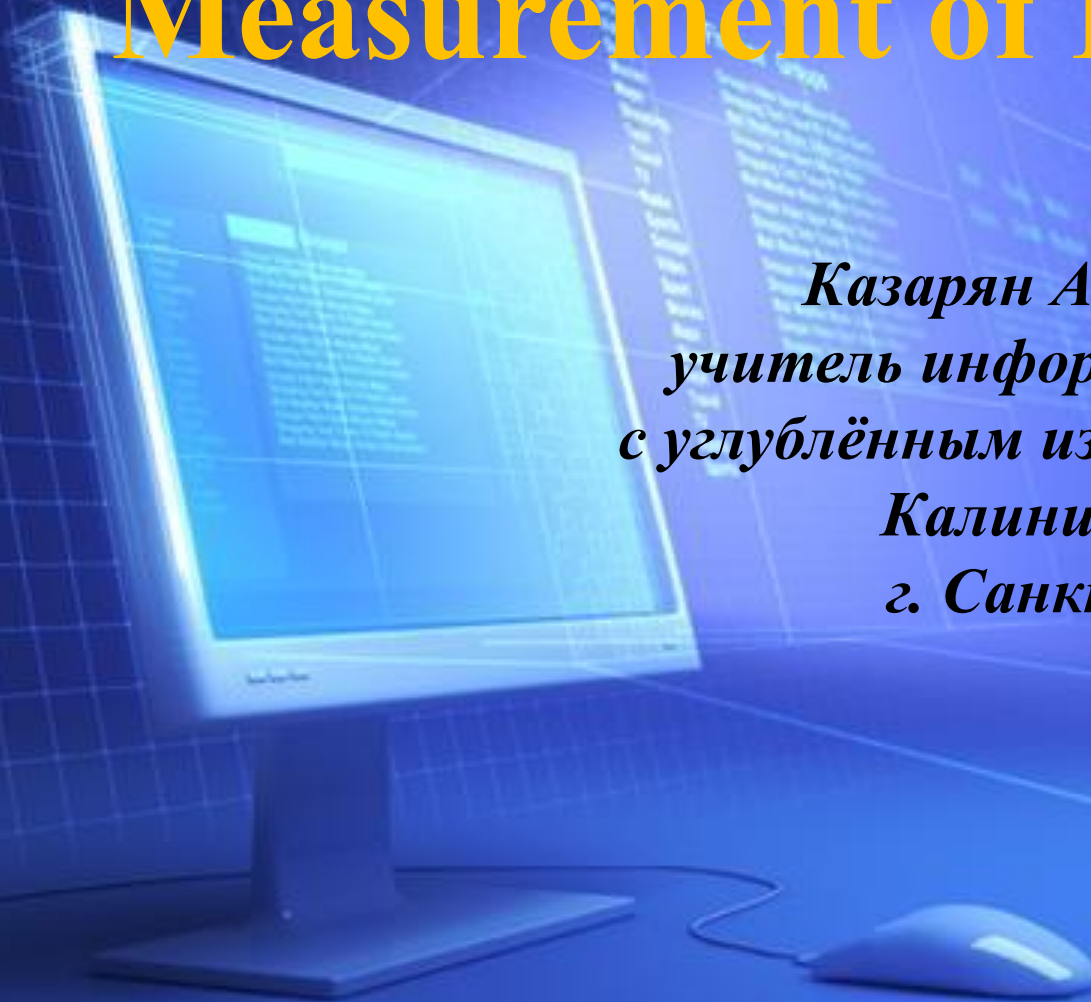


Измерение количества информации

Measurement of Information

*Казарян Анаит Рафиковна
учитель информатики школы №156
с углублённым изучением информатики
Калининского района
г. Санкт-Петербурга*



Кто владеет информацией, тот владеет миром.

Френсис Бэкон

Who owns the information, owns the world.

Francis Bacon



Что такое информация?

What is information?

- ❑ **Информация** – знания, получаемые из разных источников.
- ❑ **Information** is certain knowledge about something or someone.

Как можно измерить информацию?

Can we measure information?

- Расстояние измеряют в миллиметрах, сантиметрах, метрах, дециметрах...
- Массу измеряют в граммах, килограммах, тоннах...

Basic postulate of information theory:
information can be treated like a measurable physical quantity, such as density or mass.

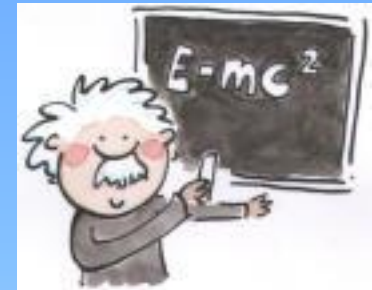
Вывод: *Для измерения информации должна быть введена своя эталонная единица.*

Как можно измерить информацию?

Can we measure information?

Пример 1:

- In winter is cold.
- In winter is cold and snowy.



Предложение 2 даёт более полезную информацию.

Пример 2:

- Tomorrow will be cold and snowy.
- Tomorrow will be warm and sunny.


Непонятно, какое из предложений даёт больше информации.

Подходы к измерению информации



содержательный
(semantic)

Количество информации в тексте связывается с содержанием текста, учитывая вероятности его символов.



алфавитный
(alphabetical)

Количество информации в тексте определяется независимо от его содержания, воспринимая текст как последовательность символов.

Алфавитный подход

Alphabetical approach

- ❑ Впервые информацию измерил американский инженер **Ральф Хартли** в 1928 году.
- ❑ First attempt to quantify information was made by **Ralph Hartley (1928)**.



Алфавитный подход

Alphabetical approach

- Вычислим количество символов в предложениях примера 1:
 1. In winter is cold. (18 символов)
 2. In winter is cold and snowy. (28 символов)
- **Вывод:** Предложение 2 содержит больше информации!

- *Вопрос: Сколько информации содержится в каждом предложении примера 1?*

Формула Хартли

- Каждый символ сообщения имеет определённый *информационный вес* – несёт *фиксированное количество информации*.
- Все символы одного алфавита имеют один и тот же вес, зависящий от мощности (количества символов) алфавита: $N=2^i$.
- Информационный объём I сообщения равен произведению количества K символов в сообщении на информационный вес i символа алфавита:

$$I = K \times i$$

K

Количество символов
в сообщении

i

Информационный вес
символа алфавита

Формула Хартли

пример 1

Английский алфавит содержит 26 букв. Для записи текста нужны пробел, точка, запятая, вопросительный знак, восклицательный знак, тире. Получается расширенный алфавит мощностью в 32 символа.

Пример: Сообщение, записанное буквами 32-символьного алфавита, содержит 18 символов. Какое количество информации оно несёт?

Решение:

$$\begin{array}{l|l} N = 32 & I = K \times i, \\ \underline{K = 18} & N = 2^i \\ I - ? & \end{array}$$

$$32 = 2^i, i = 5, I = 18 \times 5 = 90 \text{ (битов)}$$

Ответ: 90 битов ($28 \times 5 = 140$ (битов) для сообщения из 28 символов).



Содержательный подход

Semantic (probabilistic) approach

- Вычислим количество символов в предложениях примера 2:

пример 2

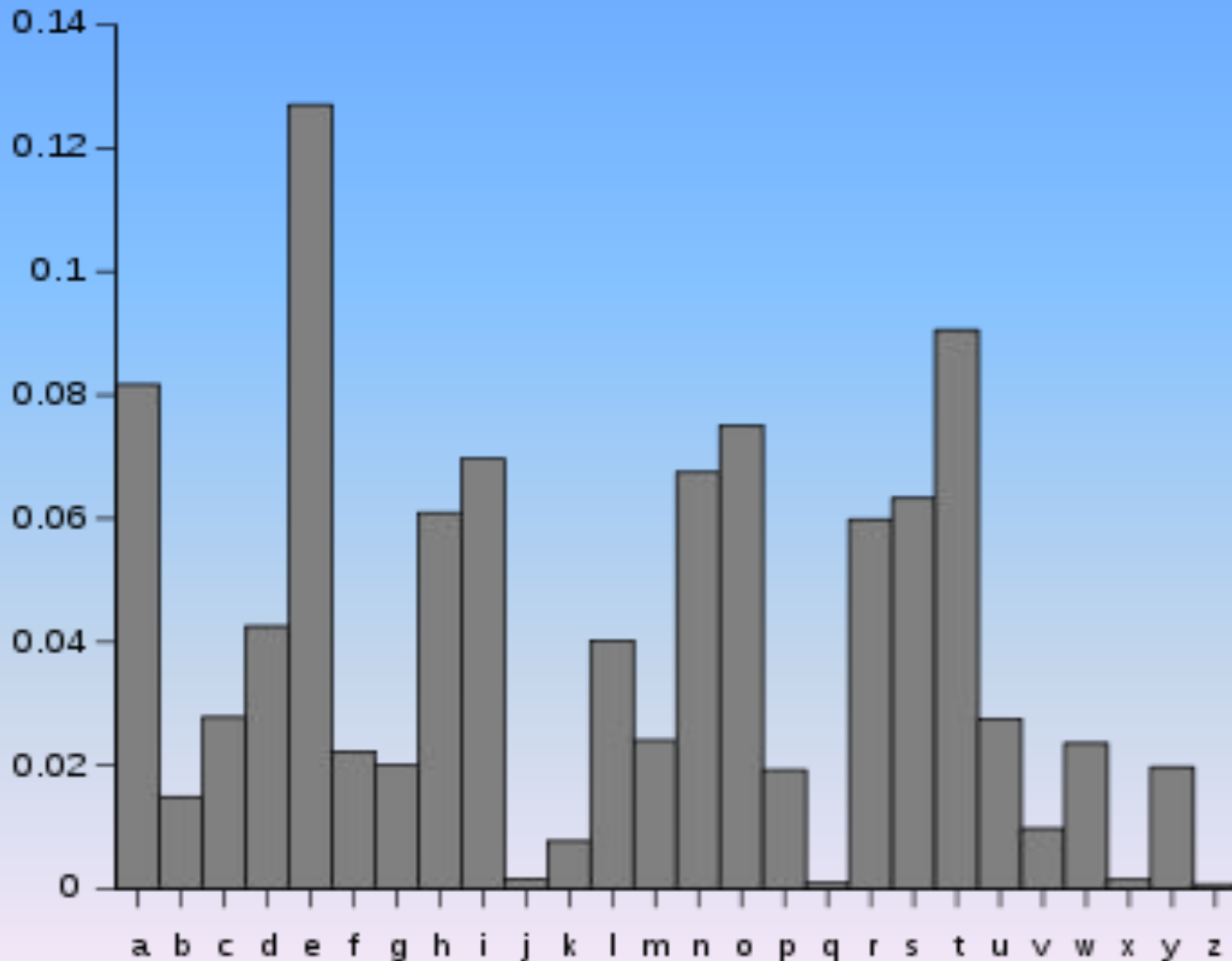
1. Tomorrow will be cold and snowy. (32 символа)
2. Tomorrow will be warm and sunny. (32 символа)

- *Вывод: по формуле Хартли предложения содержат $32 \times 5 = 160$ бит информации.*

- Но сообщение о том, что будет тепло и солнечно, более вероятно летом, а сообщение о холоде и снеге – зимой.

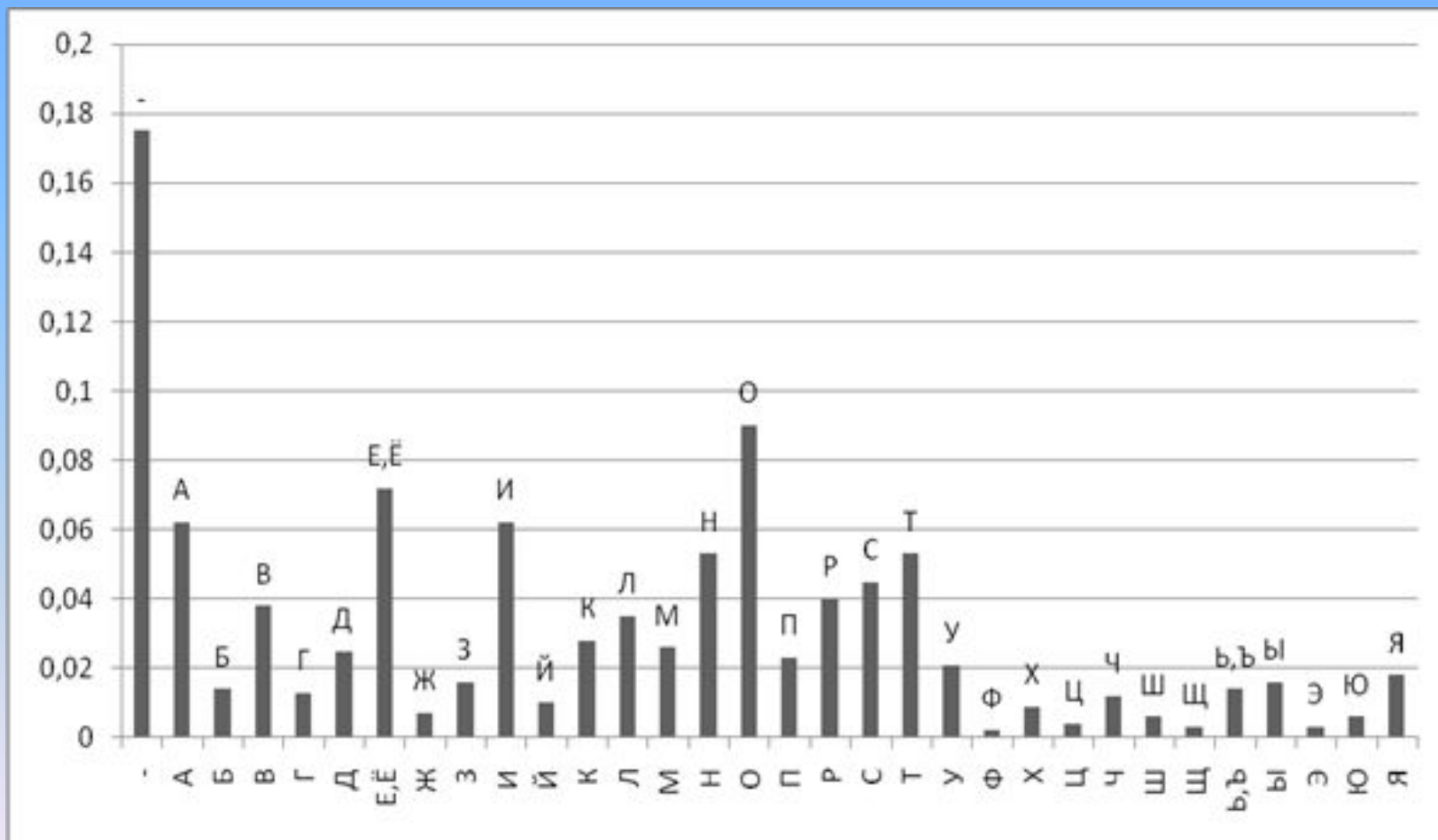
Частота букв английского языка

The frequency of English letters



Частота букв русского языка

The frequency of Russian letters



Формула Шеннона

- Информационный вес символа алфавита вычисляется по формуле:

$$I(p) = \log(1/p)$$

p

Частота появления
символа

\log

Функция, обратная
к степени

- Среднее количество информации в сообщении (*информационная энтропия*) равно

$$H = p_1 \cdot I(p_1) + p_2 \cdot I(p_2) + \dots + p_S \cdot I(p_N),$$

N - мощность (количество символов) алфавита.

буква	частота %	вес, бит	буква	частота %	вес, бит	буква	частота %	вес, бит
A	0.081	3.62	K	0.004	7.97	U	0.024	5.38
B	0.014	6.16	L	0.034	4.88	V	0.009	6.80
C	0.027	5.21	M	0.025	5.32	W	0.015	6.06
D	0.039	4.68	N	0.072	3.80	X	0.002	8.97
E	0.13	2.94	O	0.079	3.66	Y	0.019	5.72
F	0.029	5.11	P	0.02	5.64	Z	0.001	9.97
G	0.02	5.64	Q	0.0015	9.38			
H	0.052	4.27	R	0.069	3.86			
I	0.065	3.94	S	0.061	4.04			
J	0.002	8.97	T	0.105	3.25			

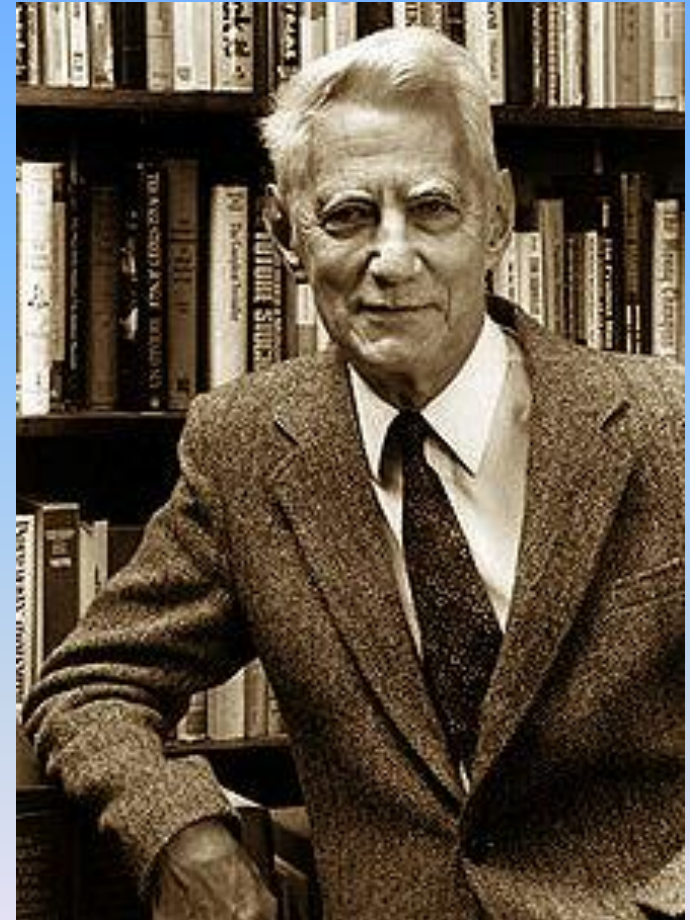
Формула Шеннона

пример 2

Пример 2:

количество информации в первом сообщении равно **3.753** бит,

количество информации во втором сообщении равно **3.856** бит.

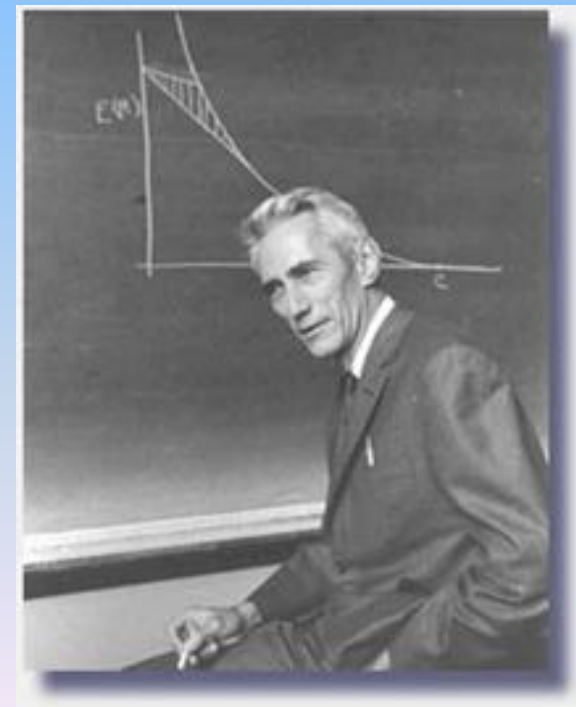


Probabilistic approach: historical notes



- ❖ Американский инженер и математик **Клод Шеннон** в 1948 году в статье «Математическая теория связи» предложил формулу для измерения количества информации в случае *различных вероятностей событий*.

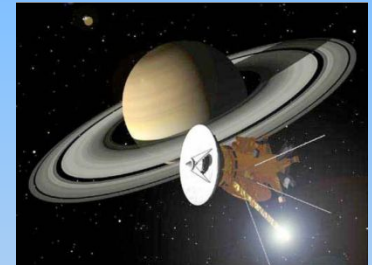
- ❖ **К. Шеннон** является основателем теории информации.
- ❖ **К. Шеннон** первым предложил использовать термин «бит» для обозначения наименьшей единицы измерения информации.



Why is the Information Theory Important?

- Thanks in large measure to Shannon's insights, digital systems have come to dominate the world of communications and information processing.

- Modems
- satellite communications
- Data storage
- Deep space communications
- Wireless technology



Задача

Информационное сообщение объёмом 720 битов состоит из 180 символов. Какова мощность алфавита, с помощью которого записано это сообщение?

Решение:

$I = 720$		$N = 2^i,$		$i = 720/180 = 4$ (бита);	
$K = 180$		$I = K \times i,$			$N = 2^4 = 16$ (СИМВОЛОВ)
<hr/> $N - ?$		$i = I/K$			

Ответ: 16 СИМВОЛОВ.

Единицы измерения информации



КОМПЬЮТЕРНЫЙ АЛФАВИТ

- *русские (РУС) буквы*
- *латинские (LAT) буквы*
- *цифры (1, 2, 3, 4, 5, 6, 7, 8, 9, 0)*
- *математические знаки (+, -, *, /, ^, =)*
- *прочие символы («», №, %, <, >, :, ;, #, &)*

Алфавит содержит 256 символов.

$$256 = 2^8 \Rightarrow i=8$$

1 байт - информационный вес символа алфавита мощностью 256.

1 байт = 8 битов

1 килобайт = 1 Кб = 1024 байта = 2^{10} байтов

1 мегабайт = 1 Мб = 1024 Кб = 2^{10} Кб = 2^{20} байтов

1 гигабайт = 1 Гб = 1024 Мб = 2^{10} Мб = 2^{20} Кб = 2^{30} байтов

1 терабайт = 1 Тб = 1024 Гб = 2^{10} Гб = 2^{20} Мб = 2^{30} Кб = 2^{40} байтов

Задача

Информационное сообщение объёмом 4 Кбайта состоит из 4096 символов. Каков информационный вес символа этого сообщения? Сколько символов содержит алфавит, с помощью которого записано это сообщение?

Решени

$$I = 4 \text{ Кб,}$$

$$K = 4096;$$

$$i - ? \quad N - ?$$

$$N = 2^i, I = K \cdot i, \quad i = I/K$$

$$I = 4 \text{ (Кб)} = 4 \times 1024 \times 8 \text{ (битов)}$$

$$i = 4 \times 1024 \times 8 / 4096 = 8 \text{ (битов)}$$

$$N = 2^8 = 256 \text{ (символов)}$$

Ответ: информационный вес символа = 8,
алфавит содержит 256 символов.

Задача

Для записи текста использовался 64-символьный алфавит. Какой объём информации в байтах содержат 10 страниц текста, если на каждой странице расположено 32 строки по 64 символа в строке?

Решение:

$$N=64, i=6, K=10 * 32 * 64= 20\ 480 \text{ (символов)}$$

$$I= K * i / 8 = 20480 * 6 / 8 = 15360 \text{ (байтов)}$$

Ответ: 15360 байтов.

Интерактивные упражнения

[Упражнение 1](#)

[Упражнение 2](#)

[Упражнение 3](#)

[Упражнение 4](#)

[Упражнение 5](#)

Выводы

- *При алфавитном подходе* измерения информации предполагается, что все символы алфавита встречаются в сообщениях, записанных с помощью этого алфавита, одинаково часто. Однако в действительности символы алфавитов в сообщениях появляются с разной частотой.
- *При вероятностном подходе* измерения информации количество информации в сообщении о некотором событии зависит от вероятности этого события. Чем меньше вероятность некоторого события, тем больше информации содержит сообщение об этом событии. Среднее количество информации достигает максимального значения при равновероятных событиях.