

*Тема: Различные подходы к
измерению количества информации.*

Урок 2

Существуют различные подходы к измерению количества информации:

Содержательный

Алфавитный

Вероятностный

Содержательный подход

***к измерению количества
информации***

Информация - это знания, которые мы получаем из внешнего мира.

Сообщение информативно

(т.е. содержит ненулевую информацию), **если оно пополняет знания человека.**

1 бит - минимальная единица измерения количества информации.

Содержательный подход

Проблема измерения информации исследована в теории информации, основатель которой - **Клод Шеннон**.

В теории информации для бита дается следующее определение:

Сообщение, уменьшающее неопределенность знания в два раза, несет *1 бит* информации.

Пример 1: Вы бросаете монету, загадывая, что выпадет: орел или решка?

Решение:

Есть два варианта возможного результата бросания монеты. Ни один из этих вариантов не имеет преимущества перед другим (равновероятны). Перед подбрасыванием монеты неопределенность знаний о результате равна двум.

После совершения действия неопределенность уменьшилась в 2 раза. Получили **1 бит** информации.

Пример 2: Студент на экзамене может получить одну из четырех оценок: **5, 4, 3, 2** с одинаковой вероятностью. После сдачи экзамена, на вопрос: «Что получил?» - ответил: «Четверку». Сколько бит информации содержится в его ответе?

Решение: можно отгадать оценку, задавая вопросы, на которые можно ответить только «да» или «нет», т.е. поиск осуществляется отбрасыванием половины вариантов.

Каждый ответ уменьшает количество вариантов в два раза и, следовательно, приносит 1 бит информации.

1 вопрос: -Оценка выше тройки? - ДА

Получен **1 бит** информации.

2 вопрос: -Ты получил пятерку? - НЕТ

(значит получил 4)

Получен еще **1 бит**.

В сумме имеем 2 бита.

Содержательный подход

Неопределенность знания о результате некоторого события (бросание монеты или игрального кубика, вытаскивание жребия и др.) - это количество возможных результатов.

Пример 3: На книжном стеллаже восемь полок. Книга может быть поставлена на любую из них. Сколько информации содержит сообщение о том, где находится книга?

1 вопрос: - Книга лежит выше четвертой полки?

- НЕТ (1, 2, 3, 4) - **1 бит**

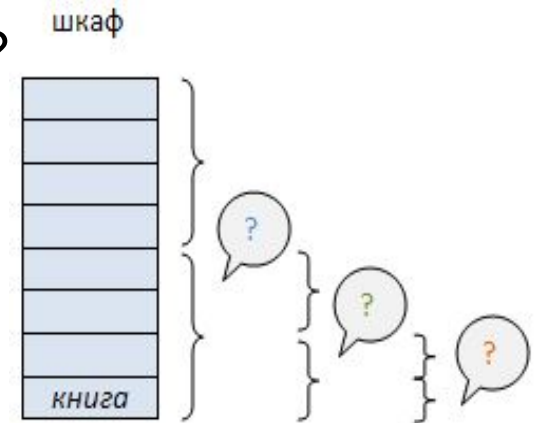
2 вопрос: - Книга лежит ниже третьей полки?

- ДА (1, 2) - **1 бит**

3 вопрос: - Книга – на второй полке?

- НЕТ (1) - **1 бит**

Книга лежит на первой полке.



Ответ: **3 бита** информации (каждый ответ уменьшал неопределенность в два раза. Всего было задано три вопроса.)

Получим формулу вычисления количества информации.

Обозначим: N – количество возможных событий (неопределенность знаний)

i - количество информации в сообщении о том, что произошло одно из N событий.

В примере с монетой $N = 2, i = 1 \quad 2^1 = 2$

В примере с оценками $N = 4, i = 2 \quad 2^2 = 4$

В примере со стеллажом $N = 8, i = 3 \quad 2^3 = 8$

Получаем формулу: $2^i = N$

Формула Хартли

$$2^i = N$$

N – число равновероятных событий

i – количество информации в сообщении

$$i = \log_2 N$$

Пример:

Шахматная доска состоит из **64** полей: **8** столбцов на **8** строк.

Какое количество бит несет сообщение о выборе одного шахматного поля?

- **Решение.**
- Поскольку выбор любой из 64 клеток равновероятен, то количество бит находится из формулы:
- $2^i = 64$,
- $i = \log_2 64 = 6$, так как $2^6 = 64$.
- Следовательно, $i = 6$ бит.

Пример:

Сколько информации содержит сообщение о том, что из колоды в **32** карты достали король пик?

- *Решение:*
- *В колоде 32 карты. В перемешанной колоде выпадение любой карты равновероятное событие.*
- $N = 32. I - ?$
- $2^I = N$
- $2^I = 32$
- $2^5 = 32$
- $I = 5 \text{ бит}$

Пример:

При игре в кости используется кубик с шестью гранями.

Сколько битов информации получает игрок при каждом бросании кубика?

Решение.

Выпадение каждой грани кубика равновероятно. Поэтому количество информации от одного результата бросания находится из уравнения: $2^i=6$.

Решение этого уравнения: $i=\log_2 6$

*Из таблицы двоичных логарифмов следует (с точностью до 3-х знаков после запятой):
 $i=2,585$ бита.*

Данную задачу также можно решить округлением i в большую сторону: $2^i=6<8=2^3, i=3$ бита.

Таблица 1.1. Количество информации в сообщении об одном из N равновероятных событий

N	i	N	i	N	i	N	i
1	0,00000	17	4,08746	33	5,04439	49	5,61471
2	1,00000	18	4,16993	34	5,08746	50	5,64386
3	1,58496	19	4,24793	35	5,12928	51	5,67243
4	2,00000	20	4,32193	36	5,16993	52	5,70044
5	2,32193	21	4,39232	37	5,20945	53	5,72792
6	2,58496	22	4,45943	38	5,24793	54	5,75489
7	2,80735	23	4,52356	39	5,28540	55	5,78136
8	3,00000	24	4,58496	40	5,32193	56	5,80735
9	3,16993	25	4,64386	41	5,35755	57	5,83289
10	3,32193	26	4,70044	42	5,39232	58	5,85798
11	3,45943	27	4,75489	43	5,42626	59	5,88264
12	3,58496	28	4,80735	44	5,45943	60	5,90689
13	3,70044	29	4,85798	45	5,49185	61	5,93074
14	3,80735	30	4,90689	46	5,52356	62	5,95420
15	3,90689	31	4,95420	47	5,55459	63	5,97728
16	4,00000	32	5,00000	48	5,58496	64	6,00000

Какое количество информации можно получить при угадывании числа из интервала от **1** до **11**?

Решение:

$$N=11$$

Чтобы найти количество информации (I), необходимо воспользоваться таблицей.

По таблице

$$I= 3,45943 \text{ бит}$$

Задано слово из **10** букв. Вы просите открыть пятую букву. Вам ее открыли. Сколько информации вы получили?

Решение:

$N=10$, следовательно, $I=\log_2 10$.

Смотрим по таблице и видим, что $I=3,32193$ бит.

Ответ: 3,3 бит

В школьной библиотеке **16** стеллажей с книгами. а каждом стеллаже **8** полок. Библиотекарь сообщил Пете, что нужная ему книга лежит на **5** стеллаже на **3** сверху полке. какое кол-во информации библиотекарь передал Пете?

Решение:

$$2^i = N$$

$$N = 16 * 8 = 128$$

$$2^i = 128$$

$$i = 7$$

Ответ: 7бит

В коробке лежат **6** разноцветных фломастеров. Какое количество информации содержит сообщение, что из коробки достали синий фломастер?

Решение:

$$N = 6,$$

следовательно,

$$I = \log_2 6.$$

Вычисляем и имеем $I = 2,58496$ бит.

***Ответ:* 2,5 бит.**

Какое количество информации несёт сообщение: «Встреча назначена на май?»

Решение:

так как месяцев в году 12, то из этого количества сообщений нужно выбрать одно.

Значит

$$N = 12, \text{ а } I = \log_2 12.$$

вычисляем и имеем $I = 3,58496$ бит.

Какое количество информации несёт сообщение о том, что встреча назначена на **20** число?

Решение:

так как дней в месяце 30 или 31,
то из этого количества сообщений нужно
выбрать одно.

Значит, $N = 30$ или 31 , $I = \log_2 30$ (или 31).

Вычисляем и имеем $I = 4,9$ бит.

Ответ: 4,9 бит.

Сообщение о том, что ваш друг живет на десятом этаже несет в себе **4** бита информации. Сколько этажей в доме?

Дано:

$$i=4 \text{ бита}$$

Решение:

$$N = 2^i$$

$$N = 2^4$$

Ответ: 16

Какое количество информации несёт сообщение о том, что встреча назначена на **20 число?**

Решение:

так как дней в месяце 30 или 31, то из этого количества сообщений нужно выбрать одно.

Значит,

$$N = 30 \text{ или } 31, I = \log_2 30 \text{ (или } 31\text{)}.$$

Вычисляем и имеем

$$I = 4,9 \text{ бит.}$$

Ответ: 4,9 бит.

*Алфавитный подход к
измерению количества
информации.*

Урок 4

Алфавитный (объёмный) подход к измерению информации позволяет определить количество информации, заключенной в тексте, записанном с помощью некоторого алфавита.

Алфавит -

**МНОЖЕСТВО ИСПОЛЬЗУЕМЫХ
СИМВОЛОВ В ЯЗЫКЕ.**

**Мощность алфавита (N) -
КОЛИЧЕСТВО СИМВОЛОВ,
ИСПОЛЬЗУЕМЫХ В АЛФАВИТЕ.**

Если допустить, что все символы алфавита встречаются в тексте с одинаковой частотой (равновероятно), то количество информации, которое несет каждый символ, вычисляется

по формуле Хартли:

$$i = \log_2 N,$$

где N - мощность алфавита

Формула Хартли задает связь между количеством возможных событий N и количеством информации i

$$N = 2^i$$

Если весь текст состоит из K символов, то при алфавитном подходе размер содержащейся в нем информации равен:

$$I = K * i,$$

где i – информационный вес одного символа в используемом алфавите.

При алфавитном подходе к измерению информации информационный объем текста зависит только от размера текста и от мощности алфавита, а не от содержания. Поэтому нельзя сравнивать информационные объемы текстов, написанных на разных языках, по размеру текста.

Найти объем информации, содержащейся в тексте из **3000** символов , написанном на русскими буквами.

Решение:

Найдем мощность алфавита:

$N = 33$ русских прописных букв + 33 русских строчных букв + 21 специальный знак = 87 СИМВОЛОВ.

Подставим в формулу и рассчитаем количество информации:

$I = \log_2 87 = 6,4$ бита.

$6,4 * 3000 = 19140$ бит.

Найти количество информации, содержащейся в немецком тексте с таким же количеством символов.

Решение:

Найдем мощность немецкого алфавита:

$N = 26$ немецких прописных буквы + 26 немецких строчных букв + 21 специальный знак = 73 символа.

Найдем информационный объем одного символа:

$I = \log_2 73 = 6,1$ бита.

Найдем объем всего текста.

$6.1 * 3000 = 18300$ бит.

Правило для измерения информации с точки зрения алфавитного подхода.

- Найти мощность алфавита – N
- Найти информационный объем одного символа – $i = \log_2 N$
- Найти количество символов в сообщении – K
- Найти информационный объем всего сообщения – $K * i$

*Считая, что каждый символ кодируется одним байтом, оцените информационный объем следующего предложения: **Белеет Парус
Одинокий В Тумане Моря Голубом!***

Решение.

Так как в предложении 44 символа (считая знаки препинания и пробелы), то информационный объем вычисляется по формуле:

$$I = 44 \cdot 1 \text{ байт} = 44 \text{ байта} = 44 \cdot 8 \text{ бит} = 352 \text{ бит}$$

а

Найти объем текста, записанного на языке, алфавит которого содержит **128** символов и **2000** символов в сообщении.

Дано: $K = 2000$, $N = 128$.

Найти: $I_T - ?$

Решение:

$I = \log_2 N = \log_2 128 = 7$ бит - объем
одного символа.

$I_T = I * K = 7 * 2000 = 14000$ бит.

Ответ: 14000 бит.

Информационный объем одного символа некоторого сообщения равен **5** битам. Каковы пределы (максимальное и минимальное значение) мощности алфавита, с помощью которого составлено это сообщение?

Решение: $N = 2^1 = 2^5 = 2^6 = 32$ – максимальное значение мощности алфавита. Если символов будет больше хотя бы на один, то для кодирования понадобится 6 бит.

Минимальное значение – 17 символов, т.к. для меньше количества символов будет достаточно 4 бит.

Ответ: 4 бита.

Информационное сообщение объемом **4** Кбайта содержит **4096** символов. Сколько символов содержит алфавит, при помощи которого было записано это сообщение?

Дано: $K = 4096$, $I_T = 4$ Кб.

Найти: N - ?

Решение:

$N = 2^I$, неизвестно I ;

$I_T = K * I$, $I = I_T / K = 4 * 1024 * 8 / 4096 = 8$ бит –
объем одного файла.

$N = 2^8 = 256$ символов – мощность
алфавита.

Ответ: алфавит содержит 256 символов.

Объем сообщения, содержащего **2048** символов, составил **1/512** часть Мбайта. Каков размер алфавита, с помощью которого записано сообщение?

Дано: $K = 2048$, $I_T = 1/512$ Мбайта.

Найти: N -?

Решение:

Выразим $I_T = 1/216$ Мбайта в битах: $1/29$ Мб = $1/29 * 2^{20} * 2^3 = 214$ бит;

$N = 21$, неизвестно I ;

$I_T = K * I$, $I_T / K = 214 / 2048 = 214 / 2^{11} = 2^3 = 8$ бит –
объем одного символа;

$N = 2^8 = 256$ символов – мощность алфавита.

Для записи текста использовался **256**-символьный алфавит. Каждая страница содержит **30** строк по **70** символов в строке. Какой объём информации содержит **5** страниц текста?

Дано: $N=256$, $X=30$ – количество строк, $Y=70$ – количество символов в строке, $M=5$ – количество страниц.

Найти: I_T - ?

Решение:

1) $I = \log_2 N = \log_2 256 = 8$ бит = 1 байт – объём одного символа;

2) $K = X \times Y \times M = 30 \times 70 \times 5 = 10500$ символов – в тексте;

3) $I_T = I \times K = 1 \times 10500 = 10500$ байт ≈ 10 Кбайт – объём всего текста.

Ответ: объём всего текста 10 Кбайт.

Для записи сообщения использовался **64**-символьный алфавит. Каждая страница содержит **30** строк. Всё сообщение содержит **8775** байтов информации и занимает **6** страниц. Сколько символов в строке?

Дано: $N=64$, $M=6$, $X=30$, $I_T=8775$ байтов,

Найти: Y - ?

Решение:

1) $K=X \times Y \times M$, $y=K/(x \times M)$ – неизвестно K ;

2) $K=I_T/I$ – неизвестно I ;

3) $I=\log_2 N - \log_2 64 = 6$ бит – объём одного символа;

4) $K=8775 \times 8 / 6 = 11700$ символов в тексте;

5) $Y=11700 / (30 \times 6) = 65$ символов в строке.

Ответ: в строке 65 символов.

Пользователь вводит текст с клавиатуры со скоростью **90** символов в минуту.

Какое количество информации будет содержать текст, который он набирал **15** минут (используется компьютерный алфавит)?

Дано: $V=90$ зн/мин $t=15$ мин, $N=256$.

Найти: I_T - ?

Решение:

1) $I_T = I \times K$;

2) $K = V \times t = 90 \times 15 = 1350$ символов содержит текст;

3) $I = \log_2 N = \log_2 256 = 8$ бит = 1 байт – объём одного символа;

4) $I = 1350 \times 1 = 1350$ байт $\times 1,3$ Кбайт – объём всего текста.

Ответ: текст содержит 1,3 Кбайта информации.

Ученик **9** класса читает текст со скоростью **250** символов в минуту. При записи текста использовался алфавит, содержащий **64** символа. Какой объём информации получит ученик, если будет непрерывно читать **20** минут?

Дано: $V=250$ сим/мин, $N=64$, $t=20$ мин.

Найти: I_T - ?

Решение:

1) $I_T = I \times K$;

2) $I = \log_2 N = \log_2 64 = 6$ бит – объём одного символа;

3) $K = V \times t = 250 \times 20 = 5000$ символов в тексте;

4) $I = 5000 \times 6 = 30000$ бит = 3750 байт $\approx 3,7$ Кбайт – объём текста.

Ответ: ученик получил 3,7 Кбайта информации.

Книга, набранная с помощью компьютера, содержит 150 страниц; на каждой странице – 40 строк, в каждой строке – 60 символов. Каков объем информации в книге?

Решение. Мощность компьютерного алфавита равна 256.

Один символ несет 1 байт информации.

($256=2^8=2^i$, $i=8$ бит=1 байт)

Значит, страница содержит $40 \times 60 = 2400$ байт информации.

Объем всей информации в книге (в разных единицах):

$2400 \times 150 = 360\,000$ байт.

$360000/1024 = 351,5625$ Кбайт.

$351,5625/1024 = 0,34332275$ Мбайт.

Подсчитайте объем информации, содержащейся в романе А. Дюма "Три мушкетера", и определите, сколько близких по объему книг можно разместить на одном лазерном диске? (в книге 590 стр., 48 строк на одной странице, 53 символа в строке)

Решение.

- 1) $590 * 48 * 53 = 1500960$ (символов).
- 2) $1500960 \text{ байт} = 1466 \text{ Кбайт} = 1,4 \text{ Мбайт}$.
- 3) На одном лазерном диске емкостью 600 Мбайт можно разместить около 428 произведений, близких по объему к роману А. Дюма "Три мушкетера".

*Вероятностный подход к
измерению количества
информации.*

Урок 4