



Измерение информации (алфавитный подход)

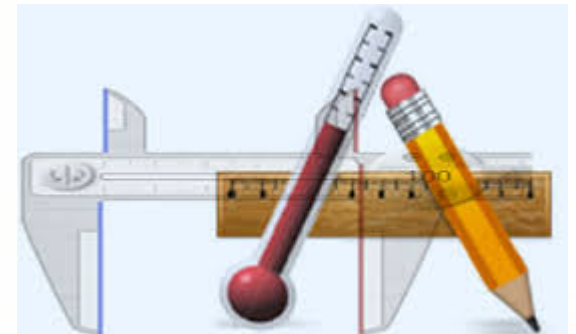
Алфавитный подход

Алфавитный подход позволяет измерять информационный объем текста на некотором языке (естественном или формальном), не связанный с содержанием этого текста.



Единицы измерения

- расстояния — метр
- массы — грамм
- времени — секунда
- информации — ?



Алфавит

Алфавит — набор букв, знаков препинания, цифр, скобок и других символов, используемых в тексте. В алфавит также следует включить и пробел.

Мощность алфавита — полное число символов алфавита. Будем обозначать эту величину буквой *N*.



Алфавит

Например, мощность алфавита из русских букв и отмеченных дополнительных символов равна **54**:

33 буквы + **10** цифр + **11** знаков препинания, скобки, пробел.



Информационный вес

Наименьшую мощность имеет алфавит, используемый в компьютере (машинный язык), его называют двоичным алфавитом, т.к. он содержит только два знака «0», «1».

Информационный вес символа двоичного алфавита принят за единицу измерения информации и называется **1 бит**.



Пример:

Информация, записанная
на машинном языке:

01110

010010

010

0111111011110

Количество
информации:

5 бит

6 бит

3 бита

13 бит



Алфавитный подход

При алфавитном подходе считают, что каждый символ текста, имеет информационный вес.

Информационный вес символа зависит от мощности алфавита.

С увеличением мощности алфавита, увеличивается информационный вес символа.



С увеличением мощности алфавита увеличивается информационный вес символов этого алфавита.

Один символ из четырехсимвольного алфавита ($N = 4$) «весит» 2 бита. Все символы такого алфавита можно закодировать комбинациями из двух цифр двоичного алфавита.

Комбинацию из нескольких знаков двоичного алфавита называют **двоичным кодом**.

Номер символа	1	2	3	4
Двоичный код	00	01	10	11

Алфавит мощностью 8 знаков можно закодировать на машинном языке с помощью трёх символов двоичного алфавита

Номер символа	1	2	3	4	5	6	7	8
Двоичный код	000	001	010	011	100	101	110	111

Следовательно, каждый символ восьмизначного алфавита весит 3 бита.

Алфавит мощностью 16 знаков кодируется четырьмя символами

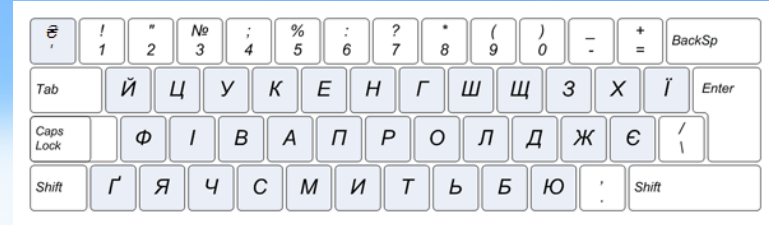
Запишем таблицу соответствия мощности алфавита (N) и количеством знаков в коде (b) - разрядностью двоичного кода.

N	2	4	8	16
b	1 бит	2 бита	3 бита	4 бита

Заметим, что $2 = 2^1$, $4 = 2^2$, $8 = 2^3$, $16 = 2^4$

В общем виде:

$$N = 2^b$$



Алфавит из которого составляется текст на компьютере (документ) состоит из 256 символов. Этот алфавит содержит символы: строчные и прописные латинские и русские буквы, цифры, знаки арифметических операций, всевозможные скобки, знаки препинания...

Из формулы $N = 2^b$ следует $256 = 2^8$

Значит, каждый символ алфавита используемого в компьютере для печати документов весит **8 бит**.

Эту величину приняли так же за единицу измерения информации и дали название байт.

8 бит = 1 байт

Задача:

Статья содержит 30 страниц, на каждой странице - 40 строк, в каждой строке 50 символов. Какой объём информации содержит статья?

- 1) На каждой странице: $50 \cdot 40 = 2000$ (символов);
- 2) Во всей статье: $2000 \cdot 30 = 60000$ (символов);
- 3) Вес каждого символа равен 1 байту, следовательно, информационный объём всей статьи: $60000 \cdot 1 = 60000$ байт
или $60000 \cdot 8 = 480000$ бит

Единицы измерения информации:

1 килобайт = 1 Кб = 2^{10} байт = 1024 байт

1 мегабайт = 1 Мб = 2^{10} Кб = 1024 Кб

1 гигабайт = 1 Гб = 2^{10} Мб = 1024 Мб

Значит, если перевести результат предыдущей задачи в более крупные единицы измерения получим:

60000 байт : 1024 $\approx$ 58,59375 Кб

60000 байт : 1024 : 1024 $\approx$ 0,057 Мб

Задание на дом

§ 4

Задание

Задачи 4, 6, 7 стр. 28