Различные подходы в определении информации

Р. Клаузиусом в 1865 была введена в термодинамику физическая величина, названная ***энтропией*** (от греч. entropía — поворот, превращение). Для всякой материальной системы энтропия — это ***мера близости системы к состоянию равновесия***. Если система находится в состоянии равновесия, в ней невозможно движение, а энтропия ее максимальна.

$S = k \ln W(E, N)$,

где k –постоянная Больцмана, $W(E, N)$ - число уровней в узком интервале энергии вблизи значения энергии E системы из N частиц. Впервые связь энтропии. с вероятностью состояния системы была установлена Л. Больцманом в 1872: возрастание энтропии системы обусловлено её переходом из менее вероятного состояния в более вероятное. Иными словами, эволюция замкнутой системы осуществляется в направлении наиболее вероятного распределения энергии по отдельным подсистемам.

Для определения явного вида функции $f(n)$ рассмотрим два независимых опыта $\boldsymbol{\alpha}$ и $\boldsymbol{\beta}$, в сложном опыте они никак не могут повлиять друг на друга. Следовательно, мера суммарной неопределенности должна быть равна сумме мер неопределенности каждого из опытов, т.е. мера неопределенности аддитивна:

$$f(n_\alpha n_\beta)=f(n_\alpha)+f(n_\beta) \qquad (1.1)$$

Можно доказать, что единственная из всех существующих классов функция ***log(n)*** такому набору свойств удовлетворяет. Таким образом: ***за меру неопределенности опыта с* n *равновероятными исходами можно принять число* log(n)**.

ИСТИНА **(True)** и ***ЛОЖЬ*** **(False)**

Единица измерения неопределенности при двух возможных равновероятных исходах опыта называется ***бит*** (происходит от английского ***binary digit***, дословно "двоичный разряд" или "двоичная единица"). Таким образом, нами установлен явный вид функции, описывающей меру неопределенности опыта, имеющего ***n*** равновероятных исходов:

$$f(n) = \log_2 n \quad (1.2).$$

Эта величина получила название ***энтропии***. В дальнейшем будем обозначать ее ***H***.

$$H_1=1/n\cdot \log_2 n =- 1/n\cdot \log_2 1/n=-p\cdot \log_2 p$$

где $p=1/n$ – вероятность любого из отдельных исходов.

Обобщая эту формулу на ситуацию, когда исходы опытов неравновероятны, получим:

$$H = -\sum_{i=1}^{n} p(X_i)\, log\, (p(X_i) \quad (1.3)$$

Именно такая удобная мера неопределенности была введена ***К. Шенноном.***

Здесь X –случайная величина, принимающая N некоторых значений, $P(X_i)$ – вероятность, того, что случайная величина принимает значение X_i.

Измерение количества информации

Информация определяется разностью между безусловной и условной энтропиями. Это уменьшение неопределенности "знания чего-то за счет того, что известно что-то".

Пусть неопределенность до опыта составляла H (***априорная неопределенность***), а после опыта – H_1 (***апостериорная неопределенность***). Очевидно, устраненная в ходе опыта неопределенность равна:

$$I=H-H_1.$$

Эта разность носит название ***количества информации,*** или ***количество устраненной неопределенности***.

В частном случае, когда неопределенность в результате опыта устраняется полностью, $I=H$.

$$I = -\sum_{i=1}^{n} p(X_i) \ \ log_2(p(X_i) \quad (1.4)$$

Здесь следует иметь в виду различный смысл количества информации и энтропии. Энтропия (неопределенность) существует ***до опыта***, тогда как информация появляется ***после опыта***.

Легко получить следствие формулы (1.4) для случая **n** равновероятных исходов. В этом случае все

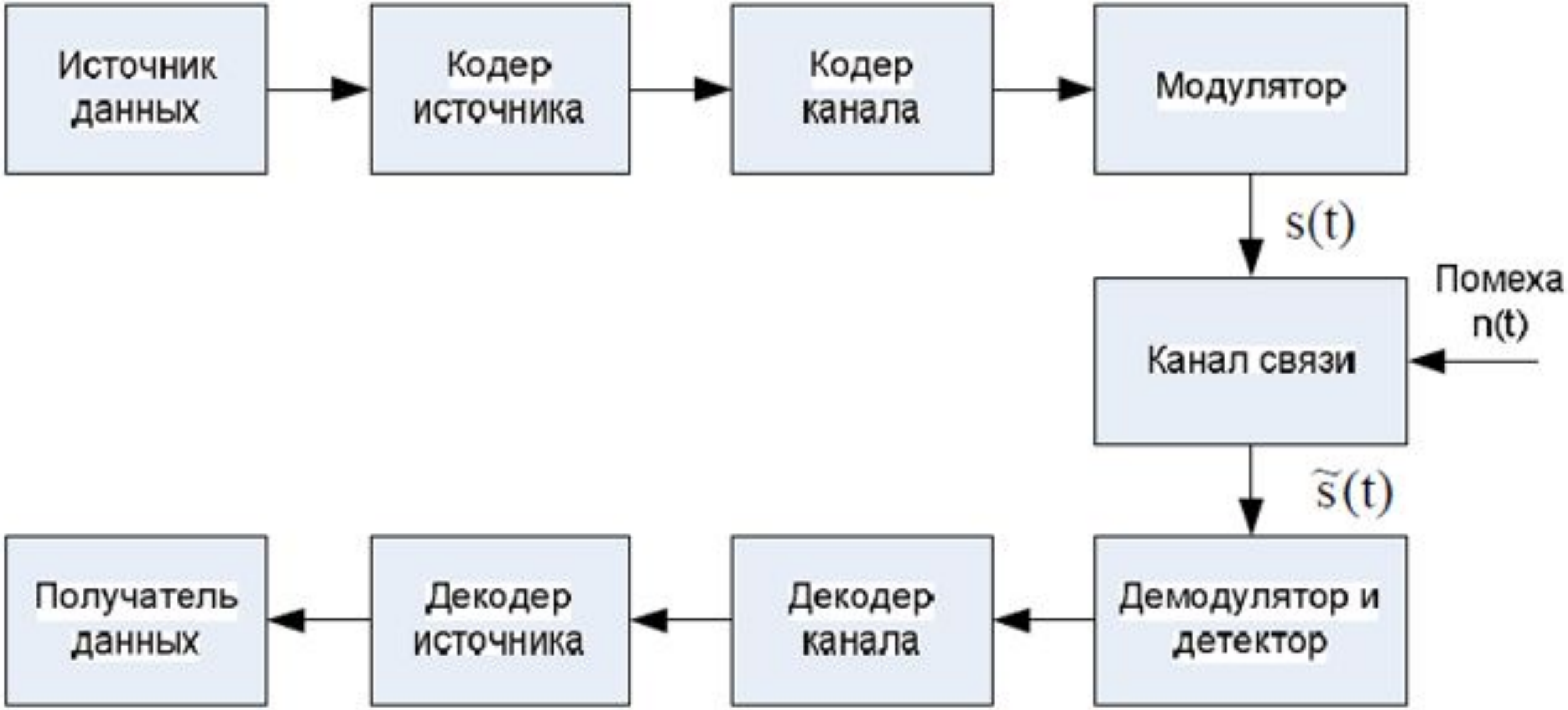
$p(A_i) = \dfrac{1}{n}$ и, следовательно, $I = \sum_{i=1}^{n} \dfrac{1}{n} \cdot \log_2 n = \log_2 n,$

$I = log_2 n$ (1.5)

Эта формула была выведена в 1928 г. американским инженером ***Р.Хартли*** и названа его именем. Ее смысл в том, что, если некоторое множество содержит **n** элементов и элемент **x** принадлежит данному множеству, то для его однозначного определения среди прочих требуется количество информации, равное $\mathbf{log_2 n}$.

Единицы измерения информации

Единицы измерения количества информации	Обозначение	Количество информации	
		число Дит	*число Бит*
байт	B	10^0	2^8
килобайт	kB	10^3	2^{10}
мегабайт	MB	10^6	2^{20}
гигабайт	GB	10^9	2^{30}
терабайт	TB	10^{12}	2^{40}
петабайт	PB	10^{15}	2^{50}
эксабайт	EB	10^{18}	2^{60}
зеттабайт	ZB	10^{21}	2^{70}
йоттабайт	YB	10^{24}	2^{80}
нонабайт	NB	10^{27}	2^{90}


Источник данных
Кодер источника
Кодер канала
Модулятор
s(t)
Помеха n(t)
Канал связи
s̃(t)
Демодулятор и детектор
Декодер канала
Декодер источника
Получатель данных