



Измерение информации.

Алфавитный (объемный) подход
Лекция №2

Основные понятия

Алфавит - это набор букв, знаков препинания, цифр, скобок и др. символов, используемых в тексте. В алфавит также следует включить и пробел, т. е. пропуск между словами.

Язык – знаковая форма предоставления информации.



Алфавитный (объемный) подход к измерению информации

- Каждый символ сообщения имеет определенный **информационный вес**;
- Каждый символ сообщения несет фиксированное **количество информации**;
- Вес одного символа зависит от **мощности алфавита**;
- Количество информации в сообщении **зависит** от числа знаков в тексте и **не зависит** от содержания текста.

Алфавитный (объемный) подход к измерению информации

N – мощность алфавита – полное число символов в алфавите.

Q- информационный объем всего сообщения.

K- количество символов в сообщении.

I-информационный объем одного символа.

$$Q=K*I$$

$$2^I=N$$

Задача

Сообщение содержит 125 символов и написано оно при помощи алфавита состоящего из 64 символов. Найдите информационный объем данного сообщения.

Дано:

$$K=125$$

$$N=64$$

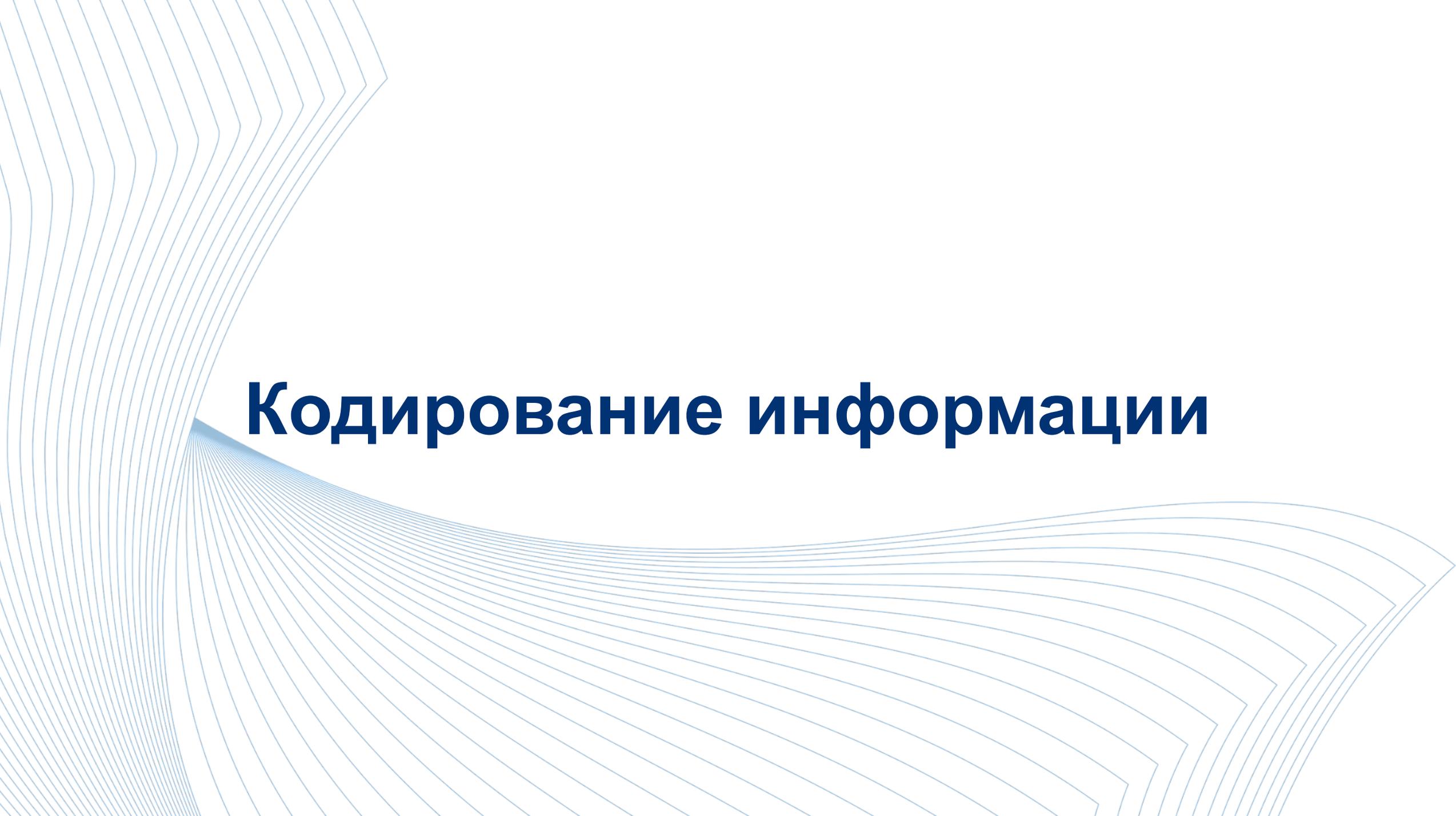
$$Q=?$$

Решение:

$$2^l=N \quad Q=K \cdot l$$

$$2^l=64 \quad Q=125 \cdot 6$$

$$l=? \quad Q=?$$

The background features a series of light blue, wavy lines that create a sense of depth and movement, resembling a stylized wave or a series of overlapping planes. The lines are most prominent on the left side and curve towards the right, framing the central text.

Кодирование информации

Кодирование – это процесс преобразования данных из исходной формы представления в коды.

Код – это набор условных символов для представления информации.

Цели использования

кодирования:

- компактное хранение, удобство при обработке и передаче информации через автоматические устройства с программным обеспечением;
- удобство при обмене данными между субъектами;
- четкое отображение информации;
- распознавание объектов и субъектов;
- шифровка конфиденциальной информации.

В компьютере вся информация кодируется при помощи двух сигналов **0** и **1**, потому, что в техническом устройстве наиболее просто реализовать два противоположных физических состояния:

- некоторый физический элемент, имеющий два различных состояния;
- намагниченность в двух противоположных направлениях; прибор
- пропускающий или нет электрический ток;
- конденсатор, заряженный или незаряженный и т.п.

Числа 0 и 1 называются битами

1 бит – это наименьшая единица измерения информации.

При помощи **1 бита** можно закодировать **2** символа
(0,1) – **2^1** .

При помощи **2^x бит** можно закодировать **4** символа
(00,01,10,11) – **2^2** .

И т.д.

Мощность компьютерного
алфавита равна **256**

$$2^l = 256$$

$$l = 8$$

1 байт = 8 бит

Единицы измерения информации:

1 Кбит = 2^{10} = 1024 бит

1 Мбит = 2^{20} = 1048576 бит

1 Гбит = 2^{30} \approx 1 млрд.бит

1 Тбит = 2^{40} бит.

1 Кбайт = 2^{10} = 1024 байта

1 Мбайт = 2^{20} = 1048576

байт

1 Гбайт = 2^{30} \approx 1 млрд.байт

1 Тбайт = 2^{40} байт.

Кодирование текстовых данных

Основные виды кодирования текста:

- **графический** – текст переводится в рисунки;
- **символьный** – преобразование происходит с помощью знаков алфавита, в котором представлен исходный текст;
- **числовой** – текст кодируется в числа.

Кодирование текстовых данных

Все символы компьютерного алфавита пронумерованы от 0 до 255.

Каждому номеру соответствует восьмиразрядный двоичный код от 00000000 до 11111111.

Этот код просто порядковый номер символа в двоичной системе счисления.

Кодирование текстовых данных

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

Международным стандартом для ПК стала таблица ASCII (Американский стандартный код для информационного обмена). На практике можно встретиться и с другой таблицей – КОИ-8 (Код обмена информацией), которая используется в глобальных компьютерных сетях.

Кодирование текстовых данных

Первые 32 кода базовой таблицы, начиная с нулевого, отдаются производителям аппаратных средств.

Начиная с 32-го по 127-й код размещены коды английского алфавита, знаков препинания, цифр и т.д.

Расширенная часть системы кодирования, определяющая значения кодов с 128-го по 255-й, используется для кодирования символов национального алфавита.

Код	Сим- вол	Код	Сим- вол	Код	Сим- вол	Код	Сим- вол
00100000	пробел	00110000	0	01000000	@	01010000	P
00100001	!	00110001	1	01000001	A	01010001	Q
00100010	"	00110010	2	01000010	B	01010010	R
00100011	#	00110011	3	01000011	C	01010011	S
00100100	\$	00110100	4	01000100	D	01010100	T
00100101	%	00110101	5	01000101	E	01010101	U
00100110	&	00110110	6	01000110	F	01010110	V
00100111	'	00110111	7	01000111	G	01010111	W
00101000	(00111000	8	01001000	H	01011000	X
00101001)	00111001	9	01001001	I	01011001	Y
00101010	*	00111010	:	01001010	J	01011010	Z
00101011	+	00111011	;	01001011	K	01011011	[
00101100	,	00111100	<	01001100	L	01011100	\
00101101	-	00111101	=	01001101	M	01011101]
00101110	.	00111110	>	01001110	N	01011110	^
00101111	/	00111111	?	01001111	O	...	

Коду 00100000 в этой таблице соответствует *пробел* — пустой промежуток величиной в один символ, который используется для отделения одного слова от другого.

Кодирование цвета

Одна из основных моделей **RGB** (red, green, blue). Этот стандарт использует всего 3 байта, по одному на каждый цвет.

Для управления яркостью вводят ещё один бит и получается модель **IRGB** (Intensity RGB).

Создание более богатой палитры осуществляется в 6-битной системе, называемой **RrGgBb**. Код 00 означает, что цвет выключен, 01 – это слабый цвет, 10 – обычный оттенок и 11 – интенсивный.

Кодирование графических данных

Растровая графика

Растровый формат можно назвать точечным.

Расположенные строго по строкам и столбцам точки имеют отдельные координаты нахождения на дисплее, цвет и уровень интенсивности.

Качество изображения напрямую зависит от количества точек – чем их больше, тем картинка качественнее.

Растровый способ кодирования подходит для фотографий.

Кодирование графических данных

Векторная графика

Векторная графика опирается на закодированные геометрические фигуры. В числовой формат приведены размеры объектов, координаты вершин, толщина контуров цвет заливки.

Векторное кодирование удобно применять при создании рекламной продукции.

Кодирование звуковой информации

Звук – это волны с постоянно меняющейся частотой и интенсивностью, вызванные колебанием частиц.

При оцифровке непрерывная акустическая волна временно превращается в прерывистую. Дискретная форма представляет собой короткие отрезки с неизменным сигналом.

Кодирование звуковой информации

Частота дискретизации – количество измерений громкости в секунду.

Глубина кодирования звука – количество данных, необходимое для преобразования прерывистых уровней громкости звукового сигнала.

От частоты дискретизации глубины кодирования звука зависит точность воспроизведения оригинального звука.

Кодирование чисел

Числа кодируются методом перевода в двоичную систему счисления.

Выделяют два способа представления чисел:

- форма с фиксированной точкой – для целых чисел;
- форма с плавающей точкой – для действительных чисел.

Задачи

- ✓ Алфавит языка оценивают в 32 буквы. Каков информационный вес одной буквы такого алфавита?
- ✓ У племени "чичевоков" в алфавите 24 буквы и 8 цифр. Знаков препинания и арифметических знаков нет. Каков информационный объем одного символа?
- ✓ Книга, набранная с помощью компьютера, содержит 150 страниц. На каждой странице — 40 строк, в каждой строке — 60 символов. Объем одного символа равен 8. Каков объем информации в

Задачи

Переведите в биты:

- 5 Гбайт
- 0,5 Кбайт
- 512 Мбайт

Переведите в байты:

- 16 Гбит
- 8 Кбит
- 4 Мбит