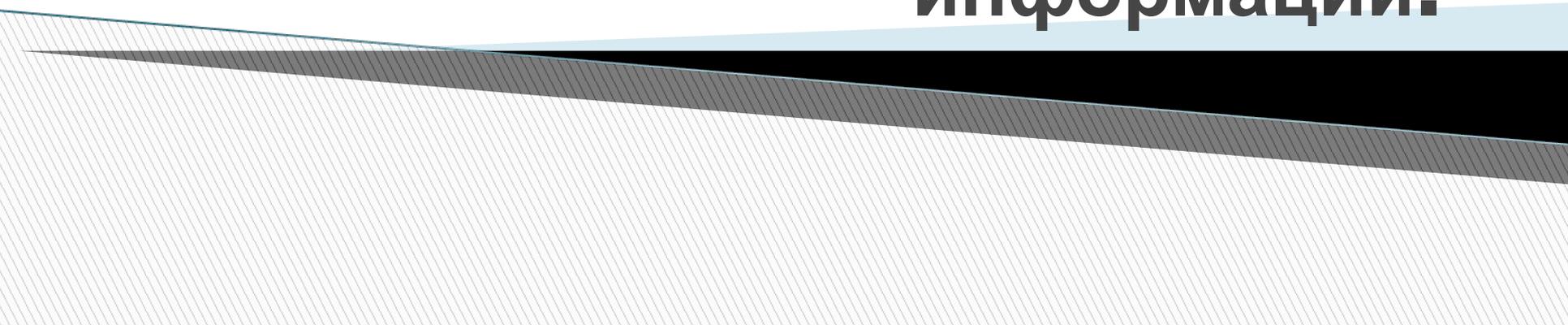


**Кодирование информации.
Двоичное кодирование.
Вероятностный и алфавитный
подход к определению
количества информации.
Алфавитный подход к
определению количества
информации.**



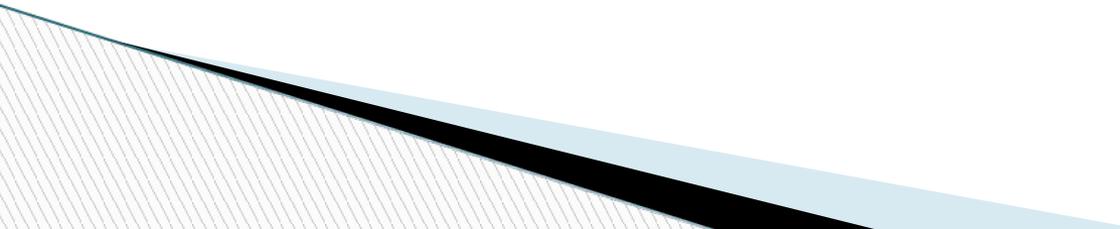
Давайте вспомним, что такое:

□ Код-

набор символов для представления информации.

□ Кодирование –

процесс представления информации в виде кода.



Одну и ту же информацию можно кодировать различными способами:

▣ **КОМПЬЮТЕР – русский язык**

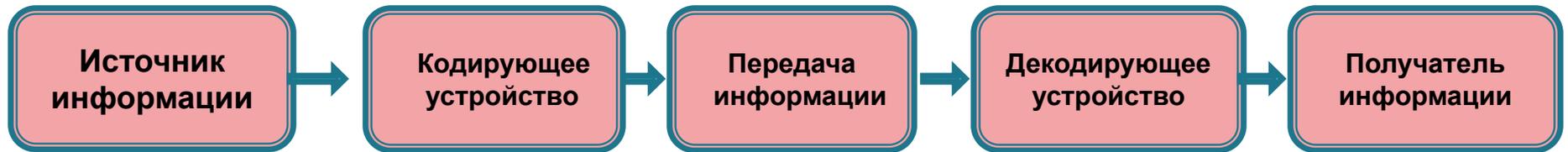


▣ **Код Морзе**

▣ **COMPUTER – английский язык**

▣ **67 79 77 80 85 84 69 82- код
ASCII**

В процессе обмена информацией кроме кодирования происходит процесс декодирования



НАПРИМЕР



Вспомним!

Кодирование чисел

- ▣ Сегодня такой способ представления информации, с помощью языка содержащего два символа 0 и 1, широко используется в технических устройствах.
- Эти два символа 0 и 1 принято называть битами (от англ. **binary digit** – двоичный знак).
- ▣ **Бит** – наименьшая единица измерения информации и обозначается двоичным числом.
- Более крупной единицей изменения объема информации принято считать 1 байт, который состоит из 8 бит.
- ▣ 1 **байт** = 8 битов.

Кодирование текстовой информации

- Так же применяется двоичное кодирование!
- Каждому символу алфавита сопоставимое определенное целое число- код символа.

СКОЛЬКО ЖЕ БИТ НЕОБХОДИМО ДЛЯ КОДИРОВАНИЯ СИМВОЛОВ ?

Подсчитаем примерное достаточное количество символов и по формуле вычислим необходимое количество бит!

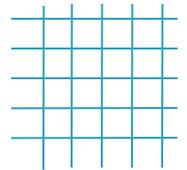
- 33 прописные буквы+33 строчные буквы + 26 английских строчных букв+26 прописных английских букв + 10 цифр + знаки припинания + скобки и знаки математических операций + специальные символы (@ ? # ? \$? %) + знаки псевдографики =**256**
- Для кодирования такого количества информации достаточно **8 бит - 256=28**
- **8 бит- 1 байт**
- С помощью 1 байта можно закодировать **256** различных символов

Кодирование графической и звуковой информации

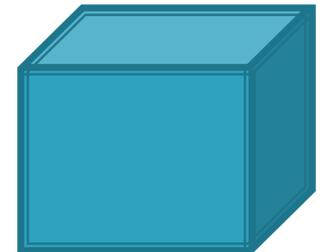
- ▣ **Изображение и звук** передают сигналы, имеющие **непрерывную аналоговую форму**.
- ▣ Для преобразования сигнала необходимо преобразовать его в дискретную форму- путем дискретизации, т.е. разбиением непрерывного графического изображения или звукового сигнала на отдельные элементы. Каждому элементу присваивается конкретное значение в виде двоичного кода

Приемы кодирования графической информации

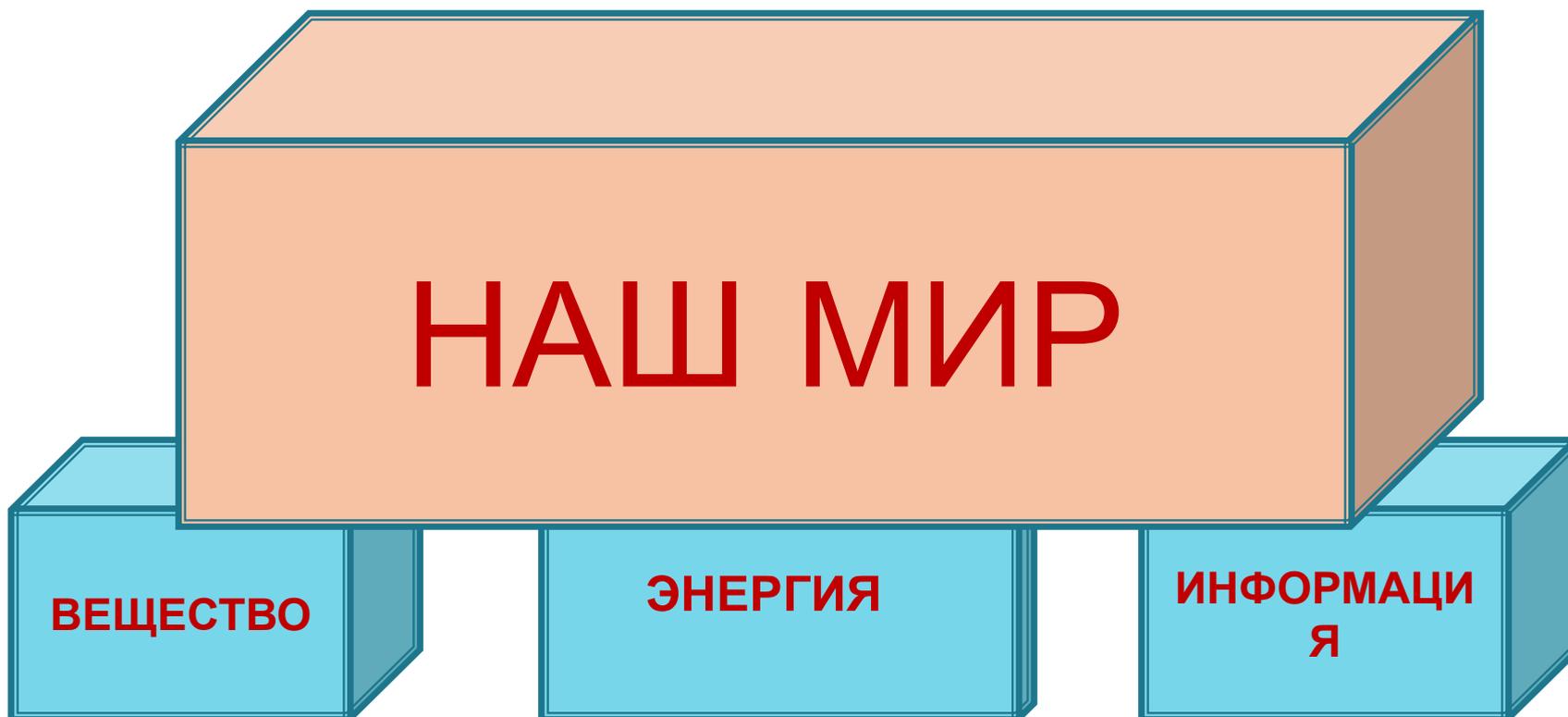
- **Растровый** – с помощью накладывания сетки-растра изображение разбивается на конечное количество элементов (точек). Каждая точка имеет свой код, цвет



- **Векторный** – кодирование объекта через координаты точек и длину радиуса, тип линии (сплошная или пунктирная), цвет и толщины.



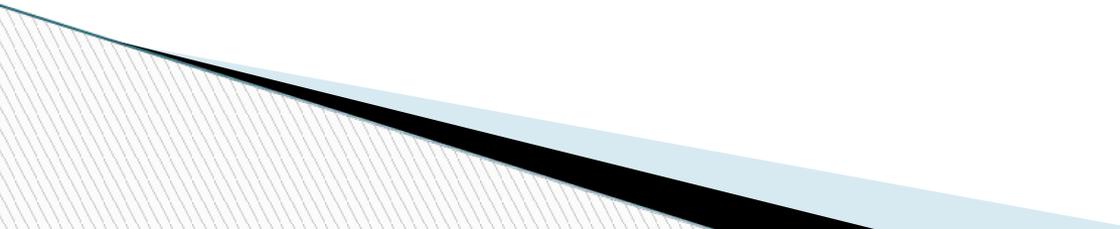
Вероятностный подход к определению информации



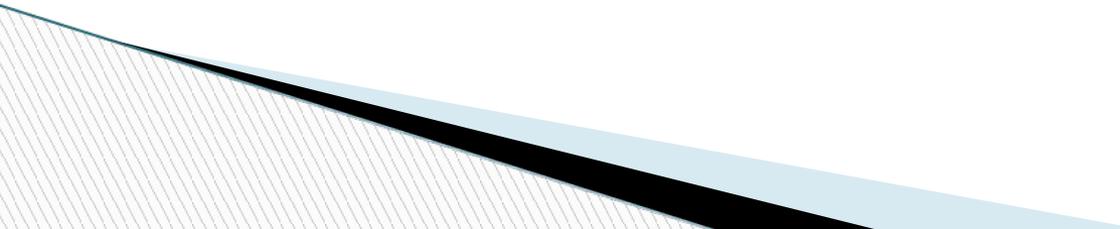
Составьте таблицу

процесс	Вещество В Энергия Э Информация И	Хранение - Х Передача - П Обработка - О
Передача факса		
Залежи газа		
Светит солнце		
Животные метят свою территорию		
Композитор пишет музыку		
Работа гидроэлектростанции		

КАК МНОГО В МИРЕ ЭНЕРГИИ, ВЕЩЕСТВА, ИНФОРМАЦИИ?

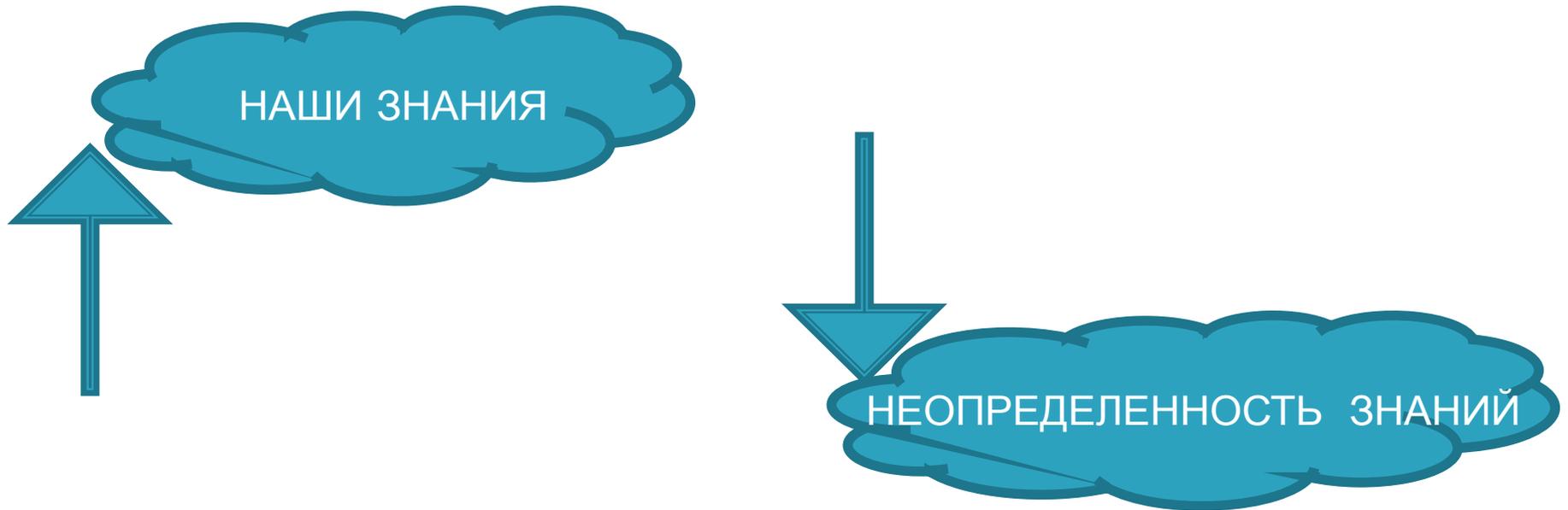
- ▣ КАК ИЗМЕРИТЬ КОЛИЧЕСТВА
ВЕЩЕСТВА?
 - ▣ КАК ИЗМЕРИТЬ КОЛИЧЕСТВО
ЭНЕРГИИ?
 - ▣ КАК ИЗМЕРИТЬ КОЛИЧЕСТВО
ИНФОРМАЦИИ?
- 

СУЩЕСТВУЕТ ДВА ПОДХОДА К ИЗМЕРЕНИЮ ИНФОРМАЦИИ

- ▣ Содержательный
(Алфавитный)
 - ▣ Вероятностный
- 

Вероятностный подход к измерению информации

- ▣ Если сообщение информативное-



- ▣ Это значит, что сообщение содержит информацию если уменьшает неопределенность знаний

Максимальное количество информации определяется следующим образом:

- **Орел или решка** – исход броска-равновероятностные события, т.к. ни одно из них не имеет преимущество перед другим – один к двум равновероятностным
- неопределенность знаний = 2
- **Шестигранный игральный кубик**
– один к шести равновероятностным
- неопределенность знаний = 6

Задание ответить письменно в тетради

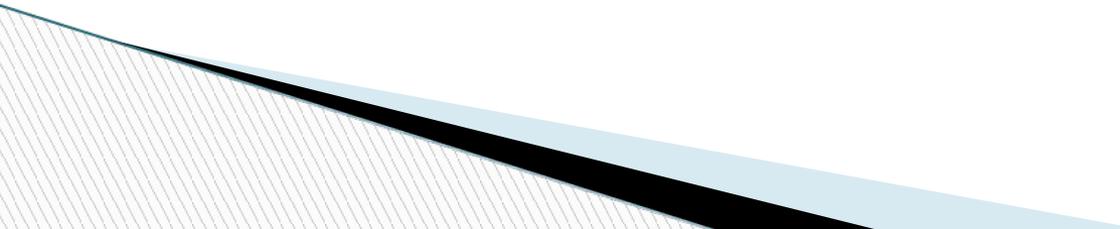
- На экзамен подготовлено 30 билетов
- Чему равно количество событий при вытягивании билетов?
- Равновероятны ли они?
- Чему равна неопределенность ученика при вытягивании билета?
- Во сколько раз уменьшится количество неопределенности знаний учащегося при вытягивании одного билета?
- Зависти ли этот показатель от номера вытянутого билета?

Максимальное количество информации определяется следующим образом:



- Рассмотрим пример с «орлом» и «орлом»
- В данном случае неопределенности знаний НЕ СУЩЕСТВУЕТ. ПОЧЕМУ ?
- Нет новой информации после броска . ПОЧЕМУ?
- Информативным сообщение после броска НЕ БУДЕТ. Почему ?
- И количество информации равно в данном случае – нулю.

Кодирование информации

- Для того, чтобы представить какую-то информацию, нам, как правило, недостаточно только одного символа или звука, и мы используем целый набор из соответствующего языка, часто даже комбинируя различные языки для полной передачи необходимой информации.
 - Таким образом, мы используем определённую систему символов, или, другими словами, кодов, для представления и передачи информации.
 - Иначе говоря, мы **кодируем информацию** при помощи того набора символов (кодов), которые для этого приняты. При этом, закодировать одну и ту же информацию мы можем с помощью различных кодов.
- 

Например,

- можно было бы убрать знак дорожного движения и поставить вместо него надпись, но тогда такая форма представления информации была бы намного менее эффективной, мы бы часто просто не успевали бы замечать, что написано на табличке и нам тяжело было бы всё успевать разглядывать, прибегая только к боковому зрению.
- **Заметьте**, что различную информацию лучше и удобнее представлять в различном виде.

Кодирование информации

- мы всегда встречаем, когда переходим на перекрёстке дорогу, у которой есть пешеходный сигнал светофора.
- **При этом каждый цвет светофора – это определённый код.** Нам доступны следующие предостережения: красный свет – движения нет, мы стоим, и зелёный – можно переходить дорогу.
- **В данном случае алфавит языка кодирования информации состоит из 2-х кодов – 2-х цветов светофора.**
- Если бы мы хотели закодировать ещё какие-либо предостережения, нам пришлось бы ввести ещё и третий цвет – например, жёлтый, как это сделано у автомобильного светофора. Но тогда наш алфавит стал состоять уже из 3-х кодов.

- В то же время, используя первоначальный набор, состоящий из 2-х цветов – красного и зелёного – мы также сможем закодировать и третье состояние светофора, соответствующего жёлтому цвету, с помощью комбинации этих двух кодов – например, с помощью 2-х горящих красных огней.
- Как видно из этого простого примера, одну и ту же информацию можно закодировать, используя различные способы и разную мощность (количество символов) языка кодирования.
- Даже если бы наш светофор, имеющий лампы только с красным и зелёным цветом, должен был бы показывать 600 различных комбинаций (предостережений, советов, режимов перехода и работы и так далее), мы это также смогли бы реализовать при помощи определённого количества так красных и зелёных ламп.
- **Однако**, естественно, чем меньше мощность языка кодирования информации, тем больше символов приходится иметь для реализации различных информационных комбинаций.

Когда создавали ВЫЧИСЛИТЕЛЬНЫЕ МАШИНЫ,

- то разработчикам в первую очередь пришла мысль закодировать информацию с помощью очень простого набора символов – нуля и единицы.
- Это было абсолютно естественно и создать такие вычислительные машины было проще всего, так как, поскольку машины эти проектировались и были электрическими, то было просто обозначить за единичку, например, присутствия сигнала, за нолик – его отсутствие, или, например, наоборот.
- Ведь намного сложнее каждый раз мерить сигнал, чтобы узнать, в какой же диапазон он попадает, чтобы ввести, например, систему кодирования из 4-х кодов 1,2,3,4.

- Кроме того, оказывается, какой бы мы язык математического представления информации (состоящий из чисел) ни возьмём, мы всегда можем представить абсолютно всю информацию без остатка с помощью этого набора чисел.
- То же относится и к нашему набору – двоичной системы счисления – 0 и 1.
- Вроде бы казалось, что всего 2 цифры не смогут отобразить всю информацию, какая только может быть в нашем материальном мире, и тем не менее, это так. Ведь компьютер способен обработать информацию любого характера – и звуковую, и текстовую, и символьную – переводя каждый вид информации в тот вид, который он понимает, т.е. в двоичный вид, используя только набор всего из двух символов – нуля и единицы.

- Итак, чтобы ввести информацию в компьютер, надо сначала её представить на понятном ему языке, т.е. перекодировать информацию в понятный для вычислительной машины вид. Как это сделать, я расскажу чуть позже.
- А сейчас, прежде всего, нужно нам самим понимать компьютерный язык представления информации.
 - **Допустим, нам нужно закодировать для компьютера фразы:**
 - «Жук ползёт по асфальту».
 - «Жук не ползёт по асфальту».
 - так сочетая 0 и 1, чтобы всю информацию представить в виде одних только единиц и нулей.

Пока пусть это будет абсолютно вся информация, которую бы мы хотели закодировать для компьютера, чтобы он нас понял.

Обратите внимание, что эта информация с точки зрения компьютера различается только тем, что здесь жук либо ползёт, либо не ползёт, поэтому можно закодировать только одно это действие вот так:

- *«**Ползёт**» присваиваем 0, а «не **ползёт**» -- 1.*
- *Теперь же, когда у нас появляется ещё и гусеница:*
- *«Гусеница ползёт по листьям»,*
- *становится важным ещё и обозначить, о ком именно идёт речь.*
- *Поэтому присваиваем жуку – 0, а гусенице 1.*
- *Получаем:*
- *«Жук **ползёт** по асфальту» =00 (жук=0 и ползёт=0)*
- *«Жук **не ползёт** по асфальту» =01 (жук=0 и не ползёт=1)*
- *«Гусеница ползёт по листьям» =10 (гусеница=1, а ползёт=0)*
- *«Гусеница не ползёт по листьям» =11 (гусеница=1 и не ползёт=1)*

Как видим, здесь мы кодируем для вычислительной машины 4 различных факта, поэтому используем 4 различные комбинации **00, 01, 10, 11**.

Напомним, что язык единиц и нулей хорошо воспринимает компьютер, а другого языка он не знает, поэтому мы и кодируем для него информацию, чтобы наша информация была понятна и ему.

Чем сложнее информация и чем больше её разнообразие, тем больше нулей и единиц нам нужно, чтобы все обозначить через них.

Возникает понятие количества информации. Оно всегда кратно степени двойки, так как **2 – это основание системы счисления (количество различных символов языка кодирования информации – 0 и 1)**, а степень связана с количеством символов.

Нужно учиться покрывать информацию, т.е. выбирать то количество кодов, которое нам необходимо.

В компьютере информация

- хранится и обрабатывается в виде последовательности нулей и единиц. Каждый символ – это единица хранения информации, меньше которой не существует.
- Такой символ – 1 или 0 – называется **битом**.
- **Бит в компьютере – это наименьшая единица измерения информации.**

- Как люди приняли и условились в компьютерах, 8 бит является минимальным количеством хранения и адресации информации в большинстве компьютеров.
- **Это означает, что мы не можем обратиться непосредственно к одному биту, но можем выбрать только минимум 8 бит за один раз.**
- Этот факт никак не связан с теорией хранения и обработки информации и является лишь следствием проектирования вычислительной машины человеком. 8 бит составляют один байт, поэтому в компьютере при измерении информации часто используют единицы измерения байты

- В компьютере мы храним и обрабатываем очень объёмные количества информации, поэтому, как и в физике, для описания этого количества приняты соответствующие приставки, кратные не 1000, а 1024 ввиду того, что $1024=2^{10}$ (вспомните, что мы используем уже не десятичную, а двоичную систему счисления), а, кроме того, число $1024\approx 1000$, что позволяет переводить информацию между приставочными единицами даже в уме с учётом лишь небольшой погрешности вычислений.
- В качестве примера можно привести текстовые файлы, занимающие после своего кодирования относительно немного места в памяти вычислительной системы, поэтому мы их измеряем в килобайтах.

- **1 Килобайт** – это 1024 байта, что примерно 1000 байт ($2^{10}=1024$, степень десять в двоичной системе счисления принято за приставку кило).
- **Обозначается как 1 Кб**
- Когда мы кодируем звук, нам нужно намного больше единиц и нулей, так как звуковая информация намного полнее и сложнее текстовой, поэтому звуковые файлы мы привыкли измерять в мегабайтах.
- **1 Мбайт** $=2^{10}=1024$ **Кбайт**, или примерно **1000 Кбайт**, что приблизительно составляет **1000000 байт** ($2^{10}*2^{10}= 2^{20}$ байт), отсюда приставка «Мега».
- **Далее, фильмы обычно занимают Гигабайты.**
- **1 Гбайт = 1024 Мбайт**, или примерно миллиард байт.
- Точно также можно расписать и для битов:
- **1 Кбит = 1024 бита**, обозначается как **1 Кбит**, **1 Мбит = 1024 Кбит = 2^{20} бит** и так далее.

Но как можно закодировать информацию?

- Чтобы компьютер нас понял, нужно предоставить ему информацию в понятном ему виде, то есть нужно закодировать информацию.
- А для этого существуют различные устройства. Например, символьную и числовую информацию можно закодировать с помощью определённого количества комбинаций.
- **Например, все буквы и символы русского алфавита мы можем закодировать различным сочетанием битов.** Устройство, которое кодирует буквенную, числовую и символьную информацию – это клавиатура. При каждом нажатии на клавишу сразу же формируются электрические сигналы, сообщаящие компьютеру, какой символ был нажат.
- **Например, мы нажали на клавишу буквы Т.** Для нас это буква Т, а для компьютера буква Т обозначится как, например, 110011 – всё зависит только от принятой таблицы кодирования.

Информационный объём сообщения и элементарная двоичная комбинаторика.

□ Каждое информационное сообщение характеризуется **определённым информационным объёмом** – количеством символов кодирования сообщения.

Информационный объём сообщения напрямую зависит от принятой системы кодирования.

- Как правило, прежде всего нас будет интересовать **информационный объём** сообщения при кодировании в двоичной системе счисления, принятой в компьютерах. Это задачи, связанные с подсчётом информации, необходимой для кодирования определённого количества печатаемых (текстовых) символов, длины звуковой дорожки или фильма, изображения с определёнными параметрами. Такие задачи решаются алгебраическим суммированием количества информации, необходимой для кодирования единицы сообщения, на количество таких единиц (например, умножением числа символов на количество бит, отводимых под каждый символ, если это количество одинаково).
- В этих задачах либо прямо ил косвенно указано, сколько символов отводится под кодирование одной информационной единицы, либо это количество необходимо определить.
- Все возможные случаи типовых задач мы рассмотрим в практической части уроков немного ниже, сейчас же нас интересует теоретический материал, который мы будем применять на практике.

Формула Хартли

- Так, что подсчёта количества бит, которое потребуется для кодирования определённого количества абсолютно равнозначных равновероятных комбинаций, необходимо воспользоваться формулой Хартли:
- $N=2^i$, где N – количество различных комбинаций, i – количество необходимых для кодирования данного количества бит.
- Если мы кодируем всего 1-м битом – 1 или 0 – то таких комбинаций $N=2^1=2$ (собственно, это 0 и 1).
- Если кодируем 2-мя битами, то комбинаций будет уже 4:
- $N=2^2=4$ (это 00, 01, 10, 11).
- Если под кодирование информации мы отведём 3 бита, то комбинаций будет 8:
- $N=2^3=8$ (это 000, 001, 010, 011, 100, 101, 110, 111), и так далее.

Обратите внимание,

- что в данном случае мы не вникаем в глубины и суть самой информации, и нам просто важно перечисление всех возможных вариантов. Это подход хорошо работает при кодировании очень большого количества информации, например, при кодировании различных цветов, символов и букв, моделей игрушек, силы звуковых колебаний, оттенков цветов и так далее.
- **Например,** если нам нужно закодировать 5 различных цветов, то 2 бита нам уже не подходят, поскольку они дадут всего 4 различных комбинации, поэтому нам придется использовать уже 3 бита, и при этом 3 каких-то комбинации останутся незадействованными (8 комбинаций всего – 5 комбинаций под различные цвета) но в то же время будут и резервом на будущее. Итак, для того, чтобы найти минимальное количество бит, которые потребуются для кодирования комбинационной информации такого рода, необходимо определить показатель степени 2-х, который заведомо будет больше или равен количеству различных комбинаций, но меньше которого всех комбинаций будет недостаточно для кодирования всех объектов или событий.

- В общем же случае, количество комбинаций вычисляется по формуле $N=s^i$, где s является мощностью применяемого языка кодирования информации, i – количество символов в каждой комбинации. Например, если у нас язык состоит из 3-х различных символов – 0, 1, 2, то его мощность равняется 3-м и количество комбинаций, каждая из которых будет состоять из 4-х символов, будет равно $N=3^4=81$.

формулы Шеннона

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

- Если же наши сведения различаются по характеру или вероятности своего наступления, то мы должны уже вникнуть в то, что же именно и с какой вероятностью произойдёт, и уже соответствующим образом закодировать и подсчитать информационный объём такой информации. В этом случае в нашем кодировании нам придётся учесть и заложить в него ещё и такие биты, которые будут точно указывать уже не на комбинации, а на вероятность происшествия кодируемых событий. Такие задачи в ЕГЭ являются немного более сложными и для своего решения требуют применения следующей формулы Шеннона для определения количества информации такого рода:
- Здесь I – это количество (объём) информации, p_i – вероятность наступления каждого отдельного события с номером i .

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

- Знак \sum является знаком суммы, который в данном случае показывает, что суммирование ведётся от первого и до некоторого события N (например, до 5). Чтобы стало ещё понятнее, запишем эту формулу для определения информационного объёма сообщения, в котором описывается 4 события с вероятностями их наступления p_1, p_2, p_3, p_4 :
- $I = - (p_1 * \log_2(p_1) + p_2 * \log_2(p_2) + p_3 * \log_2(p_3) + p_4 * \log_2(p_4))$

- Интересен тот факт, что в первом случае, когда мы просто искали необходимое количество различных комбинаций и связанное с ним количество символом, объём информационного сообщения любой сложности будет являться целым числом, а в случае с равновероятными событиями, как правило, дробным количеством бит.
- Это связано с тем, что, на самом деле, в первом случае мы часто имеем дело в описанной избыточностью информации, когда у нас остаются лишние и не задействованные комбинации, которые всё равно идут в общий счёт информационного объёма.

Открыть презентацию Решение задач на кодирование информации

- ▣ Решение заданий с 1-8 самостоятельной на уроке
- ▣ Домашнее задание с 1-3

