

Информация

Информация

1. Понятие информации
2. Классификация информации
3. Свойства информации
4. Информационные процессы
5. Системы счисления
 - 5.1. Арифметические операции в позиционных системах счисления
 - 5.2. Представление чисел в ПК
 - 5.3. Инженерный калькулятор
6. Кодирование информации
 - 6.1. Языки и алфавиты
7. Единицы измерения информации
8. Измерения количества информации
 - 6.2. Кодирование текстовой информации
 - 6.3. Кодирование графики и звука

1. Понятие информация

Понятие «**информация**» (от лат. Informatio - сведения, разъяснение, осведомление) является базовым в курсе информатики. Строгое определение этому термину дать невозможно. Тем не менее в разных науках существуют различные подходы к определению информации.

В информатике нас интересуют два определения:

(Определение Клода Шеннона) **Информация** – сведения об объектах, процессах реального мира, их свойствах, состояниях и отношениях, которые уменьшают имеющуюся в знаниях о них степень неопределенности.

Другими словами, **Информация** – это снятая неопределенность.

Другое определение:

Информация - это отражение реального (материального, предметного) мира, выражаемое в виде сигналов и знаков.

В информатике всю информацию (включая ее графические формы) мы представляем конечной совокупностью знаков (символов), из которых образуется бесконечное разнообразие информационных объектов (текстов, изображений и т.п.).

Виды информации: тексты, рисунки, фотографии, звуковые сигналы, магнитные записи и т.д.

2. Классификация информации

- 1) Информация подразделяется **по форме представления** на 2 вида:
 - **дискретная** форма представления информации - это последовательность символов, характеризующая прерывистую, изменяющуюся величину (количество дорожно-транспортных происшествий, количество тяжких преступлений и т.п.);
 - **аналоговая** или **непрерывная** форма представления информации - это величина, характеризующая процесс, не имеющий перерывов или промежутков (температура тела человека, скорость автомобиля на определенном участке пути и т.п.).
- 2) **По области возникновения** выделяют информацию:
 - **элементарную** (механическую), которая отражает процессы, явления неодушевленной природы;
 - **биологическую**, которая отражает процессы животного и растительного мира;
 - **социальную**, которая отражает процессы человеческого общества.
- 3) **По способу передачи и восприятия** различают следующие виды информации:
 - - **визуальную**, передаваемую видимыми образами и символами;
 - - **аудиальную**, передаваемую звуками;

- - *тактильную*, передаваемую ощущениями;
 - - *органолептическую*, передаваемую запахами и вкусами;
 - - *машинную*, выдаваемую и воспринимаемую средствами вычислительной техники.
- 4) Информацию, создаваемую и используемую человеком, **по общественному назначению** можно разбить на три вида:
- *личную*, предназначенную для конкретного человека;
 - *массовую*, предназначенную для любого желающего ее пользоваться (общественно-политическая, научно-популярная и т.д.) ;
 - *специальную*, предназначенную для использования узким кругом лиц, занимающихся решением сложных специальных задач в области науки, техники, экономики.
- 5) **По способам кодирования** выделяют следующие типы информации:
- *символьную*, основанную на использовании символов - букв, цифр, знаков и т. д. Она является наиболее простой, но практически применяется только для передачи несложных сигналов о различных событиях. Примером может служить зеленый свет уличного светофора, который сообщает о возможности начала движения пешеходам или водителям автотранспорта.
 - *текстовую*, основанную на использовании комбинаций символов. Здесь так же, как и в предыдущей форме, используются символы: буквы, цифры, математические знаки. Однако информация заложена не только в этих символах, но и в их сочетании, порядке следования. Так, слова КОТ и ТОК имеют одинаковые буквы, но содержат различную информацию. Благодаря взаимосвязи символов и отображению речи человека текстовая информация чрезвычайно удобна и широко используется в деятельности человека: книги, брошюры, журналы, различного рода документы, аудиозаписи кодируются в текстовой форме.
 - *графическую*, основанную на использовании произвольного сочетания в пространстве графических примитивов. К этой форме относятся фотографии, схемы, чертежи, рисунки, играющие большое значение в деятельности человек.

3. Свойства информации

Свойства информации можно рассматривать в трех аспектах: техническом - это точность, надежность, скорость передачи сигналов и т.д.; семантическом - это передача смысла текста с помощью кодов и прагматическом - это насколько эффективно информация влияет на поведение объекта.

Чтобы информация способствовала принятию на ее основе правильных решений, она должна характеризоваться такими **свойствами**, как:

1. Достоверность.

Это свойство связано с истинностью и ложностью информации. Известно, что истина объективна по содержанию (не зависит от человека), но субъективна по форме, так как является результатом деятельности человека.

Истина относительна. Она отражает предмет не полностью, ввиду ограниченности человеческих знаний. В процессе познания человеком она стремится к истине абсолютной, полностью исчерпывающей предмет познания. Достоверная информация со временем может стать недостоверной, так как она обладает свойством устаревать, то есть перестает отражать истинное положение дел.

Например, долгое время за истину принималась геоцентрическая система Птолемея (все планеты и солнце вращаются вокруг Земли). В настоящее время истинность, достоверность информации о гелиоцентричности Солнечной системы бесспорна.

Таким образом, смысл рассматриваемого свойства достоверности и заключается в том, чтобы определить насколько данная информация соответствует истинному положению дел. В противном случае недостоверная информация повлечет за собой неправильное понимание и, как следствие, принятие неправильных решений.

2. Объективность.

Данное понятие является относительным, поскольку методы получения, обработки информации, как правило, являются субъективными. Более объективной будет та информация, методы получения которой будут менее подвержены субъективным параметрам.

Так например, при наблюдении природных явлений или объектов информация будет формироваться более объективно, если рассматривать фотографию, а не рисунок, выполненный человеком.

В ходе информационного процесса (например, передачи) степень объективности информации всегда понижается. Это свойство учитывают, например, в правовых дисциплинах, где по-разному обрабатываются показания лиц, непосредственно наблюдавших события или получивших информацию косвенным путем (посредством умозаключений или со слов третьих лиц). В не меньшей степени объективность информации учитывают в исторических дисциплинах. Одни и те же события, зафиксированные в исторических документах разных стран и народов, выглядят совершенно по-разному. У историков имеются свои методы для тестирования объективности исторических данных и создания новых, более достоверных данных путем сопоставления, фильтрации и селекции исходных данных. Обратим внимание на то, что здесь речь идет не о повышении объективности данных, а о повышении их достоверности (это совсем другое свойство)

3. Достаточность, полнота.

Полнота информации во многом характеризует качество информации и определяет достаточность данных для принятия решения или создания новых знаний на основе уже имеющихся.

Чем более полными и точными будут данные, тем проще подобрать методы поиска, обработки и передачи информации.

Как неполная, так и избыточная информация сдерживает принятие решений или может повлечь ошибки.

4. Точность.

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

- **формальная** точность, измеряемая значением единицы младшего разряда числа,
- **реальная** точность, определяемая значением единицы последнего разряда числа, вероятность, которого гарантируется,
- **максимальная** точность, которую можно получить в конкретных условиях функционирования системы.
- **необходимая** точность, определяемая функциональным назначением показателя.

5. Адекватность.

Определенный уровень соответствия создаваемого с помощью полученной информации образа реальному объекту, процессу, явлению и т.п. Неадекватная информация может возникать, к примеру, при создании новой информации на основе непонятных данных.

Адекватность информации может быть выражена в трех формах:

- **Синтаксической**. Отражает формально-структурные характеристики информации и не затрагивает ее смысловое содержание. На этом уровне учитывается тип носителя и способ представления информации, скорость передачи и обработки, размеры кодов представления информации, надежность и точность преобразования этих кодов и т.п.
- **Семантической**(смысловая). Определяет степень соответствия образа объекта и самого объекта. Предполагает учет смыслового содержания информации, на этом уровне анализируются те сведения, которые отражает информация, рассматриваются смысловые связи. В информатике устанавливаются смысловые связи между кодами представления информации. Эта форма служит для формирования понятий и представлений, выявления смысла, содержания информации и ее обобщений.
- **Прагматической**(потребительская). Отражает отношение информации и ее потребителя, соответствие информации цели управления, которая на ее основе реализуется. Проявляются прагматические свойства информации только при наличии единства информации (объекта), пользователя и цели управления. Прагматический аспект рассмотрения связан с ценностью, полезностью использования информации при выработке потребителем решения для достижения своей цели. С этой точки зрения анализируются потребительские свойства информации. Эта форма адекватности непосредственно связана с практическим использованием информации, с соответствием ее целевой функции деятельности системы.

6. Актуальность и оперативность.

Это степень соответствия информации текущему моменту времени. С течением времени информация стареет. И поскольку информационные процессы растянуты во времени, то достоверная и адекватная, может стать устаревшей, что может привести к ошибочным решениям.

Оперативность – это своевременность информации. С этой характеристикой информации мы встречаемся практически ежедневно. Все СМИ стремятся оперативно получать информацию и передавать ее зрителям, слушателям и читателям. Устаревшая информация нас, как правило, не интересует.

7. Полезность, ценность.

Ценность информации зависит от того, насколько она важна для решения задачи, а также от того, насколько в дальнейшем она найдёт применение в каких-либо видах деятельности человека.

Одна и та же информация может быть очень важной для одного человека и быть абсолютно бесполезной для другого

8. Доступность.

Это свойство информации связано с возможностью ее воспринимать, понимать и использовать. Сообщение, передаваемое вам, например, азбукой Морзе, если вы ее не знаете, не несет вам никакой информации.

9. Устойчивость.

Отражает способность информации реагировать на изменение исходных данных без нарушения необходимой точности. Устойчивость обусловлена выбранной методикой ее отбора и формирования.

4. Информационные процессы

Информационные процессы – процессы, связанные с определенными операциями над информацией, а именно:

- создание;
- передача;
- хранение;
- запоминание;
- копирование;
- формализация;
- распространение;
- преобразовывание;
- обработка;
- упрощение;
- сбор;
- поиск;
- измерение;
- разрушение;
- и др.

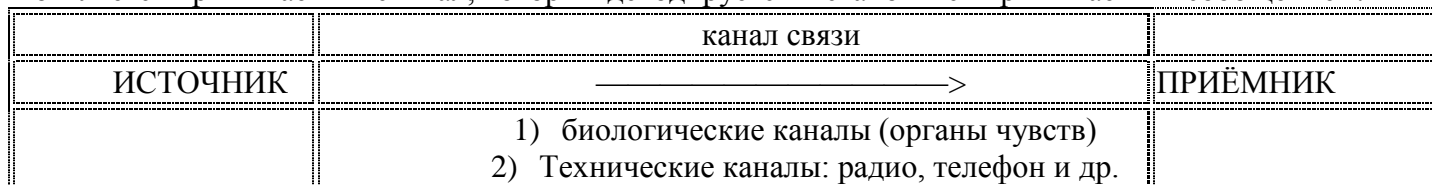
Три составляющие информационного процесса:

1. Хранение.

Для хранения информации используют носители информации, каждый из которых характеризуется объемом информации, надежностью хранения, временем доступа.

2. Передача.

Информация передается в виде сообщений от некоторого источника информации к её приёмнику посредством канала связи между ними. Источник посылает передаваемое сообщение, которое кодируется в передаваемый сигнал. Этот сигнал посылается по каналу связи. В результате в приёмнике появляется принимаемый сигнал, который декодируется и становится принимаемым сообщением.



Передача информации по каналам связи часто сопровождается воздействием помех, вызывающих искажение и потерю информации.

Например: сообщение, содержащее информацию о прогнозе погоды, передается приёмнику (телезрителю) от источника — специалиста-метеоролога посредством канала связи — телевизионной передающей аппаратуры и телевизора;

3. Обработка.

Обработка информации – получение одних информационных объектов из других информационных объектов путем выполнения некоторых алгоритмов.

Обработка является одной из основных операций, выполняемых над информацией, и главным средством увеличения объёма и разнообразия информации.

Средства обработки информации — это всевозможные устройства и системы, созданные человеком, и в первую очередь, компьютер — универсальная машина для обработки информации.

Компьютеры обрабатывают информацию путем выполнения некоторых алгоритмов.

Живые организмы и растения обрабатывают информацию с помощью своих органов и систем.

Все информационные процессы взаимосвязаны между собой.

Например, живое существо своими органами чувств (глаз, ухо, кожа, язык и т.д.) воспринимает информацию из внешнего мира, перерабатывает её в определенную последовательность нервных импульсов, передает импульсы по нервным волокнам, хранит в памяти в виде состояния нейронных структур мозга, воспроизводит в виде звуковых сигналов, движений и т.п., использует в процессе своей жизнедеятельности.

5. Системы счисления

Цифровая информация основана на применении чисел в той или иной системе исчисления.

Система счисления — это знаковая система, в которой числа записываются по определенным правилам с помощью символов некоторого алфавита, называемых цифрами.

Все СС делятся на позиционные и непозиционные.

Система счисления, в которой значение каждой цифры в произвольном месте последовательности цифр, обозначающей запись числа, не изменяется, называется **непозиционной**.

Система счисления, в которой значение каждой цифры зависит от места в последовательности цифр в записи числа, называется **позиционной**.

Римская СС является самой распространенной из **непозиционных СС**. В качестве цифр в ней используют I(1), V(5), X(10), L(50), C(100), D(500), M(1000).

Значение цифры не зависит от ее положения в числе. Например, в числе XXX (30) трижды встречается цифра X и в каждом случае обозначает одну и ту же величину — число 10, три раза по 10 — 30.

Величина числа в римской СС определяется как сумма или разность цифр в числе. Если меньшая цифра стоит слева от большей, то она вычитается, если справа — прибавляется.

Например,

1) запись десятичного числа 1998 в римской СС будет выглядеть следующим образом:
MCMXCVIII

М	С	М	Х	С	V	I	I	I
1000	100	1000	10	100	5	1	1	1
1000	+(1000-100)		+(100-10)		+5	+1	+1	+1

2) 44 → LXIV = (50-10)+(5-1)

Позиционные СС.

В позиционных СС:

- количественное значение цифр зависит от ее позиции в числе;
- позиция цифры в числе называется разрядом, разряд числа возрастает справа налево;
- основание системы равно количеству цифр (знаков в ее алфавите) и определяет во сколько раз различаются значения одинаковых цифр, стоящих в соседних позициях;
- в развернутой форме запись числа $a_n a_{n-1} \dots a_0 \dots a_{-1} \dots a_{-m}$ (где a_n, a_{n-1}, a_0 — цифры целой части числа, a_{-1}, a_{-m} — цифры дробной части) в СС с основанием h выглядит следующим образом:
 $a_n h^n + a_{n-1} h^{n-1} + \dots + a_0 h^0 + a_{-1} h^{-1} + \dots + a_{-m} h^{-m}$

Система счисления	Основание	Алфавит
Десятичная	10	0,1,...,9
Двоичная	2	0,1
Восьмеричная	8	0,1,...,7
Шестнадцатеричная	16	0,1,...,9, A(10), B(11), C(12), D(13), E(14), F(15)

Рассмотрим перевод чисел из одной СС в другую .

Чтобы перевести число из системы счисления с основанием h в десятичную надо:

1. Записать число в развернутой форме (в виде суммы произведений):

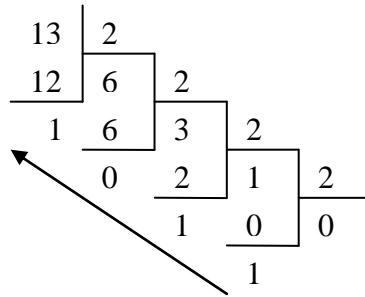
$$N = a_n h^n + a_{n-1} h^{n-1} + \dots + a_0 h^0 + a_1 h^{-1} + \dots + a_m h^{-m}$$

2. Произвести вычисления, результат этот вычисления и будет искомым числом.

Например, $23,4_8 = 2 \cdot 8^1 + 3 \cdot 8^0 + 5 \cdot 8^{-1} = 16 + 3 + 4/8 = 19,5_{10}$

Чтобы перевести целое число из 10-ой системы в систему с основанием h , надо:

1. Разделить число на h с остатком,
2. Полученное неполное частное вновь разделить с остатком и так до тех пор, пока неполное частное не станет равным нулю.
3. Записать полученную последовательность в обратном порядке – это и будет искомая запись натурального числа N в системе счисления с основанием h .



Например, $13_{10} = 1101_2$

Чтобы перевести десятичную дробную часть из 10-й системы счисления в систему счисления с основанием h :

1. Умножить исходную десятичную дробь на h , отделить целую часть произведения.
2. Новую дробную часть вновь умножить на h и так до тех пор, пока не получится нулевая дробная часть или не будет достигнута требуемая точность вычислений.
3. Записать полученные целые части произведения в прямой последовательности.

Например, $0,65_{10} = 0,1010_2$

$$0,65 \cdot 2 = 1 + 0,3$$

$$0,3 \cdot 2 = 0 + 0,6$$

$$0,6 \cdot 2 = 1 + 0,2$$

$$0,2 \cdot 2 = 0 + 0,4$$

и т.д.

Примечание. Перевод чисел, содержащих целую и дробную части, производится в два этапа. Отдельно переводится по соответствующему алгоритму целая и отдельно – дробная. В итоговой записи полученная целая часть от дробной отделяется запятой.

Например, $13,65_{10} = 1101,1010_2$

Чтобы перевести число из системы счисления t в систему с основанием h , можно:

$$N_t \rightarrow N_{10} \rightarrow N_h$$

С числами в различных системах счисления можно производить те же арифметические действия, что и с числами в десятичной системе. Однако, если напрямую это сделать затруднительно, можно сперва перевести их в десятичную, выполнить необходимые операции, после чего вновь перевести результат в СС с указанным основанием.

Например, $110_2 + 11_2 = 1001_2$

Проверка: $110_2 = 1 \cdot 2^2 + 1 \cdot 2^1 + 0 \cdot 2^0 = 4 + 2 = 6_{10}$; $11_2 = 1 \cdot 2^1 + 1 \cdot 2^0 = 2 + 1 = 3_{10}$; $6_{10} + 3_{10} = 9_{10}$; $9_{10} = 1001_2$

5.1. Арифметические операции в позиционных системах счисления

Арифметические операции во всех позиционных системах счисления выполняются по одним и тем же хорошо известным правилам.

Двоичная система:

<p>Сложение. $0+0=0$ $0+1=1$ $1+0=1$ $1+1=10$ – при сложении происходит переполнение разряда и производится перенос в старший разряд. Переполнение разряда наступает тогда, когда величина числа в нем становится равной или больше основания. Например, $\begin{array}{r} +110_2 \\ \underline{11_2} \\ 1001_2 \end{array}$</p>	<p>Вычитание. $0-0=0$ $0-1=\overline{1}1$ - при вычитании из меньшего числа большего производится заем из старшего разряда. В таблице заем обозначается 1 с чертой сверху. $1-0=1$ $1-1=0$ Например, $\begin{array}{r} -110_2 \\ \underline{11_2} \\ 11_2 \end{array}$</p>
<p>Умножение. $0*0=0$ $0*1=0$ $1*0=0$ $1*1=1$ Например, $\begin{array}{r} *110_2 \\ \underline{11_2} \\ 110_2 \\ +110_2 \\ \hline 10010_2 \end{array}$</p>	<p>Деление. $\begin{array}{r} 110_2 \overline{)11_2} \\ \underline{11} \quad 10 \\ 0 \end{array}$</p>
<p>Например, $\begin{array}{r} +37_8 \\ \underline{25_8} \\ 64_2 \end{array}$ $7+5=12$ $12-8=4$ (1) $3+2+1=6$</p>	<p>Например, $\begin{array}{r} -9C_{16} \\ \underline{78_{16}} \\ 24_{16} \end{array}$ $12-8=4$ $9-7=2$</p>

5.2. Представление чисел в ПК

1. Целые числа - формат с фиксированной запятой.

В этом случае каждому разряду ячейки памяти соответствует всегда и тот разряд числа, а "запятая" "находится" справа после младшего разряда, то есть вне разрядной сетки.

Максимальное значение для n-разрядного представления: 2^n-1

1.1. Целые неотрицательные числа.

Отводится одна ячейка памяти (например, в 8-разрядном коде).

Например, число $A_2=1011_2$ будет храниться в памяти следующим образом:

0	0	0	0	1	0	1	1
---	---	---	---	---	---	---	---

максимальное значение достигается, когда во всех разрядах 1, т.е.

$$A=1*2^7+1*2^6+1*2^5+1*2^4+1*2^3+1*2^2+1*2^1+1*2^0=255_{10}$$

Максимальное значение для n-разрядного представления: 2^n-1

1.1.1 Тип Byte (8 бит) Максимальное значение число $2^8-1=255$. Диапазон [0,255]

1.1.2 Тип Word (16 бит) Диапазон [0,65535]

1.2. Целые числа со знаком.

Отводится две ячейки памяти (16 битов), старший левый разряд отводится по знак числа (0 - положительное, 1 - отрицательное).

Прямой код – представление в компьютере положительных целых чисел с использованием формата "знак-величина".

Например, число $A_{10}=8_{10}=1000_2$ будет храниться в памяти следующим образом:

0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Максимальное положительное число (с учетом выделения одного разряда на знак) для целых чисел со знаком в n-разрядном представлении: $2^{n-1}-1$

Диапазон в 8-разрядном $[0, 2^7-1]=[0, 127]$

Дополнительный код – представление в компьютере отрицательных целых чисел с использованием формата "знак-величина".

Дополнительный код позволяет заменить арифметическую операцию вычитания операцией сложения, что существенно упрощает работу процессора и увеличивает его быстродействие.

Дополнительный код отрицательного числа А, хранящегося в n ячейках, равен: $2^{n-1}-|A|$

Для получения дополнительного кода отрицательного числа надо:

1. Модуль числа записать в прямом коде в n двоичных разрядах.
2. Получить **обратный код** числа, для этого значения всех битов инвертировать (заменить 1 0 и наоборот).
3. К полученному обратному коду прибавить 1.

Запишем дополнительный код отрицательного числа: $A_{10}=-2010_{10}=-11111011010_2$

				1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Прямой код модуля	$ -2010_{10} $			0	0	0	0	0	1	1	1	1	1	0	1	1	0	1	0
Обратный код	Инверсия			1	1	1	1	1	0	0	0	0	0	1	0	0	1	0	1
	+1	+		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Дополнительный код				1	1	1	1	1	0	0	0	0	0	1	0	0	1	1	0

Диапазон $[-2^{n-1}, 2^{n-1}-1]$

1.2.1. Короткое целое (8 бит) $[-128, 127]$

1.2.2. Целый (2 байта) $[-2^{16-1}, 2^{16-1}-1]=[-32768, 32767]$

1.2.3. Длинное целое (4 байта, 32 бита)

1.3. Вещественные числа – формат с плавающей запятой.

Положение запятой в числе может меняться.

Формат чисел с плавающей запятой базируется на экспоненциальной форме записи, в которой может быть представлено любое число.

$$A = m \cdot q^n$$

$1/n \leq m < 1$ – нормализованная мантисса числа ((правильная дробь (числитель меньше знаменателя), меньше 1)),

q – основание системы счисления,

n – порядок числа (степень основания СС).

Например, преобразуем число 471,28 в экспоненциальную форму с нормализованной мантиссой: $0,47128 \cdot 10^3$

При записи числа с плавающей запятой, выделяются разряды для хранения знака мантиссы, знака порядка, порядка и мантиссы.

1.3.1. Число обычной (одинарной) точности (4 байта – 32 бита)

8 бит – порядок и его знак, 24 – знак и мантисса

0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Знак	Порядок (макс=2 ⁷ =127)							Знак	Мантисса. Макс.знач. положительной мантиссы: 2 ²³ -1 ≈ 2 ²³ -2 ^{2,3*10} ≈ 1000 ^{2,3} =10 ^{2,3*3} ≈ 10 ⁷																												
Знак и порядок								Знак и мантисса																													
8 бит								24 бит																													

Максимальное значение числа: $2^{127}=1,70141 \dots \cdot 10^{38}$

1.3.2. Число двойной точности (8 байт – 64 бита)

(?????) 11 бит – порядок и его знак, 53 – знак и мантисса

Примеры.

Задание N 19

Соответствие между выражением: $(-8_{10} - 5_{10} =)$ и выражением в дополнительном двоичном коде:

Варианты ответов

- 1.1001₂ + 1.0110₂ =
- 1.1011₂ + 1.0101₂ =
- 1.1000₂ + 1.1011₂ =
- 1.1000₂ + 0.1010₂ =

Задание N 16

Два числа, которые обозначают одно и то же количество

Варианты ответов

- 33₁₀ и 203₄
- 110₄ и 20₁₀
- 213₄ и 25₁₀
- 25₁₀ и 37₄

Задание N 26

Соответствие между выражением: $(-12_{10} - 6_{10} =)$ и выражением в дополнительном двоичном коде:

Варианты ответов

- 1.0110₂ + 0.1010₂ =
- 1.0100₂ + 1.1010₂ =
- 0.1100₂ + 0.0110₂ =
- 1.1100₂ + 1.0110₂ =

5.3. Инженерный калькулятор

Перевод целых чисел из одной стандартной для компьютера систем счисления в другую, а также действия над числами в этих системах счисления можно производить с помощью инженерного калькулятора:

1. Запустить программу: Пуск → Все программы → Стандартные → Калькулятор
2. Вид \ Инженерный
3. Установить переключатель в одно из положений:
 - Hex - шестнадцатеричный
 - Dec - десятичный
 - Oct - восьмеричный
 - Bin - двоичный
4. Выполнить необходимые действия.

Примечание. Если необходимо перевести число из системы счисления с основанием (h=2, 8, 10, 16) в систему счисления с основанием (t=2, 8, 10, 16), то надо установить переключатель в положение h и набрать число, после чего установить переключатель в положение t.

6. Кодирование информации

6.1. Языки и алфавиты

Различают естественные языки и искусственные (формальные) языки. *Естественные* языки развивались веками и служат для общения людей между собой. *Формальные* языки разрабатываются для специальных применений. Примером формальных языков могут служить языки программирования, языки кодирования информации для ее передачи, хранения и т.п.

Каждый язык имеет свой алфавит. Под *алфавитом* понимают набор используемых символов. Под мощностью алфавита понимают количество составляющих алфавит символов.

Кодом называют совокупность знаков, предназначенных для представления информации в соответствии с определенными правилами. Такое представление называется кодированием. Кодировать информацию с целью ее передачи, хранения, преобразования.

Кодирование – это операция преобразования информации из одной формы представления (знаковой системы) другую.

Средством кодирования служит таблица соответствия знаковых систем, которая устанавливает взаимно однозначное соответствие между знаками или группами знаков двух различных знаковых систем.

Например, звук *a*, издаваемый человеком кодируется в некоторых языках буквой А. Буква А в азбуке Морзе кодируется так .- (точка-тире). В компьютере буква А латинского алфавита в десятичной системе кодируется числом 65. В свою очередь, число 65 в привычной для компьютера двоичной системе кодируется так: 0100001.

Одно и то же понятие на различных языках может кодироваться различными способами. Например, слово «стол» - это код в русском алфавите известного предмета мебели. В других языках это же предмет кодируются иначе.

В естественных языках традиционно сложились некоторые неоднозначности. Например, таким слова «коса» или «ключ» имеют несколько значений. Подобные неоднозначности в формальных языках, как правило, не допустимы.

Различны и алфавиты языков кодирования. Это латинские буквы и русские буквы, арабские и римские цифры, азбука Морзе и т.п. Количество символов в алфавитах тоже различно: русский – 33, латинский – 26, китайский – несколько тысяч иероглифов.

Количество и графическое изображение символов в формальных языках тоже может существенно различаться.

Наименьший по числу знаков алфавит имеет только один знак. Пусть это знак 1. Тогда три света светофора можно закодировать, например, так красный – 1, желтый – 11, зеленый – 111.

В процессе обмена информацией часто приходится производить операции кодирования и декодирования информации.

Например, при вводе знака алфавита в компьютер путем нажатия соответствующей клавиши на клавиатуре происходит кодирование знака, то есть перевод его в машинный код. При выводе знака на экран монитора или принтер происходит обратный процесс – декодирование, когда из машинного кода знак преобразуется в его графическое изображение.

В информатике и ВТ широко используется алфавит, имеющий два знака, так как удалось создать надежно работающие технические устройства, которые могут со 100%-ной надежностью сохранять и распознавать не более двух различных состояний (цифр):

- Электромагнитные реле (замкнуто / разомкнуто) широко использовались в первых ЭВМ;
- Участок поверхности магнитного носителя информации (намагничен / не намагничен);
- Участок поверхности лазерного диска (отражает / не отражает);
- Триггер (структурная единица ОЗУ) может устойчиво находиться в одном из двух состояний.

Принято одно состояние обозначать цифрой 0, а другое – 1. Такое кодирование называется *двоичным или бинарным*. В соответствии с этим введена и наименьшая единица информации - бит.

Бит (Bit- Binary двоичный Digit знак) – минимальное количество информации, содержащейся в сообщении вида да/нет, вкл/выкл, истина/ложь и т.п., т.е. в сообщениях, уменьшающих неопределенность в два раза.

Одного бита информации достаточно, что передать слово «да» или «нет», закодировать, например, состояние электролампочки. Кстати на многих выключателях так и пишут «1/0». Взгляд на выключатель снимает для нас неопределенность в его состоянии. При этом мы получаем количество информации, равное одному биту.

Если нужно закодировать в двоичном алфавите три цвета светофора, то потребуется минимум два бита: например, 00 – красный, 01 – желтый, 10 – зеленый. Для кодирования сторон света тоже достаточно два бита (С 00, В 01, Ю 10, З 11).

При кодировании 8 величин уже потребуется 3 бита и т.п.

7. Единицы измерения информации

Таким образом, следующей по величине единицей является байт, причем 1 байт=8 бит.

Также существуют и другие, более крупные единицы измерения информации: Кбайт, Мбайт, Гбайт и др.

В информатике система образования кратных единиц измерения информации отличается от принятых в большинстве наук. Традиционные метрические системы единиц, например, Международная система единиц СИ, в качестве множителей кратных единиц используют коэффициент 10^n (где $n=3,6,9$) и так далее, что соответствует приставкам Кило (10^3), Мега (10^6), Гига (10^9), Тера (10^{12}) и т.д.

Компьютер оперирует числами не в десятичной, а в двоичной системе счисления, поэтому в кратных единицах измерения количества информации используют коэффициент 2^n .

Соотношение единиц измерения информации:

1 байт	$=2^3$ бит	= 8бит
1 Кбайт	$=2^{10}$ байт	=1024 байт
1 Мбайт	$=2^{10}$ Кбайт	=1024 Кбайт
	$=2^{20}$ байт	=1048576 байт
1 Гбайт	$=2^{10}$ Мбайт	=1024 Мбайт
	$=2^{20}$ Кбайт	=1048576 Кбайт
	$=2^{30}$ Кбайт	=1073741824 байт

В большинстве КП каждому символу соответствует последовательность из 8 нулей и единиц. Всего существует 256 разных последовательностей из 8 нулей и единиц – это позволяет закодировать 256 разных символов, например, большие и малые русские и латинские буквы, цифры, пробел, знаки препинания, знаки арифметических действий и т.д.

В 1 символе 1 байт информации. Значит, любое сообщение можно представить в виде последовательности 0 и 1, а следовательно его можно измерить в битах или байтах.

Количество бит или байт в сообщении называется **информационным объемом сообщения**.

Например, объем фразы Это я! 6 байтов=48 бит

№1. Определить примерный информационный объем учебника, если на каждой странице 40 строк текста в среднем в строке 60 символов, а количество страниц в книге 420.

Решение: $60 \times 40 \times 420 = 1008000$ (байт) = $1008000 / 1024$ (Кбайт) = 984,375 (Кбайт)

Следует различать понятия количество и объем информации. Поясним разницу на примере. Пусть имеется большой энциклопедический словарь в 1400 страниц и такого же объема книга, на каждой странице которой все строки заполнены вопросительными знаками. Объем словаря и книги одинаков – 1440 страниц текста. Поэтому для хранения потребуется одинаковый объем памяти. Количество же информации существенно различается. Всю информацию книги можно записать одной фразой «1400 страниц с ?» но даже эта фраза для большинства имеет сомнительную ценность.

В целях экономии памяти при записи информации на хранение широко используют процедуру сжатия или архивации. Так информацию словаря можно сжать примерно в 10 раз. При этом следует помнить, что уменьшается не количество информации, а только ее объем.

8. Измерения количества информации

Какое количество информации содержится, к примеру, в тексте романа "Война и мир", в фресках Рафаэля или в генетическом коде человека? Ответа на эти вопросы наука не даёт и, по всей вероятности, даст не скоро.

А возможно ли объективно измерить количество информации? Важнейшим результатом теории информации является вывод:

В определенных, весьма широких условиях можно пренебречь качественными особенностями информации, выразить её количество числом, а также сравнить количество информации, содержащейся в различных группах данных.

В настоящее время получили распространение подходы к определению понятия "количество информации", основанные на том, что информацию, содержащуюся в сообщении, можно нестрого трактовать в смысле её новизны или, иначе, уменьшения неопределённости наших знаний об объекте.

Так, американский инженер Р. Хартли (1928 г.) процесс получения информации рассматривает как выбор одного сообщения из конечного наперёд заданного множества из N равновероятных сообщений, а количество информации I , содержащееся в выбранном сообщении, определяет как двоичный логарифм N .

$$\text{Формула Хартли: } I = \log_2 N.$$

Допустим, нужно угадать одно число из набора чисел от единицы до ста. По формуле Хартли можно вычислить, какое количество информации для этого требуется: $I = \log_2 100 \approx 6,644$. То есть сообщение о верно угаданном числе содержит количество информации, приблизительно равное 6,644 единиц информации.

Приведем другие примеры равновероятных сообщений:

1. при бросании монеты: "выпала решка", "выпал орел";

2. на странице книги: "количество букв чётное", "количество букв нечётное".

Определим теперь, являются ли равновероятными сообщения "первой выйдет из дверей здания женщина" и "первым выйдет из дверей здания мужчина". Однозначно ответить на этот вопрос нельзя. Все зависит от того, о каком именно здании идет речь. Если это, например, станция метро, то вероятность выйти из дверей первым одинакова для мужчины и женщины, а если это военная казарма, то для мужчины эта вероятность значительно выше, чем для женщины.

Для задач такого рода американский учёный Клод Шеннон предложил в 1948 г. другую формулу определения количества информации, учитывающую возможную неодинаковую вероятность сообщений в наборе.

$$\text{Формула Шеннона: } I = -\sum_{i=1}^n p_i \cdot \log_2 p_i,$$

где I – количество информации, N – количество возможных событий, p_i — вероятность i -го события

Этот подход к определению количества информации называется **вероятностным**.

Легко заметить, что если вероятности p_1, \dots, p_N равны, то каждая из них равна $1/N$, и формула Шеннона превращается в формулу Хартли.

6.2. Кодирование текстовой информации

Если каждому символу алфавита сопоставить определенное целое число (например, порядковый номер), то с помощью двоичного кода можно кодировать и текстовую информацию.

Последовательность символов называют **словом**. Восьмибитовое двоичное слово называется **байтом**. С помощью байтов можно закодировать 256 различных объектов.

Технически это выглядит очень просто, однако всегда существовали достаточно веские организационные сложности. В первые годы развития ВТ они были связаны с отсутствием необходимых стандартов, а в настоящее время вызваны, наоборот, избытием одновременно действующих и противоречивых стандартов. Для того чтобы весь мир одинаково кодировал текстовые данные, нужны единые таблицы кодирования, а это пока невозможно из-за противоречий между символами национальных алфавитов, а также противоречий корпоративного характера.

Для английского языка, захватившего де-факто нишу международного средства общения, противоречия уже сняты. Институт стандартизации США (ANSI — American National Standard Institute) ввел в действие систему кодирования ASCII (American Standard Code for Information Interchange) — стандартный код информационного обмена США. В системе ASCII закреплены две таблицы кодирования — базовая и расширенная. Базовая таблица закрепляет значения кодов от 0 до 127, а расширенная относится к символам с номерами от 128 до 255

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0		▸	␣	␣	␣	␣	␣	␣	А	Р	а	▯	▯	▯	▯	▯
1	⊕	◀	!	1	A	Q	a	q	Б	С	б	▯	▯	▯	▯	▯
2	⊙	‡	"	2	B	R	b	r	В	Т	в	▯	▯	▯	▯	▯
3	♥	‡	#	3	C	S	c	s	Г	У	г	▯	▯	▯	▯	▯
4	♦	⌘	\$	4	D	T	d	t	Д	Ф	д	▯	▯	▯	▯	▯
5	♣	Ⓢ	%	5	E	U	e	u	Е	Х	е	▯	▯	▯	▯	▯
6	♠	-	&	6	F	V	f	v	Ж	Ц	ж	▯	▯	▯	▯	▯
7	·	±	'	7	G	W	g	w	З	Ч	э	▯	▯	▯	▯	▯
8	■	↑	(8	H	X	h	x	И	Ш	и	▯	▯	▯	▯	▯
9		↓)	9	I	Y	i	y	И	Щ	й	▯	▯	▯	▯	▯
A		→	*	:	J	Z	j	z	К	Ъ	к	▯	▯	▯	▯	▯
B	⊗	←	+	;	K	[k	[Л	Ы	л	▯	▯	▯	▯	▯
C	♀	└	,	<	L	\	l	l	М	Ь	м	▯	▯	▯	▯	▯
D		↔	-	=	M]	m]	Н	Э	н	▯	▯	▯	▯	▯
E	♠	▲	.	>	N	^	n	~	О	Ю	о	▯	▯	▯	▯	▯
F	⊗	▼	/	?	O	_	o	␣	П	Я	п	▯	▯	▯	▯	▯

Таблица кодов ASCII (American Standard Code for Information Interchange)

Первые 32 кода базовой таблицы, начиная с нулевого, отданы производителям аппаратных средств (в первую очередь производителям компьютеров и печатающих устройств). В этой области размещаются так называемые управляющие коды, которым не соответствуют никакие символы языков, и, соответственно, эти коды, не выводятся ни на экран, ни на устройства печати, но ими можно управлять тем, как производится вывод прочих данных.

Начиная с кода 32 по код 127 размещены коды символов английского алфавита, знаков препинания, цифр, арифметических действий и некоторых вспомогательных символов.

Аналогичные системы кодирования текстовых данных были разработаны и в других странах.

На сегодняшний момент существует 1 кодировка латинского алфавита: ASCII и 5 кодовых таблиц для русских символов: КОИ-8, CP1251, C3866, Mac, ISO.

Одна из самых распространенных кодировок русских символов носит название КОИ-8 (код обмена информацией, восьмизначный) — ее происхождение относится ко временам действия Совета Экономической Взаимопомощи государств Восточной Европы.

Международный стандарт, в котором предусмотрена кодировка символов русского алфавита, носит название кодировки ISO (International Standard Organization — Международный институт стандартизации).

Универсальная система кодирования текстовых данных

Да недавнего времени байта было достаточно, чтобы закодировать все символы текста в русском и латинском алфавита: буквы, цифры, знаки препинания, управляющие сигналы — все то, что передается ПК с клавиатуры. С развитием информатики байт начал сдерживать возможность увеличения количества используемых символов. В настоящее время завершается переход на двух байтовой кодирование символов.

Такая система, основанная на 16-разрядном кодировании символов, получила название *универсальной* — **UNICODE**. Шестнадцать разрядов позволяют обеспечить уникальные коды для 65 536

различных символов — этого поля достаточно для размещения в одной таблице символов большинства языков планеты. Начиная с 1997 года эту кодировку поддерживает Windows&Office.

Несмотря на тривиальную очевидность такого подхода, простой механический переход на данную систему долгое время сдерживался из-за недостаточных ресурсов средств вычислительной техники (в системе кодирования UNICODE все текстовые документы автоматически становятся вдвое длиннее). Во второй половине 90-х годов технические средства достигли необходимого уровня обеспеченности ресурсами, и сегодня мы наблюдаем постепенный перевод документов и программных средств на универсальную систему кодирования. Для индивидуальных пользователей это еще больше добавило забот по согласованию документов, выполненных в разных системах кодирования, с программными средствами, но это надо понимать как трудности переходного периода.

6.3. Кодирование графики и звука

Так как вся информация в ПК обрабатывается и сохраняется в цифровом виде, то и графическая и звуковая информация должны быть представлены в цифровом виде. С этой целью ее кодируют.

Кодирование растровой графики

Весь экран дисплея делится на точки – пиксели, подсвечивание которых создает видимое отображение текста и рисунков. Различают растровую и векторную графику.

При растровом способе отображения растрового изображения приходится иметь дело со всеми точками, входящими в изображение.

Информационный объем растрового изображения (Q) определяется как произведение числа входящих в изображение точек (N) на информационный объем одной точки (q), который зависит от количества возможных цветов, т.е. $Q=Nq$

Изображение, глубина цвета	Информационный объем точки	Примечание
Черно-белое	1 бит (1- точка подсвечивается, 0 – не подсвечивается)	
Оттенки серого, между белым и черным еще 6 серых – всего 8	$q=\log_2 8 = 3$ бита	
Цветное изображение получается за счет различной яркости трех основных цветов RGB. Цветные изображения могут отображаться в различных режимах, соответственно изменяется и информационный объем точки:		
RGB, 16 цветов	$q=\log_2 16 = 4$ бита	
RGB, 256 цветов	$q=\log_2 256 = 8$ бит = 1 байт	
RGB, 65 636 цветов	$q=\log_2 65536 = 16$ бит = 2 байта	Кодирование цветной графики 16-разрядными двоичными числами называется режимом High Color .
RGB, 16 777 216 цветов	$q=\log_2 16\,777\,216 = 24$ бит = 3 байта Если для кодирования яркости каждой из основных составляющих использовать по 256 значений (восемь двоичных разрядов), как это принято для полутоновых черно-белых изображений, то на кодирование цвета одной точки надо затратить 24 разряда.	Режим представления цветной графики с использованием 24 двоичных разрядов называется полноцветным (True Color).
При кодировании CMYK (голубой (Cyan, C), пурпурный (Magenta, M) и желтый (Yellow, Y), черный цвет обозначается буквой K, потому, что буква B уже занята синим цветом) для представления полноцветной (True Color) графики надо иметь 32 двоичных разряда.		

Например, 1) объем растрового черно-белого изображения 20x30 точек: $Q=20*30*1=600$ бит=75 байт
2) объем растрового цветного изображения из 256 цветов - 20x30 точек: $Q=20*30*8=4800$ бит=600 байт

3) В процессе преобразования растрового графического файла количество цветов уменьшилось с 65536 до 256. Во сколько раз уменьшится информационный объем файла?

В 1-м случае информационный объем точки $q = \log_2 65536 = 16$ бит, во втором - $q = \log_2 256 = 8$ бит.

Ответ: в 2 раза

Информационный объем видеоизображения обуславливает расчет объема видеопамати, необходимой для хранения изображения со всего экрана монитора.

Например, для размера экрана 1280x1024 точки и режиме True Color объем видеопамати должен быть не менее $Q = 1280 * 1024 * 3$ байта = 3,75 Мбайта

Кодирование векторной графики

Векторное изображение состоит из набора элементарных деталей – графических примитивов (линии, прямоугольник, окружности, дуги и т.п.). Положение этих элементов на экране определяется координатами точек. Кроме этого задаются цвет отображаемых элементов, толщина линии и другие характеристики. Информация о векторном рисунке кодируется обычным способом, как хранятся тексты, т.е. хранится не векторное изображение, а только его координаты и характеристики. Поэтому для хранения векторных изображений требуется гораздо меньше памяти, чем для хранения растровых.

При запуске программы с векторным рисунком он создается каждый раз заново.

Кодирование звука

Звук представляет собой звуковую волну с непрерывно меняющейся амплитудой и частотой. Человек слышит звук с частотами от 16 Герц (колебаний в секунду) до 20 килогерц. До недавнего времени звуковые сигналы передавались, записывались на пластинках и магнитных лентах и воспроизводились в виде амплитудных колебаний. Но для того, чтобы ПУ мог обрабатывать звук, непрерывный звуковой сигнал должен быть превращен в последовательность электрических импульсов (двоичных 0 и 1). Он становится дискретным. Непрерывная волна разбивается на отдельные временные участки, для которых устанавливается определенная амплитуда и громкость. («гладкая кривая заменяется ступеньками»). Уровни громкости можно рассматривать как набор возможных состояний. Чем больше уровней, тем большее количество информации будет нести значение каждого уровня, тем лучше звучания. Обычная звуковая карта обеспечивает 16-битную глубину кодирования звука. Количество различных уровней можно рассчитать по формуле: $N = 2^I = 2^{16} = 65536$, где I – глубина звука.

Каждому значению амплитуды звукового сигнала присваивается 16-битный код.

Качество кодирования зависит и от количества измерений сигнала в единицу времени, т.е. частотой дискретизации. Частота от 8кГц (радиотрансляция) до 48кГц (аудио CD). Возможны также моно и стерео режимы.

Например, оценить информационный объем стереоаудиофайл длительностью звучания 1 сек при высоком качестве звука (16 битов, 48кГц).

$16 \text{ бит} * 48000 \text{ Гц} * 2 \text{ (стерео)} = 153600 \text{ бит} = 192 \text{ 000 байт} = 187,5 \text{ Кбайт}$

((Приемы и методы работы со звуковой информацией пришли в вычислительную технику наиболее поздно. В итоге методы кодирования звуковой информации двоичным кодом далеки от стандартизации. Множество отдельных компаний разработали свои корпоративные стандарты, но если говорить обобщенно, то можно выделить два основных направления.

Метод FM (Frequency Modulation) основан на том, что теоретически любой сложный звук можно разложить на последовательность простейших гармонических сигналов разных частот, каждый из которых представляет собой правильную синусоиду, а следовательно, может быть описан числовыми параметрами, то есть кодом. В природе звуковые сигналы имеют непрерывный спектр, то есть являются аналоговыми. Их разложение в гармонические ряды и представление в виде дискретных цифровых сигналов выполняют специальные устройства — **аналогово-цифровые преобразователи (АЦП)**. Обратное преобразование для воспроизведения звука, закодированного числовым кодом, выполняют **цифро-аналоговые преобразователи (ЦАП)**. При таких преобразованиях неизбежны потери информации, связанные с методом кодирования, поэтому качество звукозаписи обычно получается не вполне удовлетворительным и соответствует качеству звучания простейших электромузыкальных инструментов с окрасом, характерным для электронной музыки. В то же время данный метод кодирования

обеспечивает весьма компактный код, и потому он нашел применение еще в те годы, когда ресурсы средств вычислительной техники были явно недостаточны.

Метод таблично волнового (Wave-Table) синтеза лучше соответствует современному уровню развития техники. Если говорить упрощенно, то можно сказать, что где-то в заранее подготовленных таблицах хранятся образцы звуков для множества различных музыкальных инструментов (хотя не только для них). В технике такие образцы называют сэмплами. Числовые коды выражают тип инструмента, номер его модели, высоту тона, продолжительность и интенсивность звука, динамику его изменения, некоторые параметры среды, в которой происходит звучание, а также прочие параметры, характеризующие особенности звука. Поскольку в качестве образцов используются «реальные» звуки, то качество звука, полученного в результате синтеза, получается очень высоким и приближается к качеству звучания реальных музыкальных инструментов.))