

ИНФОРМАЦИЯ

Презентация создана
учителем математики и информатики
Ковалевой Анной Леонидовной
ГБОУ СОШ №341 г.СПб
2013-2014



Существуют несколько определений информации

- Информация (от англ. informatio – осведомление, разъяснение, изложение) – ЭТО
- связанные между собой сведения, неизвестные до их получения, которые дают представление о явлениях или объектах
- совокупность сигналов (сообщений), воспринимаемых человеком непосредственно или с помощью специальных устройств, расширяющее его знания об окружающем мире и протекающих в нем процессах
- сведения об объектах окружающего мира, которые повышают уровень осведомленности человека
- знания, которые человек получает из разных источников



Классификация информации:

- 1) ПО СПОСОБУ ВОСПРИЯТИЯ
- 2) ПО СПОСОБУ ПРЕДСТАВЛЕНИЯ
- 3) ПО ОБЩЕСТВЕННОМУ ЗНАЧЕНИЮ
- 4) ПО СТАДИИ ОБРАБОТКИ
- 5) ПО СТАБИЛЬНОСТИ
- 6) ПО МЕСТУ ВОЗНИКНОВЕНИЯ



ПО СПОСОБУ ВОСПРИЯТИЯ:

- **Визуальная** (восприятие органами зрения)
- **Аудиальная** (восприятие через органы слуха)
- **Вкусовая** (восприятие через органы ощущения вкуса)
- **Тактильная** (восприятие через органы осязания)
- **Обонятельная** (восприятие через органы обоняния)



80-90 %
информации —

здоровый человек получает при помощи зрения

8-15%
информации —

здоровый человек получает при помощи слуха

1-5%
информации —

здоровый человек получает при помощи остальных органов чувств

Основной орган чувств	Животное
Нос (обоняние)	Лисы, собаки
Уши (главная информация – звуковая)	Летучие мыши



ПО СПОСОБУ ПРЕДСТАВЛЕНИЯ:

- **ТЕКСТОВАЯ** (представляет собой совокупность букв, цифр и специальных символов, с помощью которых происходит отображение информации на бумаге или ином физическом носителе)
- **ЧИСЛОВАЯ** (числа и числовые выборки данных в различных системах счисления)
- **ГРАФИЧЕСКАЯ** (графики, схемы, диаграммы, рисунки)
- **ЗВУКОВАЯ** (человеческая речь, аудиозаписи)
- **КОМБИНИРОВАННАЯ** (синтез нескольких форм представления информации)



ФОРМЫ ПРЕДСТАВЛЕНИЯ ИНФОРМАЦИИ



ЗНАКОВАЯ

(текстовая, числовая):

- 1) Информация на естественном языке
- 2) Информация на формальном языке

ОБРАЗНАЯ

(графическая, звуковая):

- 1) Изображения
- 2) Звук



ПО ОБЩЕСТВЕННОМУ ЗНАЧЕНИЮ:

- **ЛИЧНАЯ** (информация, предназначенная для личного пользования человека)
- **СПЕЦИАЛЬНАЯ** (информация, предназначенная для использования определенной группой лиц для специальных целей)
- **МАССОВАЯ** (информация, предназначенная для широкого круга лиц, общественности)



ПО СТАДИИ ОБРАБОТКИ:

- **ПЕРВИЧНАЯ** (непосредственно полученная информация)
- **ВТОРИЧНАЯ** (результат обработки первичной информации)
- **ПРОМЕЖУТОЧНАЯ** (информация, используемая как исходные данные для последующих расчетов, для построения графиков, таблиц, диаграмм и др.)
- **РЕЗУЛЬТАТИВНАЯ** (информация, возникающая в процессе обработки первичной и промежуточной информации, используется для принятия решений или публичного представления)



ПО СТАБИЛЬНОСТИ:

- **ПЕРЕМЕННАЯ** (информация об объектах или событиях, динамически меняющаяся со временем)
- **ПОСТОЯННАЯ** (неизменная многократно используемая в течение длительного периода времени информация)



ПО МЕСТУ ВОЗНИКНОВЕНИЯ:

- **ВХОДНАЯ** (информация, поступающая в организацию, отдел или подразделение)
- **ВЫХОДНАЯ** (информация, исходящая из одной организации в другую)
- **ВНУТРЕННЯЯ** (возникает внутри рассматриваемого объекта или структуры)
- **ВНЕШНЯЯ** (возникает вне рассматриваемого объекта или структуры)



Классификация знаний



**ДЕКЛАРАТИВНЫЕ
ЗНАНИЯ**
(Я знаю, что ...)



**ПРОЦЕДУРНЫЕ
ЗНАНИЯ**
(Я знаю, как ...)

ОТЛИЧИЕ ПОНЯТИЙ ДАННЫЕ И ИНФОРМАЦИЯ:

Данные – это результаты наблюдений над объектами и явлениями, которые по каким-либо причинам не используются, а только хранятся. Как только данные начинают использоваться, они превращаются в информацию.



СВОЙСТВА ИНФОРМАЦИИ:

- **АКТУАЛЬНОСТЬ** (ценность информации в данный момент времени)
- **ДОСТОВЕРНОСТЬ** (определяется получением информации без искажений)
- **ПОНЯТНОСТЬ** (выражение информации на языке, понятном тому или иному лицу)
- **ПОЛНОТА** (степень достаточности полученной информации для принятия необходимого решения в данный момент времени)
- **ПОЛЕЗНОСТЬ** (ценность информации при решении задач конкретного вида)
- **ОБЪЕКТИВНОСТЬ** (степень независимости информации от субъективного мнения человека или группы лиц)



● **САМОСТОЯТЕЛЬНАЯ РАБОТА:** Выберите правильный ответ:

● №1. Если вы собираетесь провести выходной день на природе, то своевременной информацией для вас будет...

- а) сведения о погоде в такой же день прошлого года
- б) прогноз погоды на выходной день

● №2. Слухи, вымыслы, непроверенные гипотезы – это...

- а) достоверная информация
- б) недостоверная информация

● №3. Информация о том, как с помощью подручных средств добыть огонь, будет для вас наиболее полезной...

- а) если вы попадете на необитаемый остров
- б) в нашей повседневной жизни

● №4. О правилах дорожного движения каждый водитель автомобиля...

- а) должен обладать полной информацией
- б) может иметь неполную информацию

● №5. Волга впадает в Каспийское море – это ...

- а) достоверная информация
- б) недостоверная информация



ЗНАКИ И ЗНАКОВЫЕ СИСТЕМЫ:

- **ЗНАК** – это заменитель объекта – предмета, явления, свойства или отношения (примеры знаков: пиктограммы, ярлыки, символы и т.д.)
- **ЗНАКОВАЯ СИСТЕМА** – это множество всех входящих в нее знаков (АЛФАВИТ ЗНАКОВОЙ СИСТЕМЫ) и правила оперирования с этими знаками



ЯЗЫК как знаковая система



ЕСТЕСТВЕННЫЕ ЯЗЫКИ

(языки, используемые для общения людей – например, русский, английский, французский и т.д.)



ФОРМАЛЬНЫЕ (ИСКУССТВЕННЫЕ) ЯЗЫКИ

(языки, используемые специалистами в профессиональной деятельности – например, язык математики, химических формул, нотная грамотность, азбука Морзе, системы счисления, языки программирования и т.д.)



Язык — знаковый способ представления информации. С помощью языка информация передаётся в знаковой форме.

Знаковая система состоит из упорядоченного набора знаков (символов), который называется **алфавитом**. Полное количество символов алфавита называется **мощностью алфавита**. Например, алфавит русского языка состоит из 33 букв, латинского — из 26 букв.

Минимально возможное количество символов в алфавите равно двум. Существующие технические электронные устройства надёжно сохраняют и распознают только два различных состояния, поэтому именно такой алфавит используется в компьютере. Он называется **двоичным алфавитом**, его символы — цифры 0 и 1. С помощью этих двух символов можно представить любую информацию в компьютере.

Если для сообщения используется двоичный алфавит и длина сообщения — один знак, можно составить два различных сообщения (0 и 1). Если длина сообщения — два знака, можно сформировать $2 \times 2 = 2^2 = 4$ разных комбинации (00, 01, 10, 11). При длине сообщения три знака получим $2 \times 2 \times 2 = 2^3 = 8$ различных комбинаций (000, 001, 010, 011, 100, 101, 110, 111). И т. д.

Комбинации символов двоичного алфавита называют **двоичными кодами**. Количество знаков в коде называется **длиной кода**.



ФОРМУЛА для вычисления количества комбинаций в зависимости от мощности алфавита

Если мощность алфавита равна 2, длина кода равна L , можно составить $K = 2^L$ различных двоичных кодов.

Если мощность алфавита равна N , длина кода L , можно составить $K = N^L$ различных кодовых комбинаций.

ФОРМУЛА ХАРТЛИ: $N = p^i$

p – количество вариантов (используемых символов) МОЩНОСТЬ АЛФАВИТА

i – количество информации (длина строки символов) ДЛИНА КОДА

N – число равновероятных событий КОЛИЧЕСТВО КОМБИНАЦИЙ



- ПРИМЕР №1.
- Какой должна быть минимальная длина двоичного кода, если требуется составить 18 различных комбинаций?
- РЕШЕНИЕ:

$$2^x = 18, x \approx 5 \text{ СИМВОЛОВ}$$

- ОТВЕТ: 5 СИМВОЛОВ

Решить самостоятельно:

КАКОЙ ДОЛЖНА БЫТЬ МИНИМАЛЬНАЯ ДЛИНА ДВОИЧНОГО КОДА, ЕСЛИ ТРЕБУЕТСЯ СОСТАВИТЬ 70 РАЗЛИЧНЫХ КОМБИНАЦИЙ?



- ПРИМЕР №2.
- Световое табло состоит из лампочек. Каждая лампочка может находиться в одном из трех состояний («включено», «выключено» или «мигает»). Какое наименьшее количество лампочек должно находиться на табло, чтобы с его помощью можно было передать 18 различных сигналов?

- РЕШЕНИЕ:

$$3^x = 18, x \approx 3$$

- ОТВЕТ: 3 лампочки

Решить самостоятельно:

РЕШИТЬ ПРИМЕР №2 ПРИ УСЛОВИИ 100 СИГНАЛОВ



ШИФРОВАНИЕ ИНФОРМАЦИИ

Во все времена наиболее уязвимым местом при передаче секретной информации был канал связи. Поэтому ещё древние люди начали изобретать способы **шифрования** для защиты информации. При шифровании сообщение изменяется способом, известным источнику и приёмнику, и передаётся по каналу связи. Однако узнать истинный смысл сообщения могут только те приёмники, которые знают, как именно, по каким правилам источник искажил сообщение. Способы шифрования информации изучает **криптология**.

В настоящее время известно большое количество самых разных шифров, отличающихся сложностью, стойкостью к дешифрованию и т. д.



ШИФР ЦЕЗАРЯ

Среди древнейших шифров одним из самых известных является ШИФР ЦЕЗАРЯ. Для того, чтобы зашифровать слово шифром Цезаря, необходимо произвести замену буквы алфавита на другую, отстоящую от нее на определенное количество букв. Длина сдвига называется КЛЮЧОМ ШИФРА ЦЕЗАРЯ.

ПРИМЕР ШИФРА ЦЕЗАРЯ С КЛЮЧОМ = 3

X	Y	Z	A	B	C	...	U	V	W	X	Y	Z
A	B	C	D	E	F	...	X	Y	Z	A	B	C

Восстановление исходного сообщения из полученного называется **дешифровкой**. Для восстановления исходного сообщения надо применить шифр Цезаря с ключом -3 к полученному сообщению.



- Зашифруйте слово **WINDOWS** шифром Цезаря с ключом = 3.

РЕШЕНИЕ:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C

WINDOWS ----- **ZLQGRZV**



АЗБУКА МОРЗЕ

- Использовался для первой телеграфной машины
- (тире-длинный сигнал,
- точка-короткий сигнал,
- пауза-отсутствие сигнала)



СИМВОЛЫ АЗБУКИ МОРЗЕ

А • —	Б — • • •	В • — —	Г — — •
Д — • •	Е •	Ж • • • —	З — — • •
И • •	Й • — — —	К — • —	Л • — • •
М — — —	Н — •	О — — — —	П • — — •
Р • — •	С • • •	Т — — —	У • • —
Ф • • — •	Х • • • •	Ц — • — •	Ч — — — •
Ш — — — — —	Щ — — — • —	Ъ • — — • — •	Ы — • — —
Ь — • • —	Э • • — • •	Ю • • — — —	Я • — • —

РАСШИФРУЙТЕ СЛЕДУЮЩЕЕ СЛОВО:

— — — — — • — • — — — — — • — — — — — • — — — — —

ОТВЕТ: МОЛОДЦЫ



АЗБУКА БРАЙЛЯ

- Создана французским педагогом Луи Брайлем для слепых (буквы этого алфавита выдавлены на листе плотной бумаги в виде точек определенным образом)



A	B	C	D	E	F	G	H	I
⠠	⠠	⠠	⠠	⠠	⠠	⠠	⠠	⠠
J	K	L	M	N	O	P	Q	R
⠠	⠠	⠠	⠠	⠠	⠠	⠠	⠠	⠠
S	T	U	V	W	X	Y	Z	
⠠	⠠	⠠	⠠	⠠	⠠	⠠	⠠	

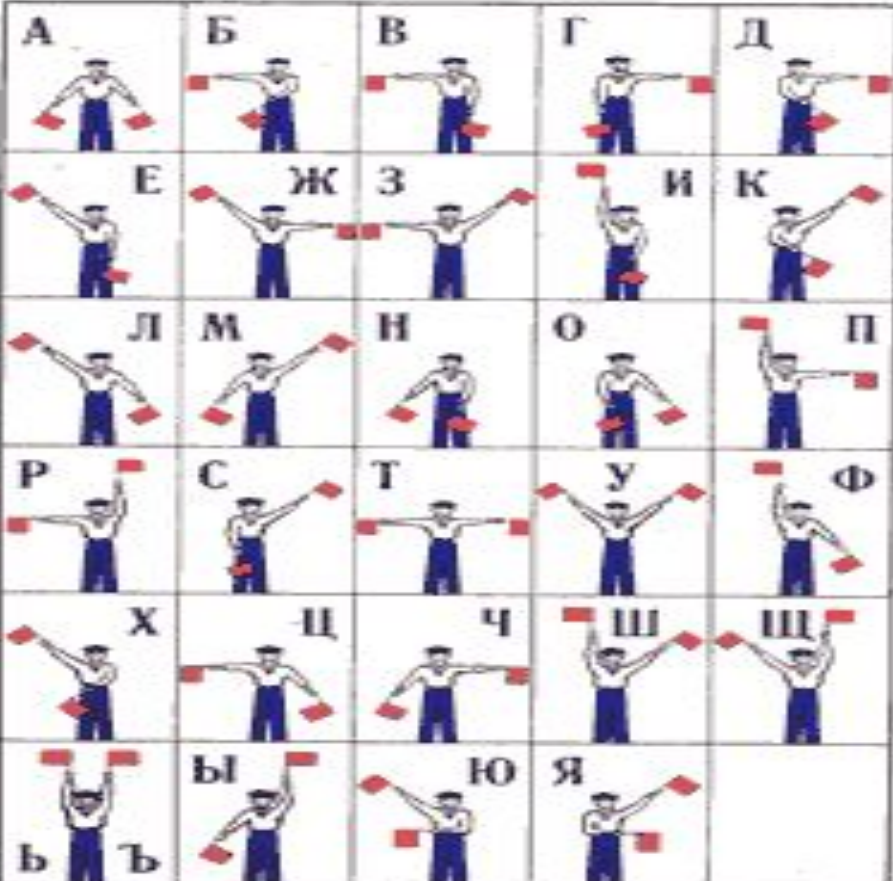
Кодирование

1. Пользуясь азбукой Брайля (рис. 43), расшифруйте слово;



ОТВЕТ: WINDOWS

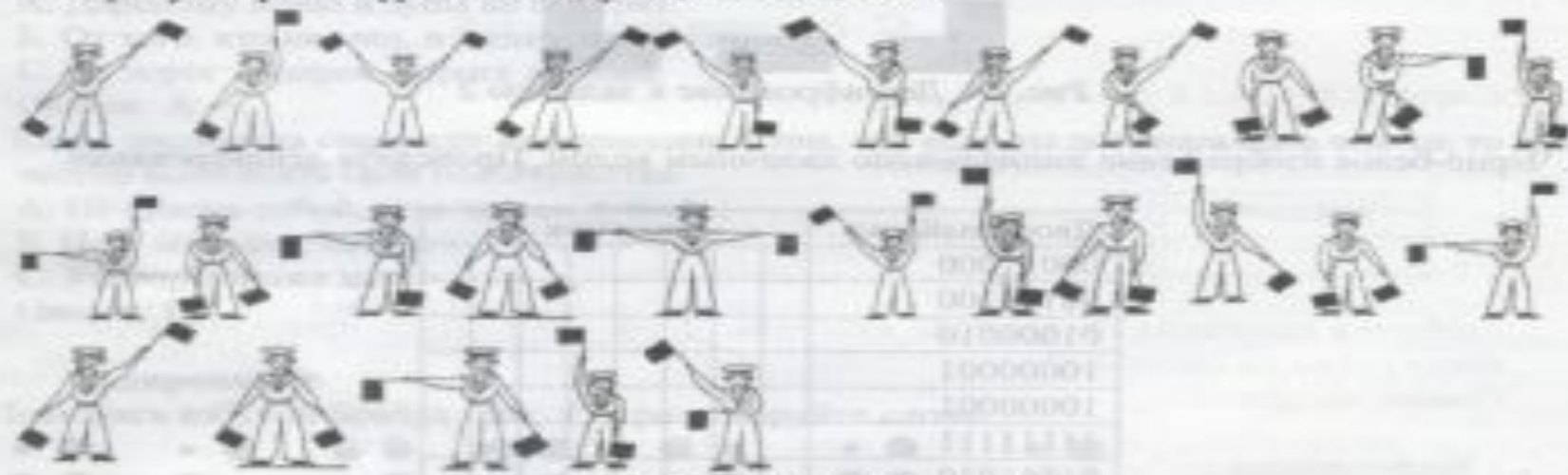




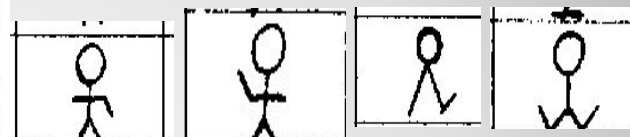
СЕМАФОРНАЯ РУССКАЯ АЗБУКА (азбука жестов)

ОТВЕТ: Мы умеем кодировать
информацию

Попробуйте расшифровать следующее сообщение:



Пляшущие человечки



ОТВЕТ: Дело



А		К		Х	
Б		Л		Ц	
В		М		Ч	
Г		Н		Ш	
Д		О		Щ	
Е		П		Ъ	
Ё		Р		Ы	
Ж		С		Ь	
З		Т		Э	
И		У		Ю	
Й		Ф		Я	

ЗАДАНИЕ

Расшифруйте изречение знаменитого математика К.Ф.Гаусса:



Шифр

	○	△	□	◇
◻	ф	т	р	к
◼	у	ц	е	м
◼	а	и	н	-

Ответ: Математика - царица наук, а арифметика - царица математики.



Самостоятельная работа:

- №1. Зашифровать слово «OFFICE» шифром Цезаря с ключом = 3.
- №2. Зашифровать слово «среда» азбукой Морзе.
- №3. Зашифровать шифром из задачи про Гаусса слова:
 - 1 ряд: «центнер»
 - 2 ряд: «муфта»
 - 3 ряд: «нитка»



ЗАДАНИЕ:

- Выполните письменно №9 и №10 из учебника (стр.24)
- №9. Вождь племени Мульти поручил своему министру разработать двоичный код и перевести в него всю важную информацию. Достаточно ли пятиразрядного двоичного кода, если алфавит, используемый племенем Мульти, содержит 26 символов?
- №10. От разведчика была получена следующая зашифрованная радиোগрамма, переданная с использованием азбуки Морзе:
- . . - . . - - . . - - - .
- При передаче радиোগраммы было потеряно разбиение на буквы, но известно, что в радиোগрамме использовались только нижеуказанные буквы. Определите текст радиোগраммы.

И	А	Н	Г	Ч
..	.-	-.	---.	----.

ОТВЕТЫ:

№9. Да, достаточно, пятиразрядный двоичный код позволяет кодировать 32 различных символа алфавита. №10. НАИГАЧ

ДОМАШНЕЕ ЗАДАНИЕ:

Составить самостоятельно задачу на кодирование информации и решить ее.



ПРОБЛЕМНАЯ СИТУАЦИЯ

Расшифруйте сообщение с помощью кода Морзе:

— — — — — — — — — —

А • —	Б — • • •	В • — —	Г — — •
Д — • •	Е •	Ж • • • —	З — — • •
И • •	Й • — — —	К — • —	Л • — • •
М — —	Н — •	О — — —	П • — — •
Р • — •	С • • •	Т —	У • • —
Ф • • — •	Х • • • •	Ц — • — •	Ч — — — •
Ш — — — —	Щ — — • —	Ъ • — — • •	Ы — • — —
Ь — • • —	Э • • — • •	Ю • • — —	Я • • • —

КАКОЙ ЖЕ ОТВЕТ?

МОМО? ОМОМ? ШМШ? Вопрос: можно ли однозначно расшифровать это сообщение?



КЛАССИФИКАЦИИ КОДОВ

- **РАВНОМЕРНЫЕ КОДЫ** и **НЕРАВНОМЕРНЫЕ КОДЫ**



Равномерный код:

Кодовые слова всех символов исходного алфавита имеют одинаковую длину
(например Код Брайля)



Неравномерный код:

Кодовые слова имеют разную длину
(например Код Морзе)



- **Особенность неравномерных кодов:**
- **НЕДОСТАТКИ:** В некоторых случаях однозначно декодировать сообщение, записанное с использованием неравномерного кода, невозможно (поскольку началом кодовых слов могут служить несколько букв). Поэтому для однозначного декодирования лучше использовать РАВНОМЕРНЫЕ КОДЫ.
- **ДОСТОИНСТВА:** Высокая скорость передачи сообщений (наиболее часто встречающиеся символы имеют самые короткие комбинации)



- **Особенность равномерных кодов:**
- **НЕДОСТАТКИ:** Сообщение, записанное таким кодом имеет большую длину (по сравнению с неравномерным кодом) и поэтому передается медленнее.
- **ДОСТОИНСТВА:** Такое сообщение всегда можно декодировать однозначно.



ПРЕФИКСНЫЕ КОДЫ

- это коды со словами переменной длины, в котором ни одно кодовое слово не является началом другого кодового слова.

Возьмем сообщение: 100111011010

Пример префиксного кода: 0 10 11

(так как сообщение с помощью него можно расшифровать однозначно)

Пример непrefиксного кода: 0 10 100 11

(нельзя расшифровать однозначно)



ЗАДАЧА:

Для кодирования некоторой последовательности, состоящей из букв А, Б, В, Г и Д, решили использовать неравномерный двоичный код, позволяющий однозначно декодировать двоичную последовательность, появляющуюся на приёмной стороне канала связи. Для букв А, Б, В и Г использовали такие кодовые слова: А – 100, Б – 101, В – 111, Г – 110.

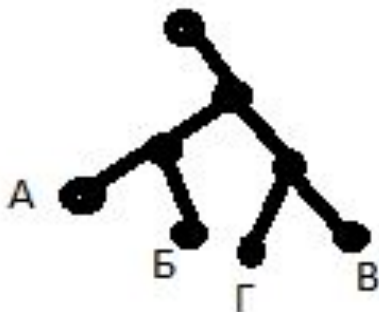
Укажите, каким кодовым словом из перечисленных ниже может быть закодирована буква Д. Код должен удовлетворять свойству однозначного декодирования. Если можно использовать более одного кодового слова, укажите кратчайшее из них.

Варианты ответов: 1 0 01 10



РЕШЕНИЕ:

Построим дерево декодирования (0 - ветвь влево, 1 - ветвь вправо)



ТОГДА:

1 и 10 – нет
(у этих точек
есть ветви)
Подходит 0 и
01 (но в
ответ пишем
самое
короткое 0)

Условие Фано

- 1) закодированное сообщение можно однозначно декодировать с начала, если выполняется *условие Фано*:
никакое кодовое слово не является началом другого кодового слова
- 2) закодированное сообщение можно однозначно декодировать с конца, если выполняется *обратное условие Фано*:
никакое кодовое слово не является окончанием другого кодового слова

Для однозначного декодирования достаточно, чтобы выполнялось условие Фано или обратное условие Фано



ПРИМЕР. Для кодирования некоторой последовательности, состоящей из букв А, Б, В, Г и Д, используется неравномерный двоичный код, позволяющий однозначно декодировать полученную двоичную последовательность. Вот этот код: А-00, Б-010, В-011, Г-101, Д-111. Можно ли сократить для одной из букв длину кодового слова так, чтобы код по-прежнему можно было декодировать однозначно? Коды остальных букв меняться не должны. Выберите правильный вариант ответа.

- 1) для буквы Б – 01 2) это невозможно
3) для буквы В – 01 4) для буквы Г – 01



РЕШЕНИЕ:

Для однозначного декодирования достаточно, чтобы выполнялось условие Фано или обратное условие Фано.

- 1) Проверяем 1 ответ: А-00 Б-01 В-011 Г-101 Д-111

Прямое условие Фано не выполнено (начало Б совпадает с началом В)

Обратное условие Фано не выполнено (конец Б совпадает с концом Г)

- 2) Проверяет 3 ответ: А-00 Б-010 В-01 Г-101 Д-111

Прямое условие Фано не выполнено (начало В совпадает с началом Б)

Обратное условие Фано не выполнено (конец В совпадает с концом Г)

- 3) Проверяет 4 ответ: А-00 Б-010 В-011 Г-01 Д-111

Прямое условие Фано не выполнено (начало Г совпадает с началом В,Б)

Обратное условие Фано **выполнено** (конец Г не совпадает ни с каким концом) **ОТВЕТ: 4**

Ваня шифрует русские слова, записывая вместо каждой буквы её номер в алфавите (без пробелов). Номера букв даны в таблице.

А	1	И	11	У	21	Э	31
Б	2	К	12	Ф	22	Ю	32
В	3	Л	13	Х	23	Я	33
Г	4	М	14	Ц	24		
Д	5	Н	15	Ч	25		
Е	6	О	16	Ш	26		
Ё	7	П	17	Щ	27		
Ж	8	Р	18	Ъ	28		
З	9	С	19	Ы	29		
И	10	Т	20	Ь	30		

Некоторые шифровки можно расшифровать не одним способом. Например, 311333 может означать «ВАЛЯ», может – «ЭЛЯ», а может – «ВААВВВ».

Даны четыре шифровки:

4224

8131

7334

4105

Только одна из них расшифровывается единственным способом. Найдите ее и расшифруйте. То, что получилось, запишите в качестве ответа.





От разведчика была получена следующая зашифрованная радиограмма, переданная с использованием азбуки Морзе:

- • • - • • - - • • - - - - •

При передаче радиограммы было потеряно разбиение на буквы, но известно, что в радиограмме использовались только следующие буквы:

И	А	Н	Г	Ч
••	•-	-•	--•	---•

Определите текст радиограммы.

- 1) ГАИГАЧ
- 2) НАИГАН
- 3) НАИГАЧ
- 4) ГАИГАН

№3

Мальчики играли в шпионов и закодировали сообщение придуманным шифром. В сообщении присутствуют только буквы из приведённого фрагмента кодовой таблицы:

А	Б	В	Г	Д	Е	Ж
10	110	12	122	0	22	102

Расшифруйте сообщение 110101210102:

Ответ:



САМОСТОЯТЕЛЬНАЯ РАБОТА (на оценку)

Задания с выбором одного ответа

Пример 4.5. От разведчика была получена следующая зашифрованная радиোগрамма, переданная с использованием азбуки Морзе:

• — — • • • • • — • — — • • • — • —

При передаче радиোগраммы было потеряно разбиение на буквы, но известно, что в радиোগрамме использовались только следующие буквы:

А	К	О	Р	Ф
—	• —	• • —	• • • —	• • • •

Определите текст радиোগраммы. В ответе укажите, сколько букв было в исходной радиোগрамме.

- | | |
|------|-------|
| 1) 7 | 3) 9 |
| 2) 8 | 4) 12 |

Пример 4.6. От разведчика была получена следующая зашифрованная радиোগрамма, переданная с использованием азбуки Морзе:

• — — • • — • — • • • — — • • — — —

При передаче радиোগраммы было потеряно разбиение на буквы, но известно, что в радиোগрамме использовались только следующие буквы:

Ф	И	В	А	Н
—	• — •	• •	• — — —	• — — •



ПОДВЕДЕНИЕ ИТОГОВ УРОКА:

- С каким понятием вы сегодня познакомились?
- Расскажите, какие бывают коды?
- Чем они отличаются?
- В чем их достоинства и недостатки?
- **ДОМАШНЕЕ ЗАДАНИЕ:**
- выучить теоретический материал по теме и составить самостоятельно задачу на выбранный вами код.



Алфавитный подход к измерению информации

- Каждый символ некоторого сообщения имеет определенный **информационный вес** – несет фиксированное количество информации.

- Информационный вес символа двоичного алфавита принят за **МИНИМАЛЬНУЮ ЕДИНИЦУ ИЗМЕРЕНИЯ ИНФОРМАЦИИ** – **1 БИТ** т.е. двоичная цифра 0 или 1 несет 1 бит. (от англ. **binary digit** – двоичная цифра)



Информационный вес символа произвольного алфавита

$$N = 2^i$$

N – мощность исходного алфавита

I – информационный вес символа алфавита
(выражается в битах)

Алфавит любого естественного или формального языка всегда можно заменить двоичным алфавитом



Примеры:

- №1. Производится бросание симметричной четырехгранной пирамидки. Какое количество информации мы получаем в зрительном сообщении о ее падении на одну из граней?

Ответы: 1) 1 бит 2) 2 бита 3) 4 бита 4) 1 байт

№2. Из непрозрачного мешочка вынимают шарики с номерами, и известно, что информационное сообщение о номере шарика несет 5 битов информации. Определите количество шариков в мешочке?

№3. Какое количество информации при игре в крестики-нолики на поле размером 4×4 клетки получит второй игрок после первого хода первого игрока?

№4 Какое количество информации несет двоичный код 10101010?

- Ответы:

- №1. - 2

Т.к. $N=4$, то $i=2$ бита

- №2. - 32

Т.к. $i = 5$ бит, то $N = 32$

- №3. 4 бита

Так как $N = 4*4 = 16$, то $i = 4$ бита

- №4. 8 бит

Т.к. одна двоичная цифра (один разряд двоичного числа) несет 1 бит, то умножаем количество цифр 8 на 1.

$$N = 2^i$$

Чем большее количество знаков содержит алфавит, тем большее количество информации несет один знак. Например, буква русского алфавита несет 5 битов, т.к.

$$2^5 = 32 < 33 \text{ букв}$$

- ВНИМАНИЕ!
- На самом деле количество информации, которое несет знак, зависит от вероятности его получения: чем менее вероятно получение знака, тем больше его информационная емкость.
- С точки зрения теории информации информационная емкость буквы «А» меньше, чем у буквы «Ф» (А встречается чаще, чем Ф)

Информационный объем сообщения

- $I = k \cdot i$

- k – количество символов в сообщении
- i – информационный вес символа
- I – информационный объем сообщения



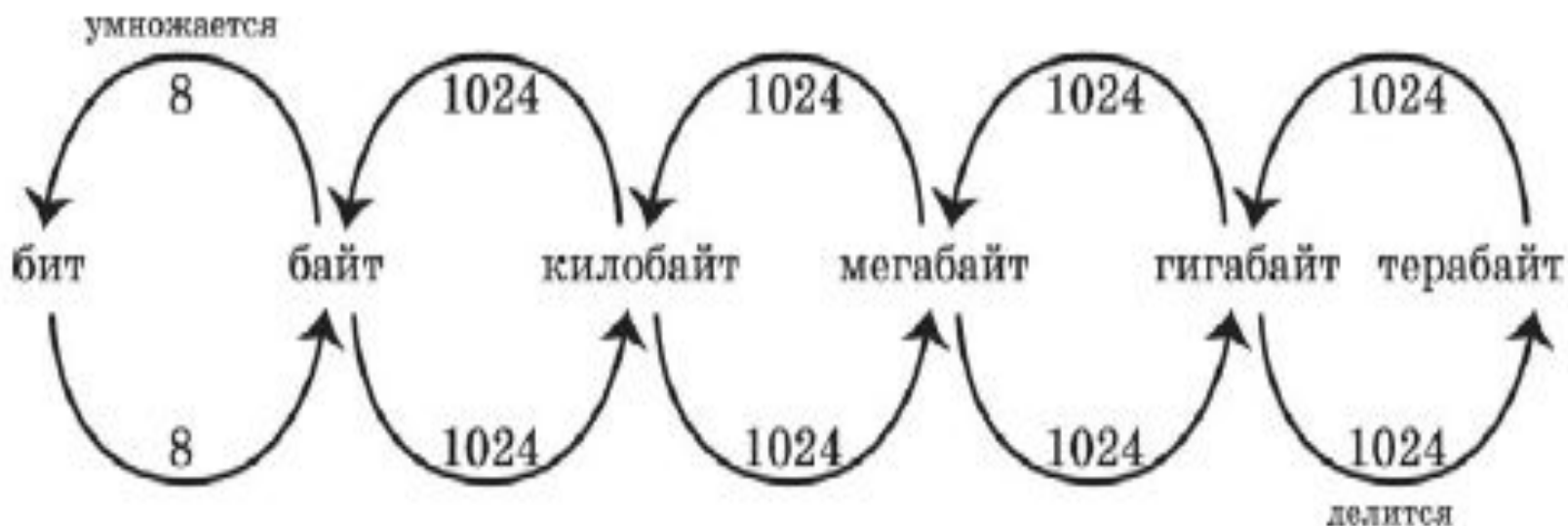
ЕДИНИЦЫ ИЗМЕРЕНИЯ ИНФОРМАЦИИ:

- 1 байт = 2^3 (8) бит (это информационный вес символа алфавита мощностью 256)
- 1 Кбайт (Килобайт) = 2^{10} (1024) байт = 2^{13} (8192) бит
- 1 Мбайт (Мегабайт) = 2^{10} (1024) Кбайт =
= 2^{20} (1.048.576) байт = 2^{23} (8.388.608) бит
- 1 Гбайт (Гигабайт) = 2^{10} (1024) Мбайт =
= 2^{20} (1.048.576) Кбайт = 2^{30} (1.073.741.824) байт =
= 2^{33} (8.589.934.592) бит
- 1 Тбайт (Терабайт) = 2^{10} (1024) Гбайт =
= 2^{20} (1.048.576) Мбайт = 2^{30} (1.073.741.824) Кбайт =
= 2^{40} (1.099.511.627.776) байт =
= 2^{43} (8.796.093.022.208) бит
- 1 Эбайт (Эксобайт) = 1024 Тбайта



Последовательность действий при переводе одних единиц измерения информации в другие приведена на схеме 1.

Схема 1



n	1	2	3	4	5	6	7	8	9	10	11	12	13
2^n	2	4	8	16	32	64	128	256	512	1024	2048	4096	8192



Пример №1.

● Алфавит племени Пульти содержит 8 символов. Каков информационный вес символа этого алфавита?

● РЕШЕНИЕ:

$$N = 2^i$$

N – мощность алфавита

i – информационный вес

$$8 = 2^i$$

символа этого алфавита

Значит $i = 3$ бита

Самостоятельная работа:

- 1) Некоторый алфавит содержит 512 символов. Каков информационный вес символа этого алфавита?
- 2) Некоторый алфавит содержит 70 символов. Каков информационный вес символа этого алфавита?



Пример №2.

Сообщение, записанное буквами 32-символьного алфавита, содержит 140 символов. Какое количество информации оно несет?

$$k=140, N=32, I=? \quad N = 2^i$$

$$32 = 2^i$$

Значит $i = 5$ бит

$$I = k \cdot i = 140 \cdot 5 = 700 \text{ битов}$$

Самостоятельная работа:

1) Сообщение, записанное буквами 16-символьного алфавита, содержит 78 символов. Какое количество информации оно несет?

2) Задан алфавит из 64 символов. Буквами данного алфавита записано предложение. Какое количество информации оно несет, если известно, что оно содержит 25 символов?



Пример №3.

- Информационное сообщение объемом 720 битов состоит из 180 символов. Какова мощность алфавита, с помощью которого записано это сообщение?
- $I=720$ битов, $k=180$, N -?

$$I = k \cdot i$$

$$i = I / k = 720 / 180 = 4 \text{ бита}$$

$$N = 2^i = 2^4 = 16 \text{ СИМВОЛОВ}$$

Самостоятельная работа:

- 1) Информационное сообщение состоит из 60 символов, объем этого сообщения 420 бит. Какова мощность алфавита, с помощью которого записано это сообщение?
- 2) Найти информационный вес символа алфавита, если известно, что сообщение из 24 символов занимает 192 бита.



Пример №4.

- Считая, что каждый символ кодируется одним байтом, определите, чему равен информационный объем следующего высказывания *Жан-Жака Руссо* в битах:
- **«Тысячи путей ведут к заблуждению, к истине – только один».**
- РЕШЕНИЕ: Считаем вручную количество символов в предложении (включая пробел, запятые, тире, кавычки, точку и т.д.). Получаем 59 символов. Тогда $I = 59 * 1 = 59$ байт = $59 * 8$ бит = 472 бит.



Пример №5.

- Для регистрации на сайте некоторой страны пользователю требуется придумать пароль. Длина пароля – ровно 11 символов. В качестве символов используются десятичные цифры и 12 различных букв местного алфавита, причём все буквы используются в двух начертаниях: как строчные, так и заглавные (регистр буквы имеет значение!).
- Под хранение каждого такого пароля на компьютере отводится минимально возможное и одинаковое целое количество байтов, при этом используется посимвольное кодирование и все символы кодируются одинаковым и минимально возможным количеством битов.
- Определите объём памяти, который занимает хранение 60 паролей.
- 1) 540 байт 2) 600 байт 3) 660 байт 4) 720 байт
- РЕШЕНИЕ:
- Мощность алфавита = $10+12+12=34$ (10 цифр, 12 заглавных и 12 строчных букв). По формуле Хартли получаем: $i=6$ бит (т.к. $2^5 = 32 < 34$, $2^6 = 64 > 34$). Тогда $I=11*6=66$ бит (на 1 пароль) ≈ 9 байт (на 1 пароль)
- И, окончательно, т.к. 60 паролей, то $60*9=540$ байт
- ОТВЕТ:1)



Пример №6.

- Объем сообщения, содержащего 4096 символов, равен 1/512 части Мбайта. Какова мощность алфавита, с помощью которого записано это сообщение?
- 1) 8 2) 16 3) 4096 4) 16384
- РЕШЕНИЕ:
- $I = 1/512$ часть Мбайта = $2^{23}/512$ бит = $2^{23}/2^9 = 2^{14}$ бит
- $i = I / k = 2^{14}$ бит / 4096 симв. = $2^{14} / 2^{12} = 2^2 = 4$ бита (на 1 символ)
- $N = 2^i = 2^4 = 16$ символов.
- ОТВЕТ: 2)



Самостоятельная работа:



№1. Считая, что каждый символ кодируется одним байтом, оцените информационный объем в битах следующего предложения:

«Мой дядя самых честных правил, Когда не в шутку занемог, Он уважать себя заставил И лучше выдумать не мог».

№2. Выберите самый маленький объем информации из приведенных ниже:

- 1) 1000 Кбайт
- 2) 1 000 000 байт
- 3) 10 000 000 бит
- 4) 1 Мбайт

№3. Считая, что каждый символ кодируется 16-ю битами, оцените информационный объем в битах следующего предложения: **Блажен, кто верует, тепло ему на свете!**

№4. Текстовый документ, состоящий из 3072 символов, хранился в 8-битной кодировке КОИ-8. Этот документ был преобразован в 16-битную кодировку Unicode. Укажите, какое дополнительное количество Кбайт потребуется для хранения документа. В ответе запишите только число.

№5. Сколько существует различных последовательностей из символов «плюс» и «минус», длиной ровно в пять символов?

- 1) 64
- 2) 50
- 3) 32
- 4) 20

№6. Два текста содержат одинаковое количество символов. Первый текст составлен в алфавите мощностью 16 символов, а второй текст – в алфавите из 256 символов. Во сколько раз количество информации во втором тексте больше, чем в первом?

- 1) 12
- 2) 2
- 3) 4
- 4) 16

ОТВЕТЫ

- №1. 864 бита
- №2. 2)
- №3. 624 бита
- №4. 3 Кбайта
- №5. 3)
- №6. 2)



Задание: Выполнить письменно из учебника №6,7,8,9 (с.29-30)

№6. Определите количество информации в сообщении из K символов алфавита мощностью N , заполняя таблицу:

N	$N = 2^i$	i (битов)	K	$I=K*i$
8			200	
16			110	
64			120	
128			100	
256			80	

№7. Племя Мульти пишет письма, Пульти пользуется 32-символьным алфавитом. Вожди племен обменялись письмами. Письмо племени Мульти содержит 120 символов, а письмо племени Пульти – 96. Сравните информационные объемы сообщений, содержащихся в письмах.

№8. Информационное сообщение объемом 650 битов состоит из 130 символов. Каков информационный вес каждого символа этого сообщения?

№9. Выразите количество информации в различных единицах, заполняя таблицу:

Бит	Байт	Кбайт
24576		
	2048	
		1,5
2^{13}		
	2^{11}	



- Домашнее задание:
- №10, 11, 12 (с.30 учебника)
- Подготовиться к контрольной работе

✓ 10. Информационное сообщение объемом 375 байтов состоит из 500 символов. Каков информационный вес каждого символа этого сообщения? Какова мощность алфавита, с помощью которого было записано это сообщение?

✓ 11. Для записи текста использовался 64-символьный алфавит. Какое количество информации в байтах содержат 3 страницы текста, если на каждой странице расположено 40 строк по 60 символов в строке?

✓ 12. Сообщение занимает 6 страниц по 40 строк, в каждой строке записано по 60 символов. Информационный объем всего сообщения равен 9000 байтов. Каков информационный вес одного символа? Сколько символов в алфавите языка, на котором записано это сообщение?



**РАЗЛИЧНЫЕ
СИСТЕМЫ
КОДИРОВАНИЯ**



Существуют стандартные таблицы кодов. Они могут использовать один или два байта для кодирования одного символа.

Широко используется таблица кодов, известная как **стандарт ASCII** (**American Standard Code for Information Interchange** — Американский стандартный код для обмена информацией), использующая один байт для кодирования одного символа. ASCII представляет собой кодировку для представления десятичных цифр, символов латинского и национального алфавитов, знаков препинания, символов арифметических операций и управляющих символов. Управляющие символы называют непечатаемыми символами, к ним относятся такие, как «перевод строки» (код символа 10), «возврат каретки» (код 13) и др.

Первая половина кодовой таблицы содержит **стандартные** символы ASCII (символы с кодами 0 — 127), они одинаковые во всех странах.



ASCII

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 2 таблицы кодирования – базовая и расширенная. Базовая закрепляет значения кодов от 0 до 127, а расширенная относится к символам с номерами от 128 до 255.



- Текст состоит из символов – букв, цифр, знаков препинания и т.д.
- Компьютер различает вводимые символы по их двоичному коду. Нажимая на определенную клавишу на клавиатуре, в компьютер поступает определенная последовательность электрических импульсов разной силы, которую можно представить в виде последовательностей нулей (0) и единиц (1).



- **0-31: 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.



- **32-127:**

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

- (Например, 48-57 это коды цифр 0-9, 65-90 это коды букв A-Z, 97-122 это коды букв a-z и т.д.)



- **128-255:**

- Коды с номерами от 128 до 255 используются для кодирования букв национального алфавита, символов национальной валюты и т.п. Поэтому в кодовых таблицах для разных языков одному и тому же коду соответствуют разные символы.
- Существует несколько вариантов расширенной таблицы ASCII
- (примеры расширения:
- CP-866-используется для ОС MS-DOS,
- CP-1251-используется для Windows,
- KOI-8- используется в сети Интернет, Unicode)



Таблицу кодов не надо запоминать, но следует помнить последовательность символов:

- 1) знаки препинания и арифметических операций;
- 2) цифры от 0 до 9;
- 3) прописные символы латинского алфавита;
- 4) строчные символы латинского алфавита.



Вторая часть кодовой таблицы (символы с кодами 128 — 255) называют **расширенными кодами ASCII**. В расширенные коды ASCII включают символы национальных алфавитов, например символы кириллицы. Но даже с учётом этих дополнительных знаков алфавиты многих языков не удаётся охватить при помощи 256 знаков. По этой причине существуют различные варианты кодировки ASCII, включающие символы разных языков.

Отсутствие согласованных стандартов привело к появлению различных кодовых таблиц (вернее, различных вторых частей кодовых таблиц) для кодирования символов кириллицы, среди которых

- международный стандарт ISO 8859;
- кодовая таблица фирмы Microsoft CP-1251 (кодировка Windows);
- кодовая таблица, применяемая в ОС Unix KOI8R и др.

По этой причине тексты на русском языке, набранные с использованием одной кодовой таблицы, невозможно прочитать при использовании другой кодовой таблицы.

В настоящее время в компьютерах широко применяется стандарт кодирования **Unicode (Юникод)**, в котором для кодирования одного символа отводятся один байт, два байта или четыре байта. Первые 128 символов Юникода совпадают с символами ASCII. Остальная часть кодовой таблицы включает символы, используемые в основных языках мира.

Универсальная система кодирования ТЕКСТОВЫХ ДАННЫХ



Если проанализировать организационные трудности, связанные с созданием **единой системы** кодирования текстовых данных, то можно прийти к выводу, что они вызваны **ограниченным набором кодов (256, так как на символ отводится 8 бит, а $2^8=256$)**. В то же время, очевидно, что если, например, кодировать символы не восьмиразрядными двоичными числами, а числами с большим количеством разрядов, то и диапазон возможных значений кодов станет намного больше.

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. **UNICODE** описывает алфавиты всех известных языков, в том числе и мертвых. Например, он содержит коды придуманных языков (язык эльфов и Мордора из «Властелина колец»)

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 о согласовании документов, выполненных в разных системах кодирования, с программными средствами**, но это надо понимать как трудности переходного периода.



Дан символ. Как определить его код?

- 1) По кодовой таблице ASCII (на бумаге)
- 2) С помощью текстового процессора Word (вставка-символ-шрифт Times New Roman- код символа)

Дан код. Как определить по нему символ?

- 1) По кодовой таблице ASCII (на бумаге)
- 2) С помощью текстового редактора Блокнот: при нажатой клавише Alt
- -ввести 3 цифры кода (если CP-866) или
- -4 цифры кода (первый – 0, если CP-1251)
- и затем отпустить Alt