел содержит этот диапазон?

18. В корзине лежат 16 шаров. Все шары разного цвета. Сколько информации несет сообщение о том, что из корзины достали белый шар?

19. Сообщение о том, что ваш друг живет на 7 этаже, несет 3 бита ин формации. Сколько этажей в доме?

20. Сообщение о том, что Петя живет в третьем подъезде, несет 2 бита информации. Сколько подъездов в доме?

Формула Хартли отвлечена от семантических и качественных, индивидуальных свойств рассматриваемой системы (качества информации, содержащейся в системе, в проявлениях системы с помощью рассматриваемых N состояний системы). Это основная положительная сторона этой формулы. Но имеется и основная отрицательная сторона: формула не учитывает различимость и различность рассматриваемых N состояний системы.

Уменьшение (увеличение) H может свидетельствовать об уменьшении(увеличении) разнообразия состояний N системы.

Существует множество ситуаций, когда возможные события имеют различные вероятности реализации. Формулу для вычисления количества информации в случае различных вероятностей событий предложил К. Шеннон в 1948 году.

Формула Шеннона. Формула Шеннона дает оценку информации независимо, отвлеченно от ее смысла.

Рассмотрим *пример*: в коробке имеется 50 шаров. Из них 40 белых и 10 черных. Очевидно, вероятность того, что при вытаскивании «не глядя» попадется белый шар больше, чем вероятность попадания черного.

Обозначим $p_ч$ – вероятность попадания при вытаскивании черного шара, $p_б$ вероятность попадания белого шара.

Тогда: $P_ч = 10/50 = 0,2$; $p_б = 40/50 = 0,8$.

Отсюда видно, что вероятность попадания белого шара в 4 раза больше, чем черного.

Из примера можно сделать вывод: *если N – это общее число возможных исходов какого-то процесса* (вытаскивание шара), *и из них интересующее нас событие* (вытаскивание белого шара) *может произойти К раз, то вероятность этого события равна K/N*.

Вероятность выражается в долях единицы. В частном случае, вероятность достоверного события равна 1 (из 50 белых шаров вытащен белый шар); вероятность невозможного события равна нулю (из 50 белых шаров вытащен черный шар).

Качественную связь между вероятностью события и количеством информации в сообщении об этом событии можно выразить так: *чем меньше вероятность некоторого события, тем больше информации содержит сообщение об этом событии.*

Количественная зависимость между вероятностью события (p) и количеством информации в сообщении о нем (I) выражается формулой:

$$I = \log_2 (1/p).$$

Пример 1. В задаче о шарах определим количество информации в сообщении о попадании белого шара и черного шара:

$i_б = \log_2(1/0,8) = \log_2(1,25) = 0,321928$; $i_ч = \log_2(1/0,2) = 1\log_2 5 = 2,321928$.

Согласно формуле Шеннона, количество информации определяется по формуле:

$$I = -\sum_{i=1}^{N} p_i log_2 p_i,$$

где I – количество информации; N – количество возможных событий; p_i - вероятность i –го события.

Пример. Пусть при бросании несимметричной четырехгранной пирамидки вероятности отдельных событий будут равны: p_1 = ½, p_2 = ¼ , p_3 = 1/8 , p_4 = 1/8. Тогда количество информации, которое

мы получим после реализации одного из них, можно рассчитать по формуле

I = – (l/2*log 2 l/2 + 1/4*log 2 1/4 + 1/8*1log 2 1/8 + l/8*log 2 l/8) = (1/2 + 2/4 + 3/8 + 3/8) битов = 14/8 битов = 1,75 бита.

Этот подход к определению количества информации называется *вероятностным.*

Для частного, но широко распространенного и рассмотренного выше случая, когда события равновероятны, величину количества информации **I** можно рассчитать по формуле:

$$I = -\sum_{i=1}^{N} \frac{1}{N} log_2 \frac{1}{N} = log_2 N$$

По формуле можно определить, например, количество информации, которое мы получим при бросании симметричной и однородной четырехгранной пирамидки:

I = log 2 4 = 2 бита.

Таким образом, при бросании симметричной пирамидки, когда события равновероятны, мы получим большее количество информации (2 бита), чем при бросании несимметричной (1,75 бита), когда события неравновероятны.

Количество информации, которое мы получаем, достигает максимального значения, если события равновероятны.

Таким образом, К. Шенноном доказана теорема о единственности меры количества информации. Для случая равномерного закона распределения плотности вероятности формула Шеннона совпадает с формулой Хартли.

Измерение ценности информации. Теперь рассмотрим один важных вопросов измерения информации – *измерение ценности информации.* Ценность информации не может быть определена независимо от конкретного процесса, в котором эти сведения используются, от рецептора информации. Приведем простой пример. Учебник физики для 11-ого класса содержит богатую информацию. Какова ее ценность? Для ученика начального класса эта ценность равна нулю, так как он еще не в состоянии эту информацию воспринять. Для ученого-физика она также равна нулю, так как все это он уже знает. Наибольшую ценность учебник представляет для учащихся 11-класса. Таким образом, ценность одной информации для различных лиц может быть различной, она

зависит от предварительного запаса информации, которым уже располагает ее приемник.

Хотя такое субъективное свойство как «ценность» (или «значимость») информации до недавнего времени непосредственно не измерялось, однако посредством информационной энтропии этот важный параметр информации стал измеряемым, что нашло применение в обработке результатов тестирования.

Ярким примером применения элементов теории информации в практической деятельности человека служит измерение параметров учения обучающегося посредством информационной энтропии. В настоящее время среди различных подходов контроля учебной деятельности учащихся тестирование считается более справедливым и объективным способом измерения их знаний. Известно, что тестирование относится к статистической форме контроля знаний и то есть имеет информативную сущность. На основе тестирования мы обычно получаем информацию о количественной характеристике знаний обучаемого. Это информация определяется формулой

$$I_i = - lnP_i , \quad P_i = L_{+i} / L$$

где P_i - вероятность правильного ответа на вопрос с номером i, L_{+i} - число правильных ответов, L - число испытуемых. Вероятность принимает значение между нулем и единицей, следовательно, логарифм от P_i , будет отрицательным числом, но из-за знака минус в формуле результат будет положительным числом - информация принимает положительные значения в интервале от нуля до бесконечности.

Более информативным будет маловероятное ($P_i \approx 0$) событие - правильные ответы к трудному вопросу. Если $P_i \approx 1$, т.е. на вопрос с номером i все испытуемые нашли правильный ответ, то информация от этого вопроса близка к нулю, т.к. ln1 = 0. Отсюда следует, что именно информация является количественной характеристикой относительной трудности, значимости вопроса; при обработке результатов, окончательном подборе вопросов теста необходимо использовать формулу.

Информация, определенная таким образом, является элементарным актом проявления знания и определяет коэффициент значимости конкретного вопроса предлагаемого теста. По единичным ответам без обоснования, доказательства правильности

его выбора нельзя судить об уровне знаний испытуемого. При тестировании об относительном уровне знаний можно судить только по среднему значению информации - информационной энтропии, определяемой по формуле

$$S = -\sum_i P_i \ln P_i,$$

где индекс суммирования i принимает значения, равные номерам вопросов, на которые даны правильные ответы.

Понятие *энтропии* является общенаучным, характеризует степень равновесия, совершенства сложного объекта, меру неопределенности (из-за его сложности) этого объекта. Относительный уровень знаний будет выше у того абитуриента (ученика), у которого информационная энтропия тестирования больше, т.е. больше мера неопределенности (трудности) вопросов, на которые получены правильные ответы.

При определении конкурсных мест по тестированию (например, на вступительных, выпускных экзаменах) учитывается коэффициент значимости каждого вопроса, путем вычисления соответствующей информации. Относительный уровень знаний каждого испытуемого определяется средней информацией или информационной энтропией.

Пример 2. Приведем пример обработки результатов конкурсного тестирования.

Воспользуемся предложенной таблицей результатов тестирования (рис.3), где указаны правильные ответы семи испытуемых (L = 7) на 5 вопросов (i = 5).

Справа таблицы представлены баллы Σ, вычисленные простым суммированием. Очевидно, что на основе простого суммирования одинаковых баллов за каждый вопрос нельзя делать окончательные выводы об уровне знаний. Суммарные баллы оказались одинаковыми у нескольких испытуемых, хотя они ответили правильно на разные вопросы (случаи L = 1, 3, 4, 7). Кому нужно отдать предпочтение ?

Вопросы с номерами i = 1, 4 оказались относительно трудными, значит, эти вопросы несут в себе большую информацию, то есть они ценнее, нежели другие вопросы теста. Напротив, второй вопрос имел нулевую ценность для этой группы тестируемых. Какое преимущество должны получить участники конкурса, которые затратили больше времени на первые трудные вопросы и правильно

ответили на них? Суммарный балл каждого столбца делится на общее число испытуемых L = 7 и находится P_i - вероятность правильного ответа на i - вопрос. После этого вычисляется информация i – вопроса по формуле $I_i = -lnP_i$. Умножая P_i на I_i , соответствующие только правильным ответам, и суммируя эти произведения по строке, находим информационную энтропию S знания каждого испытуемого. Например, информационная энтропия (средняя информация) знания первого испытуемого равна S_1=0,3579 + 0 + 0,2403 = 0,5983. Относительный уровень знаний выше у того, у кого информационная энтропия S знания больше. Поскольку испытуемые под номерами 3 и 7 имеют одинаковую информационную энтропию знания, они должны делить 4-ое и 5-ое места. Окончательное распределение конкурсных мест представлено в столбце № .

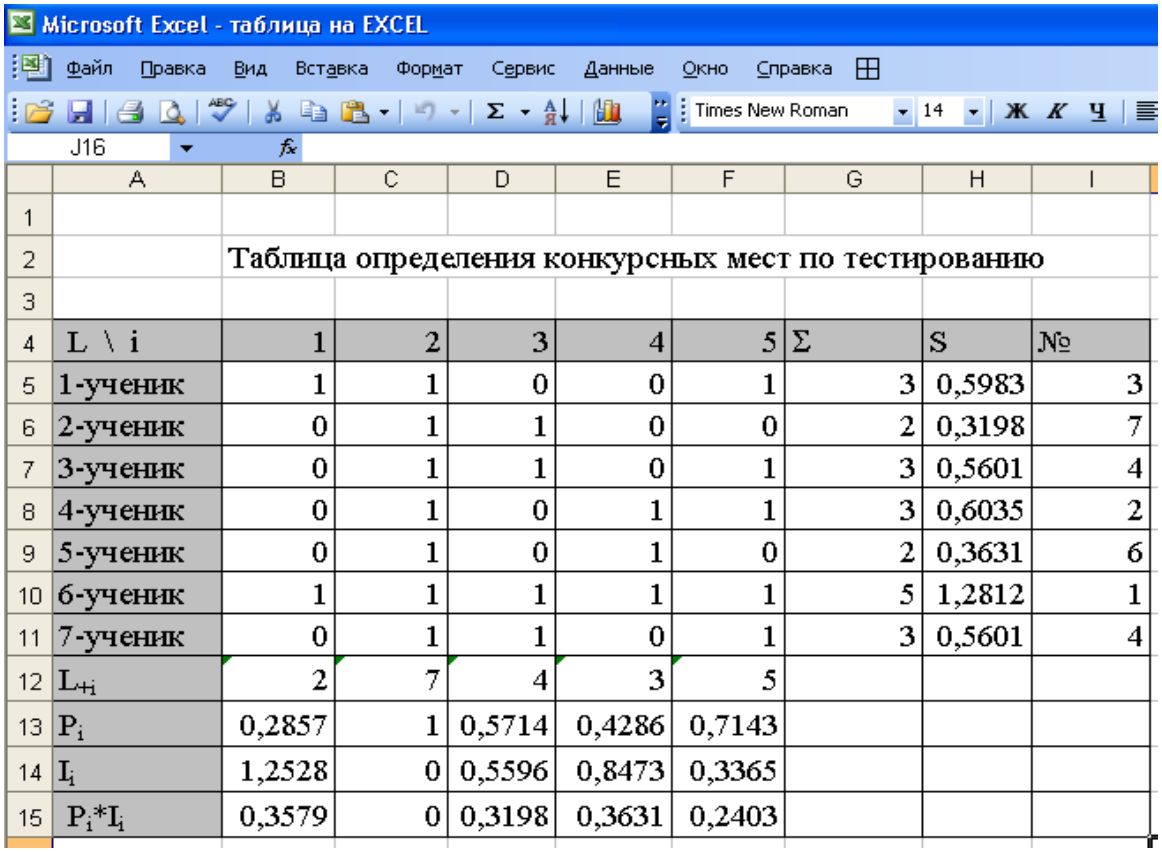


Таблица определения конкурсных мест по тестированию

L \ i	1	2	3	4	5	Σ	S	№
1-ученик	1	1	0	0	1	3	0,5983	3
2-ученик	0	1	1	0	0	2	0,3198	7
3-ученик	0	1	1	0	1	3	0,5601	4
4-ученик	0	1	0	1	1	3	0,6035	2
5-ученик	0	1	0	1	0	2	0,3631	6
6-ученик	1	1	1	1	1	5	1,2812	1
7-ученик	0	1	1	0	1	3	0,5601	4
L_{+i}	2	7	4	3	5			
P_i	0,2857	1	0,5714	0,4286	0,7143			
I_i	1,2528	0	0,5596	0,8473	0,3365			
P_i*I_i	0,3579	0	0,3198	0,3631	0,2403			

Рис. 3- Результаты тестирования обучаемых

Обработку данных тестирования целесообразно проводить с помощью приложения Microsoft Excel.

Пример 3. Дети играют во время обеда в распознавание предмета, который находится на кухне: один из игроков -ведущий - загадывает предмет, а другие игроки должны найти его, задавая различные вопросы. При этом ведущий может отвечать только «да» или «нет». Выигрывает тот ведущий, если он получит от игроков больше вопросов, чем другие ведущие. Ведущий загадал предмет «хлеб». Если больше половины предметов составляют продукты, какой из следующих вопросов больше уменьшит *неопределенность* в процессе распознания задуманного предмета: «Этот предмет металлический?», «Этот предмет съедобный?», «Этот предмет деревянный?»

Ответ: «Этот предмет съедобный?»

Пример 4. Задуман один из 16 предметов, находящихся в классе. Каким образом, следует задавать вопросы, чтобы их было как можно больше для распознания предмета?

Ответ: Нужно задавать вопросы следующим образом: «Это первый предмет?», «Это второй предмет?» и т.д.

Пример 5. Введем следующее правило для угадывания задуманного предмета: разделить совокупность предметов на две равные части и спросить, в какой части находится задуманный предмет. Сколько вопросов буде задано при таком правиле угадывания предмета для его однозначного определения? Число предметов равно 16.

Решение. Делим всю совокупность предметов на две равные части и спросить (например, один размещаем справа, а другие - слева) и спрашиваем: «В какой части – правой или левой - находится задуманный предмет?» При любом ответе («в правой» или «в левой») мы будем знать 8 предметов, среди которых находится задуманный предмет». Теперь разбиваем уже эти 8 предметов на две равные части по 4 предмета и задаем второй вопрос: «В какой части – правой или левой - находится задуманный предмет?» После получения ответа мы будем знать, среди каких четырех предметов находится задуманный. Нетрудно догадаться, что, если продолжать действовать аналогичным образом, однозначно определить задуманный предмет можно после четвертого вопроса.

Пример 6. Имеется 27 монет одинакового достоинства, среди которых одна монета - фальшивая. Фальшивая монета более

тяжелая, чем остальные, но не отличается от них по внешнему виду. Пользуясь чашечными весами без гирь нужно обнаружить фальшивую монету с помощью не более трех взвешиваний.

Решение. Если мы положим часть монет на одну чашку весов, а столько же – на другую, то результат данного взвешивания даст нам информацию не только о тех монетах, которые находятся на чашках весов, но и об оставшихся монетах. Действительно, допустим, если весы при этом взвешивании будут находиться в равновесии, тогда фальшивую монету можно искать среди не взвешенных монет. Логика подсказывает, что монеты выгоднее всего следует разложить на три равные кучки по 9 монет. Одну кучку нужно положить на левую чашку весов, вторую – на правую, а третью пока не трогать. При взвешивании возможен один из трех результатов: 1) перетянет левая чашка; 2) перетянет правая чашка; 3) весы окажутся в равновесии. В первом случае фальшивая монета лежит на левой чашке весов, во втором – на второй, а в третьем - в отложенной кучке. Таким образом, после первого взвешивания мы выделим кучку из 9 монет, среди которых находится фальшивая. Теперь разделим эти 9 монет на три части по 3 монеты в каждой. Повторим взвешивание этих частей монет по аналогии с предыдущим. После второго взвешивания мы обнаруживаем ту часть монет, в которой лежит фальшивая монета. Понятно, что третье взвешивание позволит нам обнаружить фальшивую монету.

Пример 7 Дано $3^к$ монет, среди которых одна фальшивая монета (более тяжелая по сравнению с другими монетами). Нужно определить число взвешиваний, позволяющих однозначно обнаружить фальшивую монету.

Решение. В предыдущем задании, когда нужно было найти фальшивую монету среди $27 = 3^3$ монет, оптимальное число взвешиваний было равно 3. На основе индуктивного подхода можно легко определить, что нам следует провести *к* взвешиваний

Пример 8. Найдите *вероятность* вытаскивания туза с колоды карты.

Ответ: Вероятность вытащить туза равна 4/36=1/9.

Пример 9. Выполните опыт, бросая кубик и спичечную коробку. Наблюдайте выпадение помеченной грани кубика и выпадение помеченной грани коробки. Какое из этих событий является *равновероятным*, а какое – *неравновероятным.*

Ответ: Выпадение помеченной грани кубика – равновероятное событие, выпадение помеченной грани коробки. – неравновероятное событие

Пример 10. Используя *формулу Хартли* найдите, сколько вопросов нужно задать при угадывании одного предмета из 16.

Решение. Количество информации в сообщении при вероятностном подходе в соответствии с формулой Хартли вычисляется по формуле $I = \log_2 N = \log_2 16 = 4$ *бит* (N – число равновероятных исходов события, о котором идет речь в сообщении).Таким образом, для угадывания задуманного предмета из 16 нужно задать 4 вопроса.

Пример 11. Сколько битов информации необходимо получить при угадывании предмета из совокупности в N предметов.

Решение. Из рассмотренных выше заданий мы можем написать $2^{к-1} < N \leq 2^{к}$, где к - число информации в одном сообщении. Прологарифмуем записанное выражение и получим $(к-1) < \log_2 N \leq к$.

Пример 12. Найдите количество вопросов, которые необходимо задать для угадывания одного предмета из 100.

Решение. Поскольку $2^6 < 100 \leq 2^7$, необходимо задать 7 вопросов. Можно посчитать и по другому: $I = \log_2 100 = 6{,}644$. Поскольку, мы не можем задать нецелое число вопросов, то нужно задать 7 вопросов.

Пример 13. Из шести монет две – фальшивые, причем их массы одинаковые. При каком минимальном числе взвешиваний можно обнаружить фальшивые монеты.

Решение. Несложно подсчитать, что комбинацию «две монеты из шести» можно выбрать 15 способами. Значит, чтобы выбрать один из этих способов, нужно иметь $\log_2 15 = 3{,}907$ бита информации. Три взвешивания дают $3\log_2 3 = 4{,}752$ бита информации, а два взвешивания $2\log_2 3 = 3{,}168$ бита информации. По этому двух взвешиваний может не хватить, а трех взвешиваний оказывается достаточно для нахождения этих двух фальшивых монет.

Пример 14. Опишите процедуру взвешивания, о котором говорится в предыдущей задаче.

Решение. Разделим монеты на две равные части. При взвешивании на весах возможны два случая: 1) Перетянет, допустим, правая чашка, значит, две фальшивые монеты находятся

в правой чашке; 2) весы окажутся в равновесии, следовательно, фальшивые монеты находятся в разных чашках.

В случае 1) проводим еще одно взвешивание – взвешиваем на весах две монеты, оставляя одну в стороне. Если весы находятся в равновесии, то эти две монеты и есть фальшивые. Если одна чаша весов перетягивает, то настоящая монета – на другой чаше. То есть достаточно выполнить еще одно взвешивание, значить всего – два взвешивания.

В случае 2) для каждой из двух групп монет надо провести одно взвешивание, т.е. всего – 1+2=3.

Пример 15. Некоторая система может находиться в четырех состояниях: в первом – с вероятностью 0,1, во втором и третьем – с вероятностью 0,25, в четвертом – с вероятностью 0,4. Чему равно среднее значение информации (или неопределенность выбора) в системе? Чему оно равно, если система может находиться только в состоянии номер 2. Приведите пример для этого состояния.

Решение. Используем *формулу Шенона* . $I_{ср} = - \sum_i P_i \log_2 (P_i)$

$I= -0,1\cdot \log_2 0,1 – 2 \cdot 0,25\cdot \log_2 0,25 – 0,5\cdot \log_2 0,4 = 5,644$ бита

Если система может находиться только в состоянии номер 2, то никакого выбора нет, $p_2=1$, $p_1 = p_2 = p_3 = 0$, а значит, I=0, т.е. количество информации о состоянии системы нулевое, так как это состояние точно известно.

Пример 16. Из условия предыдущей задачи можно убедиться, что сумма вероятностей всех состояний (событий) предложенной системы равна единице, т.е. 0,1+0,25+0,25+0,4=1; $(P_1 + P_2 + P_3 + P_4 = 1)$. Дать пояснение этому соотношению.

Ответ. Сумма всех вероятностей должна быть равна единице, так как какое-то событие из совокупности неравновероятных событий всегда происходит.

Пример 17. В ящике лежат три белых шара и один черный, наудачу выбирается один из шаров. Найти среднее значение информации о реализации *события*: появления белого или черного шара.

Решение. Обозначим элементарное событие, заключающееся в вытаскивании белого шара, через A_1, а выбор черного шара - через A_2. Вероятность этих событий равна $P_1= 3/4$, $P_2= 1/4$. По формуле Шеннона имеем

$$I = -3/4 \log_2(3/4) + (- 3/4 \log_2(1/4)) == 2-3/4\log_2 3= 0,82$$

Средняя информация оказалась меньше 1, это говорит о том, что при испытании выбора шаров чаще появляется белый шар, а значит неопределенность в таком испытании меньше одного бита. Если в ящике все шары были белыми, то вероятность появление белого шара P=1, тогда I = 0. При таком испытании нет определенности, всегда выбирается белый шар. В этом случае вероятность появления черного шара P=0, соответственно информация его появления $I = \infty$.

Пример 18. ДНК человека можно представить себе как некоторое слово четырех буквенного алфавит X= {A,B,C,Д}, где каждой буквой помечается звено цепи ДНК *(нуклеоид).* Каждое такое звено подразумевает состояние равновероятного события, т.е. нуклеоид может находиться в одном из этих состояний. Определите, сколько битов информации содержит примерно $1{,}5 \cdot 10^{23}$ нуклеоидов.

Решение. На один нуклеоид приходится $\log_2 4 = 2$ бита информации. Следовательно структура ДНК в организме позволяет хранить $3 \cdot 10^{23}$ бита информации.

Задачи:

1. В коробке лежат 64 цветных карандаша. Сообщение о том, что достали белый карандаш, несет 4 бита информации. Сколько белых карандашей было в корзине?

2. В ящике лежат перчатки (белые и черные). Среди них – 2 пары черных. Сообщение о том, что из ящика достали пару черных перчаток, несет 4 бита информации. Сколько всего пар перчаток было в ящике?

3. Известно, что в ящике лежат 20 шаров. Из них 10 – черных, 5 – белых,4 – желтых и 1 – красный. Какое количество информации несут сообщения о том, что из ящика случайным образом достали черный шар, белый шар, желтыйшар, красный шар?

4. За четверть ученик получил 100 оценок. Сообщение о том, что он получил четверку, несет 2 бита информации. Сколько четверок ученик получил зачетверть?

5. В корзине лежат белые и черные шары. Среди них 18 черных шаров. Сообщение о том, что из корзины достали белый шар, несет 2 бита информации. Сколько всего в корзине шаров?

6. На остановке останавливаются автобусы с разными номерами. Сообщение о том, что к остановке подошел автобус с номером №1 несет 4 бита информации. Вероятность появления на остановке автобуса с номером №2 в два раза меньше, чем вероятность появления автобуса с номером №1. Сколько информации несет сообщение о появлении на остановке автобуса с номером №2?

Кибернетический (алфавитный) подход к измерению информации. *Алфавитный подход* к измерению информации позволяет определить количество информации, заключенной в тексте. Алфавитный подход является *объективным,* т.е. он не зависит от субъекта (человека), воспринимающего текст.

Множество символов, используемых при записи текста, называется *алфавитом.* Полное количество символов в алфавите называется **мощностью** (размером) алфавита. Если допустить, что все символы алфавита встречаются в тексте с одинаковой частотой (равновероятно), то количество информации, которое несет каждый символ, вычисляется по формуле:

$i = \log_2 N$, где N – мощность алфавита.

Следовательно, в 2-х символьном алфавите каждый символ «весит» 1 бит ($\log_2 2 = 1$); в 4-х символьном алфавите каждый символ несет 2 бита информации ($\log_2 4 = 2$); в 8-ми символьном – 3 бита ($\log_2 8 = 3$) и т. д.

Один символ из алфавита мощностью 256 (2^8) несет в тексте 8 бит информации. Такое количество информации называется *байт.* Алфавит из 256 символов используется для представления текстов в компьютере.

1 байт = 8 бит. Если весь текст состоит из К символов, то при алфавитном подходе размер содержащейся в нем информации равен:

$$I = K * i,$$

где i – информационный вес одного символа в используемом алфавите. Для измерения информации используются и более крупные единицы: *1 Кбайт, (килобайт)* = 2^{10} *байт* = 1024 *байта*

1 Мбайт (мегабайт) = 2^{10} *Кбайт* = 1024 *Кбайта*

1 Гбайт (гигабайт) = 2^{10} *Мбайт* =1024 *Мбайта*

Пример. Книга, набранная с помощью компьютера, содержит 150 страниц; на каждой странице – 40 строк, в каждой строке – 60 символов. Каков объем информации в книге?

Решение. Мощность компьютерного алфавита равна 256. Один символ несет 1 байт информации. Значит, страница содержит 40 * 60 = 2400 байт информации. Объем всей информации в книге (в разных единицах):

2400 * 150 = 360 000 байт.

360000 / 1024 = 351,5625 Кбайт.

351,5625 / 1024 = 0,34332275 Мбайт

Вероятностный метод применим и для алфавитного подхода к измерению информации, заключенной в тексте. Известно, что разные символы (буквы алфавита, знаки препинания и др.) встречаются в тексте с разной частотой и, следовательно, имеют разную вероятность. Значит, измерять информационный вес каждого символа в тексте так, как это делалось раньше (в предположении равновероятности), нельзя.

Задачи

1. Алфавит некоторого племени состоит из 8 букв. Какое количество информации несет одна буква этого алфавита?
2. Сообщение, записанное буквами из 64-х символьного алфавита, содержит 20 символов. Какой объем информации оно несет?
3. Племя Мульти имеет 32-х символьный алфавит. Племя Пульти использует 64-х символьный алфавит. Вожди племен обменялись письмами. Письмо племени Мульти содержало 80 символов, а письмо племени Пульти – 70 символов. Сравните объемы информации, содержащейся в письмах.
4. Информационное сообщение объемом 1,5 Кбайта содержит 3072 символа. Сколько символов содержит алфавит, при помощи которого было записано это сообщение?
5. Объем сообщения, содержащего 2048 символов, составил 1/512 часть Мбайта. Каков размер алфавита, с помощью которого записано сообщение?
6. Сколько символов содержит сообщение, записанное с помощью 16-ти символьного алфавита, если объем его составил 1/16 часть Мбайта?
7. Сколько килобайтов составляет сообщение, содержащее 12288 битов?
8. Сколько килобайтов составит сообщение из 384 символов 16-ти символьного алфавита?
9. Для записи текста использовался 256-символьный алфавит. Каждая страница содержит 30 строк по 70 символов в строке. Какой объем информациисодержат 5 страниц текста?
10. Сообщение занимает 3 страницы по 25 строк. В каждой строке записано по 60 символов. Сколько символов в использованном алфавите, если все сообщение содержит 1125 байтов?
11. Для записи сообщения использовался 64-х символьный алфавит. Каждая страница содержит 30 строк. Все сообщение содержит 8775 байтов информации и занимает 6 страниц. Сколько символов в строке?
12. Сообщение занимает 2 страницы и содержит 1/16 Кбайта информации. На каждой станице записано 256 символов. Какова мощность использованного алфавита?

13. Сообщение, записанное буквами из 256-и символьного алфавита, содержит 60 символов. Какой объем информации оно несет?

14. Для записи текста используется 256-символьный алфавит и его объем составил 1/512 часть Мбайта. Сколько символов содержится в сообщении?

Вопросы для самопроверки

1. Дайте определения понятий «информация», «данные», «знания» – как базовых понятий в информатике. Раскройте их взаимосвязь. Приведите примеры.
2. Дайте классификацию информации по различным признакам (по способу представления, по способу восприятия, по массовому значению). Приведите примеры.
3. Приведите примеры информационных процессов в природе и технике в соответствии с универсальной схемой передачи информации.
4. Какие формы представления информации существуют? Раскройте их основные характеристики.
5. Перечислите атрибутивные свойства информации, дайте их краткую характеристику.
6. Перечислите прагматические свойства информации, дайте их краткую характеристику.
7. Перечислите динамические свойства информации, дайте их краткую характеристику.
8. В чем состоит процесс дискретизации информации и в каких случаях он используется? Приведите примеры.
9. Раскройте сущность понятия «количество информации».
10. Назовите существующие единицы измерения информации и соотношения между ними.
11. Раскройте сущность различных подходов к измерению количества информации (Мера Хартли, Мера Шеннона, закон аддитивности), приведите применяемые формулы.
12. Раскройте сущность кибернетического (алфавитного) подхода к измерению количества информации, приведите применяемые формулы.

II. КОДИРОВАНИЕ И ДЕКОДИРОВАНИЕ ИНФОРМАЦИИ

Область действия, предмет и задачи теории кодирования. Понятия код, кодирование и декодирование информации. Универсальная схема передачи информации. Понятие системы счисления, их виды. Системы счисления, применяемые в ЭВМ. Правила перевода чисел из одной системы счисления в другую. Арифметические действия в двоичной системе счисления.

Принципы представления целых и вещественных чисел в памяти компьютера. Кодирование текстовой информации. Принципы формирования графической информации в компьютере. Принципы кодирования звука.

2.1 Теоретические вопросы кодирования и декодирования информации

Область действия, предмет и задачи теории кодирования

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. С обсуждения этих вопросов и начнем освоение теории кодирования.

Абстрактный алфавит. Понятия код, кодирование, декодирование. Схема передачи сообщения в случае перекодировки. Вспомним схему передачи информации (см. рис.4). Так как информация передается в виде сообщений, то эти сообщения должны записываться с помощью некоторого набора

знаков.

Для дискретных сообщений этот набор знаков должен быть ограничен, и они должны отличаться друг от друга.

Полный набор символов, в котором определен их порядок называется *алфавитом.* Например: русский алфавит, алфавит Морзе, алфавит Бейсика, алфавит русифицированной клавиатуры IBM PC и т.п.

В канале связи сообщение, составленное из символов одного алфавита, может преобразовываться в сообщение из символов другого алфавита.

Правило, описывающее однозначное соответствие символов алфавитов при таком преобразовании, называют *кодом,* саму процедуру – *перекодировкой.* Преобразование сообщения может осуществляться в момент поступления от Источника в Канал Связи (*кодирование*) и в момент приема получателем (*декодирование*). Устройства, обеспечивающие эти операции, называются *кодировщиком и декодировщиком* соответственно.

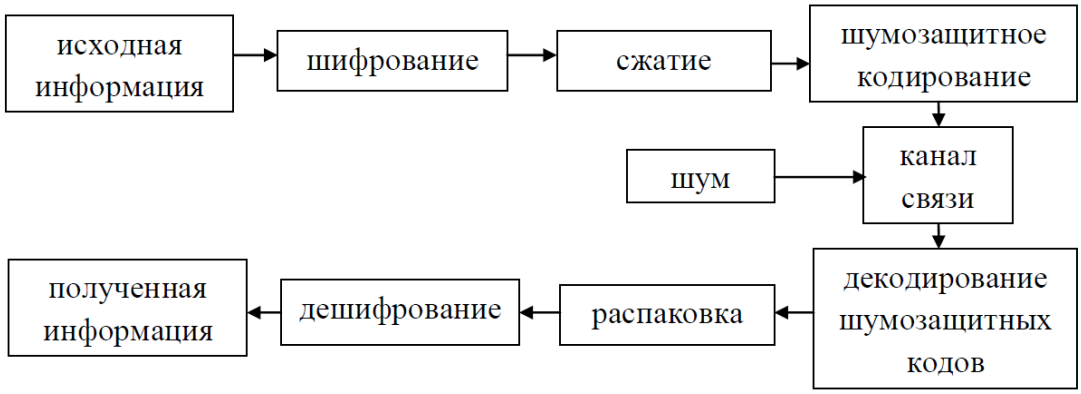


Рис.4 - Универсальная схема передачи информации в случае кодировки

При передаче сообщений по каналам связи могут возникать помехи, приводящие к искажению принимаемых знаков (например, треск в телефоннойтрубке и т.п.) – *шумы.*

Передача сообщений при наличии помех, особенно с внедрением компьютерных телекоммуникаций – это серьезная теоретическая и практическая задача. При работе с кодированной информацией можно выделить следующие основные проблемы:

- установление самого факта того, что произошло искажение;

- выяснение, в каком конкретно месте это произошло;
- исправление ошибки, хотя бы с некоторой степенью достоверности.Рассмотрим выделенные выше понятия подробнее.

Кодирование информации – процесс преобразования сигнала из формы, удобной для непосредственного использования информации, в форму, удобную для передачи, хранения или автоматической переработки (Цифровое кодирование, аналоговое кодирование, таблично-символьное кодирование, числовое кодирование). Процесс преобразования сообщения в комбинацию символов в соответствии с кодом называется *кодированием*, процесс восстановления сообщения из комбинации символов называется *декодированием.*

Для представления дискретной информации используется некоторый алфавит. Однако однозначное соответствие между информацией и алфавитом отсутствует. Другими словами, одна и та же информация может быть представлена посредством различных алфавитов. В связи с такой возможностью возникает проблема перехода от одного алфавита к другому, причём, такое преобразование не должно приводить к потере информации.

Алфавит, с помощью которого представляется информация до преобразования, называется первичным; алфавит конечного представления – вторичным.

Код – (1) правило, описывающее соответствие знаков или их сочетаний одного алфавита знакам или их сочетаниям другого алфавита; – (2) знаки вторичного алфавита, используемые для представления знаков или их сочетаний первичного алфавита.

Код – совокупность знаков (символов) и система определённых правил, при помощи которой информация может быть представлена (закодирована) в виде набора из таких символов для передачи, обработки и хранения. Конечная последовательность кодовых знаков называется словом. Наиболее часто для кодирования информации используют буквы, цифры, числа, знаки и их комбинации.

Код – набор символов, которому приписан некоторый смысл. Код является знаковой системой, которая содержит конечное число символов: буквы алфавита, цифры, знаки препинания, знаки препинания, знаки математических операций и т.д.

Кодирование – операция отожествления символов или групп символов одного кода и символов другого кода.

Кодирование информации – процесс формирования определенного представления информации.

Кодирование информации – процесс преобразования сигналов или знаков одного алфавита в знаки или сигналы другого.

Декодирование – операция, обратная кодированию, т.е. восстановление информации (восстановление в первичном алфавите).

Шифрование – разновидность кодирования.

Операции кодирования и декодирования называются обратимыми, если их последовательное применение обеспечивает возврат к исходной информации без каких-либо её 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.

Таким образом, кодирование предшествует передаче и хранению информации. При этом 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.

Любой код должен обеспечивать однозначное чтение сообщения (надежность), так и, желательно, быть экономным (использовать в среднем поменьше символов на сообщение).

Компьютер может обрабатывать только информацию, представленную в числовой форме. Вся другая информация (звуки, изображения, показания приборов и т.д.) для обработки на компьютере должна быть преобразована в числовую форму. С помощью компьютерных программ можно преобразовывать полученную информацию, в том числе - текстовую. При вводе в

компьютер каждая буква кодируется определенным числом, а при выводе на внешние устройства (экран или печать) для восприятия человеком по этим числам строятся изображения букв. Соответствие между набором букв и числами называется кодировкой символов. Как правило, все числа в компьютере представляются с помощью нулей и единиц, т.е. словами, компьютеры работают в двоичной системе счисления, поскольку при этом устройства для их обработки получаются значительно более простыми.

Кодер – программист, специализирующийся на кодировании - написании исходного кода по заданным спецификациям.

Кодер – одна из двух компонент кодека (пары кодер – декодер).

Декодер – некоторое звено, которое преобразует информацию из внешнего вида в вид, применяемый внутри узла. В программном обеспечении: модуль программы или самостоятельное приложение, которое преобразует файл или информационный поток из внешнего вида в вид, который поддерживает другое программное обеспечение.

Кодирование информации распадается на этапы:

1) Определение объёма информации, подлежащей кодированию.
2) Классификация и систематизация информации.
3) Выбор системы кодирования и разработка кодовых обозначений.
4) Непосредственное кодирование.

2.2 Использование систем счисления в теории кодировании

Понятие системы счисления. Виды систем счисления. Система счисления — это способ представления чисел и соответствующие ему правила действия над числами. Разнообразные системы счисления, которые существовали раньше и которые используются в наше время, можно разделить на непозиционные и позиционные. Знаки, используемые при записи чисел, называются цифрами.

В непозиционных системах счисления от положения цифры в записи числа не зависит величина, которую она обозначает.

Примером непозиционной системы счисления является римская система (римские цифры). В римской системе в качестве цифр используются латинские буквы:

I	**V**	**X**	**L**	**C**	**D**	**M**
1	**5**	**10**	**50**	**100**	**500**	**1000**

Пример 1. Число CCXXXII складывается из двух сотен, трех десятков и двух единиц и равно двумстам тридцати двум.

В римских числах цифры записываются слева направо в порядке убывания. В таком случае их значения складываются. Если же слева записана меньшая цифра, а справа - большая, то их значения вычитаются.

Пример 2.

VI =5+1=6, а IV=5-1=4.

Пример 3.

MCMXCVIII = 1000 + (-100 + 1000) + (-10 + 100) + 5+1+1+1= 1998.

В позиционных системах счисления величина, обозначаемая цифрой в записи числа, зависит от ее позиции. Количество используемых цифр называется *основанием* позиционной системы счисления.

Система счисления, применяемая в современной математике, является *позиционной десятичной системой*. Ее основание равно десяти, т. к. запись любых чисел производится с помощью десяти цифр: **0, 1, 2, 3, 4, 5, 6, 7, 8, 9.**

Позиционный характер этой системы легко понять на примере любого многозначного числа. Например, в числе 333 первая тройка означает три сотни, вторая — три десятка, третья — три единицы.

Для записи чисел в позиционной системе с основанием n нужно иметь *алфавит* из n цифр. Обычно для этого при $n < 10$ используют n первых арабских цифр, а при $n > 10$ к десяти арабским цифрам добавляют буквы. Алфавиты некоторых систем счисления приведены в табл. 1.

Если требуется указать основание системы, к которой относится число, то оно приписывается нижним индексом к этому числу. Например: 101101_2, 3671_8,$3B8F_{16}$.

Таблица 1 - Алфавиты некоторых систем счисления

Основание	**Название**	**Алфавит**
n=2	Двоичная	01
n=3	Троичная	012
n =8	Восьмеричная	01234567
n=16	Шестнадцатеричная	0123456789ABCDEF

В системе счисления с основанием q (q-ичная система счисления) единицами разрядов служат последовательные степени числа q. q единиц какого-либо разряда образуют единицу следующего разряда. Для записи числа в q-ичной системе счисления требуется q различных знаков (цифр), изображающих числа 0, 1, ..., q-1. Запись числа q в q-ичной системе счисления имеет вид 10.

Развернутой формой записи числа называется запись в виде: $A_q = \pm(a_{n-1} q^{n-1} + a_{n-2} q^{n-2} + \ldots + a_0 q^0 + a_{-1} q^{-1} + a_{-2} q^{-2} + \ldots + a_{-m} q^{-m})$

Здесь A_q – само число, q – основание системы счисления, a_i – цифры дан- ной системы счисления, n – число разрядов целой части числа, m – число раз- рядов дробной части числа.

Пример 1. Получить развернутую форму десятичного числа 26,387.

Решение: $26{,}387_{10} = 2 * 10^1 + 6 * 10^\circ + 3 * 10^{-1} + 8 * 10^{-2} + 7 * 10^{-3}$

Пример 2. Получить развернутую форму числа $101{,}11_2$

Решение: $101{,}11_2 = 1 * 10^{10} + 0 * 10^1 + 1 * 10^0 + 1 * 10^{-1} + 1\ 8\ 10^{-10}$.

Обратите внимание, что в любой системе счисления ее основание записывается как 10.

Перевод чисел из одной системы счисления в другую. Перевод чисел в десятичную систему счисления. Если все слагаемые в развернутой форме недесятичного числа представить в десятичной системе и вычислить полученное выражение по правилам десятичной арифметики, то получится число в десятичной системе, равное данному. По этому принципу производится перевод из недесятичной системы в десятичную.

Пример 1. Числа $15FC_{16}$ и $101{,}11_2$ перевести в десятичную систему.

Решение:

$15FC_{16} = 1* 16^3 + 5 * 16^2 + 15 * 16^1 + 12 = = 4096 + 1280 + 240 + 12 = 5628_{10}$.

$101{,}11_2 = 1 * 2^2 + 0 * 2^1 + 1 * 2^0 + 1 * 2^{-1} + 1 * 2^{-2} = = 4 + 1 + 1/2 + 1/4 = 5 + 0{,}5 + 0{,}25 = 5{,}75_{10}$

Перевод десятичных чисел в другие системы счисления Перевод целых чисел.

1) Основание новой системы счисления выразить в десятичной системе счисления и все последующие действия производить в десятичной системе счисления;

2) последовательно выполнять деление данного числа и получаемых неполных частных на основание новой системы счисления до тех пор, пока не получим неполное частное, меньшее делителя;

3) полученные остатки, являющиеся цифрами числа в новой системе счисления, привести в соответствие с алфавитом новой системы счисления;

4) составить число в новой системе счисления, записывая его, начиная с последнего частного.

Пример 2. Перевести число 37_{10} в двоичную систему.

Решение:

37	**2**				
36	**18**	**2**			
1	**18**	**9**	**2**		
	0	**8**	**4**	**2**	
		1	**4**	**2**	**2**
			0	**2**	**1**
				0	

Таким образом, $37_{10} = 100101_2$

Пример 3. Перевести десятичное число 315 в восьмеричную и в шестнадцатеричную системы:

Решение:

315	**8**	
24	**39**	**8**
75	**32**	**4**
72	**7**	
3		

315	**16**	
16	**19**	**16**
155	**16**	**1**
144	**3**	
11		

Отсюда следует: $315_{10}=473_8=13B_{16}$ **(**Напомним, что $11_{10}=B_{16}$)

Перевод дробных чисел.

1) Основание новой системы счисления выразить в десятичной системе и все последующие действия производить в десятичной системе счисления;

2) последовательно умножать данное число и получаемые дробные части произведений на основание новой системы до тех пор, пока дробная часть произведения не станет равной нулю или не будет достигнута требуемая точность представления числа в новой системе счисления;

3) полученные целые части произведений, являющиеся цифрами числа в новой системе счисления, привести в соответствие с алфавитом новой системы счисления;

4) составить дробную часть числа в новой системе счисления, начиная с целой части первого произведения.

Пример 4. Перевести десятичную дробь 0,1875 в двоичную, восьмеричную и шестнадцатеричную системы.

Решение:

В двоичную

0	1	8	7	5
*				2
0	3	7	5	0
*				2
0	7	5	0	0
*				2
1	5	0	0	0
*				2
1	0	0	0	0

В восьмеричную

0	1	8	7	5
*				8
1	5	0	0	0
*				8
4	0	0	0	0

В шестнадцатеричную

0	1	8	7	5
*			1	6
1	1	2	5	0
1	8	7	5	
3	0	0	0	0

Здесь вертикальная черта отделяет целые части чисел от дробных частей. Отсюда: $0{,}1875_{10} = 0{,}0011_2 = 0{,}14_8 = 0{,}3_{16}$.

Перевод смешанных чисел, содержащих целую и дробную части, осуществляется в два этапа. Целая и дробная части исходного числа переводятся отдельно по соответствующим алгоритмам. В итоговой записи числа в новой системе счисления целая часть отделяется от дробной запятой (точкой).

Пример 5. Перевести десятичное число 315,1875 в восьмеричную и в шестнадцатеричную системы счисления.

Решение: Из рассмотренных выше примеров следует: $315{,}1875_{10} = 473{,}14_8 = 13B{,}3_{16}$.

Системы счисления, используемые в ЭВМ (с основанием 2^n)

Для того, чтобы *целое двоичное число* записать в системе счисления с основанием $q=2^n$ (4,8,16 и т.д.), нужно:

1) данное двоичное число разбить справа налево на группы по *n* цифр в каждой;
2) если в последней левой группе окажется меньше *n* разрядов, то ее надо дополнить слева нулями до нужного числа разрядов;
3) рассмотреть каждую группу как *n* -разрядное двоичное число и записать ее соответствующей цифрой в системе счисления с основанием $q=2^n$

Для того чтобы *дробное двоичное число* записать в системе счисления с основанием $q=2^n$ нужно:

1) данное двоичное число разбить слева направо на группы по *n* цифр в каждой;
2) если в последней правой группе окажется меньше *n* разрядов, то ее надо дополнить справа нулями до нужного числа разрядов;
3) рассмотреть каждую группу как *n*-разрядное двоичное число и записать ее соответствующей цифрой в системе счисления $q=2^n$

Для того чтобы *произвольное двоичное число* записать в системе счисления с основанием $q=2^n$, нужно:

1) данное двоичное число разбить слева и справа (целую и дробную части) на группы по *n* цифр в каждой;
2) если в последних правой и левой группах окажется меньше *n* разрядов, то их надо дополнить справа и слева нулями до нужного числа разрядов;
3) рассмотреть каждую группу как *n* - разрядное двоичное число и записать ее соответствующей цифрой в системе счисления с основанием $q=2^n$

Для того чтобы *произвольное число*, записанное в системе счисления *с основанием* $q=2^n$, перевести в двоичную систему счисления, нужно каждую цифру этого числа заменить ее *n* - разрядным эквивалентом в двоичной системе счисления.

Применительно к компьютерной информации часто используются системы с основанием 8 (восьмеричная) и 16 (шестнадцатеричная), поэтому необходимо пользоваться таблицами соответствия.

Таблица 2 - Двоично-шестнадцатеричная таблица

16	**2**	**16**	**2**
0	0000	8	1000
1	0001	9	1001
2	0010	A	1010
3	0011	B	1011
4	0100	C	1100
5	0101	D	1101
6	0110	E	1110
7	0111	F	1111

Таблица 3 - Двоично–восьмеричная таблица

8	**2**
0	000
1	001
2	010
3	011
4	100
5	101
6	110
7	111

*Пример 6***.** Перевести число $15FC_{16}$ в двоичную систему.

Решение: Для решения задачи воспользуемся двоично-шестнадцатеричной таблицей. В одном столбце таблицы помещены шестнадцатеричные цифры, напротив, в соседнем столбце — равные им двоичные числа. Причем все двоичные числа записаны в четырехзначном виде (там, где знаков меньше четырех, слева добавлены нули).

А теперь проделаем следующее: каждую цифру в шестнадцатеричном числе 15FC заменим на соответствующую ей в таблице четверку двоичных знаков. Иначе говоря, перекодируем число 15FC по таблице в двоичную форму. Получается: 0001 0101 1111 1100.

Если отбросить нули слева (в любой системе счисления они не влияют на значение целого числа), то получим искомое двоичное число. Таким образом: $15FC_{16} = 1010111111100_2$.

В справедливости этого равенства можно убедиться, производя тот же перевод через десятичную систему.

Пример 7: Перевести двоичное число 1101111010111101111 в шестнадцатеричную систему.

Решение: Разделим данное число на группы по четыре цифры, начиная справа. Если в крайней левой группе окажется меньше четырех цифр, то дополним ее нулями.

0011 0111 1010 1110 1111.

А теперь, глядя на двоично-шестнадцатеричную таблицу, заменим каждую двоичную группу на соответствующую шестнадцатеричную цифру.

3 7 A E F Следовательно: $1101111010111101111_2 = 37AEF_{16}$.

Пример 8: Перевести смешанное число $1011101{,}10111_2$ в шестнадцатеричную систему.

Решение: Перевод дробных чисел производится аналогично. Группы по четыре двоичных знака выделяются от запятой как влево так и вправо. Поэтому:

$$1011101{,}10111_2 \Rightarrow 0101\ 1101,\ 1011\ 1000 \Rightarrow 5D{,}B8_{16}.$$

Связь между двоичной и восьмеричной системами устанавливается аналогично. В этом случае используется двоично-восьмеричная таблица, приведенная ниже. Каждой восьмеричной цифре соответствует тройка двоичных цифр.

Пример 9: Перевести смешанное число $1011101{,}10111_2$ в восьмеричнуюсистему.

Решение: Группы по три двоичных знака выделяются от запятой как влево, так и вправо. Затем производится перекодировка по таблице:

$1011101{,}10111_2 \Rightarrow 001\ 011\ 101,\ 101\ 110 \Rightarrow 135{,}56_8$

Арифметика в позиционных системах счисления. Арифметические операции во всех позиционных системах счисления вы- полняются по одним и тем же правилам. В основе всех арифметических операций лежат таблицы соответствия для одноразрядных чисел.

Таблица 4 - Таблицы соответствия для двоичной системы счисления

Сложение	Вычитание	Умножение
0+0=0	0-0=0	0*0=0
0+1=1	(1)0-1=1	0*1=0
1+0=1	1-0=1	1*0=0
1+1=(1)0	1-1=0	1*1=1

Сложение. В его основе лежит таблица сложения одноразрядных двоичных чисел. Важно обратить внимание на то, что при сложении двух единиц происходит переполнение разряда и производится перенос в старший разряд. Переполнение наступает тогда, когда величина числа в нем становится равной или большей основания.

Пример:

1100111.001_2

+

1110.10_2

1110101.101_2

Вычитание. В его основе лежит таблица вычитания одноразрядных двоичных чисел. Важно обратить внимание на то, что при вычитании из меньшего числа (0) большего (1) производится заем из старшего разряда.

Пример:

1100111.001_2

-

1110.10_2

1011000.101_2

Умножение. В его основе лежит таблица умножения одноразрядных двоичных чисел. Важно обратить внимание на то, что умножение многоразрядных двоичных чисел происходит с последовательным умножением множимого на цифры множителя.

Пример:

1100111.001_2

*****

1110.10_2

0000000000
1100111001
0000000000
1100111001
1100111001
1100111001

10111010111.01010

Деление. Операция деления выполняется по алгоритму, подобному алгоритму деления в десятичной системе счисления (операции умножения и вычитания)

Практическая работа

Задание №1. Перевести данное число из десятичной системы счисления в двоичную, восьмеричную и шестнадцатеричную системы счисления (с проверкой).

Задание № 2. Перевести данное число в десятичную систему счисления.

2.3 Способы кодирования различных видов информации

Представление числовой информации Структура внутренней памяти. Основные структурные единицы памяти компьютера: бит, байт, машинное слово.

Бит. Все данные и программы, хранящиеся в памяти компьютера, имеют вид двоичного кода. Один символ из двух символьного алфавита несет 1 бит информации. *Ячейка памяти, хранящая один двоичный знак, называется «бит».* В одном бите памяти хранится один бит информации.

Битовая структура памяти определяет первое свойство памяти — *дискретность.*

Байт. *Восемь расположенных подряд битов памяти образуют байт.*

В одном байте памяти хранится один байт информации. Во внутренней памяти компьютера все байты пронумерованы. Нумерация начинается от нуля. Порядковый номер байта называется его *адресом.* В компьютере адреса обозначаются двоичным кодом. Используется также шестнадцатеричная форма обозначения адреса.

Пример 1. Компьютер имеет оперативную память 2 Кбайт. Указать адрес последнего байта оперативной памяти (десятичный, шестнадцатеричный, двоичный).

Решение. Объем оперативной памяти составляет 2048 байт. Десятичный адрес (номер) последнего байта равен 2047, так как нумерация байтов памяти начинается с нуля. $2047_{10} = 7FF_{16} = 0111\ 1111\ 1111_2$.

*Машинное слово***.** Наибольшую последовательность бит, которую процессор может обрабатывать как единое целое, называют машинным словом. *Длина машинного слова может быть разной — 8, 16, 32 бита и т.д. Адрес машинного слова в памяти компьютера равен адресу младшего байта, входящего в это слово.*

Занесение информации в память, а также извлечение ее из памяти производится по адресам. Это свойство памяти называется *адресуемостью.*

Пример 2. Объем оперативной памяти компьютера равен 1 Мбайту, а адрес последнего машинного слова - 1 048 574. Чему равен размер машинного слова?

Решение: 1Мбайт = 1024 Кбайта = 1 048 576 байт. Так как нумерация байтов начинается с нуля, значит адрес последнего байта будет равен 1 048 575. Таким образом, последнее машинное слово включает в себя 2 байта с номерами 1 048 574 и 1 048 575. *Ответ:* 2 байта.

Задачи:

№1. Оперативная память компьютера содержит 163840 машинных слов, что составляет 0,625 Мбайт. Сколько бит содержит каждое машинное слово?

№2. Объем оперативной памяти компьютера составляет 1/8 часть Мбайта. Сколько машинных слов составляют оперативную память, если одно машинноеслово содержит 64 бита?

№3. Вы работаете на компьютере с 2-х байтовым машинным словом. С каким шагом меняются адреса машинных слов?

№4. Вы работаете на компьютере с 4-х байтовым машинным словом. С каким шагом меняются адреса машинных слов?

№5. Компьютер имеет объем оперативной памяти 0,5 Кбайт. Адреса машинных слов меняются с шагом 4. Сколько машинных слов составляют оперативную память компьютера?

№6. Компьютер имеет объем оперативной памяти 0,5 Кбайт. Адреса машинных слов меняются с шагом 2. Сколько машинных слов составляют оперативную память компьютера?

№7. Компьютер имеет объем оперативной памяти 1 Кбайт. Адреса машинных слов меняются с шагом 2. Сколько машинных слов составляют оперативную память компьютера?

№8. Какой объем имеет оперативная память компьютера, если 3FF - шестнадцатеричный адрес последнего байта оперативной памяти?

Для представления чисел в памяти компьютера используются два формата: *формат с фиксированной точкой* и *формат с плавающей точкой.* В формате с фиксированной точкой представляются только целые числа, в формате с плавающей точкой — вещественные числа (целые и дробные).

Целые числа. Множество целых чисел, представимых в памяти ЭВМ, ограничено. Диапазон значений зависит от размера ячеек памяти, используемых для их хранения. *В k-разрядной ячейке может храниться 2^k различных значений целых чисел.*

Пример 1. Пусть для представления целых чисел в компьютере использу- ется 16-разрядная ячейка (2 байта). Определить, каков диапазон хранимых чисел, если:

а) используются только положительные числа; б) используются как положительные так и отрицательные числа в равном количестве.

Решение. Всего в 16-разрядной ячейке может храниться 2^{16} = 65536 различных значений. Следовательно:

а) диапазон значений от 0 до 65535 (от 0 до 2^k-l);

б) диапазон значений от -32768 до 32767 (от $–2^{k-1}$ до 2^{k-1}-l).

Чтобы получить *внутреннее представление целого положительного числа N,* хранящегося в k-разрядном машинном слове, необходимо:

1) перевести число N в двоичную систему счисления;

2) полученный результат дополнить слева незначащими нулями до k разрядов.

Пример 2. Получить внутреннее представление целого числа 1607 в 2-х байтовой ячейке.

Решение. $N = 1607_{10} = 11001000111_{2.}$ Внутреннее представление этого числа в ячейке будет следующим: 0000 0110 0100 0111. Шестнадцатеричная форма внутреннего представления числа получается заменой 4-х двоичных цифр одной шестнадцатеричной цифрой: 0647.

Для записи *внутреннего представления целого отрицательного числа (- N)* необходимо:

*1)*получить внутреннее представление положительного числа N;

*2)*получить обратный код этого числа заменой 0 на 1 и 1 на 0;

*3)*к полученному числу прибавить 1.

Данная форма представления целого отрицательного числа называется *дополнительным кодом.* Использование дополнительного кода позволяет заменить операцию вычитания на операцию сложения уменьшаемого числа с дополнительным кодом вычитаемого.

Пример 3. Получить внутреннее представление целого отрицательногочисла -1607.

Решение: 1) Внутреннее представление положительного числа:

0000 0110 0100 0111

2) обратный код: 1111 1001 1011 1000

3) результат прибавления 1: 1111 1001 1011 1001 — это внутреннее двоичное представление числа – 1607. Шестнадцатеричная форма: F9B9.

Двоичные разряды в ячейке памяти нумеруются от 0 до k справа налево. Старший, k-й разряд во внутреннем представлении любого положительного числа равен нулю, отрицательного числа – единице. Поэтому этот разряд называется знаковым разрядом.

Вещественные числа. Формат с плавающей точкой использует представление вещественного числа R в виде произведения мантиссы m на основание системы счисления n в некоторой целой степени p, которую называют порядком: $R = m \times n^p$. Представление числа в форме с плавающей точкой неоднозначно. Например, справедливы следующие равенства:

$25.324 = 2.5324 \times 10^1 = 0.0025324 \times 10^4 = 2532.4 \times 10^{-2}$ и т. п.

В ЭВМ используют *нормализованное представление числа в форме с плавающей точкой.* Мантисса в нормализованном представлении должна удовлетворять условию: $0.1_p \le m < 1_p$.

Иначе говоря, мантисса меньше единицы и первая значащая цифра – не ноль.

В памяти компьютера мантисса представляется как целое число, содержащее только значащие цифры (0 целых и запятая не хранятся). Следовательно, внутреннее представление вещественного числа сводится к представлению пары целых чисел: мантиссы и порядка.

В разных типах ЭВМ применяются различные варианты представления чисел в форме с плавающей точкой. Для примера рассмотрим внутреннее представление вещественного числа в 4-х байтовой ячейке памяти.

В ячейке должна содержаться следующая информация о числе: знак числа,порядок и значащие цифры мантиссы.

±маш. Порядок	МАН	ТИС	СА
1- й байт	2-й байт	3-й байт	4-й байт

В старшем бите 1-го байта хранится знак числа: 0 обозначает плюс, 1 – минус. Оставшиеся 7 бит первого байта содержат *машинный порядок.* В следующих трех байтах хранятся значащие цифры мантиссы (24 разряда).

В семи двоичных разрядах помещаются двоичные числа в диапазоне от 0000000 до 1111111. Значит, машинный порядок изменяется в диапазоне от 0 до 127 (в десятичной системе счисления). Всего 128 значений. Порядок, очевидно, может быть как положительным так и отрицательным. Разумно эти 128 значений разделить поровну между положительными и отрицательными значениями порядка: от -64 до 63.

Машинный порядок смещен относительно математического и имеет только положительные значения. Смещение выбирается так, чтобы минимальному математическому значению порядка соответствовал нуль.

Связь между машинным порядком (Мр) и математическим (р) в рассматриваемом случае выражается формулой:

$$Мр = р + 64.$$

Полученная формула записана в десятичной системе. В двоичной системе формула имеет вид: $Мр_2 = р_2 + 100\ 0000_2$.

Для записи *внутреннего представления вещественного числа* необходимо:

1) перевести модуль данного числа в двоичную систему

счисления с 24 значащими цифрами;

2) нормализовать двоичное число;

3) найти машинный порядок в двоичной системе счисления;

4) учитывая знак числа, выписать его представление в 4-х байтовом машинном слове.

Пример 4. Записать внутреннее представление числа 250,1875 в форме сплавающей точкой.

Решение:

1. Переведем данное число в двоичную систему счисления с 24 значащи- ми цифрами: $250{,}1875_{10} = 11111010, 0011000000000000_2$.

2. Запишем в форме нормализованного двоичного числа с плавающей точкой: $0{,}111110100011000000000000 \times 10^{1000}$. Здесь мантисса, основание системы счисления ($2_{10} = 10_2$) и порядок ($8_{10} = 1000_2$) записаны в двоичной системе.

3. Вычислим машинный порядок в двоичной системе счисления: $Mp_2 = 1000 + 100\ 0000 = 100\ 1000$.

4. Запишем представление числа в 4-х байтовой ячейке памяти с учетом знака числа:

0	1001000	11111010	00110000	00000000
31	24	23		0

Шестнадцатеричная форма: 48FA3000.

Пример 5. По шестнадцатеричной форме внутреннего представления числа в форме с плавающей точкой C9811000 восстановить само число.

Решение:

1. Перейдем к двоичному представлению числа в 4-х байтовой ячейке, заменив каждую шестнадцатеричную цифру 4-мя двоичными цифрами:

1100 1001 1000 0001 0001 0000 0000 0000

1	1001001	10000001	00010000	00000000
31	24	23		0

2. Заметим, что получен код отрицательного числа, поскольку в старшем разряде с номером 31 записана 1. Получим порядок числа: $p = 1001001_2 - 1000000_2 = 1001_2 = 9_{10}$.

3. Запишем в форме нормализованного двоичного числа с плавающейточкой с учетом знака числа:

$-0{,}10000001000100000000000 \times 2^{1001}$.

4. Число в двоичной системе счисления имеет вид:

$-100000010{,}001_2$

5. Переведем число в десятичную систему счисления:

$-100000010{,}001_2 = -(1 \times 2^8 + 1 \times 2^1 + 1 \times 2^{-3}) = -258{,}125_{10}$

Практическая работа «Целые числа в памяти компьютера» Задания (для всех вариантов):

1. Получить двоичную форму внутреннего представления целого числа в 2-х байтовой ячейке.
2. Получить шестнадцатеричную форму внутреннего представления целого числа в 2-х байтовой ячейке.
3. По шестнадцатеричной форме внутреннего представления целого числав 2-х байтовой ячейке восстановить само число.

№ Варианта	номера заданий		
	1	2	3
1	1450	-1450	F67D
2	1341	-1341	F7AA
3	1983	-1983	F6D7
4	1305	-1305	F700
5	1984	-1984	F7CB
6	1453	-1453	F967
7	1833	-1833	F83F
8	2331	-2331	F6E5
9	1985	-1985	F8D7
10	1689	-1689	FA53
11	2101	-2101	F840
12	2304	-2304	FAE7
13	2345	-2345	F841
14	2134	-2134	FAC3
15	2435	-2435	FA56

Практическая работа «Вещественные числа в памяти компьютера».

Задания (для всех вариантов)

1. Получить шестнадцатеричную форму внутреннего представления числа в формате с плавающей точкой в 4-х

байтовой ячейке.

2. По шестнадцатеричной форме внутреннего представления вещественного числа в 4-х байтовой ячейке восстановить само число.

№ варианта	Номера заданий	
	1	2
1	26.28125	C5DB0000
2	-29.625	45D14000
3	91.8125	C5ED0000
4	-27.375	47B7A000
5	139.375	C5D14000
6	-26.28125	488B6000
7	27.375	C7B7A000
8	-33.75	45DB0000
9	29.625	C88B6000
10	-139.375	45ED0000
11	333.75	C6870000
12	-333.75	46870000
13	224.25	C9A6E000
14	-91.8125	49A6E000
15	33.75	48E04000

Представление символьной информации. Информатика и ее приложения интернациональны. Это связано как с объективными потребностями в единых правилах и законах хранения, передачи и обработки информации, так и с тем, что в этой сфере заметен приоритет одной страны (США).

Так как для внутреннего представления информации в ЭВМ используется двоичная система счисления (0 и 1), то кодирование «внешних» символов основывается на сопоставлении каждому из них определенной группы двоичных знаков. При этом из-за технических соображений используют двоичные группы равной длины.

Попробуем подсчитать минимальную длину такой комбинации: латинский алфавит: 26 букв (стр.) + 26 букв (прописных) + 10 цифр + 10 знаков препинания + 10 разделителей (скобки и пр.) + знаки математических операций + спец. символы (типа #,$ и т.п.) $\approx$100

$$i = \log_2 N \Rightarrow N=2^i ; 2^6 <100<2^7$$

Но для кодирования хотя бы 2-х алфавитов и этого недостаточно. (Английский + русский + перечисленные выше спец. символы).

Минимальное значение i в этом случае должно быть 8, тогда 2^8 = 256 двоичных комбинаций. Т.к. 8 двоичных символов = 1 байту, то говорят о системах «байтового» кодирования.

Наиболее распространены 2 системы:

а) EBCDIC (Extended Binary Coded Decimal Interchange Code) – тяготеет к

«большим» машинам;

б) ASCII (American Standard Code for Information Interchange) чаще используется в мини- и микроЭВМ (включая ПК), создана в 1963 г.

Изначально, это была система семибитного кодирования, ограничивалась только английским алфавитом + «символы пишущей машинки» + «управляющие символы».

Вторая версия: 8-битная расширенная кодировка, в которой первые 128 символов совпадают с исходными, а остальные отданы под буквы некоторых европейских языков.

Для представления букв русского алфавита (кириллицы) в рамках ASCII было предложено несколько версий (КОИ-7, модифицированная альтернативная кодировка).

Но и 8-битная кодировка недостаточна для кодирования всех символов, которые хотелось бы иметь в расширенном алфавите.

Все препятствия могут быть сняты при переходе на 16-битную кодировку UNICODE, допускающую 65536 кодовых комбинаций.

*Таблица кодировки***:** таблица, в которой устанавливается соответствие между символами и их порядковыми номерами в компьютерном алфавите.

Все символы компьютерного алфавита пронумерованы от 0 до 255. Каждому номеру соответствует 8-разрядный двоичный код от 00000000 до 11111111. Этот код есть порядковый номер символа в двоичной системе счисления.

Для разных типов ЭВМ используются различные таблицы кодировки. С распространением персональных компьютеров типа IBM PC международным стандартом стала таблица кодировки под названием **ASCII** (American Standard Code for Information

Interchange) – Американский стандартный код для информационного обмена (См. Приложение 4). Стандартными в этой таблице являтся только первые 128 символов, т. е. символы с номерами от нуля (двоичный код 00000000) до 127 (01111111). Сюда входят буквы латинского алфавита, цифры, знаки препинания, скобки и некоторые другие символы. Остальные 128 кодов, начиная со 128 (двоичный код 10000000) и кончая 255 (11111111), используются для кодировки букв национальных алфавитов, символов псевдо- графики и научных символов (например, символы ≥, ≤, ±). В русских национальных кодировках в этой части таблицы размещаются символы русского алфавита.

Принцип последовательного кодирования алфавита: в кодовой таблице ASCII латинские буквы (прописные и строчные) располагаются в алфавитном порядке. Расположение цифр также упорядочено по возрастанию значений. Данное правило соблюдается и в других таблицах кодировки. Благодаря этому и в машинном представлении для символьной информации сохраняется понятие «алфавитный порядок».

Представление графической информации. *Компьютерная графика* — раздел информатики, предметом которого является работа на компьютере с графическими изображениями (рисунками, чертежами, фотографиями, видеокадрами и пр.).

Растровое представление. *Пиксель* — наименьший элемент изображения на экране (точка на экране). *Растр* — прямоугольная сетка пикселей на экране. *Разрешающая способность экрана* — размер сетки растра, задаваемого в виде произведения M × N, где M – число точек по горизонтали, N – число точекпо вертикали (число строк). *Видеоинформация* — информация об изображении, воспроизводимом на экране компьютера, хранящаяся в компьютерной памяти. *Видеопамять* — оперативная память, хранящая видеоинформацию во время ее воспроизведения в изображение на экране. *Графический файл* — файл, хранящий информацию о графическом изображении.

Число цветов, воспроизводимых на экране дисплея (К), и число бит, отводимых в видеопамяти под каждый пиксель (N), связаны формулой: $\mathbf{K = 2^N}$**.** Величину N называют *битовой глубиной.*

Страница — раздел видеопамяти, вмещающий информацию об одном образе экрана (одной «картинке» на экране). В видеопамяти могут размещаться одновременно несколько страниц.

Пример 1. На экране с разрешающей способностью 640 x 200 высвечиваются только двухцветные изображения. Какой минимальный объем видеопамяти необходим для хранения изображения?

Решение. Так как битовая глубина двухцветного изображения равна 1, а видеопамять, как минимум, должна вмещать одну страницу изображения, то объем видеопамяти равен $640 \times 200 \times 1 = 128000$ бит = 16000 байт.

Все многообразие красок на экране получается путем I смешивания трех базовых цветов: *красного, синего* и *зеленого.* Каждый пиксель на экране состоит из трех близко расположенных элементов, светящихся этими цветами. Цветные дисплеи, использующие такой принцип, называются RGB (Red, Green, Blue) мониторами.

Код цвета пикселя содержит информацию о доле каждого базового цвета. Если все три составляющие имеют одинаковую интенсивность (яркость), то из их сочетаний можно получить 8 различных цветов (2^3). Следующая таблица показывает кодировку 8-цветной палитры с помощью трехразрядного двоичного кода. В ней наличие базового цвета обозначено единицей, а отсутствие нулем.

*Пример 2***.** Из смешения каких цветов получается розовый цвет?

Решение: Глядя на таблицу 5, видим, что код розового цвета – 101. Это значит, что розовый цвет получается смешением красной и синей красок.

Таблица 5 - Двоичный код восьмицветной палитры

К (R)	**З (G)**	**С (B)**	**Цвет**
0	0	0	Черный
0	0	1	Синий
0	1	0	Зеленый
0	1	1	Голубой
1	0	0	Красный
1	0	1	Розовый

1	1	0	Коричневый
1	1	1	Белый

Векторное представление. При векторном подходе изображение рассматривается как совокупность простых элементов: прямых линий, дуг, окружностей, эллипсов, прямоугольников, закрасок и пр., которые называются *графическими примитивами***.** Графическая информация – это данные, однозначно определяющие все графические примитивы, составляющие рисунок.

Положение и форма графических примитивов задаются в *системе графических координат,* связанных с экраном. Обычно начало координат расположено в верхнем левом углу экрана. Сетка пикселей совпадает с координатной сеткой. Горизонтальная ось X направлена слева направо; вертикальная ось Y – сверху вниз.

Отрезок прямой линии однозначно определяется указанием координат его концов; окружность – координатами центра и радиусом; многоугольник - координатами его углов, закрашенная область – граничной линией и цветом закраски и пр.

Задачи:

1. Черно-белое (без градаций серого) растровое графическое изображение имеет размер 10*10 точек. Какой объем памяти займет это изображение?

2. В процессе преобразования растрового графического изображения количество цветов увеличилось с 16 до 256. Во сколько раз увеличился объем, занимаемый им в памяти?

3. Для хранения изображения размером 64 × 32 точек выделен 1 Кбайт памяти. Определите, какое максимальное число цветов допустимо использовать в этом случае.

4. 256-цветный рисунок содержит 120 байт информации. Из скольких точек он состоит?

Представление звуковой информации. Физическая природа звука – колебания в определенном диапазоне частот, передаваемые звуковой волной через воздух (или другую упругую среду). Процесс преобразования звуковых волн в двоичный код в памяти компьютера включает следующие этапы:

1. Звуковая волна **2.** Микрофон **3.** Переменный электрический ток **4.** Аудиоадаптер **5.** Двоичный код **6.** Память эвм

Следовательно, процесс воспроизведения звуковой информации, сохраненной в памяти ЭВМ следующий:

1. Память эвм **2.** Двоичный код **3.** Аудиоадаптер

4. Переменный электрический ток **5.** Динамик **6.** Звуковая волна

Аудиоадаптер (звуковая плата) – специальное устройство, подключаемое к компьютеру, предназначенное для преобразования электрических колебаний звуковой частоты в числовой двоичный код при вводе звука и для обратного преобразования (из числового кода в электрические колебания) при воспроизведении звука.

Качество компьютерного звука определяется характеристиками аудиоадаптера: частотой дискретизации и разрядностью.

Частота дискретизации — это количество измерений входного сигнала за секунду. Частота измеряется в герцах (Гц).

Разрядность регистра — число бит в регистре аудиоадаптера. Разрядность определяет точность измерения входного сигнала.

Звуковой файл — файл, хранящий звуковую информацию в числовой двоичной форме.

Пример: Определить размер (в байтах) цифрового аудиофайла, время звучания которого составляет 10 секунд при частоте дискретизации 22,05 кГц и разрешении 8 бит. Файл сжатию не подвержен.

Решение: Формула для расчета размера (в байтах) цифрового аудиофайла (монофоническое звучание):

Размер (байт) = (частота дискретизации в Гц) × (время записи в секундах) × (разрешение в битах) / 8. Таким образом, размер файла вычисляется так:

22050 × 10 × 8 / 8 = 220 500 байт.

Задачи:

1. Определить объем памяти для хранения цифрового аудиофайла, время звучания которого составляет две минуты при частоте дискретизации 44,1 кГци разрешении 16 бит.

2. Одна минута записи цифрового аудиофайла занимает на диске 1,3 Мб, разрядность звуковой платы – 8. С какой частотой дискретизации записан звук?

3. Объем свободной памяти на диске – 0,01 Гб, разрядность звуковой платы – 16. Какова длительность звучания цифрового аудиофайла, записанного с частотой дискретизации 44100 Гц?

4. Рассчитайте время звучания моноаудиофайла, если при 16-битном кодировании и частоте дискретизации 32 кГц его объем равен 700 Кбайт.

5. Оцените информационный объем моноаудиофайла (в килобайтах) длительностью звучания 1 мин. если «глубина» кодирования и частота дискретизации звукового сигнала равны соответственно: 16 бит и 8 кГц.

Вопросы для самопроверки

1. Что такое «код»? Приведите примеры.
2. Что такое «кодирование»? Приведите примеры.
3. Что такое «декодирование»? Приведите примеры.
4. Нарисуйте универсальную схему передачи информации в случае кодирования. Охарактеризуйте назначение используемых в схеме устройств.
5. Что такое система счисления?
6. Что такое алфавит системы счисления?
7. Какие системы счисления называют непозиционными? Приведите примеры.
8. Какие системы счисления называют позиционными? Приведите примеры.
9. Что такое основание системы счисления?
10. По какому правилу формируется алфавит позиционной системы счисления?
11. Запишите алфавиты двоичной, восьмеричной и шестнадцатеричной системсчисления.
12. Почему в вычислительной технике за основу взята система счисления поосновании 2?
13. Сформулируйте правило перевода чисел из любой позиционной системысчисления в десятичную.
14. Сформулируйте правило перевода целых чисел из десятичной системысчисления в любую позиционную.
15. Сформулируйте правило перевода дробных чисел из десятичной системысчисления в любую позиционную.
16. Сформулируйте правило перевода смешанных чисел из двоичной системысчисления в восьмеричную, шестнадцатеричную.
17. Сформулируйте правило перевода смешанных чисел из двоичной системысчисления в восьмеричную, шестнадцатеричную
18. Сформулируйте правило перевода смешанных чисел из восьмеричной,шестнадцатеричной системы счисления в двоичную.
19. Сформулируйте правило перевода смешанных чисел из восьмеричной системы счисления в шестнадцатеричную и наоборот.
20. Назовите основные структурные единицы памяти компьютера.
21. Какие форматы используются для представления чисел в памяти компьютера? В каком формате представляются целые числа в памяти ЭВМ?
22. Как получить дополнительный код целого числа? В каком случае

он используется?

23. В каком формате представляются действительные числа в памяти ЭВМ? Какая таблица кодировки используется для кодирования текстовой информации в памяти компьютера? Особенности и принцип построения таблицы кодировки.

24. В чем заключается растровое представление графической информации?

25. Что такое растр? Пиксель? Битовая глубина?

26. Приведите формулу для вычисления объема графического файла.

27. В чем заключается векторное представление графической информации? Чтотакое графические примитивы?

28. Опишите процесс кодирование звуковой информации.

29. Опишите процесс декодирования звуковой информации.

30. Какими характеристиками аудиоадаптера определяется качество компьютерного звука?

31. Приведите формулу для вычисления объема звукового файла.

III. ТЕХНОЛОГИЯ ПЕРЕДАЧИ ДАННЫХ ПО КАНАЛАМ СВЯЗИ

Понятие канала связи, виды каналов связи; Критерии эффективности канала связи. Виды кодирования информации; Теоремы Шеннона и Котельникова; Алгоритмы и недостатки системы эффективного кодирования; Преимущества блочного кодирования; Основные понятия помехоустойчивого кодирования; Сущность и алгоритмы основных методов сжатия информации

3.1 Виды каналов связи и источники информации

На современном этапе развития средства коммуникации и связи играют важную роль для обеспечения эффективной передачи информации, которая может осуществляться вручную либо механически при помощи автоматизированных систем по различным каналам связи.

Первый способ передачи информации и до настоящего времени имеет широкое распространение. При этом информация передается либо при помощи курьера, либо по почте. К достоинствам этого способа можно отнести полную достоверность и конфиденциальность передаваемой информации, контроль за ее получением (при почтовой рассылке в пунктах регистрации прохождения), минимальные издержки, не требующие никаких капитальных затрат. Главными недостатками такого подхода являются невысокая скорость передачи информации и неоперативность в получении ответов.

Второй способ значительно увеличивает скорость передачи информации, повышает оперативность принятия решений, но при этом увеличиваются капитальные и текущие издержки. При грамотной организации производственного процесса на предприятии этот способ передачи информации в конечном итоге существенно повышает экономическую эффективность функционированияпредприятия.

Вспомним универсальную схему передачи информации (см. рис. 1 и рис. 3) из которой следует, что для передачи информации необходимы: источник информации, потребитель информации,

приемо-передающие устройства, между которыми могут существовать каналы связи.

При ручном или механическом способе передачи информации на каждом этапе принимают участие люди, при автоматизированной передаче могут использоваться различные электронные приборы и устройства. Одной из проблем, возникающей при автоматизированной передаче информации, является качество передачи информации, которое значительно снижается из-за возникающих в каналах связи и в приемо-передающих устройствах помех. Для снижения последних, улучшения качества передаваемой информации и обеспечения ее достоверности в приемо-передающие устройства встраиваются специальные схемы. Чем меньше помех, тем качественнее работают автоматизированные системы.

Таким образом, *информационный канал (канал связи)* – это совокупность устройств, объединенных линиями связи, предназначенных для передачи информации от источника информации (начального устройства канала) до ее приемника (конечного устройства канала).

Линии связи обеспечивают прохождение информационных сигналов между устройствами канала. Информация обычно передается при помощи *сигнала:* электрического тока (по проводам), света (по оптоволокну), электромагнитных волн радиодиапазона (в пространстве) и, редко, звука (в плотной среде: атмосфере, воде и т.п.) и прочих.

Устройства канала — это, как правило, репитеры, просто передающие усиленным принятый сигнал (пример, радиорелейные линии).

К устройствам канала иногда относят и кодеры/декодеры, но в только тех случаях, когда кодирование/декодирование происходит с высокой скоростью, не требующей ее специального учета, как замедляющего фактора; обычно же кодеры/декодеры относят к источникам или приемникам информации.

Каналы связи являются основным звеном любой системы передачи информации.

Таблица 6 - **Классификация каналов связи**

Признак классификации	**Характеристики каналов связи**
Физическая природа передаваемого сигнала	Механические, акустические, оптические и электрические. В свою очередь, каналы связи могут быть проводными (электрические провода, кабели, световоды) и беспроводными, использующие электромагнитные волны, распространяющиеся в эфире (радиоканалы, инфракрасные каналы и т.д.)
Способ передачи информации	**Симплексные** передают информацию в одном направлении. **Дуплексные** передают информацию одновременно и в прямом, и обратном направлении. **Полудуплексные** осуществляют попеременную передачу информации либо в прямом, либо в обратном направлении.
Форма представления передаваемой информации	**Аналоговые** представляют информацию в непрерывной форме – в виде непрерывного сигнала какой-либо физической природы. **Цифровые** представляют информацию в цифровой (прерывной – дискретной, импульсной) форме сигналовкакой-либо физической природы
Время существования	**Коммутируемые** – временные, создаются только на время передачи информации. По окончании передачи информации и разъединении уничтожаются. **Некоммутируемые** – создаются на длительное время с определенными постоянными характеристиками. Их еще называют выделенными.

Скорость передачи информации	**Низкоскоростные** (50 – 200 бит/с) используются в телеграфных каналах связи. **Среднескоростные** (от 300 – 9600 бит/с) используются в телефонных (аналоговых) каналах связи. Новые стандарты могут использовать скорость от 14 – 56 кбит/с. Для передачи информации по низко- и среднескоростным каналам используются проводные линии связи. **Высокоскоростные** (свыше 56 кбит/с) называют широкополосными. Для передачи информации используются специальные кабели: - экранированные (Shielded Twisted Pair – STP) - неэкранированные (Unshielded Twisted Pair – UTP) с витыми парами из медных проводов; - коаксиальные (Coaxial Cable – CC), - оптоволоконные (Fiber Optic Cable – FOC), - радиоканалы.

Поскольку существует множество различных каналов связи, то передаваемую информацию необходимо представить в виде, соответствующем данному каналу. Такое преобразование обычно связано с модуляцией сигналов.

Модуляция – изменение какого-либо параметра сигнала в канале связи (модулируемого сигнала) в соответствии с текущими значениями передаваемых данных (т.е. моделирующего сигнала). Обратное преобразование модулированного сигнала в модулирующий называется *демодуляция.* Для этих целей существуют специальные устройства – *модемы.* Название «модем» состоит из двух составляющих: первый слог обозначает *модулятор* – устройство прямого преобразования сигнала, второй слог – *демодулятор* – устройство обратного пре- образования сигнала.

В современных модемах чаще всего используются следующие виды модуляции:

- *частотная* (FSK – Frequency Shift Keying); фазовая (PSK – Phase Shift Keying);

- *квадратурная амплитудная* (QAM – Quadrature Amplitude Modulation). При передаче сигналов одним из важнейших параметров является *помехоустойчивость*. Первые два вида модуляции являются весьма помехоустойчивыми, так как при передаче искажается обычно лишь амплитуда сигнала. В последнем виде модуляции для защищенности от помех применяют более помехоустойчивый способ – квадратурную амплитудную модуляцию.

Любое преобразование и передача данных по каналам связи осуществляются в соответствии с принятыми протоколами передачи информации.

Протокол передачи данных – это совокупность правил, которые определяют формат данных и процедуры передачи их по каналу связи, в которых, как правило, указываются способ модуляции, соединение с каналом, представление данных и т.д. Все это делается для повышения достоверности передаваемых данных.

Все модемы имеют определенные стандарты передачи данных, которые устанавливаются Международным институтом телекоммуникаций (ITU – International Telecommunication Union). Обычно стандарт включает несколько протоколов передачи данных. Одним из наиболее эффективных стандартов является стандарт V.34. Он выполняет тестирование канала связи, определяя при этом наиболее эффективный режим работы модема.

В случае передачи большого потока информации, когда она представлена в виде файла, для ее передачи необходимо использовать специальные протоколы, которые осуществляют процедуры разбиения информации на блоки, автоматическое обнаружение и исправление ошибок, повторную пересылку неверно принятых блоков информации, восстановление передачи после обрыва и т.п.

Технические характеристики канала определяются принципом действия входящих в него устройств, видом сигнала, свойствами и составом физической среды, в которой распространяются сигналы, свойствами применяемого кода.

Любая система передачи информации характеризуется такими показателями как помехоустойчивость, эффективность и надежность. Кроме этого для характеристики системы передачи

информации необходимо иметь представление о пропускной способности канала связи.

Помехоустойчивость канала характеризует его способность обеспечивать передачу сигналов в условиях помех. Помехи принято делить на внутренние (представляет собой тепловые шумы аппаратуры) и внешние (они многообразны и зависят от среды передачи). Помехоустойчивость канала зависит от аппаратных и алгоритмических решений по обработке принятого сигнала, которые заложены в приемо-передающее устройство. Помехоустойчивость передачи сигналов через канал может быть повышена за счет кодирования и специальной обработки сигнала.

Термин «эффективность» означает «экономичность». Эффективность можно оценивать в денежных единицах (стоимость строительства каналокилометра линии связи, срок окупаемости, эксплутационные расходы и т.д.). Хотя это – важные показатели эффективности, мы будем определять *эффективность систем связи*, пользуясь их техническими характеристиками. Например, если какая-то система связи обеспечивает передачу информации по каналу тональной частоты со скоростью 1200 бит/с, а вторая система со скоростью 9600 бит/с, с нашей точки зрения вторая система считается более эффективной (с точки зрения использования пропускной способности канала связи), однако, стоимость второй системы может быть более высокой, чем первой. Да и с точки зрения технических показателей вторая система может считаться менее эффективной, чем первая, если сравнение производить не по пропускной способности канала (или полосе частот канала связи), а, например, по мощности сигнала, передаваемого по линии связи. Во втором случае может потребоваться большая мощность сигнала и с этой точки зрения вторая система может оказаться менее эффективной и с точки зрения её технических показателей.

Современные системы связи должны обеспечивать достаточно высокую скорость передачи информации при заданной полосе частот, минимальной мощности сигнала и минимальной вероятности искажений.

Для оценки эффективности систем связи наиболее часто пользуются тремяпоказателями эффективности:

- β-эффективность (показывает, как используется мощность

сигналапри передаче информации с заданной скоростью);

- γ-эффективность (показывает, как используется полоса частот кана-ла связи);

- η-эффективность (показывает, как используется пропускная спо-собность канала связи), определяемые формулами:

$$\beta = \frac{R}{P_c/N_0} \qquad \gamma = \frac{R}{F_k} \qquad \eta = \frac{R}{C}$$

В этих формулах используются следующие величины:

R – скорость передачи информации;

P_c/N_0 – отношение мощности сигнала к спектральной плотности мощности помехи;

F_k – полоса пропускания канала связи;

C – пропускная способность канала связи.

Все перечисленные показатели эффективности являются безразмерными величинами и определяются в предположении, что в канале связи обеспечивается достаточно малая (заранее заданная) вероятность искажения сигналов (при передаче дискретных сигналов) или заданное отношение мощности сигнала к мощности помехи (при передаче непрерывных сигналов).

Надежность коммуникационной системы связана с помехоустойчивостью и эффективностью и определяется средним временем безотказной работы. Для вычислительных сетей среднее время безотказной работы должно быть достаточно большим и составлять, как минимум, несколько тысяч часов. При анализе надежности необходимо оговаривать надежность передачи сообщений и надежность связи в целом. *Надежность передачи* – это вероятность правильной передачи сообщения при условии правильной работы аппаратуры (т.е. ошибки при передаче обуславливаются исключительно шумами). *Надежность связи* – это вероятность правильного приема сообщения с учетом влияния помех и общей надежности аппаратуры. Таким образом, надежность передачи и надежность связи – это взаимозависимые параметры.

Пропускная способность канала (линии) связи характеризует его потенциальные возможности и определяется максимальной скоростью передачи информации. Измеряется пропускная способность в битах в секунду. Практически так же используются основные производные: Кбит/сек, Мбит/сек, Гбит/сек.

Но не все так просто, как кажется на первый взгляд. Если применяется сразу технология модуляции и кодирования различных параметров сигнала, то такая *пропускная способность канала (линии) связи* будет измеряться в бодах (показывает, сколько произошло изменений параметра сигнала за секунду). В этом случае должно быть предельно понятно, что чем выше частота сигнала при заданном кодировании, то тем больше данных можно пропустить по каналу (линии) связи, т.е. пропускная способность будет выше.

Для *определения пропускной способности* канала (линии) связи в расчет берется взаимосвязь между возможной пропускной способностью и полосой пропускания канала (линии) связи. Причем для определения и расчета в данномслучае не важен способ физического кодирования.

Для расчета пропускной способности канала (линии) связи используется следующая формула (закон Шеннона-Хартли):

$$C=B*log_2(1+Ps/Pn)$$

В этой формуле используются следующие параметры:

C – максимально возможная пропускная способность канала (линии) связи;

B – ширина полосы пропускания (интервалы частот, используемых в каналах связи);

Ps/Pn – соотношение существующего сигнала к шуму.

Из этого соотношения видно, что хотя теоретического предела пропускной способности линии с фиксированной полосой пропускания не существует, на практике такой предел имеется. Действительно, повысить скорость передачи в линии можно за счет увеличения мощности передатчика или же уменьшения мощности шума (помех) на линии связи. Обе эти составляющие поддаются изменению с большим трудом. Повышение мощности передатчика ведет к значительному увеличению его габаритов и стоимости. Снижение уровня шума требует применение специальных кабелей с хорошими защитными экранами, что весьма дорого, а также снижение шума в передатчике и промежуточной аппаратуре, чего достичь весьма не просто. К тому же влияние мощностей полезного сигнала и шума на пропускную способность ограничено логарифмической зависимостью, которая растет далеко не так быстро, как прямо пропорциональная. Так, при достаточно

типичном исходном отношении мощности сигнала к мощности шума повышение мощности передатчика в 2 раза даст только 15% увеличение пропускной способности линии.

Из расчета пропускной способности по закону Шеннона-Хартли можно сделать вывод, что надо использовать более широкий кабель, либо соотношение сигнала к шуму сделать в разы больше (или увеличить наш сигнал, или уменьшить внешние шумы). Например, рассмотрим обычный телефонный канал с тональной частотой, в котором максимальная пропускная способность может быть 33 Кбит в секунду. При условии, что для расчета пропускной способности канала (линии) связи мы использовали максимальные значения ширины пропускания (B = 3.1 кГц) и соотношения сигнала к шуму (Ps/Pn = 30 Дб).

3.2 Кодирование информации при передаче по дискретному каналу. Вопросы криптографии

Необходимость кодирования информации возникла задолго до появления компьютеров. Речь, азбука и цифры – есть не что иное, как система моделирования мыслей, речевых звуков и числовой информации. В технике потребность кодирования возникла сразу после создания телеграфа, но особенно важной она стала с изобретением компьютеров.

Область действия теории кодирования распространяется на передачу данных по реальным (или зашумленным) каналам, а предметом является обеспечение корректности переданной информации. Иными словами, она изучает, как лучше упаковать данные, чтобы после передачи сигнала из данных можно было надежно и просто выделить полезную информацию. Иногда теорию кодирования путают с шифрованием, но это неверно: криптография решает обратную задачу, ее цель - затруднить получение информации из данных.

С необходимостью кодирования данных впервые столкнулись более полутораста лет назад, вскоре после изобретения телеграфа. Каналы были дороги и ненадежны, что сделало актуальной задачу минимизации стоимости и повышения надёжности передачи телеграмм. Проблема ещё более обострилась в связи с прокладкой трансатлантических кабелей. С 1845 вошли в употребление

специальные кодовые книги; с их помощью телеграфисты вручную выполняли «компрессию» сообщений, заменяя распространенные последовательности слов более короткими кодами. Тогда же для проверки правильности передачи стали использовать контроль чётности, метод, который применялся для проверки правильности ввода перфокарт ещё и в компьютерах первых поколений. Для этого во вводимую колоду последней вкладывали специально подготовленную карту с контрольной суммой. Если устройство ввода было не слишком надежным (или колода - слишком большой), то могла возникнуть ошибка. Чтобы исправить её, процедуру ввода повторяли до тех пор, пока подсчитанная контрольная сумма не совпадала с суммой, сохраненной на карте. Эта схема неудобна, и к тому же пропускает двойные ошибки. С развитием каналов связи потребовался более эффективный механизм контроля.

Первым теоретическое решение проблемы передачи данных по зашумленным каналам предложил Клод Шеннон. Работая в *Bell Labs*, Шеннон написал работу «Математическая теория передачи сообщений» (1948), где показал, что если пропускная способность канала выше энтропии источника сообщений, то сообщение можно закодировать так, что оно будет передано без излишних задержек. В одной из теорем Шеннон доказал, что при наличии канала с достаточной пропускной способностью сообщение может быть передано с некоторыми временными задержками. Кроме того, он показал возможность достоверной передачи при наличии шума в канале. Формула $C = W \log ((P+N)/N)$, высечена на скромном памятнике Шеннону, установленном в его родном городе в штате Мичиган.

Труды Шеннона дали пищу для множества дальнейших исследований в области теории информации, но практического инженерного приложения они не имели. Переход от теории к практике стал возможен благодаря усилиям Ричарда Хэмминга, коллеги Шеннона по *Bell Labs*, получившего известность за открытие класса кодов «коды Хэмминга». Существует легенда, что к изобретению своих кодов Хэмминга подтолкнуло неудобство в работе с перфокартами на релейной счетной машине в середине сороковых годов. Ему давали время для работы на машине в выходные дни, когда не было операторов, и ему самому

приходилось возиться с вводом. Хэмминг предложил коды, способные корректировать ошибки в каналах связи, в том числе и в магистралях передачи данных в компьютерах, прежде всего между процессором и памятью. Коды Хэмминга показали, как можно практически реализовать возможности теоремы Шеннона. Хэмминг опубликовал свою статью в 1950, хотя во внутренних отчетах его теория кодирования датируется 1947. Поэтому некоторые считают, что отцом теории кодирования следует считать Хэмминга, а не Шеннона.

Ричард Хэмминг (1915 – 1998) получил степень бакалавра в Чикагском университете в 1937. В 1939 г. он получил степень магистра в Университете Небраски, а степень доктора по математике – в Университете Иллинойса. В 1945 Хэмминг начал работать в рамках Манхэттенского проекта. В 1946 поступил на работу в Bell Telephone Laboratories, где работал с Шенноном. В 1976 получил кафедру в военно-морской аспирантуре в Монтерей в Калифорнии. Труд, сделавший его знаменитым, фундаментальное исследование кодов обнаружения и исправления ошибок, Хэмминг опубликовал в 1950. В 1956 г. он принимал участие в работе над IBM 650. Его работы заложили основу языка программирования, который позднее эволюционировал в языки программирования высокого уровня. В знак признания заслуг Хэмминга в области информатики институт IEEE учредил медаль за выдающиеся заслуги в развитии информатики и теории систем, которую назвал его именем.

Хэмминг первым предложил «коды с исправлением ошибок» (*Error- Correcting Code, ECC).* Современные модификации этих кодов используются во всех системах хранения данных и для обмена между процессором и оперативной памятью. Один из их вариантов, коды Рида-Соломона применяются в компакт-дисках, позволяя воспроизводить записи без скрипов и шумов, вызванных царапинами и пылинками. Существует множество версий кодов, построенных «по мотивам» Хэмминга, они различаются алгоритмами кодирования и количеством проверочных битов. Особое значение подобные коды приобрели в связи с развитием дальней космической связи с межпланетными станциями.

Среди новейших кодов *ECC* следует назвать коды *LDPC (Low-Density Parity-check Code).* Вообще-то они известны лет тридцать,

но особый интерес к ним обнаружился именно в последние годы, когда стало развиваться телевидение высокой чёткости. Коды *LDPC* не обладают 100-процентной достоверностью, но вероятность ошибки может быть доведена до желаемой, и при этом с максимальной полнотой используется пропускная способность канала. К ним близки «турбокоды» (*Turbo Code*), они эффективны при работе с объектами, находящимися в условиях далекого космоса и ограниченной пропускной u1089 способности канала.

В историю теории кодирования прочно вписано имя В. А. Котельникова. В 1933 г. в «Материалах по радиосвязи к I Всесоюзному съезду по вопросам технической реконструкции связи» он опубликовал работу «О пропускной способности «эфира» и «проволоки». Имя Котельникова входит в название одной из важнейших теорем теории кодирования, определяющей условия, при которых переданный сигнал может быть восстановлен без потери информации. Эту теорему называют по-разному, в том числе «теоремой *WKS*» (аббревиатура *WKS* взята от *Whittaker, Kotelnikov, Shannon*). В некоторых источниках используют и *Nyquist-Shannon sampling theorem*, и *Whittaker-Shannon sampling theorem*, а в отечественных вузовских учебниках чаще всего встречается просто «теорема Котельникова». На самом же деле теорема имеет более долгую историю. Ее первую часть в 1897 доказал французский математик Э. Борель. Свой вклад в 1915 внес Э. Уиттекер. В 1920 г. японец К. Огура опубликовал поправки к исследованиям Уиттекера, а в 1928 американец Гарри Найквист уточнил принципы оцифровки и восстановления аналогового сигнала.

Таким образом, современная теория кодирования опирается на следующие теоремы.

Первая теорема Шеннона декларирует возможность создания системы эффективного кодирования дискретных сообщений:

При отсутствии помех передачи всегда возможен такой вариант кодирования сообщения, при котором избыточность кода будет сколь угодно близкой к нулю.

Вторая теорема Шеннона гласит, что при наличии помех в канале, всегда можно найти такую систему кодирования, при которой сообщения будут переданы с заданной достоверностью.

Эти теоремы не дают конкретного метода построения кода, но указывают на пределы достижимого в создании помехоустойчивых кодов.

Теорема Котельникова: Для сигналов с ограниченным спектром, где F наибольшая частота в спектре сигнала, чтобы восстановить все свойства сигнала, достаточно взять значения сигнала через равные промежутки времени T - такие, чтобы выполнялось условие $T \leq \frac{1}{2F}$.

Чем выше частота дискретизации, тем точнее происходит перевод непрерывной информации в дискретную. Но с ростом этой частоты растет и размер дискретных данных, получаемых при таком переводе, и, следовательно, сложность их обработки, передачи и хранения. Однако для повышения точности дискретизации необязательно безграничное увеличение ее частоты. Эту частоту разумно увеличивать только до 7 предела, определяемого теоремой о выборках, называемой также теоремой Котельникова или законом Найквиста (Nyquist).

Примером использования этой теоремы являются лазерные компакт-диски, звуковая информация на которых хранится в цифровой форме. Чем выше будет частота дискретизации, тем точнее будут воспроизводиться звуки и тем меньше их можно будет записать на один диск, но ухо обычного человека способно различать звуки с частотой до 20 КГц, поэтому точно записывать звуки с большей частотой бессмысленно. Согласно теореме о выборках, частоту дискретизации нужно выбрать не меньшей 40 КГц (в промышленном стандарте на компакт-диске используется частота 44.1 КГц).

При преобразовании дискретной информации в непрерывную, определяющей является скорость этого преобразования: чем она выше, с тем более высокочастотными гармониками получится непрерывная величина. Но чем большие частоты встречаются в этой величине, тем сложнее с ней работать. Например, обычные телефонные линии предназначены для передачи звуков частотой до 3 КГц.

Все используемые в теории кодирования *коды делятся на два больших класса:*

1) Коды с исправлением ошибок (имеют целью восстановить с вероятностью, близкой к единице, посланное сообщение);

2) Коды с обнаружением ошибок (имеют целью выявить с вероятностью, близкой к единице, наличие ошибок).

В зависимости от целей различают следующие виды кодирования:

1)кодирование по образцу

Данный вид кодирования применяется для представления дискретного сигнала на том или ином машинном носителе. Большинство кодов, используемых в информатике для кодирования по образцу, имеют одинаковую длину и используют двоичную систему для представления кода (и, возможно, шестнадцатеричную как средство промежуточного представления). Используются кодовые таблицы (Например, ASCII).

2)криптографическое кодирование, или шифрование, – используется, когда нужно защитить информацию от несанкционированного доступа.

Криптография (тайнопись) — это раздел математики, в котором изучаются и разрабатываются системы изменения письма с целью сделать его непонятным для непосвященных лиц. Известно, что еще в V веке до нашей эры тайнопись использовалась в Греции. В современном мире, где все больше и больше услуг предоставляется через использование информационных технологий, проблема защиты информации методами криптографии имеет первостепенное значение. Сегодня большая часть обмена информацией проходит по компьютерным сетям и часто (в бизнесе, военным и прочее) нужно обеспечивать конфиденциальность такого обмена. Теоретические основы классической криптографии впервые были изложены Клодом Шенноном в конце 1940-х годов. В качестве символов кодирования могут использоваться как символы произвольного алфавита,так и двоичные коды.

Существуют различные методы криптографии.

а) *простейшая система шифрования* – это замена каждого знака письма на другой знак по выбранному правилу (называется: *простая замена или подстановка***).** Юлий Цезарь, например, заменял в своих секретных письмах первую букву алфавита на четвертую, вторую – на пятую, последнюю – на третью и т.п., т.е. A на D, B на E, Z на C и т.п.

Недостаток: Шифры простой замены легко поддаются расшифровке, при знании исходного языка сообщения, т.к. каждый

письменный язык характеризуется частотой встречаемости своих знаков. Например, в английском языке чаще всего встречается буква E, а в русском – О. Таким образом, в шифрованном подстановкой сообщении на русском языке самому частому знаку будет с большой вероятностью соответствовать буква О. Вероятность будет расти с ростом длины сообщения.

б) усовершенствованные *шифры-подстановки* используют возможность заменять символ исходного сообщения на любой символ из заданного для него множества символов, что позволяет выровнять частоты встречаемости различных знаков шифра, но подобные шифры удлиняют сообщение и замедляют скорость обмена информацией.

В шифрах-перестановках знаки сообщения специальным образом переставляются между собой, например, записывая сообщение в строки заданной длины и беря затем последовательность слов в столбцах в качестве шифра. Сообщение «ТЕОРИЯИНФОРМАЦИИ», используя строки длины 4, будет в шифрованном таким методом виде выглядеть как «ТИФАЕЯОЦОИРИРНМИ», потому что при шифровании использовался следующий прямоугольник:

ТЕОР
ИЯИН
ФОРМ
АЦИИ.

Шифры-перестановки в общем случае практически не поддаются дешифровке. Для их дешифровки необходимо знать дополнительную информацию. Крупный недостаток подобных шифров в том, что если удастся каким-то образом расшифровать хотя бы одно сообщение, то в дальнейшем можно расшифровать и любое другое. Модификацией шифров-перестановок являются шифры-перестановки со словом-ключом, которое определяет порядок взятия слов-столбцов.

в) **системы с ключевым словом или просто ключом,** известные с XVI века, широко применяются до сих пор. Их особенностью является два уровня секретности. Первый уровень – это собственно способ составления кода, который постоянно известен лицам, использующим данный шифр. Второй уровень – это ключ, который посылается отдельно от основного сообщения

по особо защищенным каналам и без которого расшифровка основного сообщения невозможна.

Наиболее простой способ использования ключа хорошего шифра следующий: под символами сообщения записывается раз за разом ключ, затем номера соответствующих знаков сообщения и ключа складываются. Если полученная сумма больше общего числа знаков, то от нее отнимается это общее число знаков. Полученные числа будут номерами символов кода. С ростом длины ключа трудоемкость дешифровки подобного шифра стремительно растет. Если в качестве ключа использовать случайную последовательность, то получится не раскрываемый шифр. Проблема этого шифра - это способ передачи ключа.

В информационных сетях использование традиционных систем шифрования с ключом затрудненно необходимостью иметь специальный особо защищенный способ для передачи ключа. В 1976 году У. Диффи (Diffie W.) и М. Хеллман (Hellman M.) - инженеры-электрики из Станфордского университета, а также студент Калифорнийского университета Р. Меркль (Merkle R.), предложили новый принцип построения криптосистем, не требующий передачи ключа принимающему сообщение и сохранения в тайне метода шифрования. На идеях Диффи и Хеллмана основаны следующие системы: без передачи ключей, с открытым ключом и электронная подпись - все они в свою очередь основаны на математическом фундаменте теории чисел.

3) помехозащитное (помехоустойчивое) кодирование

*а) простейший код для борьбы с шумом – это контроль четности***,** он, в частности, широко используется в модемах. Кодирование заключается в добавлении к каждому байту девятого бита таким образом, чтобы дополнить количество единиц в байте до заранее выбранного для кода четного (even) или нечетного (odd) значения. Используя этот код, можно лишь обнаруживать большинство ошибок.

б) простейший код, исправляющий ошибки, – это тройное повторение каждого бита. Если с ошибкой произойдет передача одного бита из трех, то ошибка будет исправлена, но если случится двойная или тройная ошибка, то будут получены неправильные данные. Часто коды для исправления ошибок используют совместно с кодами для обнаружения ошибок. При тройном

повторении для повышения надежности три бита располагают не подряд, а на фиксированном расстоянии друг от друга.

Использование тройного повторения естественно значительно снижает скорость передачи данных.

4) эффективное, или оптимальное, кодирование – используется для устранения избыточности информации, т.е. снижения ее объема, например, в архиваторах. Для кодирования символов исходного алфавита используют двоичные коды переменной длины: чем больше частота символа, тем короче его код.

3.3 Сжатие и архивация информации

Сущность и методы эффективного кодирования. Метод Шеннона-Фано. Метод Хаффмана.

Эффективность кода определяется средним числом двоичных разрядов для кодирования одного символа – $I_{ср}$ по формуле:

$$I_{ср} = \sum_{s=1}^{k} n_s * f_s \ ,$$

где k – число символов исходного алфавита;

n_s – число двоичных разрядов для кодирования символа s;

f_s – частота символа *s*.

Существуют два классических метода эффективного кодирования: метод Шеннона-Фано и метод Хаффмена. Входными данными для обоих методов является заданное множество исходных символов для кодирования с их частотами; результат – эффективные коды.

Метод сжатия по алгоритму Шеннона-Фано. При использовании этого метода выполняются следующие шаги:

1) упорядочить множество исходных символов по убыванию их частот;

2) список символов разделить на две части (назовем их первой и второй частями) так, чтобы суммы частот обеих частей были точно или примерно равны. В случае, когда точного равенства достичь не удается, разница между суммами должна быть минимальна;

3) кодовым комбинациям первой части дописать 1, кодовым комбинациям второй части дописать 0;

4) проанализировать первую часть: если она содержит только один символ,работа с ней заканчивается, – считается, что код для ее

символов построен;

5) продолжить построение кода второй части. Если в ней символов больше одного, переходят к шагу 1) и процедура повторяется, если она содержит только один символ, работа с ней заканчивается. Код построен.

Пример 1. Даны символы a, b, c, d с частотами f_a= 0,5; f_b=0,25; f_c=0,125; f_d=0,125. Построить эффективный код методом Шеннона-Фано.

Решение:

1) Сведем исходные данные в таблицу, упорядочив символы по невозрастанию их частот:

Исходные символы	**Частоты символов**	**Формируемый код**
a	0,5	
b	0,25	
c	0,125	
d	0,125	

2) Список символов разделим на две части (назовем их первой и второй частями) так, чтобы суммы частот обеих частей (назовем их Σ1 и Σ2) были точно или примерно равны. Следовательно, первая линия деления проходит под символом *a*: соответствующие суммы Σ1 и Σ2 равны между собойи равны 0,5.

3) Формируемым кодовым комбинациям дописывается 1 для верхней (первой) части и 0 для нижней.

Исходные символы	**Частоты символов**	**Формируемый код**
A	0,5	1
B	0,25	0
C	0,125	0
D	0,125	0

4) Так как верхняя часть списка содержит только один элемент (символ *a*), работа с ней заканчивается, а эффективный код для этого символа считается сформированным (в таблице, приведенной

выше, эта часть списка частот символов выделена заливкой).

5) Второе деление выполняется под символом *b*: суммы частот Σ1 и Σ2 вновь равны между собой и равны 0,25. Тогда кодовой комбинации символов верхней части дописывается 1, а нижней части – 0. Таким образом, к полученным на первом шаге фрагментам кода, равным 0, добавляются новые символы:

Исходные символы	Частоты символов	Формируемый код
A	0,5	1
B	0,25	01
C	0,125	00
D	0,125	00

Поскольку верхняя часть нового списка содержит только один символ (*b*), формирование кода для него закончено (соответствующая строка таблицы вновь выделена заливкой). Третье деление проходит между символами *c* и *d*: к кодовой комбинации символа *c* приписывается 1, коду символа *d* приписывается 0:

Исходные символы	Частоты символов	Формируемый код
A	0,5	1
B	0,25	01
C	0,125	001
D	0,125	000

Поскольку обе оставшиеся половины исходного списка содержат по од ному элементу, работа со списком в целом заканчивается.

Таким образом, получили коды:

a – 1, *b* – 01, *c* – 001, *d* – 000.

Определим эффективность построенного кода по формуле:

Icp = 0,5*1 + 0,25*2 + 0,125*3 + 0,125*3 = 1,75 бит/символ.

Поскольку при кодировании четырех символов кодом постоянной длины требуется два двоичных разряда, сэкономлено 0,25 двоичного разряда в среднем на один символ.

Задания:

№1. Даны символы A, B, C с частотами f_a=0,4, f_b=0,2, f_c=0,4. Построить эффективный код методом Шеннона-Фано. Определить эффективность кода.

№2. Дано сообщение «Скоро сессия!». Построить эффективный код методом Шеннона-Фано. Определить эффективность кода.

№3. Дано сообщение «Красная краска». Построить эффективный код методом Шеннона-Фано. Определить эффективность кода.

Сжатие по алгоритму Хаффмана. Один из первых алгоритмов эффективного кодирования информации был предложен Д.А. Хаффманом в 1952 году. Идея алгоритма состоит в следующем: зная вероятности символов в сообщении, можно описать процедуру построения кодов переменной длины, состоящих из целого количества битов. Символам с большей вероятностью ставятся в соответствие более короткие коды. Классический алгоритм Хаффмана на входе получает таблицу частот встречаемости символов в сообщении. Далее на основании этой таблицы строится дерево кодирования Хаффмана (H-дерево).

Сжатие данных по Хаффману применяется при сжатии фото- и видеоизображений (JPEG, стандарты сжатия MPEG), в архиваторах (PKZIP, LZH и др.), в протоколах передачи данных MNP5 и MNP7.

Сжимая файл по алгоритму Хаффмана необходимо выполнить следующую последовательность действий:

1) прочитать файл полностью и подсчитать сколько раз встречается каждый символ из расширенного набора ASCII;
2) подсчитать частоту вхождения каждого символа;
3) просмотреть таблицу кодов ASCII и сформировать мнимую компоновку между кодами по убыванию. То есть, не меняя местонахождение каждого символа из таблицы в памяти отсортировать таблицу ссылок на них по убыванию.

*Пример***:** Дан файл длиной в 100 байт и состоящий из 6 различных символов.

Решение:

1) Допустим, мы подсчитали вхождение каждого из символов в файл иполучили следующее:

Символ	A	B	C	D	E	F
Число вхождений	10	20	30	5	25	10

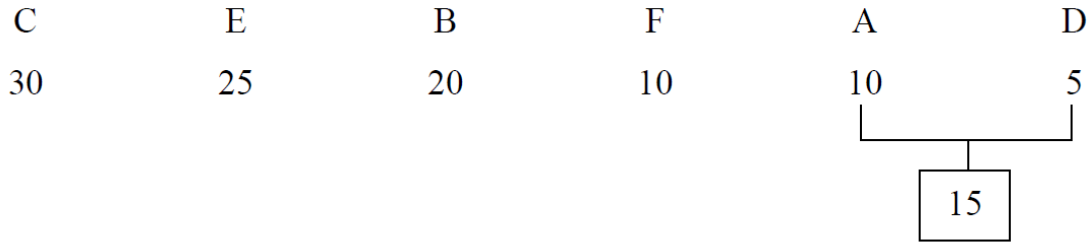
2) Числа, представленные в таблице п.1 и будем называть

частотой вхождения для каждого символа.

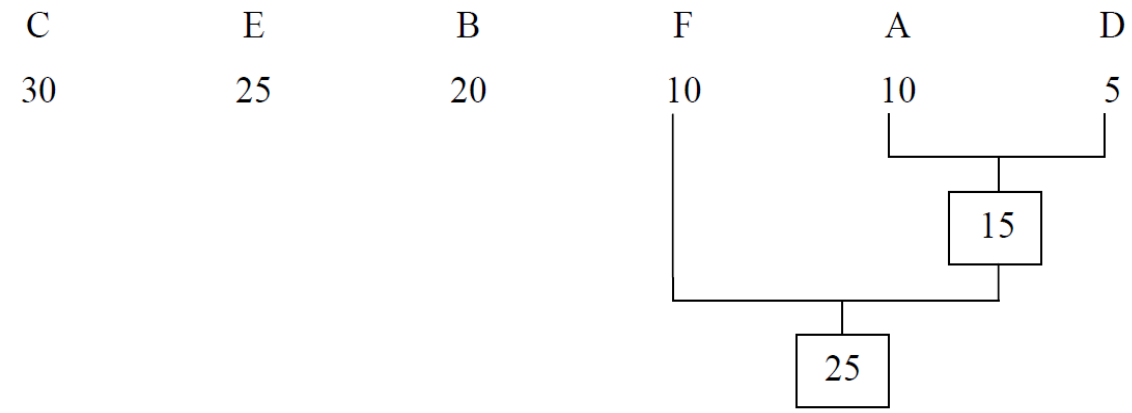
3) Сформируем мнимую компоновку между кодами по убыванию:

Символ	C	E	B	F	A	D
Число вхождений	30	25	20	10	10	5

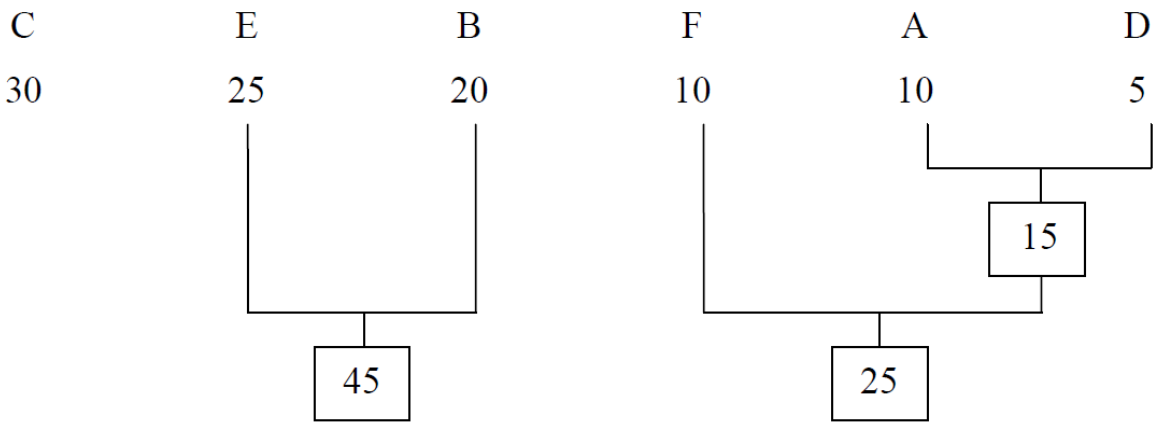
4) Возьмем из последней таблицы символы с наименьшей частотой. В нашем случае это D (5) и какой либо символ из F или A (10), можно взять любой из них, например A. Сформируем из «узлов» D и A новый «узел», частота вхождения для которого будет равна сумме частот D и A (5+10=15):



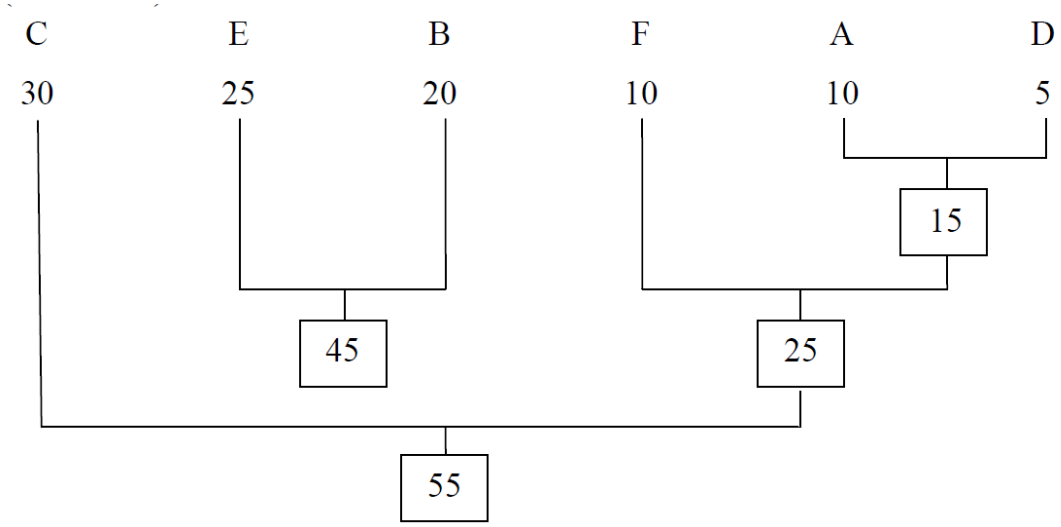
Номер в рамке – сумма частот символов D и A. Теперь мы снова ищем два символа с самыми низкими частотами вхождения. Исключая из просмотра D и A и рассматривая вместо них новый «узел» с суммарной частотой вхождения. Самая низкая частота теперь у F и нового «узла». Снова сделаем операцию слияния узлов: (10+15=25):



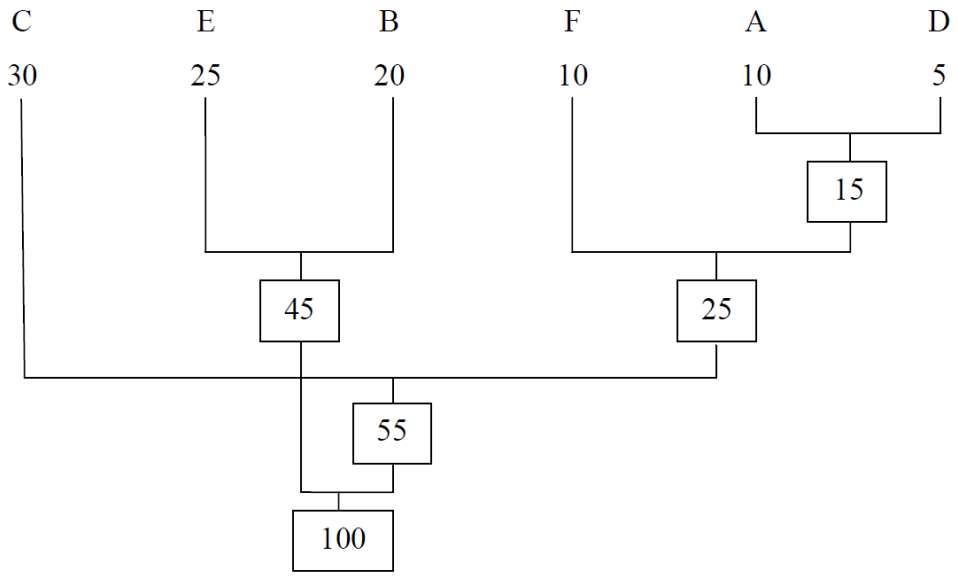
Рассматриваем таблицу снова для следующих двух символов B и E(20+25=45) .



Далее объединяем в новый узел C и узел 25, т.к. они имеют меньшие частоты (30+25=55):



И, наконец, объединяем между собой узлы 45 и 55 (45+55=100):



Теперь, когда наше дерево создано можно кодировать файл. Мы должны всегда начинать из корня (в нашем случае 100). Кодируя первый символ (лист дерева C) мы прослеживаем вверх по дереву все повороты ветвей и если делаем левый поворот, то запоминаем 0, а если правый поворот, то запоминаем 1. Так для C, мы будем идти вправо к 55 (запомним 1), затем влево к самому символу (запомним 0). Код Хаффмана для нашего символа C – 10. Для следующего символа (E) у нас получается – влево (запомним 0), влево (запомним 0), следовательно код для символа E – 00 и т.д.

Выполнив выше сказанное для всех символов получим: C = 10 (2 бита)

E = 00 (2 бита)
B = 01 (2 бита)
F = 110 (3 бита)
A = 1110 (4 бита)
D = 1111 (4 бита)

Каждый символ изначально представлялся 8-ю битами (согласно таблице кодировки ASCII), и так как в результате применения алгоритма Хаффмана мы уменьшили число битов необходимых для представления каждого символа, следовательно, уменьшили размер выходного файла. Сжатие складывается следующим образом:

Символ	Частота	Первоначально (бит)	Уплотненные биты	Уменьшено на
C	30	30*8=240	30*2=60	180
E	25	25*8=200	25*2=50	150
B	20	20*8=160	20*2=40	120
F	10	10*8=80	10*3=30	50
A	10	10*8=80	10*4=40	40
D	5	5*8=40	5*4=20	20
Всего:		800	240	560

Первоначальный размер файла составлял 100 байт (800 бит);

Размер сжатого файла: 30 байт – 240 бит; 240 составляет 30% из 800, так что мы сжали этот файл на 70%.

Для восстановления первоначального файла, необходимо иметь декодирующее дерево, так как деревья будут различны для разных файлов. Следовательно, нужно сохранять дерево вместе с файлом. Это превращается в итоге в увеличение размеров выходного файла и в нашем случае сжатие будет приблизительно 20% (максимально идеализированный алгоритм Хаффмана может достигать сжатия в 33%).

Сущность подстановочного или словарно-ориентированного метода сжатия информации. Методы Лемпела-Зива. Методы Шеннона-Фэно и Хаффмена называются статистическими методами. Более практический характер носят словарные алгоритмы. Их преимущество перед статистическими теоретически объясняется тем, что они позволяют кодировать последовательности символов разной длины. Неадаптивные статистические алгоритмы тоже можно использовать для таких последовательностей, но в этом случае их реализация становится весьма ресурсоемкой.

Рассмотрим сущность словарно-ориентированного метода сжатия на основе алгоритма LZ77. Алгоритм LZ77 разработан израильскими математиками Авраамом Лемпелом (**L**empel) и Якобом Зивом (**Z**iv) и был опубликован в 1977 г. Многие программы сжатия информации используют ту или иную модификацию LZ77. Одной из причин популярности алгоритмов LZ является их исключительная простота при высокой эффективности сжатия.

Основная идея LZ77 состоит в том, что второе и последующие вхождения некоторой строки символов в сообщении заменяются ссылками на ее первое вхождение.

LZ77 использует уже просмотренную часть сообщения как словарь. Чтобы добиться сжатия, он пытается заменить очередной фрагмент сообщения на указатель в содержимое словаря.

LZ77 использует «скользящее» по сообщению окно, разделенное на две неравные части. Первая, большая по размеру, включает уже просмотренную часть сообщения. Вторая, намного меньшая, является буфером, содержащим еще незакодированные символы входного потока. Обычно размер окна составляет несколько килобайт, а размер буфера - не более ста байт. Алгоритм пытается найти в словаре (большей части окна) фрагмент, совпадающий с содержимым буфера.

Алгоритм LZ77 выдает коды, состоящие из трех элементов:

1) смещение в словаре относительно его начала подстроки, совпадающей сначалом содержимого буфера;
2) длина этой подстроки;
3) первый символ буфера, следующий за подстрокой.

Классический алгоритм Лемпеля-Зива – LZ77 формулируется следующим образом: «если в прошедшем ранее выходном потоке уже встречалась подобная последовательность байт, причем запись о ее длине и смещении от текущей позиции короче чем сама эта последовательность, то в выходной файл записывается ссылка (смещение, длина), а не сама последовательность». Так фраза

«КОЛОКОЛ_ОКОЛО_КОЛОКОЛЬНИ» закодируется как «КОЛО(-4,3)_(-5,4)О_(-14,7)ЬНИ».

Пример.

Закодировать по алгоритму LZ77 строку «КРАСНАЯ КРАСКА».

Словарь (8)	Буфер (5)	Код
«……..»	«КРАСН»	<0,0, ‘К’>
«…….К»	«РАСНА»	<0,0,’Р’>
«……КР»	«АСНАЯ»	<0,0,’А’>
«…..КРА»	«СНАЯ »	<0,0,’С’>
«….КРАС»	«НАЯ К»	<0,0,’Н’>
«…КРАСН»	«АЯ КР»	<5,1,’Я’>
«.КРАСНАЯ»	« КРАС»	<0,0,’ ’>
«КРАСНАЯ.»	«КРАСК»	<0,4,’К ’>
«АЯ КРАСК»	«А….»	<0,0,’А ’>

В последней строчке, буква «А» берется не из словаря, т.к. она последняя.

Длина кода вычисляется следующим образом: длина подстроки не может быть больше размера буфера, а смещение не может быть больше, чем *(размер словаря – 1)* . Следовательно, длина двоичного кода смещения будет округленным в большую сторону **$\log_2$ (размер словаря)**, а длина двоичного кода для длины подстроки будет округленным в большую сторону **$\log_2$ (размер буфера+1)**. А символ кодируется 8 битами (например, ASCII+).

Следовательно, в приведенном примере длина полученного кода равна$9 \times (3+3+8) = 126$ бит, против $14 \times 8 = 112$ бит исходной длины строки.

Декодирование кодов LZ77 проще их получения, т.к. не нужно осуществлять поиск в словаре.

Недостатки LZ77:

- с ростом размеров словаря скорость работы алгоритма-кодера пропорционально замедляется;
- кодирование одиночных символов очень неэффективно;
- невозможность кодирования подстрок, отстоящих друг от друга на расстоянии, большем длины словаря;
- длина подстроки, которую можно закодировать, ограничена размером буфера.

Кодирование одиночных символов можно сделать эффективным, отказавшись от ненужной ссылки на словарь для них. Кроме того, в некоторые модификации LZ77 для повышения степени сжатия добавляется возможность для кодирования идущих

подряд одинаковых символов. Если механически чрезмерно увеличивать размеры словаря и буфера, то это приведет к снижению эффективности кодирования, т.к. с ростом этих величин будут расти и длины кодов для смещения и длины, что сделает коды для коротких подстрок недопустимо большими. Кроме того, резко увеличится время работы алгоритма- кодера.

В 1978 г. авторами LZ77 был разработан алгоритм LZ78, лишенный названных недостатков.

LZ78 не использует «скользящее» окно, он хранит словарь из уже про- смотренных фраз. При старте алгоритма этот словарь содержит только одну пустую строку (строку длины нуль). Алгоритм считывает символы сообщения до тех пор, пока накапливаемая подстрока входит целиком в одну из фраз словаря. Как только эта строка перестанет соответствовать хотя бы одной фразе словаря, алгоритм генерирует код, состоящий из индекса строки в словаре, которая до последнего введенного символа содержала входную строку, и символа, нарушившего совпадение. Затем в словарь добавляется введенная подстрока. Если словарь уже заполнен, то из него предварительно удаляют менее всех используемую в сравнениях фразу.

Ключевым для размера получаемых кодов является размер словаря во фразах, потому что каждый код при кодировании по методу LZ78 содержит номер фразы в словаре. Из последнего следует, что эти коды имеют постоянную длину, равную округленному в большую сторону двоичному логарифму размера словаря +8 (это количество бит в байт-коде расширенного ASCII).

Пример. Закодировать по алгоритму LZ78 строку «КРАСНАЯ КРАСКА», используя словарь длиной 16 фраз.

Входная фраза (в словарь)	Код	Позиция словаря
		0
«К»	<0, ‘К’>	1
«Р»	<0,’Р’>	2
«А»	<0,’А’>	3
«С»	<0,’С’>	4
«Н»	<0,’Н’>	5
«АЯ»	<3,’Я’>	6
« »	<0,’ ’>	7
«КР»	<1,’Р’>	8
«АС»	<3,’С’>	9
«КА»	<1,’А’>	10

Указатель на любую фразу такого словаря – это число от 0 до 15, для его кодирования достаточно четырех бит.

В 1984 г. Уэлчем (Welch) был путем модификации LZ78 создан алгоритм LZW. Алгоритм, названный в честь своих создателей Лемпеля, Зива и Велча (Lempel, Ziv и Welch), не требует вычисления вероятностей встречаемости символов или кодов. Основная идея состоит в замене совокупности байтов в исходном файле ссылкой на предыдущее появление той же совокупности.

Алгоритмы LZ77, LZ78 разработаны математиками и могут использоваться свободно. Алгоритм LZW является запатентованным и, таким образом, представляет собой интеллектуальную собственность. Его безлицензионное использование особенно на аппаратном уровне может повлечь за собой неприятности.

Любопытна история патентования LZW. Заявку на LZW подали почти одновременно две фирмы - сначала IBM и затем Unisys, но первой была рассмотрена заявка Unisys, которая и получила патент. Однако, еще до патентования LZW был использован в широко известной в мире Unix программе сжатия данных compress.

Процесс сжатия с использованием LZW выглядит следующим образом. Последовательно считываются символы входного потока и происходит проверка, существует ли в созданной таблице строк такая строка. Если такая строка существует, считывается

следующий символ, а если строка не существует, в поток заносится код для предыдущей найденной строки, строка заносится в таблицу, а поиск начинается снова.

Например, если сжимают байтовые данные (текст), то строк в таблице окажется 256 (от «0» до «255»). Для кода очистки и кода конца информации используются коды 256 и 257. Если используется 10-битный код, то под коды для строк остаются значения в диапазоне от 258 до 1023. Новые строки формируют таблицу последовательно, т. е. можно считать индекс строки ее кодом.

Пример 1: Рассмотрим пример сжатия сообщения **«abacabadabacabae»** с использованием данного алгоритма.

Сначала создадим начальный словарь единичных символов. В стандартной кодировке ASCII имеется 256 различных символов, поэтому, для того, чтобы все они были корректно закодированы (если нам неизвестно, какие символы будут присутствовать в исходном файле, а какие – нет), начальный размер кода будет равен 8 битам. Если нам заранее известно, что в исходном файле будет меньшее количество различных символов, то вполне разумно уменьшить коли- чество бит. Чтобы инициализировать таблицу, мы установим соответствие кода 0 соответствующему символу с битовым кодом 00000000, тогда 1 соответствует символу с кодом 00000001, и т.д., до кода 255. На самом деле мы указали, что каждый код от 0 до 255 является корневым.

Больше в таблице не будет других кодов, обладающих этим свойством. По мере роста словаря, размер групп должен расти, с тем, чтобы учесть новые элементы. 8-битные группы дают 256 возможных комбинации бит, поэтому, когда в словаре появится 256-е слово, алгоритм должен перейти к 9-битным группам. При появлении 512-ого слова произойдет переход к 10-битным группам, что дает возможность запоминать уже 1024 слова и т.д.

В нашем примере алгоритму заранее известно о том, что будет использоваться всего 5 различных символов, следовательно, для их хранения будет использоваться минимальное количество бит, позволяющее нам их запомнить, то есть 3 (8 различных комбинаций).

Символ	Битовый код
a	000
b	001
c	010
d	011
e	100

- **Шаг 1:** Согласно изложенному выше алгоритму, мы добавим к изначально пустой строке «a» и проверим, есть ли строка «a» в таблице. Поскольку мы при инициализации занесли в таблицу все строки из одного символа, то строка «a» есть в таблице.
- **Шаг 2:** Далее мы читаем следующий символ «b» из входного потока и проверяем, есть ли строка «ab» в таблице. Такой строки в таблице пока нет. Добавляем в таблицу <5> «ab». В поток: <0>;
- **Шаг 3:** «ba» – нет. В таблицу: <6> «ba». В поток: <1>;
- **Шаг 4:** «ac» – нет. В таблицу: <7> «ac». В поток: <0>;
- **Шаг 5:** «ca» – нет. В таблицу: <8> «ca». В поток: <2>;
- **Шаг 6:** «ab» – есть в таблице; «aba» – нет. В таблицу: <9> «aba». В поток: <5>;
- **Шаг 7:** «ad» – нет. В таблицу: <10> «ad». В поток: <0>;
- **Шаг 8:** «da» – нет. В таблицу: <11> «da». В поток: <3>;
- **Шаг 9:** «aba» – есть в таблице; «abac» – нет. В таблицу: <12> «abac». Впоток: <9>;
- **Шаг 10:** «ca» – есть в таблице; «cab» – нет. В таблицу: <13> «cab». В поток: <8>;
- **Шаг 11:** «ba» – есть в таблице; «bae» - нет. В таблицу: <14> «bae». В поток: <6>;
- **Шаг 12:** И, наконец, последняя строка «e», за ней идет конец сообщения,поэтому мы просто выводим в поток <4>.

Текущая строка	Текущий символ	Следующий символ	Вывод		Словарь	
			Код	Биты		
ab	a	b	0	000	5:	ab
ba	b	a	1	001	6:	ba
ac	a	c	0	000	7:	ac
ca	c	a	2	010	8:	ca
ab	a	b	-	-	-	-
aba	b	a	5	101	9:	aba
ad	a	d	0	000	10:	ad
da	d	a	3	011	11:	da
ab	a	b	-	-	-	-
aba	b	a	-	-	-	-
abac	a	c	9	1001	12:	abac
ca	c	a	-	-	-	-
cab	a	b	8	1000	13:	cab
ba	b	a	-	-	-	-
bae	a	e	6	0110	14:	bae
e	e	-	4	0100	-	-

Итак, мы получаем закодированное сообщение «0 1 0 2 5 0 3 9 8 6 4». Каждый символ исходного сообщения был закодирован группой из трех бит, сообщение содержало 16 символов, следовательно, длина сообщения составляла$3 \times 16 = 48$ бит.

Закодированное же сообщение так же сначала кодировалось трехбитными группами, а при появлении в словаре восьмого слова – четырехбитными, итого длина сообщения составила 7 * 3 + 4 * 4 = 37 бит, что на 11 бит короче исходного.

Пример 2: Выполнить алгоритм сжатия следующей последовательности «45, 55, 55, 151, 55, 55, 55».

- **Шаг 1.** Поместим в выходной поток сначала код очистки <256>, потом добавим к изначально пустой строке «45» и проверим, есть ли строка

- «45» в таблице. Поскольку мы при инициализации занесли в таблицу все строки из одного символа, то строка «45» – есть в таблице.

- **Шаг 2.** Читаем следующий символ «55» из входного потока и проверяем, есть ли строка «45, 55» в таблице. Такой строки в

таблице пока нет. Мы заносим в таблицу строку «45, 55» (с первым свободным кодом 258) и записываем в поток код <45> «45, 55» - нет, т.е. добавляем в таблицу <258> «45, 55». В поток: <45>;

- **Шаг 3.** «55, 55» – нет. В таблицу: <259> «55, 55». В поток: <55>;
- **Шаг 4.** «55, 151» – нет. В таблицу: <260> «55, 151». В поток: <55>;
- **Шаг 5.** «151, 55» – нет. В таблицу: <261> «151, 55». В поток: <151>;
- **Шаг 6.** «55, 55» – есть в таблице;
- **Шаг 7.** «55, 55, 55» - нет. В таблицу: «55, 55, 55» <262>. В поток: <259>;

Текущая строка	**Текущий символ**	**Следующий символ**	**Вывод**		**Словарь**	
			Код	**Биты**		
ASCII+					0-255	
			256	100000000	256	
45,55	45	55	**45**	101101	258	45,55
55,55	55	55	**55**	110111	259	55,55
55,151	55	151	**55**	110111	260	55,151
151,55	151	55	**151**	10010111	261	151,55
55,55	55	55	**-**	-	-	-
55,55,55	55	55	**259**	100000011	262	55,55,55

Последовательность кодов для данного примера, попадающих в выходной поток: <256>, <45>, <55>, <55>, <151>, <259>. (Получено: 44 бита, а в исходном сообщении было: 7 × 8 = 56 бит.)

При переполнении словаря, т.е. когда необходимо внести новую фразу в полностью заполненный словарь, из него удаляют либо наиболее редко используемую фразу, либо все фразы, отличающиеся от одиночного символа.

Практическая работа «Методы Лемпела-Зива» Задания (для всех вариантов):

1. Используя алгоритм закодировать сообщение «ПРОГРАММНЫЕ ПРОДУКТЫ ФИРМЫ MICROSOFT» используя алгоритм LZ77 (допустим, что размер окна составляет 20 символов, из них словаря – 12 символов, а буфера – 8).
2. Используя алгоритмы:
   - LZ77 (словарь – 12 байт, буфер – 4 байта);

- LZ78 (словарь – 16 фраз);
- LZW (словарь – ASCII+ и 16 фраз)

закодировать следующие сообщения и вычислить длины кодов в битах:

1) «AABCDAACCCCDBB»
2) «КИБЕРНЕТИКИ»
3) «СИНЯЯ СИНЕВА СИНИ»

Вопросы для самопроверки

1. Что такое канал связи? Раскройте сущность основных характеристик канала связи.
2. Какое назначение и цели эффективного кодирования?
3. Что такое эффективность кода? Как определить эффективность кода?
4. В чем состоит основная идея алгоритма Шеннона-Фано?
5. В чем состоит основная идея алгоритма Хаффмана?
6. В чем состоит основная идея алгоритмов Лемпела-Зива?
7. В чем отличие статистических и словарных методов кодирования?
8. В каких случаях используются изученные методы эффективного коди рования?

ЛИТЕРАТУРА

1 Информатика. Энциклопедический словарь для начинающих. Под ред. Д.А. Поспелова. – М.: Педагогика-Пресс, 1994. -240 с.

2 Шрайберг Я.Л., Гончаров М.В. Справочное руководство по основам информатики и вычислительной техники: – М.: Финансы и статистика, 1995. -212 с.

3 Терминологический словарь по основам информатики и вычислительной техники. Под ред. А.П. Ершова. – Москва.: Просвещение, 1991. – 159 с.

4 Большая Советская Энциклопедия. - М.: Советская энциклопедия. 1980. – 1600 с.

5 Информатика. Базовый курс: учебник для вузов / Симонович С.В. и др. – СПб.: Издательство Питер, 1999. – 640 с.

6 Информатика: учебное пособие / А.В. Могилев, Н.И. Пак, Е.К. Хеннер; Под ред. Е.К. Хеннера. – М., 1999. – 816 с.

7 Гуров И.П. Основы теории информации и передачи сигналов. – СПб.: BHV-Санкт-Петербург, 2000. – 97 с.: ил.

8 Информатика. Задачник-практикум в 2 т. / Под ред. И.Г. Семакина, Е.К. Хеннера: Том 1. – М.: Лаборатория Базовых Знаний, 2001. – 304 с.

9 Лидовский В.В.: Теория информации: учебное пособие. – М.: Компания Спутник+, 2004. – 111 с.

10 Практикум по информатике: учеб. пособие / А.В. Могилев, Н.И. Пак, Е.К. Хеннер; Под ред. Е.К.Хеннера. – М.: Издательский центр «Академия», 2001. – 608 с.

11 Цымбал В.П. Теория информации и кодирование: учебник. – 4-е изд. перераб. и доп. – Киев: Вища школа, 1992. – 263 с.

12 Хохлов Г.И. Основы теории информации: учебное пособие. - М.: Издательский центр ≪Академия≫, 2008. - 176с.

13 Яглом А.М., Яглом. И.М. Вероятность и информация - М.: Наука, 1973. - 512 с.

ПРИЛОЖЕНИЯ

Приложение 1

Ответы к практической работе «Целые числа в памяти компьютера»

№ варианта	Номера задачи		
	1	2	3
1	0000 0101 1010 1010	FA56	-2435
2	0000 0101 0011 1101	FAC3	-2134
3	0000 0111 1011 1111	F841	-2345
4	0000 0101 0001 1001	FAE7	-2304
5	0000 0111 1100 0000	F840	-2101
6	0000 0101 1010 1101	FA53	-1689
7	0000 0111 0010 1001	F8D7	-1985
8	0000 1001 0001 1011	F6E5	-2304
9	0000 0111 1100 0001	F83F	-1833
10	0000 0110 1001 1001	F967	-1453
11	0000 1000 0011 0101	F7CB	-1984
12	0000 1001 0000 0000	F700	-1305
13	0000 1001 0010 1001	F6D7	-1983
14	0000 1000 0101 0110	F7AA	-1341
15	0000 1001 1000 0011	F67D	-1450

Приложение 2

Ответы к практической работе «Вещественные числа в памяти компьютера»

№ варианта	**Номера заданий**	
	1	**2**
1	45D14000	-27.375
2	C5ED0000	26.28125
3	47B7A000	-29.625
4	C5DB0000	91.8125
5	488B6000	-26.28125
6	C5D14000	139.375
7	45DB0000	-91.8125
8	C6870000	27.375
9	45ED0000	-139.375
10	C88B6000	29.625
11	49A6E000	-33.75
12	C9A6E000	33.75
13	48E04000	-333.75
14	C7B7A000	333.75
15	46870000	224.25

Приложение 3

Количество информации об одном из N равновероятных событий $i=\log_2 N$			
N	I	N	I
1	0,00000	**33**	5,04439
2	1,00000	**34**	5,08746
3	1,58496	**35**	5,12928
4	2,00000	**36**	5,16993
5	2,32193	**37**	5,20945
6	2,58496	**38**	5,24793
7	2,80735	**39**	5,28540
8	3,00000	**40**	5,32193
9	3,16993	**41**	5,35755
10	3,32193	**42**	5,39232
11	3,45943	**43**	5,42626
12	3,58496	**44**	5,45943
13	3,70044	**45**	5,49185
14	3,80735	**46**	5,52356
15	3,90689	**47**	5,55459
16	4,00000	**48**	5,58496
17	4,08746	**49**	5,61471
18	4,16993	**50**	5,64386
19	4,24793	**51**	5,67243
20	4,32193	**52**	5,70044
21	4,39232	**53**	5,72792
22	4,45943	**54**	5,75489
23	4,52356	**55**	5,78136
24	4,58496	**56**	5,80735
25	4,64386	**57**	5,83289
26	4,70044	**58**	5,85798
27	4,75489	**59**	5,88264
28	4,80735	**60**	5,90689
29	4,85798	**61**	5,93074
30	4,90689	**62**	5,95420
31	4,95420	**63**	5,97728
32	5,00000	**64**	6,00000

Приложение 4

Таблица кода ASCII

Стандартная часть кода								
32		00100000	**64**		01000000	**96**	‘	01100000
33	!	00100001	**65**	A	01000001	**97**	a	01100001
34	«	00100010	**66**	B	01000010	**98**	b	01100010
35	#	00100011	**67**	C	01000011	**99**	c	01100011
36	$	00100100	**68**	D	01000100	**100**	d	01100100
37	%	00100101	**69**	E	01000101	**101**	e	01100101
38	&	00100110	**70**	F	01000110	**102**	f	01100110
39	‘	00100111	**71**	G	01000111	**103**	g	01100111
40	(	00101000	**72**	H	01001000	**104**	h	01101000
41	)	00101001	**73**	I	01001001	**105**	i	01101001
42	*	00101010	**74**	J	01001010	**106**	j	01101010
43	+	00101011	**75**	K	01001011	**107**	k	01101011
44	,	00101100	**76**	L	01001100	**108**	l	01101100
45	-	00101101	**77**	M	01001101	**109**	m	01101101
46	.	00101110	**78**	N	01001110	**110**	n	01101110
47	/	00101111	**79**	O	01001111	**111**	o	01101111
48	0	00110000	**80**	P	01010000	**112**	p	01110000
49	1	00110001	**81**	Q	01010001	**113**	q	01110001
50	2	00110010	**82**	R	01010010	**114**	r	01110010
51	3	00110011	**83**	S	01010011	**115**	s	01110011
52	4	00110100	**84**	T	01010100	**116**	t	01110100
53	5	00110101	**85**	U	01010101	**117**	u	01110101
54	6	00110110	**86**	V	01010110	**118**	v	01110110
55	7	00110111	**87**	W	01010111	**119**	w	01110111
56	8	00111000	**88**	X	01011000	**120**	x	01111000
57	9	00111001	**89**	Y	01011001	**121**	y	01111001
58	:	00111010	**90**	Z	01011010	**122**	z	01111010
59	;	00111011	**91**	[	01011011	**123**	{	01111011
60	<	00111100	**92**	\	01011100	**124**	|	01111100
61	=	00111101	**93**	]	01011101	**125**	}	01111101
62	>	00111110	**94**	^	01011110	**126**	~	01111110
63	?	00111111	**95**	_	01011111	**127**		01111111

Альтернативная часть кода								
128	А	10000000	**138**	Л	10001011	**148**	Ц	10010110
129	Б	10000001	**139**	М	10001100	**149**	Ч	10010111
130	В	10000010	**140**	Н	10001101	**150**	Ш	10011000
131	Г	10000011	**141**	О	10001110	**151**	Щ	10011001
132	Д	10000100	**142**	П	10001111	**152**	Ъ	10011010
133	Е	10000101	**143**	Р	10010000	**153**	Ы	10011011
134	Ж	10000110	**144**	С	10010001	**154**	Ь	10011100
135	З	10000111	**145**	Т	10010010	**157**	Э	10011101
136	И	10001000	**156**	У	10010011	**158**	Ю	10011110
155	Й	10001001	**146**	Ф	10010100	**159**	Я	10011111
137	К	10001010	**147**	Х	10010101			

Мукушев Базарбек Агзашулы
Турдина Айжан Базарбековна

Основы теории информации

Сдано в набор 24.02.2022 Подписано к печати 13.06.2022
Формат 60х84 1/6 Заказ № 2267
Объем усл. печ. л. Тираж 10 экз.

Типография Казахского агротехнического университета им.С.Сейфуллина, 2022
010011, г. Нур-Султан, пр.Жеңіс 62 а, тел. 39 39 17