ИНФОРМАТИКА

2. Количественная оценка информации

Единицы измерения информации

За единицу измерения информации принят **1 бит**.

Алфавитный подход

1 бит – это количество информации, которое можно передать в сообщении, состоящем из одного двоичного знака (0 или 1).

Содержательный подход

1 бит – это количество информации, уменьшающее неопределённость знаний в два раза.

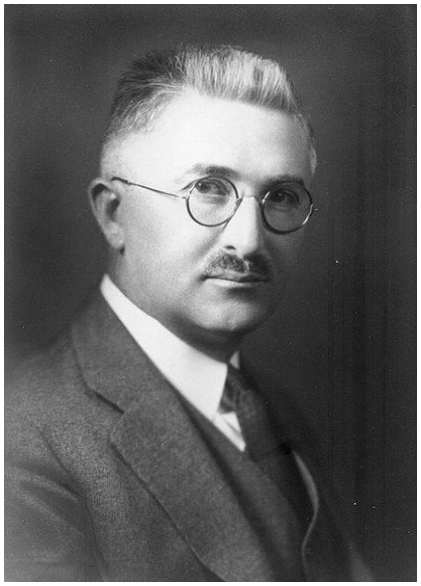
1 ***байт*** = 8 бит
1 Кб (***килобайт***) = 2^{10} байт = 1024 байт
1 Мб (***мегабайт***) = 2^{20} байт = 1024 Кб
1 Гб (***гигабайт***) = 2^{30} байт = 1024 Мб
1 Тб (***терабайт***) = 2^{40} байт = 1024 Гб
1 Пб (***петабайт***) = 2^{50} байт = 1024 Тб
1 Эб (***эксабайт***) = 2^{60} байт = 1024 Пб
1 Зб (***зеттабайт***) = 2^{70} байт = 1024 Эб
1 Йб (***йоттабайт***) = 2^{80} байт = 1024 Зб



1) Сколько в 2^{18} килобайтах содержится гигабайт?
2) Сколько в 2^{20} килобитах содержится мегабайт ?

Ответ:
1) 2^{18} Кб = 0,25 Гб
2) 2^{20} килобит = 125 Мб
(1 килобит = 1000 бит)

Формула Хартли



Ральф Винтон Лайон Хартли (США)

Первую попытку определения количества информации в ***1928*** году сделал Ральф Хартли.

Формула Хартли – логарифмическая мера информации, определяющая количество информации, содержащееся в сообщении.

$$I = K \cdot \log_2 N$$



I – количество информации в сообщении в би

K – длина сообщения (количество символов в сообщении),

N – мощность алфавита (количество символов в используемом алфавите).



Может ли для передачи информации быть использован алфавит, содержащий только 1 символ?



Формула Хартли

Обозначим через ***i*** количество информации в одном символе алфавита мощности ***N***:

$$i = \log_2 N$$

Тогда мощность алфавита можно оценить по формуле:

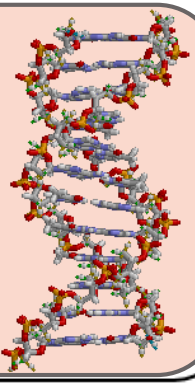
$$N = 2^i$$

Количество возможных вариантов сообщений длиной ***K***:

$$m = N^K$$

Пример. Цепь ДНК состоит из нуклеотидов четырех типов: *Аденин*, *Гуанин*, *Тимин*, *Цитозин*.

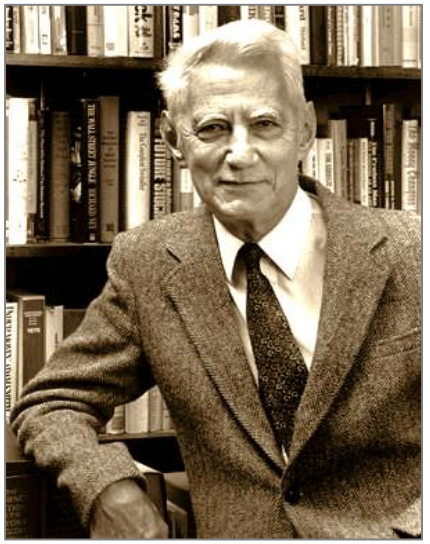
1) Какова мощность алфавита ДНК и сколько информации несет в себе каждый нуклеотид?
2) Сколько разнообразных геномов у простейшего вируса, если длина его ДНК около 5000 нуклеотидов?
3) В человеческой ДНК примерно 10^9 нуклеотидов. Каково разнообразие?



Формула Шеннона



В формуле Хартли не учтено, что разные символы могут иметь разную вероятность появления.



Клод Элвуд Шеннон (США)

В ***1948*** году Клод Шеннон предложил вероятностный подход, в котором напрямую соотнёс энтропию и информацию:

$$I = -\sum_{i=1}^{N} p_i \log_2 p_i$$

I – количество информации в одном событии в битах,

N – количество возможных событий (символов),

p_i – вероятность i-го события.

бит

дит

нит

Замечания:

1. Следует учитывать, что $p_1 + p_2 + \ldots + p_N = 1$.
2. Если $p_1 = p_2 = \ldots = p_N$, то получится формула Хартли.

Формула Шеннона

При исследовании объекта стоит учитывать следующие факты:

1. Чем больше ***N***, тем более разнообразен объект, а значит тем больше ***I***.

2. Если суть объекта выражается только одной возможной ситуацией, которая заведомо произойдет (т.е. $p_1=1, p_2=0,\ldots, p_N=0$), то $I = 0$ (никакой новой информации об объекте мы не получим).

3. Максимальное значение информации I_{max} при заданном значении N может быть получено при равновероятностном проявлении $p_1=p_2=\ldots=p_N = 1/N$ (чем менее известен объект, тем больше о нём информации мы можем получить, то есть он обладает большей информацией).



В ящике лежат **20** шаров: **10** — черных, **4** — белых и **6** — красных. Какое количество информации несёт сообщение о цвете вынутого шара?

$$I = -\sum_{i=1}^{3} p_i \log_2 p_i = -\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{5}\log_2\frac{1}{5} + \frac{3}{10}\log_2\frac{3}{10}\right) = \textbf{1,49 бит}$$

СПАСИБО

ЗА

ВНИМАНИЕ!