

- На *синтаксическом уровне* для оценки количества информации используют вероятностные методы, которые принимают во внимание только вероятностные свойства информации и не учитывают другие (смысловое содержание, полезность, актуальность и т. д.). Разработанные в середине XX в. математические и, в частности, вероятностные методы позволили сформировать подход к оценке количества информации как к мере уменьшения неопределенности знаний. Такой подход, называемый также вероятностным, постулирует принцип: если некоторое сообщение приводит к уменьшению неопределенности наших знаний, то можно утверждать, что такое сообщение содержит информацию. При этом сообщения содержат информацию о каких-либо событиях, которые могут реализоваться с различными вероятностями. Формулу для определения количества информации для событий с различными вероятностями и получаемых от дискретного источника информации предложил американский ученый К. Шеннон в 1948 г. Согласно этой формуле количество информации может быть определено следующим образом:

Формула Шеннона

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

де I – количество информации; N – количество возможных событий (сообщений); p_i – вероятность отдельных событий (сообщений); Σ – математический знак суммы чисел (греческая буква сигма).

Знак суммы

$$\sum_{i=1}^n a_i = a_1 + a_2 + \dots + a_n$$

● Пример

- Требуется определить, какое количество информации можно получить после реализации одного из шести событий. Вероятность первого события составляет 0,15; второго – 0,25; третьего – 0,2; четвертого – 0,12; пятого – 0,12; шестого – 0,1, т. е. $P_1 = 0,15$; $P_2 = 0,25$; $P_3 = 0,2$; $P_4 = 0,18$; $P_5 = 0,12$; $P_6 = 0,1$.



● *Решение.*

$$I = -(P_1 \cdot \log_2 P_1 + P_2 \cdot \log_2 P_2 + P_3 \cdot \log_2 P_3 + P_4 \cdot \log_2 P_4 + \\ + P_5 \cdot \log_2 P_5 + P_6 \cdot \log_2 P_6) = -(0,15 \cdot \log_2 0,15 + 0,25 \cdot \log_2 0,25 + \\ + 0,2 \cdot \log_2 0,2 + 0,18 \cdot \log_2 0,18 + 0,12 \cdot \log_2 0,12 + 0,1 \cdot \log_2 0,1) \text{ бит.}$$

$$I = 2,52 \text{ бит.}$$

Для количественного определения (оценки) любой физической величины необходимо определить единицу измерения, которая в теории измерений носит название меры. Как уже отмечалось, информацию перед обработкой, передачей и хранением необходимо подвергнуть кодированию. Кодирование производится с помощью специальных алфавитов (знаковых систем). В информатике, изучающей процессы получения, обработки, передачи и хранения информации с помощью вычислительных (компьютерных) систем, в основном используется двоичное кодирование, при котором используется знаковая система, состоящая из двух символов 0 и 1.

Исходя из вероятностного подхода к определению количества информации эти два символа двоичной знаковой системы можно рассматривать как два различных возможных события, поэтому за единицу количества информации принято такое количество информации, которое содержит сообщение, уменьшающее неопределенность знания в два раза (до получения событий их вероятность равна 0,5, после получения – 1, неопределенность уменьшается соответственно: $1/0,5 = 2$, т. е. в 2 раза). Такая единица измерения информации называется битом (от англ. слова *binary digit* – двоичная цифра). Таким образом, в качестве меры для оценки количества информации на синтаксическом уровне, при условии двоичного кодирования, принят один бит.

- Следующей по величине единицей измерения количества информации является байт, представляющий собой последовательность, составленную из восьми бит, т. е.
- $1 \text{ байт} = 2^3 \text{ бит} = 8 \text{ бит}$.
- В информатике также широко используются кратные байту единицы измерения количества информации, однако в отличие от метрической системы мер, где в качестве множителей кратных единиц применяют коэффициент 10^n , где $n = 3, 6, 9$ и т. д., в кратных единицах измерения количества информации используется коэффициент 2^n . Выбор этот объясняется тем, что компьютер в основном оперирует числами не в десятичной, а в двоичной системе счисления.

Кратные байту единицы измерения количества информации вводятся следующим образом:

1 Килобайт (Кбайт) = 2^{10} байт = 1024 байт,

1 Мегабайт (Мбайт) = 2^{10} Кбайт = 1024 Кбайт,

1 Гигабайт (Гбайт) = 2^{10} Мбайт = 1024 Мбайт,

1 Терабайт (Тбайт) = 2^{10} Гбайт = 1024 Гбайт,

1 Петабайт (Пбайт) = 2^{10} Тбайт = 1024 Тбайт,

1 Экзабайт (Эбайт) = 2^{10} Пбайт = 1024 Пбайт.

● Единицы измерения количества информации, в названии которых есть приставки «кило», «мега» и т. д., с точки зрения теории измерений не являются корректными, поскольку эти приставки используются в метрической системе мер, в которой в качестве множителей кратных единиц используется коэффициент 10^n , где $n = 3, 6, 9$ и т. д. Для устранения этой некорректности международная организацией *International Electrotechnical Commission*, занимающаяся созданием стандартов для отрасли электронных технологий, утвердила ряд новых приставок для единиц измерения количества информации: киби (*kibi*), меби (*mebi*), гиби (*gibi*), теби (*tebi*), пети (*peti*), эксби (*exbi*). Однако пока используются старые обозначения единиц измерения количества информации, и требуется время, чтобы новые названия начали широко применяться.

- Количество информации, которое несет один знак алфавита, тем больше, чем больше знаков входит в этот алфавит. Количество знаков, входящих в алфавит, называется мощностью алфавита. Количество информации (информационный объем), содержащееся в сообщении, закодированном с помощью знаковой системы и содержащем определенное количество знаков (символов), определяется с помощью формулы:

- $V=I \cdot K,$

- где V – информационный объем сообщения; $I = \log_2 N$, информационный объем одного символа (знака); K – количество символов (знаков) в сообщении; N – мощность алфавита (количество знаков в алфавите).

$$K_h = \pm(a_{n-1}h^{n-1} + a_{n-2}h^{n-2} + \dots + a_0h^0 + a_{-1}h^{-1} + a_{-2}h^{-2} + \dots + a_{-m}h^{-m}),$$

или в рекуррентной форме:

$$K_h = \pm \sum_{i=-m}^{n-1} a_i h^i,$$

где K – представляемое число; h – основание системы счисления; a – разрядный коэффициент, $a = 0, 1, 2, 3, \dots, h-1$, т. е. цифры, принадлежащие алфавиту данной системы счисления; i – номер разряда, позиция; n – число целых разрядов числа; m – число дробных разрядов числа.

- В десятичной системе счисления формула может быть записана следующим образом:

$$K_{10} = \pm \sum_{i=-m}^{n-1} a_i 10^i,$$

где $a = 0, 1, 2, 3, \dots, 9$.

- В двоичной:

$$K_2 = \pm \sum_{i=-m}^{n-1} a_i 2^i,$$

где $a = 0, 1$.

- В восьмеричной:

$$K_8 = \pm \sum_{i=-m}^{n-1} a_i 8^i,$$

где $a = 0, 1, 2, 3, \dots, 7$.

- Например, число $1473_8 = 1 \cdot 8^3 + 4 \cdot 8^2 + 7 \cdot 8^1 + 3 \cdot 8^0$.

$$K_{16} = \pm \sum_{i=-m}^{n-1} a_i 16^i,$$

где $a = 0, 1, 2, 3..., 9, A, B, C, D, E, F$.

В шестнадцатеричной системе счисления используется шестнадцать цифр, из которых десять цифр арабские (0, 1, 2, 3, 4, 5, 6, 7, 8, 9), а остальные цифры (10, 11, 12, 13, 14, 15) обозначаются буквами латинского алфавита (A = 10, B = 11, C = 12, D = 13, E = 14, F = 15).

Например, число $33B_{16} = 3 \cdot 16^2 + 3 \cdot 16^1 + B \cdot 16^0$.

- Преобразование чисел, представленных в двоичной, восьмеричной и шестнадцатеричной системах счисления, в десятичную осуществляется достаточно просто. Для этого необходимо записать число в развернутой форме.

Например:

$$101,11_2 = 1 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0 + 1 \cdot 2^{-1} + 1 \cdot 2^{-2} = \\ = 4 + 1 + 0,5 + 0,25 = 5,75_{10};$$

$$97,5_8 = 9 \cdot 8^1 + 7 \cdot 8^0 + 5 \cdot 8^{-1} = 72 + 7 + 0,625 = 79,625_{10};$$

$$19F_{16} = 1 \cdot 16^2 + 9 \cdot 16^1 + F \cdot 16^0 = 256 + 144 + 15 = 416_{10}. (F = 15)$$

Арифметические операции в двоичной системе счисления основаны на таблицах сложения, вычитания и умножения одноразрядных двоичных чисел.

При сложении двух единиц происходит переполнение разряда и производится перенос единицы в старший разряд, при вычитании 0–1 производится заем из старшего разряда, в таблице «Вычитание» этот заем обозначен 1 с чертой над цифрой.

Сложение			Вычитание			Умножение		
$0 + 0$	$=$	0	$0 - 0$	$=$	0	$0 \cdot 0$	$=$	0
$0 + 1$	$=$	1	$0 - 1$	$=$	1	$0 \cdot 1$	$=$	0
$1 + 0$	$=$	1	$1 - 0$	$=$	1	$1 \cdot 0$	$=$	0
$1 + 1$	$=$	10	$1 - 1$	$=$	0	$1 \cdot 1$	$=$	1

Ниже приведены примеры выполнения арифметических операций над числами, представленными в различных системах счисления:

$$\begin{array}{r}
 1010_2 \\
 + \frac{101_2}{1111_2} \\
 \hline
 \end{array}
 \begin{array}{r}
 1010_2 \\
 - \frac{101_2}{0101_2} \\
 \hline
 \end{array}
 \times
 \begin{array}{r}
 1010_2 \\
 \frac{101_2}{1010} \\
 \hline
 \end{array}
 +
 \begin{array}{r}
 1010 \\
 \hline
 110010_2
 \end{array}
 \begin{array}{r}
 1010_2 \overline{) 101_2} \\
 \underline{101} \\
 0
 \end{array}
 +
 \begin{array}{r}
 131_8 \\
 \frac{65_8}{216_8} \\
 \hline
 \end{array}
 -
 \begin{array}{r}
 D2_{16} \\
 \frac{63_{16}}{6F_{16}} \\
 \hline
 \end{array}$$