

Алфавитный подход к определению информации



Подходы к измерению информации

Подходы к измерению информации

Содержательный (вероятностный)

рассматривает
информацию с точки
зрения человека

Через
неопределенность
знаний с учетом
вероятности событий

Алфавитный

рассматривает
информацию для
технического
устройства

Через количество
символов с учетом
информационного
веса символов

Содержательный подход к измерению информации рассматривает информацию с точки зрения человека, как уменьшение неопределенности наших знаний.

Однако любое техническое устройство не воспринимает содержание информации. Поэтому в вычислительной технике используется другой подход к определению количества информации. Он называется **алфавитным подходом**

При алфавитном подходе к определению количества информации отвлекаются от содержания информации и рассматривают информационное сообщение как последовательность знаков определенной знаковой системы



Алфавитный подход к измерению информации

Все множество используемых в языке символов будем традиционно называть алфавитом.

Обычно под алфавитом понимают только буквы, но поскольку в тексте могут встречаться знаки препинания, цифры, скобки, то мы их тоже включим в алфавит. В алфавит также следует включить и пробел, т.е. пропуск между словами.



Алфавит - множество символов, используемых при записи текста

Мощность (размер) алфавита - полное количество символов в алфавите
обозначать эту величину буквой N

Алфавитный подход к измерению информации

Предположим, что каждый символ появляется с одинаковой вероятностью может быть любым символом алфавита. (В действительности это не совсем так, но для упрощения примем такое предположение). В каждой позиции текста может появиться любой из N символов. Тогда, согласно формуле:

$$2^i = N$$

N

МОЩНОСТЬ АЛФАВИТА

число символов в алфавите (его размер)

i

ИНФОРМАЦИОННЫЙ ВЕС СИМВОЛА

количество информации в одном символе

Алфавитный подход к измерению информации

Чтобы найти количество информации во всем тексте, нужно посчитать число символов в нем и умножить на количество информации в одном символе

$$I = K \times i$$

K

ЧИСЛО СИМВОЛОВ В СООБЩЕНИИ

I

КОЛИЧЕСТВО
СООБЩЕНИИ

ИНФОРМАЦИИ

В

Вывод: При алфавитном подходе к измерению информации количество информации зависит не от содержания, а от размера текста и мощности алфавита

!!! Алфавитный подход является *объективным*, т.к. он не зависит от субъекта (человека), воспринимающего текст

Алфавитный подход к измерению информации

Кодировка текста в компьютере

Алфавит	Мощность	Количество бит на символ
Двоичный (0, 1)	2	1 бит
Кодировка ASCII	256	8 бит = 1 байт
Кодировка Windows	256	8 бит = 1 байт
Кодировка Unicode	65536	16 бит = 2 байта
Восьмеричный (0 – 7)	8	3 бита
Шестнадцатеричный (0 – 9, A, B,C, D, E, F)	16	4 бита

Алфавитный подход к измерению информации

Задача

Сколько бит информации содержит слово «компьютер», если оно набрано на компьютере с помощью кодировки Windows.

Решение

В кодировке Windows 256 символов, значит, на 1 символ приходится 8 бит или 1 байт информации.

слове 9 символов, поэтому оно содержит 9×8 бит = 72 бит информации.

Ответ: 72 бит

Алфавитный подход к измерению информации

Задача

Сколько информации содержит предложение «Скажите, пожалуйста, который час?». Слово набрано на компьютере при помощи кодировки Unicode.

Решение

В кодировке Unicode 1 символ содержит 2 байта информации. В предложении 27 букв, 3 знака препинания и 3 пробела – всего 33 символа.

Т.к. на 1 символ в кодировке Unicode приходится 2 байта, то предложение содержит $2 \text{ байта} * 33 = 66 \text{ байт}$ информации.

Ответ: 66 байт

Алфавитный подход к измерению информации

ЗАДАЧА

Книга, подготовленная с помощью компьютера, содержит 150 страниц. На каждой странице – 40 строк, в каждой строке – 60 символов (включая пробелы между словами). Каков объем информации в книге?

РЕШЕНИЕ

Мощность компьютерного алфавита равна 256, поэтому один символ несет 1 байт информации. Значит, страница книги содержит $40 \times 60 = 2400$ байт информации.

$$[\text{кол-во символов в строке}] \times [\text{кол-во строк}] = [\text{информационный объем страницы}]$$

Объем всей информации в книге (в разных единицах):

$$[\text{информационный объем страницы}] \times [\text{кол-во страниц}] = [\text{информационный объем книги}]$$

$$2400 \times 150 = \mathbf{360\ 000\ \text{байт}} / 1024 = \mathbf{351,5625\ \text{Кбайт}} / 1024 = \mathbf{0,34332275\ \text{Мбайт}}$$

Алфавитный подход к измерению информации

ЗАДАЧА два текста содержат **одинаковое количество символов**. Первый текст составлен в алфавите мощностью **32 символа**, второй мощностью **64 символа**. Во сколько раз отличается количество информации в этих текстах?

РЕШЕНИЕ

Информационный объем текста равен произведению числа символов на информационный вес одного символа:

$$I = K \times i$$

Т.к. оба текста имеют одинаковое число символов (K), то различие объемов определяется только разницей в информативности символов алфавита. Найдем i_1 для первого алфавита и i_2 для второго алфавита:

$$2^{i_1} = 32; \text{ отсюда } i_1 = 5 \text{ бит}$$

$$2^{i_2} = 64; \text{ отсюда } i_2 = 6 \text{ бит}$$

Следовательно, информационные объемы первого и второго текстов:

$$I_1 = K \times 5 \text{ битов}; \quad I_2 = K \times 6 \text{ битов};$$

Отсюда следует, что количество информации во втором тексте, больше чем в первом в $6/5$ раз или в $1,2$ раза

