

Измерение информации

Лабораторная работа №2

Единицы измерения

1 бит (*bit*) – *минимальная единица измерения*

1 байт (*byte*) = 2^3 = 8 бит

1 Кб (килобайт) = 1024 байта = 2^{10} байта

1 Мб (мегабайт) = 1024 Кб = 2^{20} байта

1 Гб (гигабайт) = 1024 Мб = 2^{30} байта

1 Тб (терабайт) = 1024 Гб = 2^{40} байта

1 Пб (петабайт) = 1024 Тб = 2^{50} байта

Перевод в другие единицы

Пример

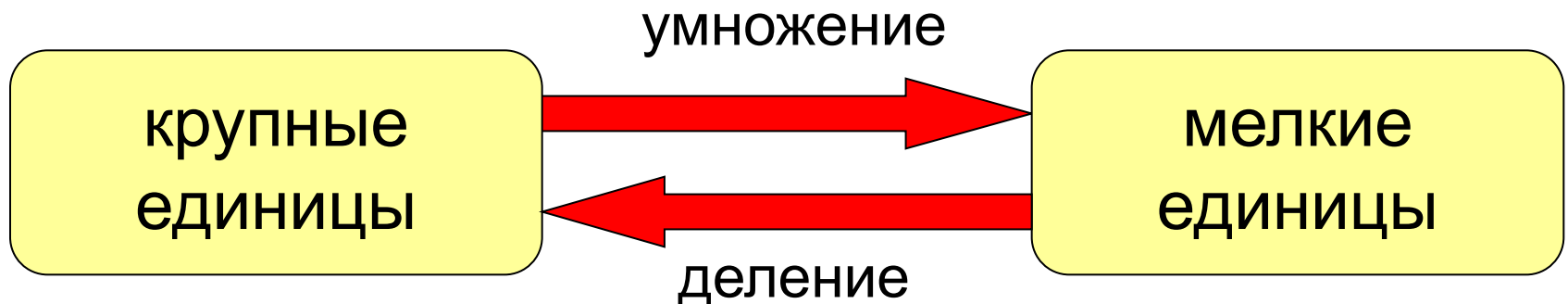
5 Кбайт = $5 * 1024$ байт = 5120 байт

15 байт = $15 * 8$ бит = 120 бит

2048 Кбайт = $2048 / 1024$ Мбайт = 2 Мбайта

1024 Мбайт = $1024/1024$ Гбайт = 1 Гбайт

3 Мбайта = $3 * 1024$ Кбайт = 3072 Кбайта



Задание 1

Сравните (расставьте знаки «=», «>», «<»)

3 байта 24 бита

1000 байт 1 Кб

220 байт 0,25 Кб

1 Мб 1500 Кб

8192 бита 1 Кб

Задание 2

Впишите недостающее число:

8 байт = ? бит

1,5 Кбайт = ? байт

512 бит = ? байт

2 Мбайта = ? Кбайт

4 бита = ? байт

3072 Кбайта = ? Мбайт

Задание 3

*Сравните информационные объемы
(расставьте знаки «=», «>», «<»)*

- 1) 1024 Кбайт ? 2^{20} байт
- 2) 1024 Кбайт ? 2 Мбайт
- 3) 1024 Кбайт ? 2^{10} бит
- 4) 1024 Кбайт ? 1 Гбайт

Задание 4

Расположите объемы памяти

**1 Кбайт; 2 байта; 10 бит; 20 бит; 1010
байт**

в порядке убывания

Алфавитный подход к измерению информации

Информация – последовательность знаков, независимо от содержания

Алфавит – конечное множество различных символов (7, 2, «, а, <, Л, ...)

Мощность алфавита N – количество символов в алфавите

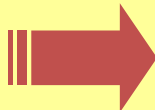
Информационный вес одного символа $i = \log_2 N$ бит ($N = 2^i$)

Информационный объём сообщения из K символов

□ **кодировка ASCII** (*American Standard Code for Information Interchange*)
 $I = K \times i = K \times \log_2 N$ бит

$256 = 2^8$  8 бит (1 байт) на 1 символ

□ **Кодировка Unicode** (включает символы различных письменностей)

$65536 = 2^{16}$  16 бит (2 байт) на 1 символ



Пример. Определить приблизительный информационный объем «Советского энциклопедического словаря»

Будем производить приблизительный расчет информационного объема текста содержащегося в этой книге (без учета графических изображений) для этого найдем страницу равномерно заполненную текстом:

1. На странице три столбца.
2. В столбце 100 строк.
3. В строке (судя по средней строке) 40 символов.
4. В книге 1640 страниц.
5. Т.о., в книге $3 \cdot 100 \cdot 40 \cdot 1640 = 19680000$ символов.
6. Будем считать, что информационный объем одного символа

= 1 Byte, т.о. информационный объем всей книге составляет

19680000 Byte.

7. Осуществим перевод в более крупные единицы измерения:

$$19680000 / 1024 = 19218,75 \text{ Kbyte}$$

Ответ. Приблизительный информационный объем СЭС равен 18,7 MByte.



Алфавитный подход к измерению информации

Задание 5. Сколько учебников (200 страниц, на каждой 60 строк по 64 символа, 1 символ занимает 8 бит) может поместиться на дискете емкостью 1,44 Мбайт; на винчестере в 6,4 Гбайт?

Задание 6. Емкость одного условного печатного листа равна приблизительно 32 Кбайт (1 символ занимает 8 бит), скорость печати - 64 символа в секунду. Сколько минут потребуется для распечатки текста одной газеты (2 усл. п.л.) на матричном принтере (без учета смены бумаги)? Ответ округлить до целого числа

Задание 7. Сколько различных символов, закодированных байтами, содержится в сообщении

110010100111100011111011110010101111101 ?

Задание 8. Сколько пар одинаковых символов, закодированных полубайтами, содержится в сообщении 101110011011100010011100

?

Алфавитный подход к измерению информации

Пример. Чему равно количество информации в слове «Информатика» (используется 32-значный алфавит)?

- считаем все символы $K = 11$
- 1 символ занимает $i = \log_2 32 = 5$ бит
- $I = 11 * 5 = 55$ бит

Пример. Сообщение содержит 4096 символов. Объём сообщения составил 1/512 Мбайт. Чему равна мощность алфавита?

- Переведём объём сообщения в биты
 $1/512 \text{ Мбайт} = 1/512 * 1\,024 * 1\,024 * 8 = 16\,384$ бит
- 1 символ занимает $i = 16384 / 4096 = 4$ бит
- Мощность алфавита равна $N = 2^i = 2^4 = 16$

Алфавитный подход к измерению информации

Задание 9. Определить информационный вес символа при использовании 16-символьного алфавита

Задание 10. Указать в килобайтах объем текстовой информации в сообщении на 40 страницах (на странице 40 строк по 80 символов в каждой) при использовании:

а) кодировки ASCII; б) кода UNICODE

Задание 11. Сообщение, записанное буквами 32-символьного алфавита, содержит 140 символов. Какое количество информации (в битах) оно несёт?

Алфавитный подход к измерению информации

Задание 12. Сколько бит в памяти надо выделить для хранения предложения «Привет, Вася!» при использовании кода UNICODE ?

Задание 13. Сколько символов содержит сообщение, записанное с помощью 16-символьного алфавита, если его объём составляет 1/16 Кбайт?

Задание 14. При перекодировке сообщения из кода UNICODE в код ASCII объём сообщения изменился на 1/512 Мбайт. Сколько символов содержит сообщение?

Алфавитный подход к измерению информации

Задание 15. Информационное сообщение объёмом 720 битов состоит из 180 символов. Какова мощность алфавита, с помощью которого записано это сообщение?

Задание 16. Племя Мульти пишет письма, пользуясь 16-символьным алфавитом. Племя Пульти пользуется 32-символьным алфавитом. Вожди племён обменялись письмами. Письмо племени Мульти содержит 100 символов, а письмо племени Пульти - 50. Сравните информационные объёмы сообщений, содержащихся в письмах.

Вероятностный подход к измерению информации

Вероятность события равна отношению числа случаев благоприятствующих ему (m) к общему числу всех равновозможных несовместных случаев (n)

$$P(A) = \frac{m}{n}$$

Пример. В лотерее разыгрывается 4 диска, 5 книг, 6 ручек
Какова вероятность выиграть диск?

- Общее число исходов $n = 4 + 5 + 6 = 15$
- Число благоприятных исходов $m = 4$
- Вероятность $P(A) = \frac{m}{n} = \frac{4}{15} = 0,26$

Задание 17. В коробке 5 синих, 4 красных и 3 зелёных карандаша. Берут 1 карандаш. Какова вероятность, что взят зелёный карандаш?

Формула Хартли

Количество информации i , содержащееся в сообщении об одном из N **равновероятных** исходов некоторого события, определяется из уравнения

$$2^i = N \text{ или } i = \log_2 N \quad \text{- формула Хартли}$$

Пример. Определить количество информации на один символ в сообщении на русском языке, считая вероятности символов одинаковыми

Решение. В русском языке 34 символа (33 буквы + символ пробела)
Информация о символе («это буква А») – исчерпывающая
Считая вероятности символов одинаковыми по формуле *Хартли*

$$I = H = \log_2 34 = 5,09 \text{ бит}$$

Задание 18. Какое количество информации содержится в сообщении об оценке, полученной учеником за контрольную работу?

Формула Хартли

Задание 19

Считая возможные исходы равновероятными событиями, определите количество информации в сообщениях:

- a. На игральном кубике выпала цифра 3.
- b. В следующем году ремонт в школе начнётся в феврале.
- c. Я приобрёл абонемент в бассейн на среду.
- d. Из 30 студентов группы старостой назначили Марию Петрову.

Формула Хартли

Пример. Определить количество информации в сообщении «это не буква Щ» (вероятности символов считать одинаковыми)

Решение. Информация в сообщении не является исчерпывающей
По формуле *Хартли* энтропия системы до получения сообщения

$$H_0 = \log_2 34 = 5,09 \text{ бит}$$

После получения сообщения $H_1 = \log_2 33 = 5,04 \text{ бит}$

Количество информации $I = H_0 - H_1 = 0,05 \text{ бит}$

Задание 20. В поезде 8 вагонов. Какое количество информации несет сообщение о том, что билет куплен
а) в 3-ий вагон? б) не в 7-й вагон?

Формула Хартли

Пример. На остановке останавливаются автобусы трех маршрутов: №1, №2 и №5, которые ходят с одинаковой частотой. Определить количество информации в сообщениях:

- а) «Пришел автобус №1» б) «Это не автобус №2»

Решение.

Сообщения «Пришел автобус №...» равновероятны. По формуле Хартли:

- а) $I = H_0 = \log_2 3 = 1,58 \text{ бит}$ б) $I = H_0 - H_1 = \log_2 3 - \log_2 2 = 1,58 - 1 = 0,58 \text{ бит}$

Задание 21. На остановке останавливаются автобусы двух маршрутов: №1 и №5, которые ходят с одинаковой частотой. Определить количество информации в сообщениях:

- а) «Пришел автобус №1» б) «Это не автобус №5»

Формула Шеннона

Количество информации для событий с различными вероятностями

$$i = \log_2 \frac{1}{p}$$

Средняя информативность сообщения в случае событий с различными вероятностями:

$$I = - \sum_{i=1}^N p_i \log_2 p_i \quad - \text{ формула Шеннона}$$

Формула Шеннона

Пример. За год по математике ученик получил 100 оценок: 60 – «5», 25 – «4», 10 – «3», 5 – «2». Определить количество информации в сообщениях об оценках.

■ Определим вероятности и количество информации каждой оценки:

$$\text{«5»} - p_1 = 60/100 = 0,6$$

$$I_5 = \log_2 1/0,6 = 0,737 \text{ бита}$$

$$\text{«4»} - p_2 = 25/100 = 0,25$$

$$I_4 = \log_2 1/0,25 = 2 \text{ бита}$$

$$\text{«3»} - p_3 = 10/100 = 0,1$$

$$I_3 = \log_2 1/0,1 = 3,322 \text{ бита}$$

$$\text{«2»} - p_4 = 5/100 = 0,05$$

$$I_2 = \log_2 1/0,05 = 4,322 \text{ бита}$$

Пример. При бросании несимметричной четырёхгранной пирамидки вероятности падения на каждую из граней составляют: $p_1=1/2$; $p_2=1/4$; $p_3=1/8$; $p_4=1/8$. Определить количество информации, которое будет получено после реализации одного из случаев

■ Используем формулу Шеннона

$$I = - \sum_{i=1}^N p_i \log_2 p_i = - \left(1/2 \cdot \log_2 1/2 + 1/4 \cdot \log_2 1/4 + 1/8 \cdot \log_2 1/8 + 1/8 \cdot \log_2 1/8 \right) = 14/8 \text{ бит} = 1,75 \text{ бит.}$$

Формула Шеннона

Пример. На остановке останавливаются автобусы трех маршрутов: №1, №2 и №5, причем №1 – вдвое чаще, чем №2 и №5. Определить количество информации в сообщениях:

а) «Пришел автобус №1» б) «Это не автобус №2» в) «Это не автобус №1»

Решение. Сообщения «Пришел автобус №...» неравновероятны: $p(№1)=0,5$; $p(№2)=0,25$; $p(№5)=0,25$. По формуле Шеннона:

а) $I = H_0 = -(0,5 \log_2 0,5 + 0,25 \log_2 0,25 + 0,25 \log_2 0,25) = 1,5 \text{ бит}$

б) $I = H_0 - H_1 = 1,5 - (- (2/3 \log_2 2/3 + 1/3 \log_2 1/3)) = 1,5 - 0,92 = 0,58 \text{ бит}$

в) $I = H_0 - H_1 = 1,5 - (- (1/2 \log_2 1/2 + 1/2 \log_2 1/2)) = 1,5 - 1 = 0,5 \text{ бит}$

Задание 22. На остановке останавливаются автобусы двух маршрутов: №1 и №5, причем №1 – вдвое чаще, чем №5. Определить количество информации в сообщениях:

а) «Пришел автобус №1» б) «Пришел автобус №5» в) «Это не автобус №5»

Формула Шеннона

Задание 23. В коробке 5 синих, 4 красных и 3 зелёных карандаша. Берут 1 карандаш. Определить количество информации в сообщении, что взят зелёный карандаш.

Задание 24. Синоптики подсчитали, что в течении 100 лет 10 марта было 34 дождливых дня, снег выпал 28 раз и 38 дней было без осадков. Определить количество информации в сообщениях, что 10 марта текущего года:

- Будет снег
- Будет дождь
- Осадков не будет

Задание 25. В корзине лежат белые и 18 чёрных шаров. Сообщение о том, что из корзины достали белый шар, несёт 2 бита информации. Сколько всего шаров в корзине?