

Кодирование

Язык и алфавит

Язык — система знаков и правил для передачи и хранения информации.

Кодирование — представление информации в форме, удобной для хранения, передачи, автоматической обработки.



Код — правило, по которому сообщение преобразуется в последовательность знаков.

Что из перечисленного ниже является кодом?

- сигналы светофора;
- флаги, вывешенные на улицах в праздник;
- правила записи шахматных партий.

Пиктограммы - изображения, похожие на то, что они обозначают.

Идеограммы - специально придуманные знаки для обозначения сложных понятий.





Дорожный знак «Главная
дорога»
№1



Дорожный знак
«Пешеходный переход»
№2

Как вы полагаете, какой из этих дорожных знаков пиктограмма, а какой — идеограмма?

Правильный ответ :

№1 – идеограмма.

№2 – пиктограмма.

Алфавит — упорядоченный набор знаков,
которые используются в языке.

Формальные и неформальные ЯЗЫКИ

РУССКИЙ (ОПРЕДЕЛЕН АВТОМАТИЧЕСКИ) ▼ ↔ САМОАНСКИЙ РУССКИЙ АНГЛИЙ ▼

Стыдно быть несчастливым	×	Maasiasi e le fiafia	☆
-----------------------------	---	----------------------	---

ОПРЕДЕЛИТЬ ЯЗЫК САМОАНСКИЙ ▼ ↔ САМОАНСКИЙ РУССКИЙ АНГЛИЙ ▼

Maasiasi e le fiafia	×	Стыд не весело	☆
----------------------	---	----------------	---

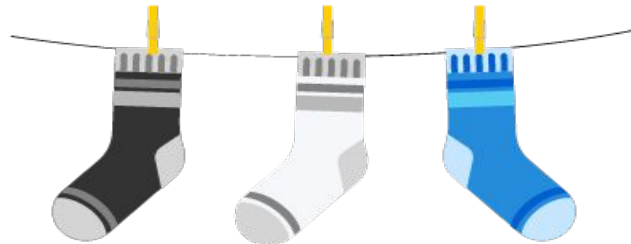
Естественные языки, которые постепенно возникли в ходе истории человечества, а не создавались искусственно, не подчинены строгим правилам.

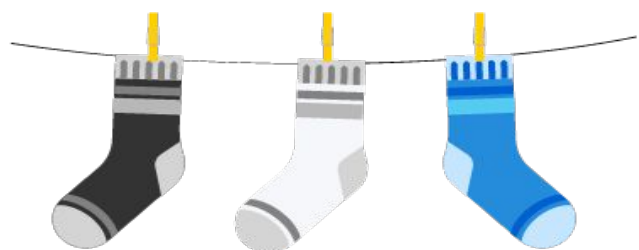
Формальный язык — язык, в котором значение каждого слова или знака, правила построения предложений и понимания их смысла однозначны.

Критерий сравнения	Естественные	Формальные
Возникновение	Сформировались постепенно в ходе развития человечества	Созданы специально для решения конкретных задач
Применение	Для повседневного общения, обмена информацией	В различных сферах деятельности
Соответствие слов и понятий	Однозначного соответствия нет: синонимы (разные слова с одинаковым смыслом), многозначные слова, омонимы (одинаковые слова с разным смыслом)	Каждое слово или знак имеет чётко определённый смысл
Правила построения сообщений	Правила неоднозначны, есть множество исключений, допускаются различные варианты передачи одной и той же информации	Строгие правила следования слов или знаков

Сообщения и их количество

Сообщение — любая последовательность символов алфавита.





Если использовать:

1 носок - $3^1=3$ (Ч или Б или С);

2 носка - $3 \cdot 3 = 3^2 = 9$ (ЧЧ, ЧБ, ЧС, БЧ, ББ, БС, СЧ, СБ, СС);

...

4 носка - $3 \cdot 3 \cdot 3 \cdot 3 = 3^4 = 81$;

....

Мощность алфавита — количество
символов в нём.

Зная мощность алфавита **M** и длину сообщений
L, можно найти количество возможных
сообщений:

$$N = M^L$$



$$5 \text{ носков} - 3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 = 3^5 = 243$$

Вопрос:

Для передачи сигналов используют табло из трёх лампочек, каждая из которых может быть в одном из двух состояний: включена или выключена.

Сколько различных сообщений может передавать такое табло?

$$N = M^L$$

где **N** - число возможных сообщений;

M - мощность алфавита;

L - длина сообщений.

$$8 = 2^3$$

Вопрос:

А если первая лампочка — индикатор исправности табло, и поэтому она обязательно должна гореть?

$$4=2^2$$

Вопрос.

Для передачи сигналов используют табло из трёх лампочек, каждая из которых может быть в **трех состояний**:
выключена, горит и мерцает.

Сколько различных сообщений может передавать такое табло?

$$27=3^3$$

Вопрос.

Предположим, табло из лампочек с тремя состояниями должно передавать **50** различных сообщений.

Сколько лампочек в нём должно быть?

Домашнее задание:

В алфавите придуманного языка всего пять буквы: *E, P, O, K, Y*.

Ответьте на следующие вопросы.

1. Какое максимальное количество трёхсимвольных слов может быть в этом языке?
2. Сколько существует различных четырёхсимвольных слов, у которых в начале E?
3. Сколько существует различных трёхсимвольных слов, если по правилам языка буква Y не может быть первой?

У прибора **три** одинаковых индикатора, меняющих цвет.
Каждая комбинация цветов отражает состояние прибора.

1. Каким должно быть количество цветов индикатора, чтобы индикаторы отражали **27** различных состояний прибора?
2. А если прибор усовершенствовали и теперь индикаторы должны отражать **70** состояний?

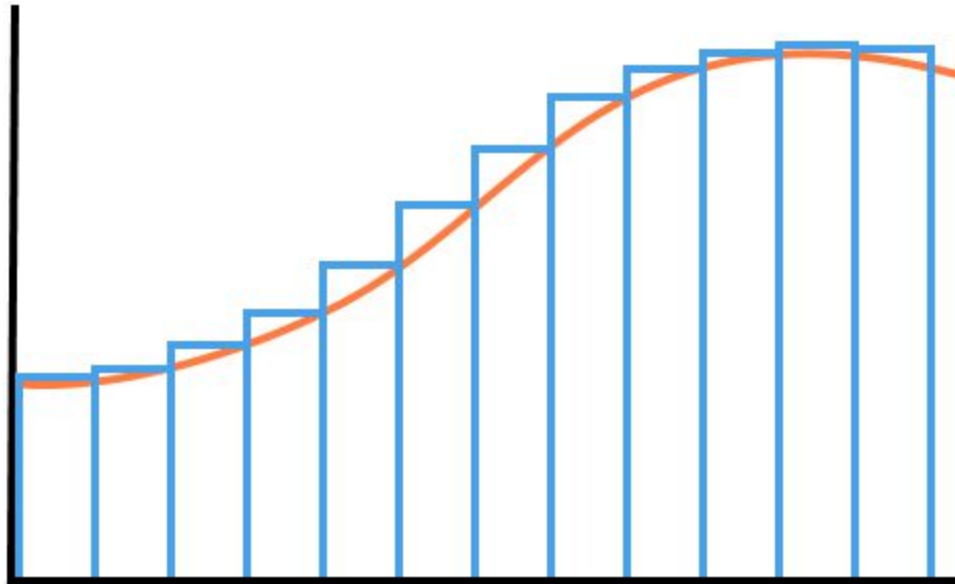
Представление информации в компьютере

Кодирование информации для хранения в памяти компьютера мы разобьем на два этапа:

- преобразование информации в числовую;
- кодирование числа с помощью алфавита из двух знаков: 0 и 1.

Дискретизация

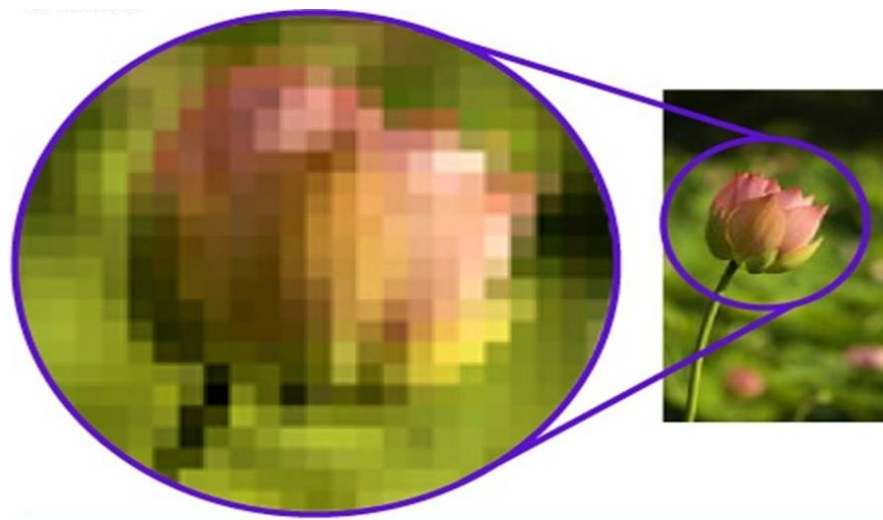
Дискретизация — представление непрерывной величины в виде множества отдельных значений.



Вопрос.

Предположим, мы записываем температуру с точностью до 1 градуса каждые 15 минут. Годится ли дискретизация с такими параметрами в таких ситуациях:

- для управления отоплением в системе «Умный дом»;
- для контроля за состоянием больного;
- для записи параметров двигателя для чёрного ящика самолёта.



Равномерный код и длина кодового слова

Декодирование — восстановление исходного сообщения из кода.

Равномерный код — код, в котором все кодовые слова имеют одинаковую длину.

Пример:

Прибор фиксирует и сохраняет в памяти данные о направлении ветра. Различает он восемь основных направлений: *С, Ю, З, В и СЗ, СВ, ЮЗ, ЮВ.*

Память прибора состоит из элементов с двумя состояниями: *0 и 1.*

Такой элемент мы будем называть **битом**.

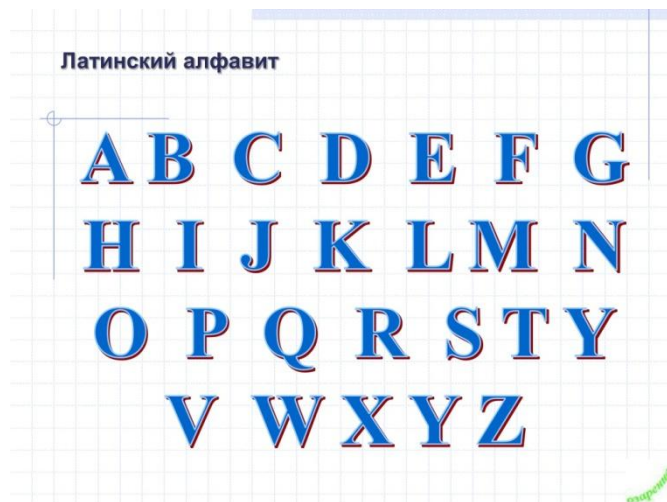
Из неравенства $2^L \geq 8$

находим длину кодового слова: это **3**.

Присвоим коды : *С — 000, СВ — 001, В — 010,
ЮВ — 011, Ю — 100, ЮЗ — 101, З — 110, СЗ — 111.*

Задание. Декодируйте цепочку записей показаний прибора:
111110101111000111111.

111	110	101	111	000	111	111
СЗ	3	ЮЗ	СЗ	С	СЗ	СЗ



4 бита: $2^4=16$

5 бита: $2^5=32$

Задание.

В детском саду маркируют шкафчики с помощью картинок-наклеек. Но вот беда: в группе 22 малыша, а картинки есть всего 4 типов. Воспитатели решили наклеивать по несколько картинок на шкафчик, на все — одинаковое количество, чтобы дети не обиделись.

Сколько картинок нужно наклеить на каждый шкафчик, чтобы у всех детей получились разные «кодовые слова»?

Если по 2 наклейки: $4*4=4^2=16$. Этого мало!

По 3 наклейки: $4*4*4=4^3=64$. Хватит на всех!

Неравномерный код и условие Фано

Неравномерный код — код с кодовыми словами различной длины.

Кодируем сообщения, состоящие из десятичных цифр.

*Цифр - **10** (от 0 до 9)*

$2^3 \leq \mathbf{10} \leq 2^4$, соответственно кодировать 1 символ нужно как минимум 4 битами.

Итак,

***0** кодируем как **0000**;*

***1** – как **0001**;*

...

***0** кодируем как **0**;*

***1** – как **1**;*

***2** – как **10**;*

***3** – как **11**; и т.д.*

Закодируем число: **3021**

При равномерном:

0011 0000 0010 0001

При неравномерном:

11 0 10 1

Условие Фано: ни одно кодовое слово не совпадает с началом другого кодового слова.

*Нужно закодировать данные о весе человека:
нормальный, ниже нормы, выше нормы.*

При равномерном коде нам потребовались бы **2** бита для каждого значения, так как

$$2^1 < 3 \leq 2^2$$

т.е. нам пришлось бы закодировать так:

00 - ниже нормы

01 - нормальный

10 - выше нормы

Обозначим нормальный вес кодом **0**. Тогда коды для недостаточного и избыточного веса должны начинаться с **1**, например, **10** и **11**.

Т.е.:

0 - нормальный вес

10 - ниже нормы

11 - выше нормы

Критерий	Равномерный код	Неравномерный код
Длина кодового слова	Одинаковая у всех	Может быть различной, обычно более частым символам соответствуют более короткие кодовые слова
Однозначное декодирование	Возможно всегда	Возможно, если соблюдается условие Фано или есть символы-разделители
Сложность разработки кода	Обычно сводится к нумерации символов исходного алфавита в какой-то системе счисления	Коды, дающие экономию памяти, довольно сложно разработать
Сложность декодирования	Простой поиск в кодовой таблице	Последовательный посимвольный разбор кода, в ходе которого определяется длина кодового слова
Применение	Основной способ представления информации в памяти компьютера	Некоторые исторически сложившиеся коды, а также ситуации, когда требуется минимизировать объём занимаемой информацией памяти
Связанные с кодом типичные задачи в курсе информатики	<ul style="list-style-type: none"> •Задачи на кодирование и декодирование •Определение длины кодового слова •Определение информационного объёма сообщения 	<ul style="list-style-type: none"> •Проверка выполнения условия Фано •Декодирование сообщений •Дополнение кода в соответствии с условием Фано •Построение кода, минимизирующего объём занимаемой памяти •Оценка среднего информационного объёма сообщения

Измерение количества информации

Бит — наименьшая единица измерения информации, одна двоичная цифра, элемент, принимающий значения **0** или **1**.

Байт — единица измерения информации, равная **8** бит.

$$1 \text{ Кбайт (килобайт)} = 1024 \text{ байт} = 2^{10} \text{ байт} = 2^{13} \text{ бит}$$

$$1 \text{ Мбайт (мегабайт)} = 1024 \text{ Кбайт} = 2^{20} \text{ байт} = 2^{23} \text{ бит}$$

$$1 \text{ Гбайт (гигабайт)} = 1024 \text{ Мбайт}$$

$$1 \text{ Тбайт (терабайт)} = 1024 \text{ Гбайт}$$

Вопрос.

Подсчитайте, сколько битов в терабайте.

1 Тбайт — это $2^{43} = 8796093022208$ бит!

1 Тбайт = 1024 Гбайт =

=1024*1024 Мбайт =

=1024*1024*1024 Кбайт =

=1024*1024*1024*1024 байт =

=1024*1024*1024*1024*8 бит

	1	2	3	4	5
1	A	B	C	D	E
2	F	G	H	I/J	K
3	L	M	N	O	P
4	Q	R	S	T	U
5	V	W	X	Y	Z

Квадрат Полибия

Вопрос. Рука сигнальщика может быть в одном из шести положений. Каким может быть максимальное количество сигналов в семафорной азбуке?

Morse code

A	• —	U	• • —
B	— • • •	V	• • • —
C	— • — •	W	• — —
D	— • •	X	— • • —
E	•	Y	— • — —
F	• • — •	Z	— — • •
G	— — •		
H	• • • •		
I	• •		
J	• — — —		
K	— • —	1	• — — — —
L	• — • •	2	• • — — —
M	— —	3	• • • — —
N	— •	4	• • • • —
O	— — —	5	• • • • •
P	• — — •	6	— • • • •
Q	— — • —	7	— — • • •
R	• — •	8	— — — • •
S	• • •	9	— — — — •
T	—	0	— — — — —

Азбука Морзе

Декодирование равномерного кода по кодовой таблице

Робот перемещается по координатной плоскости. Его система команд — перемещение на 1 в направлениях вверх, вниз, вправо и влево. Для кодирования применяется двухбитовый код, коды команд приведены в таблице.

Вверх на 1	00	Вправо на 1	10
Вниз на 1	01	Влево на 1	11

Робот начал движение от начала координат и выполнил последовательность команд, закодированную следующим образом: 1111001000111101010111. Определите его текущие координаты.

Вверх на 1	00	Вправо на 1	10
Вниз на 1	01	Влево на 1	11

1111001000111101010111 =
 1111 00 10 00 11 11 01 01 01 11

5 команд «влево» (11)

1 команда «вправо» (10)

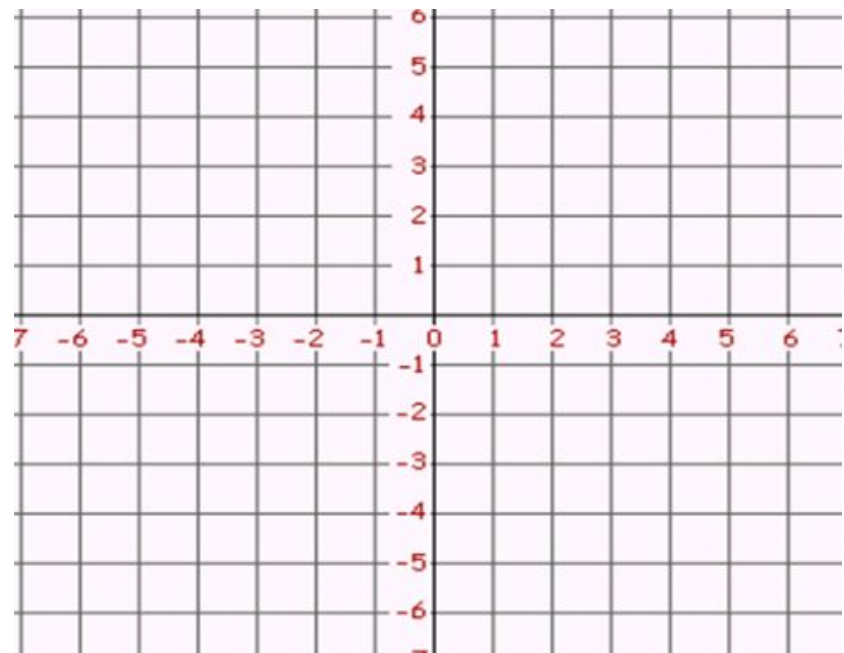
По **оси X** сдвинулся на **минус 4**

2 команды «вверх» (00)

3 команды «вниз» (01)

По **оси Y** смещение на **минус 1**

Ответ: (-4; -1)



Декодирование неравномерного кода при выполнении условия Фано

В исходном алфавите 6 букв: Е, К, Н, О, П, С.

Гласные кодируются двухбитовыми комбинациями:

00 — Е, 01 — О.

Коды согласных трёхбитовые, известно, что

100 — К, 101 — Н, 111 — С.

Декодировать сообщение: **1100011101100**.

Решение:

П — 110.

110 00 111 01 100

П Е С О К