

# Информация и информационные процессы

## Оглавление

1. Информация.....	2
2. Основные информационные процессы.....	5
3. Измерение информации: содержательный подход.....	9
4. Измерение информации: алфавитный подход.....	11
5. Кодирование информации.....	13
6. Представление (кодирование) чисел.....	17
7. Двоичное кодирование текстовой информации .....	20
8. Двоичное кодирование изображений.....	22
9. Двоичное кодирование звука. Представление видеоинформации.....	26

# 1. Информация

*Понятие информации. Виды информации. Роль информации в живой природе и в жизни людей. Язык как способ представления информации: естественные и формальные языки.*

Общепринятого определения **информации** не существует.

Слово «информация» происходит от латинского слова **informatio**, что в переводе означает сведение, разъяснение, ознакомление.

В наиболее общем случае под «информацией» понимаются сведения (данные), которые воспринимаются живым существом или устройством и сообщаются (получаются, передаются, преобразуются, сжимаются, разжимаются, теряются, находятся, регистрируются) с помощью знаков.

Понятие «информация» является базовым в курсе информатики, невозможно дать его определение через другие, более «простые» понятия. В геометрии, например, невозможно выразить содержание базовых понятий «точка», «луч», «плоскость» через более простые понятия. Содержание основных, базовых понятий в любой науке должно быть пояснено на примерах или выявлено путем их сопоставления с содержанием других понятий.

В случае с понятием «информация» проблема его определения еще более сложная, так как оно является общенаучным понятием. Данное понятие используется в различных науках (информатике, кибернетике, биологии, физике и др.), при этом в каждой науке понятие «информация» связано с различными системами понятий.

*Информация по-разному трактуется и изучается в различных областях знания.*

*Информация в физике.* В физике мерой беспорядка, хаоса для термодинамической системы является энтропия системы, тогда как информация (антиэнтропия) является мерой упорядоченности и сложности системы. По мере увеличения сложности системы величина энтропии уменьшается, и величина информации увеличивается. Процесс увеличения информации характерен для открытых, обменивающихся веществом и энергией с окружающей средой, саморазвивающихся систем живой природы (белковых молекул, организмов, популяций животных и так далее).

Таким образом, в физике информация рассматривается как антиэнтропия или энтропия с обратным знаком.

*Информация в биологии.* В биологии, которая изучает живую природу, понятие «информация» связывается с целесообразным поведением живых организмов. Такое поведение строится на основе получения и использования организмом информации об окружающей среде.

Понятие «информация» в биологии используется также в связи с исследованиями механизмов наследственности. Генетическая информация передается по наследству и хранится во всех клетках живых организмов. Гены представляют собой сложные молекулярные структуры, содержащие информацию о строении живых организмов. Последнее обстоятельство позволило проводить научные эксперименты по клонированию, то есть созданию точных копий организмов из одной клетки.

*Информация в кибернетике.* В кибернетике (науке об управлении) понятие «информация» связано с процессами управления в сложных системах (живых организмах или технических устройствах). Жизнедеятельность любого организма или нормальное функционирование технического устройства зависит от процессов управления, благодаря которым поддерживаются в необходимых пределах значения их параметров. Процессы управления включают в себя получение, хранение, преобразование и передачу информации.

*Социально значимые свойства информации.* Человек - существо социальное, для общения с другими людьми он должен обмениваться с ними информацией, причем обмен информацией всегда производится на определенном языке — русском, английском и так далее. Участники дискуссии должны владеть тем языком, на котором ведется общение, тогда информация будет понятной всем участникам обмена информацией.

Информация должна быть полезной, тогда дискуссия приобретает практическую ценность. Бесплезная информация создает информационный шум, который затрудняет восприятие полезной информации. Примерами передачи и получения бесплезной информации могут служить некоторые конференции и чаты в Интернете.

Широко известен термин «средства массовой информации» (газеты, радио, телевидение), которые доводят информацию до каждого члена общества. Такая информация должна быть достоверной и актуальной. Недостоверная информация вводит членов общества в заблуждение и может быть причиной возникновения социальных потрясений. Неактуальная информация бесплезна и поэтому

никто, кроме историков, не читает прошлогодних газет.

Для того чтобы человек мог правильно ориентироваться в окружающем мире, информация должна быть полной и точной. Задача получения полной и точной информации стоит перед наукой. Овладение научными знаниями в процессе обучения позволяют человеку получить полную и точную информацию о природе, обществе и технике.

Люди, обмениваясь между собой информацией, постоянно должны задавать себе вопросы: понятна, актуальна и полезна ли она для окружающих, достоверны ли полученные сведения. Это позволит лучше понять друг друга, найти правильное решение в любой ситуации. Вы постоянно анализируете свойства информации, часто не придавая этому значения. В повседневной жизни от свойств информации часто зависят жизнь и здоровье людей, экономическое развитие общества.



В наши дни человечество накопило огромное количество информации! Подсчитано, что общая сумма человеческих знаний до недавнего времени удваивалась каждые 50 лет. Сейчас объем информации удваивается через каждые два года. Представьте себе колоссальную библиотеку, содержащую эти сведения! От умения человека правильно воспринимать и обрабатывать информацию зависит во многом его способность к познанию окружающего мира.



Мир вокруг нас полон всевозможных образов, звуков, запахов, и всю эту информацию доносят до сознания человека его органы чувств: зрение, слух, обоняние, вкус и осязание. С их помощью человек формирует свое первое представление о любом предмете, живом существе, произведении искусства, явлении и пр.

- Глазами люди воспринимают зрительную информацию;
- Органы слуха доставляют информацию в виде звуков;
- Органы обоняния позволяют ощущать запахи;
- Органы вкуса несут информацию о вкусе еды;
- Органы осязания позволяют получить тактильную информацию.

Виды информации, которые человек получает с помощью органов чувств, называют органолептической информацией. Практически 90% информации человек получает при помощи органов зрения, примерно 9% — посредством органов слуха и только 1% — при помощи остальных органов чувств.

Воспринимая информацию с помощью органов чувств, человек стремится зафиксировать ее так, чтобы она стала понятной и другим, представляя ее в той или иной форме.

Музыкальную тему композитор может наиграть на пианино, а затем записать с помощью нот. Образы, навеянные все той же мелодией, поэт может воплотить в виде стихотворения, хореограф выразить танцем, а художник — в картине.

Человек выражает свои мысли в виде предложений, составленных из слов. Слова, в свою очередь, состоят из букв. Это — алфавитное представление информации.

Форма представления одной и той же информации может быть различной. Это зависит от цели, которую вы перед собой поставили.

### *Язык как знаковая система*

Для обмена информацией с другими людьми человек использует естественные языки (русский, английский, китайский и др.), то есть информация представляется с помощью естественных языков. В основе языка лежит алфавит, то есть набор символов (знаков), которые человек различает по их начертанию. В основе русского языка лежит кириллица, содержащая 33 знака, английский язык использует латиницу (26 знаков), китайский язык использует алфавит из десятков тысяч знаков (иероглифов).

Язык – это определенная система символов и правил представления информации.

Алфавит языка – множество используемых символов.

Последовательности символов алфавита в соответствии с правилами грамматики образуют основные объекты языка — слова. Правила, согласно которым образуются предложения из слов данного языка, называются синтаксисом. Необходимо отметить, что в естественных языках грамматика и синтаксис языка формулируются с помощью большого количества правил, из которых существуют исключения, так как такие правила складывались исторически.

Синтаксис – правила записи языковых конструкций (текста на языке).

Семантика – смысловая сторона языковых конструкций.

Наряду с естественными языками были разработаны формальные языки (системы счисления, язык алгебры, языки программирования и др.). Основное отличие формальных языков от естественных состоит в наличии строгих правил грамматики и синтаксиса.

Например, системы счисления можно рассматривать как формальные языки, имеющие алфавит (цифры) и позволяющие не только именовать и записывать объекты (числа), но и выполнять над ними арифметические операции по строго определенным правилам.

Естественные языки – это исторически сложившиеся языки национальной речи (русский, английский, китайский и др.).

Формальные языки – это искусственные языки для профессионального применения (язык математики, язык химических формул, нотная запись, язык программирования).

Некоторые языки используют в качестве знаков не буквы и цифры, а другие символы, например химические формулы, ноты, изображения элементов электрических или логических схем, дорожные знаки, точки и тире (код азбуки Морзе) и др.

Знаки могут иметь различную физическую природу. Например, для представления информации с использованием языка в письменной форме используются знаки, которые являются изображениями на бумаге или других носителях, в устной речи в качестве знаков языка используются различные звуки (фонемы), а при обработке текста на компьютере знаки представляются в форме последовательностей электрических импульсов (компьютерных кодов).

## 2. Основные информационные процессы.

*Основные информационные процессы: хранение, передача и обработка информации.*

Информационные процессы - процессы сбора, обработки, накопления, хранения, поиска и распространения информации.

Слово процесс обозначает некоторое событие, происходящее во времени: учебный процесс, процесс роста, процесс горения... Всякий процесс связан с какими-то действиями, выполняемыми человеком, силами природы, техническими устройствами, а также вследствие их взаимодействия.

У каждого процесса есть объект воздействия: ученик, растение, горючее... Очевидно, что в информационных процессах объектом воздействия является информация.

Информационные технологии (ИТ, от англ. *information technology*, IT) — широкий класс дисциплин и областей деятельности, относящихся к технологиям управления и обработки данных вычислительной техникой. Обычно под информационными технологиями понимают компьютерные технологии. В частности, ИТ имеют дело с использованием компьютеров и программного обеспечения для хранения, преобразования, защиты, обработки, передачи и получения информации. Специалистов по компьютерной технике и программированию часто называют ИТ-специалистами.

Слово «технология» происходит от греческого «*techne*», что означает «искусство», «мастерство», «умение». Т.е. технология — действия для достижения поставленной цели. В ходе реализации технологии материального производства получается какой-либо материальный продукт. Целью информационных технологий является информационный продукт.

### **Сбор информации.**

Приходится признать, что органы чувств — наш главный инструмент познания мира — не самые совершенные приспособления. Не всегда они точны и не всякую информацию способны воспринять. Не случайно о грубых, приблизительных вычислениях говорят: «на глаз». Если бы не было специальных приборов, то вряд ли человечеству удалось бы проникнуть в тайны живой клетки или отправить к Марсу и Венере космические зонды.

Вся деятельность человека связана с различными действиями с информацией, и помогают ему в этом разнообразные технические устройства.

Любое научное знание начинается тогда, когда мы можем оценить полученную информацию, сравнить ее, а значит измерить. Поэтому для получения недоступной обычными органами чувств информации широко используются специальные технические устройства.

Одно из древнейших сооружений, используемое для получения астрономической информации, находится в Англии недалеко от города Солсбери. Это Стоунхендж — «висячие камни». Он был построен примерно во II веке до н. э. Стоунхендж состоит из поставленных вертикально каменных столбов, расположенных концентрическими кольцами. На вертикальных камнях лежат горизонтальные перекладины, своего рода арки. В 1963 году с помощью новейших методов исследования было установлено, что каменные арки дают направления на крайние положения Солнца и Луны, а 56 белых лунок помогают предсказать время Солнечного и Лунного затмений.



Одно из древнейших устройств — весы. С их помощью люди получают информацию о массе объекта. Еще один наш старый знакомый — термометр — служит для измерения температуры окружающей его среды.

### **Обработка информации.**

Приобретая жизненный опыт, наблюдая мир вокруг себя, иначе говоря — накапливая все больше и больше информации, человек учится делать выводы. В древности люди говорили, что человек познает с помощью органов чувств и осмысливает познанное разумом.

Один раз дотронувшись до горячего чайника или утюга мы запоминаем это на всю жизнь. Каждый раз, случайно коснувшись горячей поверхности, мы отдергиваем руку, потому что у всех нас в детстве был свой «горячий чайник». Если проанализировать, почему так происходит, то можно сделать вывод о преобразовании (обработке) информации. Прикоснувшись к горячей поверхности, мы получили информацию при помощи органов осязания. Нервная система передала ее в мозг, где на основе

имеющегося опыта был сделан вывод об опасности. Сигнал от мозга был послан в мышцы рук, которые мгновенно сократились. Аналогичные процессы обработки информации происходят и в тот момент, когда при первых же аккордах знакомой мелодии сразу улучшается настроение или появляются слезы.

Все это примеры неосознанной обработки информации, которая ведется как бы «помимо нас», неосознанно.

Можно привести много примеров осознанной обработки информации. В этом случае человек создает новую информацию, опираясь на поступающие сведения — так называемую входную информацию — и на запас имеющихся у него знаний и опыта.

Например, на уроках химии школьник изучает правила и законы (приобретает определенные знания и навыки). Когда учитель предлагает очередную задачу (входная информация), ученик обдумывает последовательность решения, вспоминая, какие из изученных правил ему необходимо применить. Наконец, он находит ответ. Эта новая информация, созданная учеником в результате обработки входной информации, называется выходной.

Таким образом, выходная информация всегда является результатом мыслительной деятельности человека по обработке входной информации. Можно сказать, что человек постоянно занимается обработкой входной информации, преобразуя ее в выходную.

Входная информация – это информация, которую получает человек или устройства.

Выходная информация – это информация, которая получается после обработки человеком или устройством.



Обработка информации — процесс планомерного изменения содержания или формы представления информации.

Обработка информации производится в соответствии с определенными правилами некоторым субъектом или объектом (например, человек или автоматическое устройство). Будет называть его *исполнителем обработки информации*.

Входная информация, представленная в символьной форме (знаки, буквы, цифры, сигналы), называются *входными данными*. В результате обработки исполнителем получают *выходные данные*.

### **Передача информации.**

Развитие человечества не было бы возможно без обмена информацией. С давних времен люди из поколения в поколение передавали свои знания, извещали об опасности или передавали важную и срочную информацию, обменивались сведениями. Например, в Петербурге в начале XIX века была весьма развита пожарная служба. В нескольких частях города были построены высокие каланчи, с которых обзоревались окрестности. Если случался пожар, то на башне днем поднимался разноцветный флаг (с той или иной геометрической фигурой), а ночью зажигалось несколько фонарей, число и расположение которых означало часть города, где произошел пожар, а также степень его сложности.



В 1792 году во Франции Клод Шапп создал систему передачи визуальной информации, которая получила название «Оптический телеграф». В простейшем виде это была цепь типовых строений, с расположенными на кровле шестами с подвижными поперечинами, которая создавалась в пределах видимости одно от другого. Шесты с подвижными поперечинами — семафоры — управлялись при помощи тросов специальными операторами изнутри строений. Шапп создал специальную таблицу кодов, где каждой букве алфавита соответствовала определенная фигура, образуемая Семафором, в зависимости от положений поперечных брусьев относительно опорного шеста. Система Шаппа позволяла передавать сообщения на скорости два слова в минуту и быстро распространилась в Европе. В Швеции цепь станций оптического телеграфа действовала до 1880 года.

В любом процессе передачи или обмене информацией существует ее *источник* и *получатель*, а сама информация передается по *каналу связи* с помощью *сигналов*: механических, тепловых, электрических и др. В обычной жизни для человека любой звук, свет являются сигналами, несущими смысловую нагрузку. Например, сирена — это звуковой сигнал тревоги; звонок телефона — сигнал, чтобы взять трубку; красный свет светофора — сигнал, запрещающий переход дороги.



В качестве источника информации может выступать живое существо или техническое устройство. От него информация попадает в кодирующее устройство, которое предназначено для преобразования исходного сообщения в форму, удобную для передачи. С такими устройствами вы встречаетесь постоянно: микрофон телефона, лист бумаги и т. д. По каналу связи информация попадает в декодирующее устройство получателя, которое преобразует закодированное сообщение в форму, понятную получателю. Одни из самых сложных декодирующих устройств — человеческие ухо и глаз.



В процессе передачи информация может утрачиваться, искажаться. Это происходит из-за различных помех, как на канале связи, так и при кодировании и декодировании информации. С такими ситуациями вы встречаетесь достаточно часто: искажение звука в телефоне, помехи при телевизионной передаче, ошибки телеграфа, неполнота переданной информации, неверно выраженная мысль, ошибка в расчетах. Вопросами, связанными с методами кодирования и декодирования информации, занимается специальная наука — криптография.

При передаче информации важную роль играет форма представления информации. Она может быть понятна источнику информации, но недоступна для понимания получателя. Люди специально договариваются о языке, с помощью которого будет представлена информация для более надежного ее сохранения.

### **Хранение информации.**

Человеческий разум является самым совершенным инструментом познания окружающего мира. А память человека — великолепным устройством для хранения полученной информации.

Чтобы информация стала достоянием многих людей, необходимо иметь возможность ее хранить не только в памяти человека. В процессе развития человечества существовали разные способы хранения информации, которые совершенствовались с течением времени: узелки на веревках, зарубки на палках, берестяные грамоты, письма на папирусе, бумаге.

Наконец, был изобретен типографский станок, и появились книги. Поиск надежных и доступных способов хранения информации идет и по сей день.

Сегодня мы используем для хранения информации самые различные материалы: бумагу, фото- и киноплёнку, магнитную аудио- и видеоленту, магнитные и оптические диски. Все это — носители информации.

Носитель информации — материальный объект, предназначенный для хранения и передачи информации.

Материальная природа носителей информации может быть различной: молекулы ДНК, которые

хранят генетическую информацию; бумага, на которой хранятся тексты и изображения; магнитная лента, на которой хранится звуковая информация.

### ***Поиск информации.***

Просто сохранить информацию недостаточно. Нужно уметь ее пользоваться. А для того чтобы воспользоваться нужной информацией в нужный момент необходимо уметь ее быстро найти.

Поиск информации — это извлечение хранимой информации. Существуют ручной и автоматизированный методы поиска информации в хранилищах.

Методы поиска информации:

- непосредственное наблюдение;
- общение со специалистами по интересующему вас вопросу;
- чтение соответствующей литературы;
- просмотр теле-, видеопрограмм;
- прослушивание радиопередач и аудиокассет;
- работа в библиотеках, архивах;
- запрос к информационным системам, базам и банкам компьютерных данных;
- другие методы.

Для того чтобы собрать наиболее полную информацию и повысить вероятность принятия правильного решения, необходимо использовать разнообразные методы поиска информации.

В процессе поиска информации вам встретится как самая полезная, так и бесполезная, как достоверная, так и ложная, объективная и субъективная информация, но чтобы не утонуть в море информации, учитесь отбирать только полезную для решения стоящей перед вами задачи. Не уподобляйте свою голову мусорному ящику, куда сваливают все без разбора.

Для ускорения процесса получения наиболее полной информации по вопросу стали составлять каталоги (алфавитный, предметный и др.).

### ***Защита информации.***

В жизни человека информация играет очень важную роль. От нее зависит принятие решений, влияющих на развитие общества.

Для предотвращения потери информации разрабатываются различные механизмы ее защиты, которые используются на всех этапах работы с ней.

Для защиты информации используют различные способы защиты:

- контроль доступа;
- разграничение доступа;
- дублирование каналов связи;
- криптографическое преобразование информации с помощью шифров.



### 3. Измерение информации: содержательный подход.

*Измерение информации: содержательный подход. Единицы измерения информации.*

Вопрос: «Как измерить информацию?» очень непростой. Ответ на него зависит от того, что понимать под информацией. Но поскольку определять информацию можно по-разному, то и способы измерения тоже могут быть разными.

**Содержательный подход к измерению информации.**

Для человека информация — это знания человека. Рассмотрим вопрос с этой точки зрения.

Получение новой информации приводит к расширению знаний. Если некоторое сообщение приводит к уменьшению неопределенности нашего знания, то можно говорить, что такое сообщение содержит информацию.

Отсюда следует вывод, что сообщение информативно (т.е. содержит ненулевую информацию), если оно пополняет знания человека. Например, прогноз погоды на завтра — информативное сообщение, а сообщение о вчерашней погоде неинформативно, т.к. нам это уже известно.

Нетрудно понять, что информативность одного и того же сообщения может быть разной для разных людей. Например: « $2 \times 2 = 4$ » информативно для первоклассника, изучающего таблицу умножения, и неинформативно для старшеклассника.

Но для того чтобы сообщение было информативно оно должно еще быть понятным. Быть понятным, значит быть логически связанным с предыдущими знаниями человека. Определение «значение определенного интеграла равно разности значений первообразной подынтегральной функции на верхнем и на нижнем пределах», скорее всего, не пополнит знания и старшеклассника, т.к. оно ему не понятно. Для того, чтобы понять данное определение, нужно закончить изучение элементарной математики и знать начала высшей.

Получение всяких знаний должно идти от простого к сложному. И тогда каждое новое сообщение будет в то же время понятным, а значит, будет нести информацию для человека.

*Сообщение несет информацию для человека, если содержащиеся в нем сведения являются для него новыми и понятными.*

Очевидно, различать лишь две ситуации: «нет информации» — «есть информация» для измерения информации недостаточно. Нужна единица измерения, тогда мы сможем определять, в каком сообщении информации больше, в каком — меньше.

Единица измерения информации была определена в науке, которая называется теорией информации. Эта единица носит название «бит». Ее определение звучит так:

*Сообщение, уменьшающее неопределенность знаний в два раза, несет 1 бит информации.*

Например, после сдачи зачета или выполнения контрольной работы ученик мучается неопределенностью, он не знает, какую оценку получил. Наконец, учитель объявляет результаты, и он получает одно из двух информационных сообщений: «зачет» или «незачет», а после контрольной работы одно из четырех информационных сообщений: «2», «3», «4» или «5».

Информационное сообщение об оценке за зачет приводит к уменьшению неопределенности знания в два раза, так как получено одно из двух возможных информационных сообщений. Информационное сообщение об оценке за контрольную работу приводит к уменьшению неопределенности знания в четыре раза, так как получено одно из четырех возможных информационных сообщений.

*Неопределенность знаний о некотором событии — это количество возможных результатов события.*

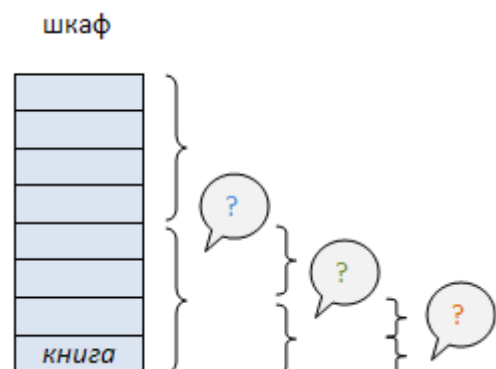
Рассмотрим еще один пример.

На книжном стеллаже восемь полок. Книга может быть поставлена на любую из них. Сколько информации содержит сообщение о том, где находится книга?

Применим метод половинного деления. Зададим несколько вопросов уменьшающих неопределенность знаний в два раза.

Задаем вопросы:

- Книга лежит выше четвертой полки?



- Нет.

- Книга лежит ниже третьей полки?

- Да .

- Книга — на второй полке?

- Нет.

- Ну теперь все ясно! Книга лежит на первой полке!

Каждый ответ уменьшал неопределенность в два раза.

Всего было задано три вопроса. Значит набрано 3 бита информации. И если бы сразу было сказано, что книга лежит на первой полке, то этим сообщением были бы переданы те же 3 бита информации.

Если обозначить возможное количество событий, или, другими словами, неопределенность знаний  $N$ , а буквой  $I$  количество информации в сообщении о том, что произошло одно из  $N$  событий, то можно записать формулу:

$$2^I = N$$

*Количество информации, содержащееся в сообщении о том, что произошло одно из  $N$  равновероятных событий, определяется из решения показательного уравнения:  $2^I = N$ .*

## 4. Измерение информации: алфавитный подход

*Измерение информации: алфавитный подход к измерению информации. Единицы измерения информации.*

А теперь познакомимся с другим способом измерения информации. Этот способ не связывает количество информации с содержанием сообщения, и называется он алфавитным подходом.

При алфавитном подходе к определению количества информации отвлекаются от содержания информации и рассматривают информационное сообщение как последовательность знаков определенной знаковой системы.

Проще всего разобраться в этом на примере текста, написанного на каком-нибудь языке. Для нас удобнее, чтобы это был русский язык.

Все множество используемых в языке символов будем традиционно называть алфавитом. Обычно под алфавитом понимают только буквы, но поскольку в тексте могут встречаться знаки препинания, цифры, скобки, то мы их тоже включим в алфавит. В алфавит также следует включить и пробел, т.е. пропуск между словами.

Полное количество символов алфавита принято называть мощностью алфавита. Будем обозначать эту величину буквой  $N$ . Например, мощность алфавита из русских букв и отмеченных дополнительных символов равна 54.

Представьте себе, что текст к вам поступает последовательно, по одному знаку, словно бумажная ленточка, выползающая из телеграфного аппарата. Предположим, что каждый появляющийся на ленте символ с одинаковой вероятностью может быть любым символом алфавита. В действительности это не совсем так, но для упрощения примем такое предположение.

В каждой очередной позиции текста может появиться любой из  $N$  символов. Тогда, согласно известной нам формуле, каждый такой символ несет  $I$  бит информации, которое можно определить из решения уравнения:  $2^I = 54$ . Получаем:  $I = 5.755$  бит.

Вот сколько информации несет один символ в русском тексте! А теперь для того, чтобы найти количество информации во всем тексте, нужно посчитать число символов в нем и умножить на  $I$ .

Посчитаем количество информации на одной странице книги. Пусть страница содержит 50 строк. В каждой строке — 60 символов. Значит, на странице умещается  $50 \times 60 = 3000$  знаков. Тогда объем информации будет равен:  $5,755 \times 3000 = 17265$  бит.

*При алфавитном подходе к измерению информации количество информации зависит не от содержания, а от размера текста и мощности алфавита.*

При использовании двоичной системы (алфавит состоит из двух знаков: 0 и 1) каждый двоичный знак несет 1 бит информации. Интересно, что сама единица измерения информации «бит» получила свое название от английского сочетания «**binary digit**» - «двоичная цифра».

Применение алфавитного подхода удобно прежде всего при использовании технических средств работы с информацией. В этом случае теряют смысл понятия «новые — старые», «понятные — непонятные» сведения. Алфавитный подход является объективным способом измерения информации в отличие от субъективного содержательного подхода.

Удобнее всего измерять информацию, когда размер алфавита  $N$  равен целой степени двойки. Например, если  $N=16$ , то каждый символ несет 4 бита информации потому, что  $2^4 = 16$ . А если  $N=32$ , то один символ «весит» 5 бит.

Ограничения на максимальный размер алфавита теоретически не существует. Однако есть алфавит, который можно назвать достаточным. С ним мы скоро встретимся при работе с компьютером. Это алфавит мощностью 256 символов. В алфавит такого размера можно поместить все практически необходимые символы: латинские и русские буквы, цифры, знаки арифметических операций, всевозможные скобки, знаки препинания....

Поскольку  $256 = 2^8$ , то один символ этого алфавита «весит» 8 бит. Причем 8 бит информации — это настолько характерная величина, что ей даже присвоили свое название — *байт*.

**1 байт = 8 бит.**

Сегодня очень многие люди для подготовки писем, документов, статей, книг и пр. используют компьютерные текстовые редакторы. Компьютерные редакторы, в основном, работают с алфавитом размером 256 символов.

В этом случае легко подсчитать объем информации в тексте. Если 1 символ алфавита несет 1 байт информации, то надо просто сосчитать количество символов; полученное число даст информационный объем текста в байтах.

Пусть небольшая книжка, сделанная с помощью компьютера, содержит 150 страниц; на каждой странице — 40 строк, в каждой строке — 60 символов. Значит страница содержит  $40 \times 60 = 2400$  байт информации. Объем всей информации в книге:  $2400 \times 150 = 360\,000$  байт.

В любой системе единиц измерения существуют основные единицы и производные от них.

Для измерения больших объемов информации используются следующие производные от байта единицы:

1 килобайт = 1Кб =  $2^{10}$  байт = 1024 байта.

1 мегабайт = 1Мб =  $2^{10}$  Кб = 1024 Кб.

1 гигабайт = 1Гб =  $2^{10}$  Мб = 1024 Мб.

Название	Условное обозначение	Соотношение с другими единицами
Килобит	Кбит	1 Кбит = 1024 бит = $2^{10}$ бит $\approx$ 1000 бит
Мегабит	Мбит	1 Мбит = 1024 Кбит = $2^{20}$ бит $\approx$ 1 000 000 бит
Гигабит	Гбит	1 Гбит = 1024 Мбит = $2^{30}$ бит $\approx$ 1 000 000 000 бит
Килобайт	Кбайт (Кб)	1 Кбайт = 1024 байт = $2^{10}$ байт $\approx$ 1000 байт
Мегабайт	Мбайт (Мб)	1 Мбайт = 1024 Кбайт = $2^{20}$ байт $\approx$ 1 000 000 байт
Гигабайт	Гбайт (Гб)	1 Гбайт = 1024 Мбайт = $2^{30}$ байт $\approx$ 1 000 000 000 байт

Прием-передача информации могут происходить с разной скоростью. Количество информации, передаваемое за единицу времени, есть скорость передачи информации или скорость информационного потока.

Очевидно, эта скорость выражается в таких единицах, как бит в секунду (бит/с), байт в секунду (байт/с), килобайт в секунду (Кбайт/с) и т.д.

## 5. Кодирование информации

**Кодирование и декодирование информации. Равномерный и неравномерный код.**

Для обмена информацией с другими людьми человек использует естественные языки. Наряду с естественными языками были разработаны формальные языки для профессионального применения их в какой-либо сфере. Представление информации с помощью какого-либо языка часто называют кодированием.

**Код** — набор символов (условных обозначений) для представления информации. **Код** — система условных знаков (символов) для передачи, обработки и хранения информации(со общения).

**Кодирование** — процесс представления информации (сообщения) в виде кода.

Все множество символов, используемых для кодирования, называется *алфавитом кодирования*. Например, в памяти компьютера любая информация кодируется с помощью двоичного алфавита, содержащего всего два символа: 0 и 1.

**Декодирование** — процесс обратного преобразования кода к форме исходной символической системы, т.е. получение исходного сообщения. Например: перевод с азбуки Морзе в письменный текст на русском языке.

В более широком смысле декодирование — это процесс восстановления содержания закодированного сообщения. При таком подходе процесс записи текста с помощью русского алфавита можно рассматривать в качестве кодирования, а его чтение — это декодирование.

Вы встречаетесь с кодированием информации при переходе дороги по сигналам светофора. Код определяет цвета светофора — красный, желтый, зеленый, основу естественного языка, на котором общаются люди, тоже положен код. Только в этом случае он называется алфавитом. При разговоре этот код передается звуками, при письме — буквами. Одну и ту же информацию можно представить с помощью различных кодов. Например, запись разговора можно зафиксировать посредством русских букв или специальных стенографических значков.

Пример sms-сообщения на телефонах с русским языком и без него:

«Здравствуй, Саша!»

«Zdravstvui, Sacha!»

Приведенные примеры иллюстрируют следующее важное правило: *для кодирования одной и той же информации могут быть использованы разные способы; их выбор зависит от ряда обстоятельств: цели кодирования, условий, имеющихся средств*. Если надо записать текст в темпе речи — используем стенографию; если надо передать текст за границу — используем английский алфавит; если надо представить текст в виде, понятном для грамотного русского человека, — записываем его по правилам грамматики русского языка.

Еще одно важное обстоятельство: *выбор способа кодирования информации может быть связан с предполагаемым способом ее обработки*. Покажем это на примере представления чисел — количественной информации. Используя русский алфавит, можно записать число "тридцать пять". Используя же алфавит арабской десятичной системы счисления, пишем «35». Второй способ не только короче первого, но и удобнее для выполнения вычислений. Какая запись удобнее для выполнения расчетов: "тридцать пять умножить на сто двадцать семь" или "35 x 127"? Очевидно — вторая.

Однако если важно сохранить число без искажения, то его лучше записать в текстовой форме. Например, в денежных документах часто сумму записывают в текстовой форме: "триста семьдесят пять руб." вместо "375 руб.". Во втором случае искажение одной цифры изменит все значение. При использовании текстовой формы даже грамматические ошибки могут не изменить смысла.

В некоторых случаях возникает потребность засекречивания текста сообщения или документа, для того чтобы его не смогли прочитать те, кому не положено. Это называется защитой от несанкционированного доступа. В таком случае секретный текст шифруется. В давние времена шифрование называлось тайнописью. Шифрование представляет собой процесс превращения открытого текста в зашифрованный, а дешифрование — процесс обратного преобразования, при котором восстанавливается исходный текст. Шифрование — это тоже кодирование, но с засекреченным методом, известным только источнику и адресату. Методами шифрования занимается наука под названием криптография.

Люди всегда искали способы быстрого обмена сообщениями. Для этого посылали гонцов, использовали почтовых голубей. У народов существовали различные способы оповещения о

надвигающейся опасности: барабанный бой, дым костров, флаги и т. д. Огни, зажигаемые на возвышенных участках местности, или же дым от костров должен оповестить о приближении врагов либо грядущем стихийном бедствии. Некоторые племена научились передавать определенные сообщения, манипулируя отраженным солнечным светом при помощи системы зеркал.

В 1792 году во Франции Клод Шапп создал систему передачи визуальной информации, которая получила название «Оптический телеграф». В простейшем виде это была цепь типовых строений, с расположенными на кровле шестами с подвижными поперечинами, которая создавалась в пределах видимости одно от другого. Шесты с подвижными поперечинами — семафоры — управлялись при помощи тросов специальными операторами изнутри строений. Шапп создал специальную таблицу кодов, где каждой букве алфавита соответствовала определенная фигура, образуемая Семафором, в зависимости от положений поперечных брусьев относительно опорного шеста. Система Шаппа позволяла передавать сообщения на скорости два слова в минуту и быстро распространилась в Европе. В Швеции цепь станций оптического телеграфа действовала до 1880 года.

По мере развития техники появлялись разные способы кодирования информации. Первым техническим средством передачи информации на расстояние стал телеграф, изобретенный в 1837 году американцем Сэмюэлем Морзе. Телеграфное сообщение — это последовательность электрических сигналов, передаваемая от одного телеграфного аппарата по проводам к другому телеграфному аппарату. Изобретатель Сэмюэль Морзе изобрел удивительный код (Азбука Морзе, код Морзе, «Морзянка»), который служит человечеству до сих пор. Информация кодируется тремя «буквами»: длинный сигнал (тире), короткий сигнал (точка) и отсутствие сигнала (пауза) для разделения букв. Таким образом, кодирование сводится к использованию набора символов, расположенных в строго определенном порядке.

Самым знаменитым телеграфным сообщением является сигнал бедствия "SOS" (Save Our Souls - спасите наши души). Вот как он выглядит: «•••---•••»

Русский символ	Латинский символ	Код Морзе
А	A	• -
Б	B	- •••
В	W	• - -
Г	G	- - •
Д	D	- ••
Е	E	•
Ж	V	••• -
З	Z	- - ••
И	I	••
Й	J	• - - -
К	K	- • -
Л	L	• - ••
М	M	- -
Н	N	- •
О	O	- - -
П	P	• - - •
Р	R	• - •
С	S	•••
Т	T	-
У	U	•• -
Ф	F	•• - •
Х	H	••••
Ц	C	- •••

Ч	ö	----•
Ш	ch	-----
Щ	Q	---•-
Ь	X	-••-
Ы	Y	-•---
Э	é	••-••
Ю	ü	••---
Я	ä	•-••-
1		•- - - -
2		••- - -
3		•••- -
4		••••-
5		•••••
6		•••••
7		- - •••
8		- - - ••
9		- - - - •
0		- - - - -
Точка		••••••
Запятая		•-••••-
/		-•••••
?		••-•••
!		- - ••• - -
@		•- - - - •

Характерной особенностью азбуки Морзе является переменная длина кода разных букв, поэтому код Морзе называют *неравномерным кодом*. Буквы, которые встречаются в тексте чаще, имеют более короткий код, чем редкие буквы. Это сделано для того, чтобы сократить длину всего сообщения. Но из-за переменной длины кода букв возникает проблема отделения букв друг от друга в тексте. Поэтому для разделения приходится использовать паузу (пропуск). Следовательно, телеграфный алфавит Морзе является троичным, т.к. в нем используются три знака: точка, тире, пропуск.

7 мая 1895 года российский ученый Александр Степанович Попов на заседании Русского Физико-Химического Общества продемонстрировал прибор, названный им "грозоотметчик", который был предназначен для регистрации электромагнитных волн. Этот прибор считается первым в мире аппаратом беспроводной телеграфии, радиоприемником. В 1897 году при помощи аппаратов беспроводной телеграфии Попов осуществил прием и передачу сообщений между берегом и военным судном. В 1899 году Попов сконструировал модернизированный вариант приемника электромагнитных волн, где прием сигналов (азбукой Морзе) осуществлялся на головные телефоны оператора. В 1900 году благодаря радиостанциям, построенным на острове Гогланд и на российской военно-морской базе в Котке под руководством Попова, были успешно осуществлены аварийно-спасательные работы на борту военного корабля "Генерал-адмирал Апраксин", севшего на мель у острова Гогланд. В результате обмена сообщениями, переданным методом беспроводной телеграфии, экипажу российского ледокола Ермак была своевременно и точно передана информация о финских рыбаках, находящихся на оторванной льдине в Финском заливе.

*Равномерный телеграфный код* был изобретен французом Жаном Морисом Бодо в конце XIX века. В нем использовалось всего два разных вида сигналов. Не важно, как их назвать: точка и тире, плюс и минус, ноль и единица. Это два отличающихся друг от друга электрических сигнала. Длина кода всех символов одинаковая и равна пяти. В таком случае не возникает проблемы отделения букв друг от друга: каждая пятерка сигналов — это знак текста. Поэтому пропуск не нужен.

Код Бодо — это первый в истории техники способ двоичного кодирования, информации. Благодаря

этой идее удалось создать буквопечатающий телеграфный аппарат, имеющий вид пишущей машинки. Нажатие на клавишу с определенной буквой вырабатывает соответствующий пятиимпульсный сигнал, который передается по линии связи.

В современных компьютерах для кодирования текста также применяется равномерный двоичный код.

Вся информация, которую обрабатывает компьютер должна быть представлена двоичным кодом с помощью двух цифр 0 и 1. Эти два символа принято называть двоичными цифрами или битами. С помощью двух цифр 0 и 1 можно закодировать любое сообщение. Это явилось причиной того, что в компьютере обязательно должно быть организовано два важных процесса: кодирование и декодирование.

Кодирование – преобразование входной информации в форму, воспринимаемую компьютером, т.е. двоичный код.

Декодирование – преобразование данных из двоичного кода в форму, понятную человеку.

С точки зрения технической реализации использование двоичной системы счисления для кодирования информации оказалось намного более простым, чем применение других способов. Действительно, удобно кодировать информацию в виде последовательности нулей и единиц, если представить эти значения как два возможных устойчивых состояния электронного элемента:

0 – отсутствие электрического сигнала;

1 – наличие электрического сигнала.

Эти состояния легко различать. Недостаток двоичного кодирования – длинные коды. Но в технике легче иметь дело с большим количеством простых элементов, чем с небольшим числом сложных.

Вам приходится постоянно сталкиваться с устройством, которое может находиться только в двух устойчивых состояниях: включено/выключено. Конечно же, это хорошо знакомый всем выключатель. А вот придумать выключатель, который мог бы устойчиво и быстро переключаться в любое из 10 состояний, оказалось невозможным. В результате после ряда неудачных попыток разработчики пришли к выводу о невозможности построения компьютера на основе десятичной системы счисления. И в основу представления чисел в компьютере была положена именно двоичная система счисления.

Способы кодирования и декодирования информации в компьютере, в первую очередь, зависит от вида информации, а именно, что должно кодироваться: числа, текст, графические изображения или звук.



## 6. Представление (кодирование) чисел

Системы счисления. Виды систем счисления: позиционные и непозиционные. Двоичная система счисления. Перевод числе из одной системы счисления в другую.

Для записи информации о количестве объектов используются числа. Числа записываются с помощью набора специальных символов.

Система счисления — способ записи чисел с помощью набора специальных знаков, называемых цифрами.

Системы счисления подразделяются на *позиционные* и *непозиционные*.

В позиционных системах счисления величина, обозначаемая цифрой в записи числа, зависит от её положения в числе (позиции).

В непозиционных системах счисления величина, которую обозначает цифра, не зависит от положения в числе.

*Непозиционные системы счисления.*

Каноническим примером фактически *непозиционной системы* счисления является *римская*, в которой в качестве цифр используются латинские буквы: I обозначает 1, V - 5, X - 10, L - 50, C - 100, D - 500, M - 1000. Натуральные числа записываются при помощи повторения этих цифр. При этом, если большая цифра стоит перед меньшей, то они складываются (принцип сложения), если же меньшая — перед большей, то меньшая вычитается из большей (принцип вычитания). Последнее правило применяется только во избежание четырёхкратного повторения одной и той же цифры. Римские цифры появились около 500 до нашей эры у этрусков.

Например, II = 1 + 1 = 2, здесь символ I обозначает 1 независимо от места в числе.

Для правильной записи больших чисел римскими цифрами необходимо сначала записать число тысяч, затем сотен, затем десятков и, наконец, единиц.

Пример: число 1988. Одна тысяча M, девять сотен CM, восемьдесят LXXX, восемь VIII. Запишем их вместе: MCMLXXXVIII.

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

*Позиционные системы счисления.*

В позиционных системах счисления величина, обозначаемая цифрой в записи числа, зависит от её положения в числе (позиции). Количество используемых цифр называется основанием системы счисления.

Например, 11 – это одиннадцать, а не два:  $1 + 1 = 2$  (сравните с римской системой счисления). Здесь символ 1 имеет различное значение в зависимости от позиции в числе.

Самой первой такой системой, когда счетным "прибором" служили пальцы рук, была *пятеричная*. Некоторые племена на филиппинских островах используют ее и в наши дни, а в цивилизованных странах ее реликт, как считают специалисты, сохранился только в виде школьной пятибалльной шкалы оценок.

Следующей после пятеричной возникла *двенадцатеричная* система счисления. Возникла она в древнем Шумере. Некоторые учёные полагают, что такая система возникла у них из подсчёта фаланг на руке большим пальцем.

Широкое распространение получила двенадцатеричная система счисления в XIX веке. . На ее широкое использование в прошлом явно указывают названия числительных во многих языках, а также сохранившиеся в ряде стран способы отсчета времени, денег и соотношения между некоторыми единицами измерения. Год состоит из 12 месяцев, а половина суток состоит из 12 часов. Элементом двенадцатеричной системы в современности может служить счёт дюжинами. Первые три степени числа 12 имеют собственные названия: 1 дюжина = 12 штук; 1 гросс = 12 дюжин = 144 штуки; 1 масса = 12 гроссов = 144 дюжины = 1728 штук. Английский фунт состоит из 12 шиллингов.

Следующая позиционная система счисления была придумана еще в Древнем Вавилоне, причем вавилонская нумерация была *шестидесятеричная*, т.е. в ней использовалось шестьдесят цифр! В более позднее время использовалась арабами, а также древними и средневековыми астрономами. Шестидесятеричная система счисления, как считают исследователи, являет собой синтез уже вышеупомянутых пятеричной и двенадцатеричной систем.

В настоящее время наиболее распространены *десятичная*, *двоичная*, *восьмеричная* и

шестнадцатеричная системы счисления. Двоичная, восьмеричная (в настоящее время вытесняется шестнадцатеричной) и шестнадцатеричная система часто используется в областях, связанных с цифровыми устройствами, программировании и вообще компьютерной документации. Современные компьютерные системы оперируют информацией представленной в цифровой форме. Числовые данные преобразуются в двоичную систему счисления.

Система счисления	Основание	Алфавит цифр
Десятичная	10	0, 1, 2, 3, 4, 5, 6, 7, 8, 9
Двоичная	2	0, 1
Восьмеричная	8	0, 1, 2, 3, 4, 5, 6, 7
Шестнадцатеричная	16	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F

Десятичная система счисления — позиционная система счисления по основанию 10. Предполагается, что основание 10 связано с количеством пальцев рук у человека. Наиболее распространённая система счисления в мире. Для записи чисел используются символы 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, называемые арабскими цифрами.

Двоичная система счисления — позиционная система счисления с основанием 2. Используются цифры 0 и 1. Двоичная система используется в цифровых устройствах, поскольку является наиболее простой и удовлетворяет требованиям:

- Чем меньше значений существует в системе, тем проще изготовить отдельные элементы.
- Чем меньше количество состояний у элемента, тем выше помехоустойчивость и тем быстрее он может работать.
- Простота создания таблиц сложения и умножения — основных действий над числами

Восьмеричная система счисления — позиционная система счисления с основанием 8. Для представления чисел в ней используются цифры 0 до 7. Восьмеричная система часто использовалась в областях, связанных с цифровыми устройствами. Характеризуется лёгким переводом восьмеричных чисел в двоичные и обратно.

Шестнадцатеричная система счисления — позиционная система счисления по основанию 16. Обычно в качестве шестнадцатеричных цифр используются десятичные цифры от 0 до 9 и латинские буквы от A до F для обозначения цифр от 10 до 15. Широко используется в низкоуровневом программировании и вообще в компьютерной документации, поскольку в современных компьютерах минимальной единицей памяти является 8-битный байт, значения которого удобно записывать двумя шестнадцатеричными цифрами.

Укажем соответствие десятичной, двоичной, восьмеричной и шестнадцатеричной систем счисления в таблице.

p=10	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
p=2	0	1	10	11	100	101	110	111	1000	1001	1010	1011	1100	1101	1110	1111	1000 0
p=8	0	1	2	3	4	5	6	7	10	11	12	13	14	15	16	17	20
p=16	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	10

При одновременной работе с несколькими системами счисления для их различения основание системы обычно указывается в виде нижнего индекса, который записывается в десятичной системе:

$123_{10}$  — это число 123 в десятичной системе счисления;

$1111011_2$  — то же число, но в двоичной системе.

Двоичное число 1111011 можно расписать в виде:  $1111011_2 = 1 \cdot 2^6 + 1 \cdot 2^5 + 1 \cdot 2^4 + 1 \cdot 2^3 + 0 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0$ .

Двоичная система счисления обладает такими же свойствами, что и десятичная, только для представления чисел используются не 10 цифр, а всего две. Соответственно и разряд числа называют не десятичным, а двоичным.

*Перевод чисел из одной системы счисления в другую*

Чтобы перевести число из позиционной системы счисления с основанием  $p$  в **десятичную**, надо представить это число в виде суммы степеней  $p$  и произвести указанные вычисления в десятичной системе счисления.

Например, переведем число  $1011_2$  в десятичную систему счисления. Для этого представим это число в виде степеней двойки и произведем вычисления в десятичной системе счисления.

$$1011_2 = 1 \cdot 2^3 + 0 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0 = 1 \cdot 8 + 0 \cdot 4 + 1 \cdot 2 + 1 \cdot 1 = 8 + 0 + 2 + 1 = 11_{10}$$

Рассмотрим еще один пример. Переведем число  $52,74_8$  в десятичную систему счисления.

$$52,74_8 = 5 \cdot 8^1 + 2 \cdot 8^0 + 3 \cdot 8^{-1} + 4 \cdot 8^{-2} = 5 \cdot 8 + 2 \cdot 1 + 7 \cdot 1/8 + 4 \cdot 1/49 = 40 + 2 + 0,875 + 0,0625 = 42,9375_{10}$$

Перевод **из десятичной** системы счисления в систему счисления с основанием  $p$  осуществляется последовательным делением десятичного числа и его десятичных частных на  $p$ , а затем выписыванием последнего частного и остатков в обратном порядке.

Переведем десятичное число  $20_{10}$  в двоичную систему счисления (основание системы счисления  $p=2$ ). В итоге получили  $20_{10} = 10100_2$ .

$$\begin{array}{r|l}
 20 & 2 \\
 \hline
 10 & 2 \\
 \hline
 5 & 2 \\
 \hline
 4 & 2 \\
 \hline
 2 & 2 \\
 \hline
 1 & 2 \\
 \hline
 0 & 2 \\
 \hline
 \end{array}$$

Числа в компьютере хранятся и обрабатываются в двоичной системе счисления. Последовательность нулей и единиц называют двоичным кодом. Специфической особенности представления чисел в памяти компьютера рассмотрим на других уроках по теме «системы счисления».

## 7. Двоичное кодирование текстовой информации

Кодирование текстовой информации. Таблица кодировки. Измерение количества информации в тексте.

Начиная с 60-х годов, компьютеры все больше стали использовать для обработки текстовой информации и в настоящее время большая часть ПК в мире занято обработкой именно текстовой информации.

Традиционно для кодирования одного символа используется количество информации = 1 байту (1 байт = 8 битов).

*Для кодирования одного символа требуется один байт информации.*

Учитывая, что каждый бит принимает значение 1 или 0, получаем, что с помощью 1 байта можно закодировать 256 различных символов. ( $2^8 = 256$ )

Кодирование заключается в том, что каждому символу ставится в соответствие уникальный двоичный код от 00000000 до 11111111 (или десятичный код от 0 до 255).

Важно, что присвоение символу конкретного кода – это вопрос соглашения, которое фиксируется кодовой таблицей.

Таблица, в которой всем символам компьютерного алфавита поставлены в соответствие порядковые номера (коды), называется таблицей кодировки.

Для разных типов ЭВМ используются различные кодировки. С распространением IBM PC международным стандартом стала таблица кодировки **ASCII (American Standard Code for Information Interchange)** – Американский стандартный код для информационного обмена.

Стандартной в этой таблице является только первая половина, т.е. символы с номерами от 0 (00000000) до 127 (01111111). Сюда входят буква латинского алфавита, цифры, знаки препинания, скобки и некоторые другие символы.

Остальные 128 кодов используются в разных вариантах. В русских кодировках размещаются символы русского алфавита.

В настоящее время существует 5 разных кодовых таблиц для русских букв (КОИ8, CP1251, CP866, Mac, ISO).

В настоящее время получил широкое распространение новый международный стандарт **Unicode**, который отводит на каждый символ два байта. С его помощью можно закодировать 65536 ( $2^{16} = 65536$ ) различных символов.

Сегодня очень многие люди для подготовки писем, документов, статей, книг и пр. используют компьютерные текстовые редакторы. Компьютерные редакторы, в основном, работают с алфавитом размером 256 символов.

В этом случае легко подсчитать объем информации в тексте. Если 1 символ алфавита несет 1 байт информации, то надо просто сосчитать количество символов; полученное число даст информационный объем текста в байтах.

Пусть небольшая книжка, сделанная с помощью компьютера, содержит 150 страниц; на каждой странице — 40 строк, в каждой строке — 60 символов. Значит страница содержит  $40 \times 60 = 2400$  байт информации. Объем всей информации в книге:  $2400 \times 150 = 360\,000$  байт.

*Обратите внимание!* Цифры кодируются по стандарту ASCII в двух случаях – при вводе-выводе и когда они встречаются в тексте. Если цифры участвуют в вычислениях, то осуществляется их преобразование в другой двоичный код (см. урок «представление чисел в компьютере»).

Возьмем число 57.

При использовании в тексте каждая цифра будет представлена своим кодом в соответствии с таблицей ASCII. В двоичной системе это – 0011010100110111.

При использовании в вычислениях, код этого числа будет получен по правилам перевода в двоичную систему и получим – 00111001.

Таблица стандартной части ASCII

символ	10-й код	2-й код	символ	10-й код	2-й код	символ	10-й код	2-й код	символ	10-й код	2-й код
	32	00100000	8	56	00111000	P	80	01010000	h	104	01101000
!	33	00100001	9	57	00111001	Q	81	01010001	i	105	01101001
"	34	00100010	:	58	00111010	R	82	01010010	j	106	01101010
#	35	00100011	;	59	00111011	S	83	01010011	k	107	01101011
\$	36	00100100	<	60	00111100	T	84	01010100	l	108	01101100
%	37	00100101	=	61	00111101	U	85	01010101	m	109	01101101
&	38	00100110	>	62	00111110	V	86	01010110	n	110	01101110
'	39	00100111	?	63	00111111	W	87	01010111	o	111	01101111
(	40	00101000	@	64	01000000	X	88	01011000	p	112	01110000
)	41	00101001	A	65	01000001	Y	89	01011001	q	113	01110001
*	42	00101010	B	66	01000010	Z	90	01011010	r	114	01110010
+	43	00101011	C	67	01000011	[	91	01011011	s	115	01110011
,	44	00101100	D	68	01000100	\	92	01011100	t	116	01110100
-	45	00101101	E	69	01000101	]	93	01011101	u	117	01110101
.	46	00101110	F	70	01000110	^	94	01011110	v	118	01110110
/	47	00101111	G	71	01000111	_	95	01011111	w	119	01110111
0	48	00110000	H	72	01001000	`	96	01100000	x	120	01111000
1	49	00110001	I	73	01001001	a	97	01100001	y	121	01111001
2	50	00110010	J	74	01001010	b	98	01100010	z	122	01111010
3	51	00110011	K	75	01001011	c	99	01100011	{	123	01111011
4	52	00110100	L	76	01001100	d	100	01100100		124	01111100
5	53	00110101	M	77	01001101	e	101	01100101	}	125	01111101
6	54	00110110	N	78	01001110	f	102	01100110	~	126	01111110
7	55	00110111	O	79	01001111	g	103	01100111	□	127	01111111

 Таблица расширенного кода ASCII (один из вариантов) (*Кодировка Windows-1251 (синоним CP1251)*)

символ	10-й код	2-й код	символ	10-й код	2-й код	символ	10-й код	2-й код	символ	10-й код	2-й код
Ъ	128	10000000		160	10100000	А	192	11000000	а	224	11100000
Г	129	10000001	Ѹ	161	10100001	Б	193	11000001	б	225	11100001
л	130	10000010	ѹ	162	10100010	В	194	11000010	в	226	11100010
і	131	10000011	Ј	163	10100011	Г	195	11000011	г	227	11100011
„	132	10000100	о	164	10100100	Д	196	11000100	д	228	11100100
…	133	10000101	Ґ	165	10100101	Е	197	11000101	е	229	11100101
†	134	10000110	і	166	10100110	Ж	198	11000110	ж	230	11100110
‡	135	10000111	§	167	10100111	З	199	11000111	з	231	11100111
€	136	10001000	Є	168	10101000	И	200	11001000	и	232	11101000
‰	137	10001001	©	169	10101001	Й	201	11001001	й	233	11101001
Љ	138	10001010	€	170	10101010	К	202	11001010	к	234	11101010
‹	139	10001011	«	171	10101011	Л	203	11001011	л	235	11101011
Њ	140	10001100	–	172	10101100	М	204	11001100	м	236	11101100
К	141	10001101	-	173	10101101	Н	205	11001101	н	237	11101101
Ъ	142	10001110	®	174	10101110	О	206	11001110	о	238	11101110
Ц	143	10001111	Ї	175	10101111	П	207	11001111	п	239	11101111
ђ	144	10010000	°	176	10110000	Р	208	11010000	р	240	11110000
‘	145	10010001	±	177	10110001	С	209	11010001	с	241	11110001
’	146	10010010	І	178	10110010	Т	210	11010010	т	242	11110010
“	147	10010011	і	179	10110011	У	211	11010011	у	243	11110011
”	148	10010100	г	180	10110100	Ф	212	11010100	ф	244	11110100
•	149	10010101	и	181	10110101	Х	213	11010101	х	245	11110101
–	150	10010110	¶	182	10110110	Ц	214	11010110	ц	246	11110110
—	151	10010111	·	183	10110111	Ч	215	11010111	ч	247	11110111
□	152	10011000	ё	184	10111000	Ш	216	11011000	ш	248	11111000
™	153	10011001	№	185	10111001	Щ	217	11011001	щ	249	11111001
љ	154	10011010	€	186	10111010	Ъ	218	11011010	ъ	250	11111010
›	155	10011011	»	187	10111011	Ы	219	11011011	ы	251	11111011
њ	156	10011100	ј	188	10111100	Ь	220	11011100	ь	252	11111100
ќ	157	10011101	ѕ	189	10111101	Э	221	11011101	э	253	11111101
ћ	158	10011110	ѕ	190	10111110	Ю	222	11011110	ю	254	11111110
џ	159	10011111	ї	191	10111111	Я	223	11011111	я	255	11111111

## 8. Двоичное кодирование изображений

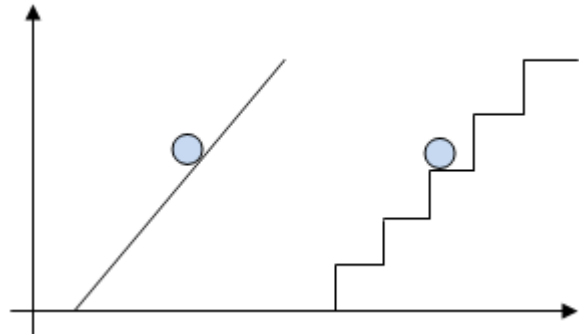
Дискретизация. Виды изображений: растровое и векторное. Цветовые модели. Графические форматы файлов.

### Аналоговый и дискретный способ кодирования

Человек способен воспринимать и хранить информацию в форме образов (зрительных, звуковых, осязательных, вкусовых и обонятельных). Зрительные образы могут быть сохранены в виде изображений (рисунков, фотографий и так далее), а звуковые — зафиксированы на пластинках, магнитных лентах, лазерных дисках и так далее.

Информация, в том числе графическая и звуковая, может быть представлена в аналоговой или дискретной форме. При аналоговом представлении физическая величина принимает бесконечное множество значений, причем ее значения изменяются непрерывно. При дискретном представлении физическая величина принимает конечное множество значений, причем ее величина изменяется скачкообразно.

Приведем пример аналогового и дискретного представления информации. Положение тела на наклонной плоскости и на лестнице задается значениями координат X и Y. При движении тела по наклонной плоскости его координаты могут принимать бесконечное множество непрерывно изменяющихся значений из определенного диапазона, а при движении по лестнице — только определенный набор значений, причем меняющихся скачкообразно.



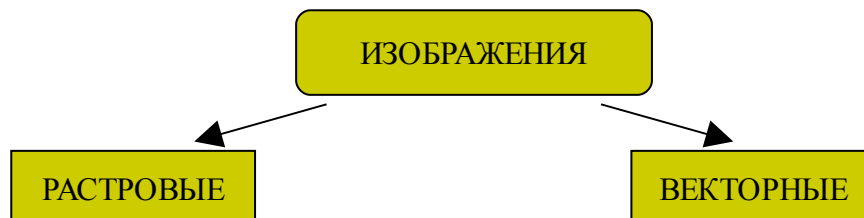
Примером аналогового представления графической информации может служить, например, живописное полотно, цвет которого изменяется непрерывно, а дискретного — изображение, напечатанное с помощью струйного принтера и состоящее из отдельных точек разного цвета. Примером аналогового хранения звуковой информации является виниловая пластинка (звуковая дорожка изменяет свою форму непрерывно), а дискретного — аудиокомпакт-диск (звуковая дорожка которого содержит участки с различной отражающей способностью).

Преобразование графической и звуковой информации из аналоговой формы в дискретную производится путем дискретизации, то есть разбиения непрерывного графического изображения и непрерывного (аналогового) звукового сигнала на отдельные элементы. В процессе дискретизации производится кодирование, то есть присвоение каждому элементу конкретного значения в форме кода.

**Дискретизация** — это преобразование непрерывных изображений и звука в набор дискретных значений в форме кодов.

### Кодирование изображений

Создавать и хранить графические объекты в компьютере можно двумя способами — как растровое или как векторное изображение. Для каждого типа изображений используется свой способ кодирования.



### Кодирование растровых изображений

Растровое изображение представляет собой совокупность точек (пикселей) разных цветов. **Пиксель** — минимальный участок изображения, цвет которого можно задать независимым образом.

В процессе кодирования изображения производится его пространственная дискретизация. Пространственную дискретизацию изображения можно сравнить с построением изображения из мозаики (большого количества маленьких разноцветных стекол). *Изображение разбивается на*

отдельные маленькие фрагменты (точки), причем каждому фрагменту присваивается значение его цвета, то есть код цвета (красный, зеленый, синий и так далее).

Для черно-белого изображения информационный объем одной точки равен одному биту (либо черная, либо белая – либо 1, либо 0).

Для четырех цветного – 2 бита.

Для 8 цветов необходимо – 3 бита.

Для 16 цветов – 4 бита.

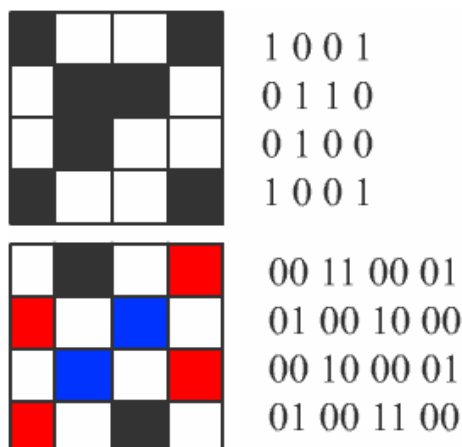
Для 256 цветов – 8 бит (1 байт).

Качество изображения зависит от количества точек (чем меньше размер точки и, соответственно, больше их количество, тем лучше качество) и количества используемых цветов (чем больше цветов, тем качественнее кодируется изображение).

Для представления цвета в виде числового кода используются две обратных друг другу цветовые модели: **RGB** или **CMYK**. Модель RGB используется в телевизорах, мониторах, проекторах, сканерах, цифровых фотоаппаратах... Основные цвета в этой модели: красный (**Red**), зеленый (**Green**), синий (**Blue**). Цветовая модель CMYK используется в полиграфии при формировании изображений, предназначенных для печати на бумаге.

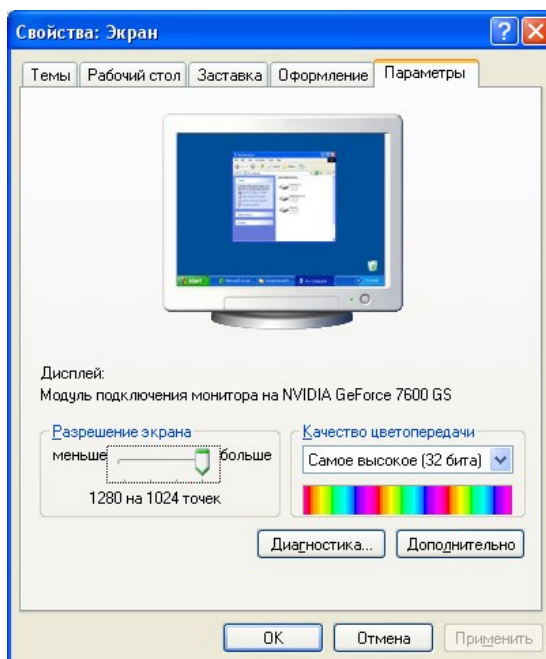
Цветные изображения могут иметь различную глубину цвета, которая задается количеством битов, используемых для кодирования цвета точки.

Если кодировать цвет одной точки изображения тремя битами (по одному биту на каждый цвет RGB), то мы получим все восемьразличных цветов.



R	G	B	Цвет
1	1	1	белый
1	1	0	желтый
1	0	1	пурпурный
1	0	0	красный
0	1	1	голубой
0	1	0	зеленый
0	0	1	синий
0	0	0	черный

На практике же, для сохранения информации о цвете каждой точки цветного изображения в модели RGB обычно отводится 3 байта (т.е. 24 бита) - по 1 байту (т.е. по 8 бит) под значение цвета каждой составляющей. Таким образом, каждая RGB-составляющая может принимать значение в диапазоне от 0 до 255 (всего  $2^8=256$  значений), а каждая точка изображения, при такой системе кодирования может быть окрашена в один из 16 777 216 цветов. Такой набор цветов принято называть True Color (правдивые цвета), потому что человеческий глаз все равно не в состоянии различить большего разнообразия.



Для того чтобы на экране монитора формировалось изображение, информация о каждой точке (код цвета точки) должна храниться в видеопамяти компьютера. Рассчитаем необходимый объем видеопамяти для одного из графических режимов. В современных компьютерах разрешение экрана обычно составляет 1280x1024 точек. Т.е. всего  $1280 * 1024 = 1310720$  точек. При глубине цвета 32 бита на точку необходимый объем видеопамяти:  $32 * 1310720 = 41943040$  бит = 5242880 байт = 5120 Кб = 5 Мб.

Растровые изображения очень чувствительны к масштабированию (увеличению или уменьшению). При уменьшении растрового изображения несколько соседних точек преобразуются в одну, поэтому теряется различимость мелких деталей изображения. При увеличении изображения увеличивается размер каждой точки и появляется ступенчатый эффект, который можно увидеть невооруженным глазом.

*Кодирование векторных изображений*

Векторное изображение представляет собой совокупность графических примитивов (точка, отрезок, эллипс...). Каждый примитив описывается математическими формулами. Кодирование зависит от прикладной среды.

Достоинством векторной графики является то, что файлы, хранящие векторные графические изображения, имеют сравнительно небольшой объем.

Важно также, что векторные графические изображения могут быть увеличены или уменьшены без потери качества.

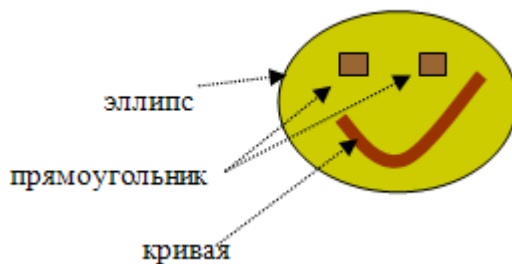
*Графические форматы файлов*

Форматы графических файлов определяют способ хранения информации в файле (растровый или векторный), а также форму хранения информации (используемый алгоритм сжатия).

Наиболее популярные растровые форматы:

- BMP
- GIF
- JPEG
- TIFF
- PNG

Bit MaP image (BMP) — универсальный формат растровых графических файлов, используется в операционной системе Windows. Этот формат поддерживается многими графическими редакторами, в том числе редактором Paint. Рекомендуется для хранения и обмена данными с другими





приложениями.

Tagged Image File Format (TIFF) — формат растровых графических файлов, поддерживается всеми основными графическими редакторами и компьютерными платформами. Включает в себя алгоритм сжатия без потерь информации. Используется для обмена документами между различными программами. Рекомендуется для использования при работе с издательскими системами.

Graphics Interchange Format (GIF) — формат растровых графических файлов, поддерживается приложениями для различных операционных систем. Включает алгоритм сжатия без потерь информации, позволяющий уменьшить объем файла в несколько раз. Рекомендуется для хранения изображений, создаваемых программным путем (диаграмм, графиков и так далее) и рисунков (типа аппликации) с ограниченным количеством цветов (до 256). Используется для размещения графических изображений на Web-страницах в Интернете.

Portable Network Graphic (PNG) — формат растровых графических файлов, аналогичный формату GIF. Рекомендуется для размещения графических изображений на Web-страницах в Интернете.

Joint Photographic Expert Group (JPEG) — формат растровых графических файлов, который реализует эффективный алгоритм сжатия (метод JPEG) для отсканированных фотографий и иллюстраций. Алгоритм сжатия позволяет уменьшить объем файла в десятки раз, однако приводит к необратимой потере части информации. Поддерживается приложениями для различных операционных систем. Используется для размещения графических изображений на Web-страницах в Интернете.

## 9. Двоичное кодирование звука. Представление видеоинформации

Кодирование звука: глубина кодирования и частота дискретизации. Представление видеоинформации. Понятие мультимедиа.

### Двоичное кодирование звука

Использование компьютера для обработки звука началось позднее, нежели чисел, текстов и графики.

Звук – волна с непрерывно изменяющейся амплитудой и частотой. Чем больше амплитуда, тем он громче для человека, чем больше частота, тем выше тон.

Звуковые сигналы в окружающем нас мире необычайно разнообразны. Сложные непрерывные сигналы можно с достаточной точностью представлять в виде суммы некоторого числа простейших синусоидальных колебаний. Причем каждое слагаемое, то есть каждая синусоида, может быть точно задана некоторым набором числовых параметров – амплитуды, фазы и частоты, которые можно рассматривать как код звука в некоторый момент времени.

В процессе кодирования звукового сигнала производится его временная дискретизация – непрерывная волна разбивается на отдельные маленькие временные участки и для каждого такого участка устанавливается определенная величина амплитуды.

Таким образом непрерывная зависимость амплитуды сигнала от времени заменяется на дискретную последовательность уровней громкости (см. рисунок).

Каждому уровню громкости присваивается его код. Чем большее количество уровней громкости будет выделено в процессе кодирования, тем большее количество информации будет нести значение каждого уровня и тем более качественным будет звучание.

Качество двоичного кодирования звука определяется глубиной кодирования и частотой дискретизации.

Частота дискретизации – количество измерений уровня сигнала в единицу времени.

Количество уровней громкости определяет глубину кодирования. Современные звуковые карты обеспечивают 16-битную глубину кодирования звука. При этом количество уровней громкости равно  $N = 2^l = 2^{16} = 65536$ .

### Представление видеоинформации

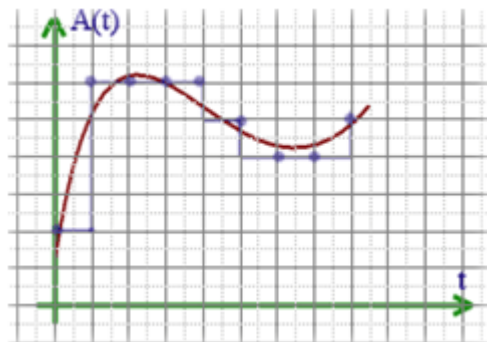
В последнее время компьютер все чаще используется для работы с видеоинформацией. Простейшей такой работой является просмотр кинофильмов и видеоклипов. Следует четко представлять, что обработка видеоинформации требует очень высокого быстродействия компьютерной системы.

Что представляет собой фильм с точки зрения информатики? Прежде всего, это сочетание звуковой и графической информации. Кроме того, для создания на экране эффекта движения используется дискретная по своей сути технология быстрой смены статических картинок. Исследования показали, что если за одну секунду сменяется более 10-12 кадров, то человеческий глаз воспринимает изменения на них как непрерывные.

Казалось бы, если проблемы кодирования статической графики и звука решены, то сохранить видеоизображение уже не составит труда. Но это только на первый взгляд, поскольку, как показывает разобранный выше пример, при использовании традиционных методов сохранения информации электронная версия фильма получится слишком большой. Достаточно очевидное усовершенствование состоит в том, чтобы первый кадр запомнить целиком (в литературе его принято называть ключевым), а в следующих сохранять лишь отличия от начального кадра (разностные кадры).

Существует множество различных форматов представления видеоданных.

В среде Windows, например, уже более 10 лет (начиная с версии 3.1) применяется формат Video for Windows, базирующийся на универсальных файлах с расширением AVI (Audio Video Interleave – чередование аудио и видео).



Более универсальным является мультимедийный формат Quick Time, первоначально возникший на компьютерах Apple.

Все большее распространение в последнее время получают системы сжатия видеоизображений, допускающие некоторые незаметные для глаза искажения изображения с целью повышения степени сжатия. Наиболее известным стандартом подобного класса служит MPEG (Motion Picture Expert Group), который разработан и постоянно развивается созданным в 1988 году Комитетом (группой экспертов) международной организации ISO/IEC (International Standards Organization/International Electrotechnical Commission) по стандартам высококачественного сжатия движущихся изображений. Методы, применяемые в MPEG, непросты для понимания и опираются на достаточно сложную математику.

Большее распространение получила технология под названием DivX (происходит от сокращения слов Digital Video Express). Благодаря DivX удалось достигнуть степени сжатия, позволившей вписать качественную запись полнометражного фильма на один компакт-диск – сжать 4,7 Гб DVD-фильма до 650 Мб.

### **Мультимедиа**

Мультимедиа (multimedia, от англ. multi - много и media - носитель, среда) - совокупность компьютерных технологий, одновременно использующих несколько информационных сред: текст, графику, видео, фотографию, анимацию, звуковые эффекты, высококачественное звуковое сопровождение.

Под словом «мультимедиа» понимают воздействие на пользователя по нескольким информационным каналам одновременно. Можно еще сказать так: мультимедиа – это объединение изображения на экране компьютера (в том числе и графической анимации и видеокадров) с текстом и звуковым сопровождением.

Наибольшее распространение системы мультимедиа получили в области обучения, рекламы, развлечений.