

Информация и информационные процессы

- § 1. Количество информации
- § 2. Передача данных
- § 3. Сжатие данных
- § 4. Информация и управление
- § 5. Информационное общество

Информация и информационные процессы

§ 1. Количество информации

Формула Хартли (1928)

$$N = 2^I$$

$$I = \log_2 N$$

I – количество информации в битах
 N – количество вариантов



Ральф
Хартли

Пример:

В аэропорту стоит 10 самолетов, из них один летит в Санкт-Петербург. Оценить количество информации в сообщении «В Санкт-Петербург летит второй самолет»?

$$I = \log_2 10 = \frac{\ln 10}{\ln 2} = \frac{\lg 10}{\lg 2} = 3,322 \text{ бита}$$

Алфавитный подход

N – мощность алфавита

Информационный объём

символа:

$$i = \log_2 N$$

вверх до целого
числа

сообщения длиной L :

$$I = L \cdot \log_2 N$$

Пример: сообщение длиной 100 символов закодировано с помощью алфавита из 50 знаков.

$$i = \log_2 50 \approx 5,644 \text{ бита}$$

6 битов

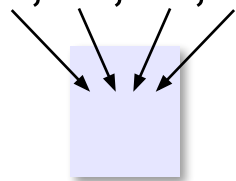
$$I = 100 \cdot \log_2 50 \approx 564,4 \text{ бита}$$

600 битов

Количество различных сообщений

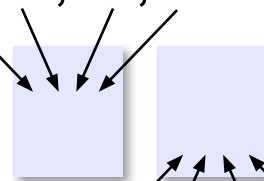
алфавит: А, Б, В, Г

А, Б, В, Г



всего: 4

А, Б, В, Г



всего: $4 \cdot 4 = 4^2 = 16$

А, Б, В, Г для **каждого** варианта

N – мощность алфавита

L – длина сообщения

Q – количество различных сообщений

$$Q = N^L$$

Информация и вероятность

Доля символов в русских текстах:

из 1000
символов
около 175
пробелов

вероятность p
появления символа

	0,175	Я	0,018
О	0,090	Ы	0,017
Е	0,072	З	0,016
А	0,063	Ь	0,015
И	0,062	Б	0,014
Т	0,053	Г	0,013
Н	0,052	Ч	0,012
С	0,045	Й	0,010
р	0,040	Х	0,009
В	0,038	Ж	0,007
Л	0,035	Ю	0,006
К	0,028	Ш	0,005
М	0,026	Ц	0,004
Д	0,025	Щ	0,003
П	0,023	Э	0,002
У	0,021	Ф	0,001

Вероятность

Вероятность события – число от 0 до 1, показывающее, как часто случается это событие в большой серии одинаковых опытов.

$$0 \leq p \leq 1$$

$$x^2 < 0$$

$$p = 0$$

событие **никогда** не происходит
(нет неопределенности)

$$p = 0,5$$

событие происходит в половине случаев (**есть неопределенность**)



$$p = 1$$

событие происходит **всегда**
(нет неопределенности)

$$x^2 \geq 0$$

Вероятность

N – количество испытаний

m – сколько раз произошло событие

$$p = \frac{m}{N}$$

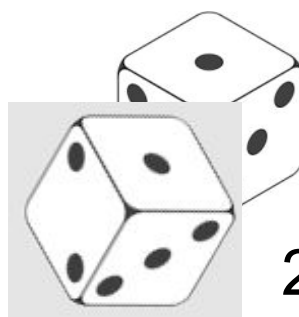


ровно 2:

$$p = \frac{1}{6}$$

чётное: $p = \frac{3}{6} = \frac{1}{2}$

меньше 3: $p = \frac{2}{6} = \frac{1}{3}$



2 и 2:

$$p = \frac{1}{36}$$

2 чётных: $p = \frac{3 \cdot 3}{36} = \frac{1}{4}$

оба меньше 3: $p = \frac{4}{36} = \frac{1}{9}$

Вероятность и информация

Чем более неожиданно событие, тем больше получено информации.

$p = 1$...AAAAAAAAAAAAAAAAAAAA

получили букву «А»: $I = 0$

$p \rightarrow 0$...**V**AAAAAAAAAAAAAAAAAAAA

получили букву «В»: $I \rightarrow \infty$

В 10 опытах будет получено в 10 раз больше информации, чем в одном (**аддитивность**).



Определили свойства количества информации!

Вероятность и информация

$f(p) = -K \cdot \log_2 p$ при $K = 1 \Rightarrow$ информация в битах

Если событие имеет вероятность p , то количество информации в битах, полученное в сообщении об этом событии, равно

$$I = -\log_2 p = \log_2 \frac{1}{p}$$

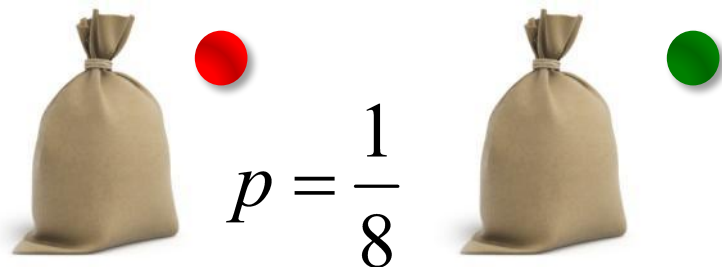
$$p = 1 \Rightarrow I = \log_2 1 = 0$$

$$p \rightarrow 0 \Rightarrow I \rightarrow \log_2 \infty = \infty$$

Вероятность и информация

Аддитивность:

по 8 шариков разного цвета



$$p = \frac{1}{8}$$

всего $8 \cdot 8 = 64$

варианта

$$p = \frac{1}{64}$$

$$I_1 = I_2 = \log_2 \frac{1}{p} = \log_2 8 = 3 \text{ бита}$$

$$I = I_1 + I_2 = 6 \text{ битов}$$

$$I = \log_2 64 = 6 \text{ битов}$$



Аддитивность выполняется!

Связь с формулой Хартли

N равновероятных событий $\Rightarrow p = \frac{1}{N}$

$$I = \log_2 \frac{1}{p} = \log_2 N$$

совпадает с
формулой Хартли

Если вероятности разные:



«**Васе достался зелёный шарик**».

$$p = \frac{6}{8} = \frac{3}{4}$$

$$I = \log_2 \frac{4}{3} \approx 0,415 \neq 0,5$$

Формула Шеннона

Количество полученной информации равно уменьшению неопределенности.

$$I = \Delta H = H_{\text{нач}} - H_{\text{кон}}$$



Как вычислить H ?

Неопределённость знаний об источнике данных (N событий, вероятности p_i):

$$H = \sum_{i=1}^N p_i \cdot \log_2 \frac{1}{p_i} = p_1 \cdot \log_2 \frac{1}{p_1} + \dots + p_N \cdot \log_2 \frac{1}{p_N}$$



Клод Шеннон


информационная **энтропия**

Формула Шеннона

«Идёт ли сейчас снег?» (1 – да, 2 – нет)

зимой: $p_1 = \frac{1}{2}$  Как вычислить p_2 ?

$$p_2 = 1 - p_1 = \frac{1}{2}$$

 Сумма вероятностей всех событий, составляющих полную систему, равна 1!

$$H = \frac{1}{2} \cdot \log_2 2 + \frac{1}{2} \cdot \log_2 2 = \log_2 2 = 1 \text{ би/т}$$

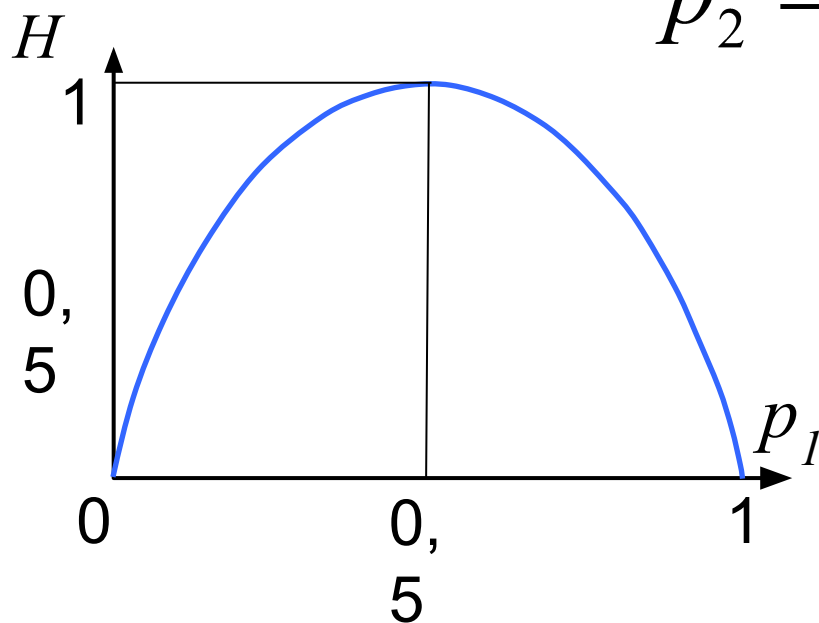
летом: $p_1 = 0,0001$, $p_2 = 0,9999$

$$H = 0,0001 \cdot \log_2 \frac{1}{0,0001} + 0,9999 \cdot \log_2 \frac{1}{0,9999} \approx 0,0015 \text{ би/т}$$

Когда неопределённость наибольшая?

Система двух событий:

$$p_2 = 1 - p_1$$



Неопределенность
максимальна, когда все
события равновероятны.

совпадает с
формулой Хартли!

$$p_1 = p_2 = \dots = p_N = \frac{1}{N}$$

$$H = \sum_{i=1}^N \frac{1}{N} \log_2 N = \log_2 N$$

Задача 1

В некоторой стране автомобильный номер длиной 7 символов составляется из латинских заглавных букв и десятичных цифр в любом порядке. Каждый символ кодируется одинаковым и минимально возможным количеством битов, а каждый номер – одинаковым и минимально возможным количеством байтов. Определите объем памяти, необходимый для хранения 20 автомобильных номеров.

Домашняя работа

1. В некоторой стране автомобильный номер состоит из 7 символов. В качестве символов используют 18 различных букв и десятичные цифры в любом порядке. Каждый такой номер в компьютерной программе записывается минимально возможным и одинаковым целым количеством байтов, при этом используют посимвольное кодирование и все символы кодируются одинаковым и минимально возможным количеством битов. Определите объем памяти, отводимый этой программой для записи 60 номеров. (Ответ дайте в байтах.)

Домашняя работа

1. В скачках участвуют 20 лошадей. Специальное устройство регистрирует прохождение каждой лошадыю финиша, записывая ее номер с использованием минимально возможного количества бит, одинакового для каждой лошади. Каков информационный объем сообщения, записанного устройством, если до финиша добрались только 15 из 20 участвовавших в скачках лошадей? (Ответ дайте в битах.)
2. В корзине лежат черные и белые шары. Среди них 18 шаров черного цвета. Сообщение о том, что достали шар белого цвета, несет 2 бита информации. Сколько всего шаров в корзине?
3. При регистрации в компьютерной системе каждому пользователю выдаётся пароль, состоящий из 15 символов и содержащий только символы из набора И,Н,Ф, О, Р, М, А, Т, К. Каждый такой пароль в компьютерной программе записывается минимально возможным и одинаковым целым количеством байт (при этом используют посимвольное кодирование и все символы кодируются одинаковым и минимально возможным количеством бит). Определите объём памяти в байтах, отводимый этой программой для записи 25 паролей.

Решение задач

1. При регистрации в компьютерной системе каждому объекту присваивается идентификатор, состоящий из 60 символов и содержащий только десятичные цифры и символы из 250-символьного специального алфавита. В базе данных для хранения каждого идентификатора отведено одинаковое и минимально возможное целое число байт.

При этом используется посимвольное кодирование идентификаторов, все символы кодируются одинаковым и минимально возможным количеством бит. Определите объём памяти (в Кбайт), необходимый для хранения 65 536 идентификаторов. В ответе запишите только целое число – количество Кбайт.

2. В велокроссе участвуют 119 спортсменов. Специальное устройство регистрирует прохождение каждым из участников промежуточного финиша, записывая его номер с использованием минимально возможного количества бит, одинакового для каждого спортсмена. Каков информационный объем в битах сообщения, записанного устройством, после того как промежуточный финиш прошли 70 велосипедистов?

3. При регистрации в компьютерной системе каждому пользователю выдаётся пароль, состоящий из 6 символов и содержащий только символы из 7-буквенного набора Н, О, Р, С, Т, У, Х. В базе данных для хранения сведений о каждом пользователе отведено одинаковое целое число байт, при этом для хранения сведений о 100 пользователях используется 1400 байт. Для каждого пользователя хранятся пароль и дополнительные сведения. Для хранения паролей используют посимвольное кодирование, все символы кодируются одинаковым и минимально возможным количеством бит. Сколько бит отведено для хранения дополнительных сведений о каждом пользователе?

Задача 4 - 8

4. Вероятность появления символа @ в некотором сообщении текста равна 0,125. Сколько битов информации несет сообщение о том, что очередной символ текста - @?
5. В садке у рыбака сидят 2 окуня, 4 плотвы и 10 гуппи. Не смотря в садок, рыбак вытаскивает наугад одну рыбу. Какова вероятность того что это будет плотва?
6. В корзине лежат 8 черных шаров и 24 белых. Какова вероятность вытащить черный шар? Сколько битов информации несет сообщение о том, что достали черный шар?
7. В коробке 64 цветных карандаша. Сообщение о том что достали белый карандаш, несет 4 бита информации. Сколько белых карандашей было в коробке?
8. В корзине лежат 32 клубка шерсти, из них 4 красных. Сколько битов информации несет сообщение о том, что достали клубок красной шерсти?

Задача 9

Объем сообщения, содержащего 4096 символов, равен 1/512 Мб. Какова мощность алфавита, с помощью которого записано это сообщение?

Домашняя работа №3

1. При регистрации в компьютерной системе каждому пользователю выдаётся пароль, состоящий из 15 символов и содержащий только символы Ш, К, О, Л, А (таким образом, используется 5 различных символов). Каждый такой пароль в компьютерной системе записывается минимально возможным и одинаковым целым количеством байт (при этом используют посимвольное кодирование и все символы кодируются одинаковым и минимально возможным количеством бит). Укажите объём памяти в байтах, отводимый этой системой для записи 30 паролей. В ответе запишите только число, слово «байт» писать не нужно.
2. Каждый сотрудник предприятия получает электронный пропуск, на котором записаны личный код, состоящий из двух частей. Первая часть кода содержит 15 символов, каждый из которых может быть одной из 26 заглавных латинских букв. Вторая часть кода содержит 8 символов, каждый из которых может быть одной из десятичных цифр. При этом в базе данных сервера формируется запись, содержащая этот код и дополнительную информацию о пользователе. Для представления кода используют посимвольное кодирование, все символы в пределах одной части кода кодируют одинаковым минимально возможным для этой части количеством битов, а для кода в целом выделяется минимально возможное целое количество байтов. Для хранения данных о 35 пользователях потребовалось 3150 байт. Сколько байтов выделено для хранения дополнительной информации об одном пользователе? В ответе запишите только целое число – количество байтов.
3. В зоопарке 32 обезьяны живут в двух вольерах, А и Б. Одна из обезьян заболела. Сообщение «Заболевшая обезьяна живет в вольере А» содержит 4 бита информации. Сколько обезьян живут в вольере Б?