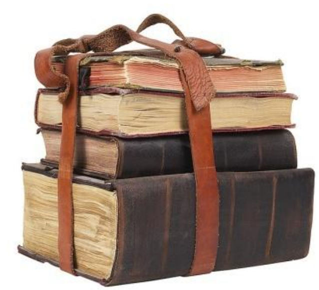
ПОНЯТИЕ ИНФОРМАЦИИ И ПОДХОДЫ К ЕЕ КОЛИЧЕСТВЕННОЙ ОЦЕНКЕ

Определение информации

Термин "информация" происходит от латинского слова "Informatio" – разъяснение, изложение, осведомленность.

Информация → Данные → Знания

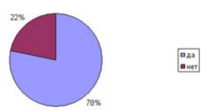


Свойства информации

- Релевантность

- Полнота

- Своевременность (актуальность)



- Достоверность

- Доступность

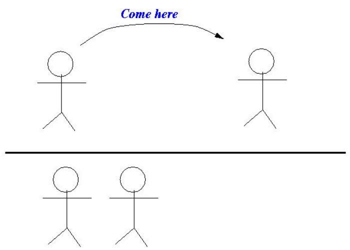
- Защищенность

- Эргономичность

- Адекватность

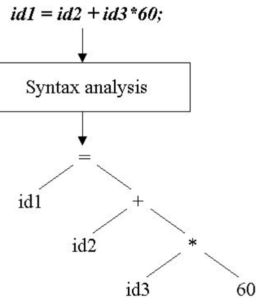
Аспекты информации

- прагматический



- семантический

- синтаксический



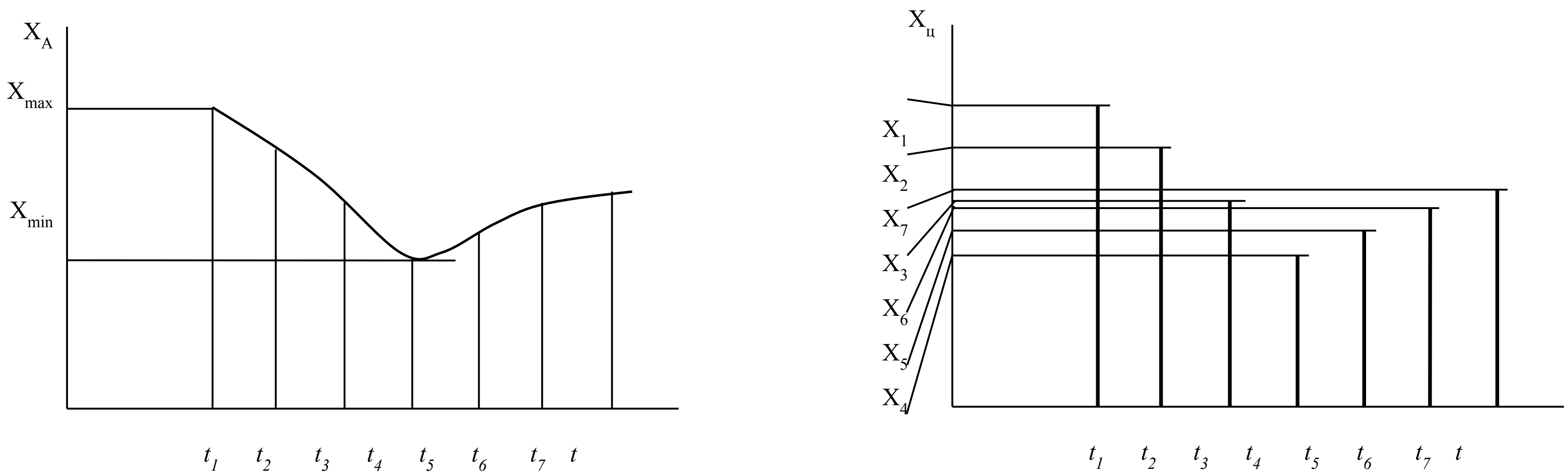
Тезаурус

Семантические связи между словами или другими смысловыми элементами языка отражают словарь – ***тезаурус***. Он состоит из двух частей: **списка слов** и **устойчивых словосочетаний**, которые сгруппированы по смыслу, и **некоторого ключа**, т е. алфавитного словаря, позволяющего расположить слова и словосочетания в определенном порядке.

Тезаурус имеет особое значение в системах хранения информации, в которые могут вводиться семантические отношения, в основном подчинения, что позволяет на логическом уровне осуществлять организацию информации в виде отдельных записей, массивов и их комплексов.

Структурная мера информации

- Элементарная единицы сообщений – символ
- Символы, собранные в группы – слова

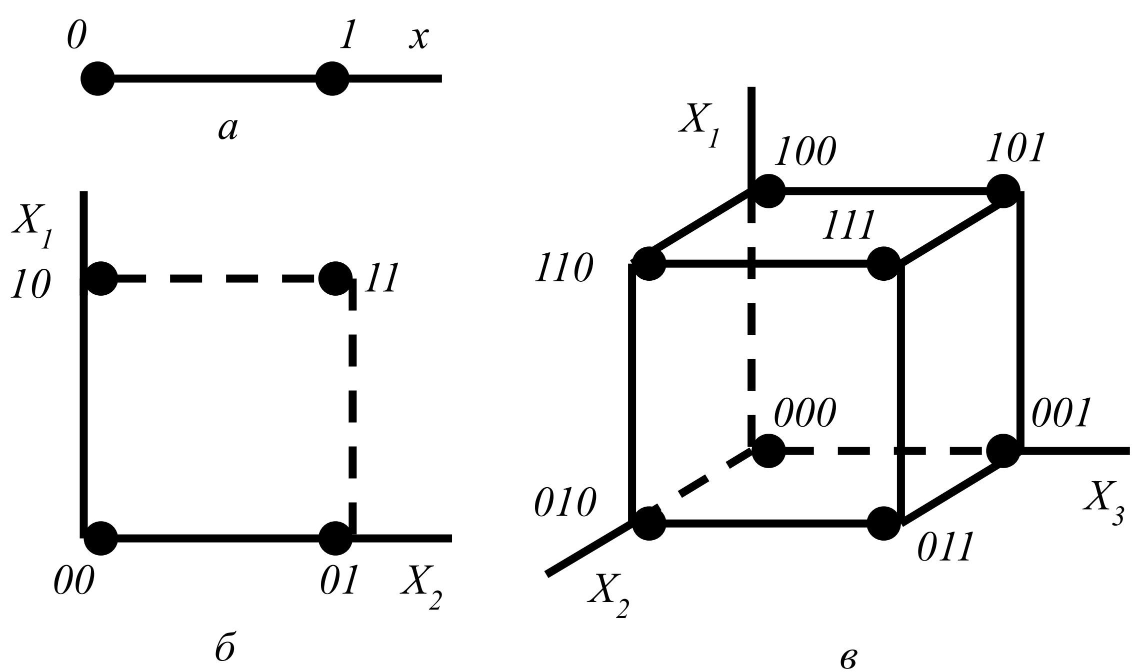


Функция, представленная в непрерывном и дискретном виде

Учитывается только дискретное строение сообщения, количество содержащихся в нём информационных элементов, связей между ними.

Структурная мера информации

<u>Геометрическая мера</u> предполагает измерение параметра геометрической модели информационного сообщения (длины, площади, объема и т.п.) в дискретных единицах. Например, геометрической моделью информации может быть линия единичной длины (рисунок ***а*** – одноразрядное слово, принимающее значение 0 или 1), квадрат (рисунок ***б*** — двухразрядное слово) или куб (рисунок ***в*** — трехразрядное слово).



Структурная мера информации

Аддитивная мера (***мера Хартли***), в соответствии с которой количество информации измеряется в двоичных единицах — **битах** (наиболее распространена). Вводятся понятия **глубины** q и **длины** n числа.

Глубина $\boldsymbol{q}$ ***числа*** – количество символов (элементов), принятых для представления информации. В каждый момент времени реализуется только один какой-либо символ.

Длина $\boldsymbol{n}$ ***числа*** – количество позиций, необходимых и достаточных для представления чисел заданной величины.

Структурная мера информации

При заданных глубине и длине числа количество чисел, которое можно представить, $N = q^n$. Величина N неудобна для оценки информационной емкости. Введем логарифмическую меру, позволяющую, вычислять количество информации — *бит*:

$$I(g)=\log_2 N = n \log_2 q$$

Следовательно, 1 бит информации соответствует одному элементарному событию, которое может произойти или не произойти.

Структурная мера информации

Количество информации при этом эквивалентно количеству двоичных символов 0 или 1. При наличии нескольких источников информации общее количество информации

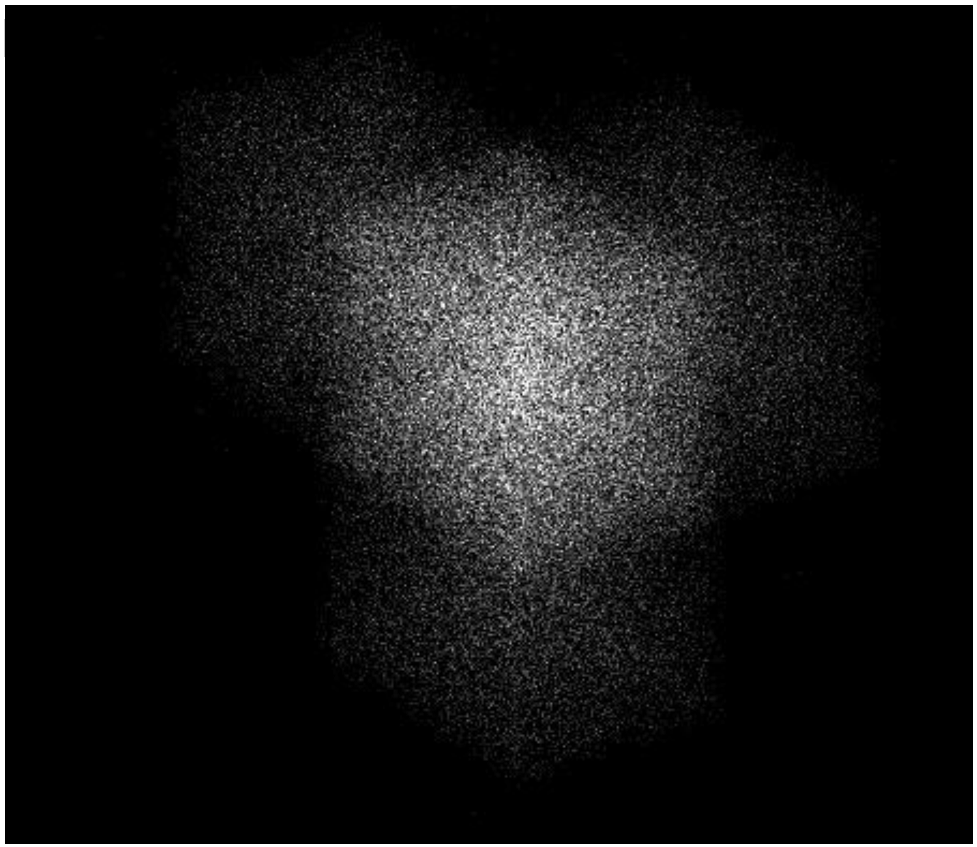
$$I(q_1, q_2, \ldots, q_k) = I(q_1) + I(q_2) + \ldots + I(q_k),$$

где $I(q_k)$ – количество информации от источника k.

Логарифмическая мера информации позволяет измерять количество информации и используется на практике.

Статистическая мера информации

Энтропия – количественная мера неопределенности и, следовательно, информатив



Иными словами, энтропия – это мера хаоса…

Статистическая мера информации

Энтропия – количественная мера неопределенности и, следовательно, информативности.

Пусть имеется N возможных исходов опыта, из них k разных типов, а i-й исход повторяется n_i раз и вносит информацию, количество которой оценивается как I_i

Тогда средняя информация, доставляемая одним опытом,

$$I_{cp} = (n_1 I_1 + n_2 I_2 + \ldots + n_k I_k)/N$$

Статистическая мера информации

Количество информации в каждом исходе связано с его вероятностью p_i и выражается в двоичных единицах (битах) как $I_i = \log_2 (1/p_i) = = -\log_2 p_i$. Тогда

$$I_{cp} = [n_1(-\log_2 p_1) + \ldots + n_k(-\log_2 p_k)]/N$$

Выражение можно записать также в виде

$$I_{cp} = (-\log_2 p_1) + (-\log_2 p_2) + \ldots + (-\log_2 p_1)$$

Отношения n/N представляют собой частоты повторения исходов, а следовательно, могут быть заменены их вероятностями $n_i/N = p_i$, поэтому их средняя информация в битах

$$I_{cp} = p_1(-\log_2 p_1) + \ldots + p_k(-\log_2 p_k)],$$

или

$$I_{cp} = -\log_2 p_i = H$$

Статистическая мера информации

Полученную величину называют **энтропией** и обозначают обычно буквой H. Энтропия обладает следующими свойствами:

1. Энтропия всегда неотрицательна
2. Энтропия равна нулю в том крайнем случае, когда одно из p_i равно единице, а все остальные — нулю
3. Энтропия имеет наибольшее значение, когда все вероятности равны между собой: $p_1 = p_2 = ... = p_k$ - $1/k$. При этом $H = -\log_2 (1/k) = \log_2 k$
4. Энтропия объекта AB, состояния которого образуются совместной реализацией состояний A и B, равна сумме энтропии исходных объектов A и B, т. е. H(AB) = H(A) + H(B).

Статистическая мера информации

Максимальное значение энтропии достигается при p=0.5, когда **два состояния равновероятны**. При вероятностях p = 0 или p = 1, что соответствует полной невозможности или полной достоверности события, энтропия равна нулю.

Количество информации только тогда равно энтропии, когда неопределенность ситуации снимается полностью. В общем случае нужно считать, что *количество информации есть уменьшение энтропии* вследствие опыта или какого-либо другого акта познания. **Если неопределенность снимается полностью, то информация равна энтропии**: $I = H$.

Статистическая мера информации

В случае неполного разрешения имеет место частичная информация, являющаяся разностью между начальной и конечной энтропией: $I = H_1 – H_2$.

Наибольшее количество информации получается тогда, когда полностью снимается неопределенность, причем эта неопределенность была наибольшей – вероятности всех событий были одинаковы. Это соответствует максимально возможному количеству информации I^1, оцениваемому мерой Хартли:

$$I^1 = \log_2 N = \log_2(1/p) = -\log_2 p\ ,$$

где N — число событий; p — вероятность их реализации в условиях равной вероятности событий.

Статистическая мера информации

Абсолютная избыточность информации $D_{\text{абс}}$ представляет собой разность между максимально возможным количеством информации и энтропией:

$$D_{\text{абс}} = I^1 - H\text{, или } D_{\text{абс}} = H_{\max} - H\text{.}$$

Пользуются также понятием относительной избыточности

$$D = (H_{\max} - H)/H_{\max}$$

Семантическая мера информации

- Содержательность события
- Логическое количество информации
- Мера целесообразности информации

Семантическая мера информации

Содержательность события выражается через функцию меры $m(i)$ – содержательности его отрицания. Оценка содержательности основана на математической логике, в которой логические функции истинности $m(i)$ и ложности $m(\square i)$ имеют формальное сходство с функциями вероятностей события $p(i)$ и антисобытия $q(i)$ в теории вероятностей.

Как и вероятность, содержательность события изменяется в пределах $0 \le m(i) \le 1$.

Логическое количество информации I_{nf}, сходное со статистическим количеством информации, вычисляется по выражению:

$$I_{nf} = \log_2 [1/m(i)] = -\log_2 m(\square i)$$

Семантическая мера информации

Мера целесообразности информации определяется как изменение вероятности достижения цели при получении дополнительной информации.

Полученная информация может быть пустой, т. е. не изменять вероятности достижения цели, и в этом случае ее мера равна нулю. В других случаях полученная информация может изменять положение дела в худшую сторону, т.е. уменьшить вероятность достижения цели, и тогда она будет дезинформацией, измеряющейся отрицательным значением количества информации. Наконец, в благоприятном случае получается добротная информация, которая увеличивает вероятность достижения цели и измеряется положительной величиной количества информации.

Мера целесообразности в общем виде может быть аналитически выражена в виде соотношения

$$I_{цел} = \log_2 p_1 - \log_2 p_0 = \log_2$$

где p_0 и p_1 – начальная (до получения информации) и конечная (после получения информации) вероятности достижения цели.

Преобразование информации

Дискретные сообщения состоят из конечного множества элементов, создаваемых источником последовательно во времени.

Непрерывные сообщения задаются какой-либо физической величиной, изменяющейся во времени. Получение конечного множества сообщений за конечный промежуток времени достигается путем ***дискретизации*** (по времени) и ***квантования*** (по уровню).

Преобразование информации

Разновидности сигналов, которые описываются функцией $x(t)$.

1. Непрерывная функция непрерывного аргумента. Значения, которые могут принимать функция $x(t)$ и аргумент t, заполняют промежутки ($x_{\min}$, $x_{\max}$) и ($-T$, T) соответственно.
2. Непрерывная функция дискретного аргумента. Значения функции $x(t)$ определяются лишь на дискретном множестве значений аргумента t_i, $i=0\pm1\pm2$, … Величина $x(t_i)$ может принимать любое значение в интервале ($x_{\min}$, $x_{\max}$).

Преобразование информации

Разновидности сигналов, которые описываются функцией $x(t)$.

3. Дискретная функция непрерывного аргумента. Значения, которые может принимать функция $x(t)$, образуют дискретный ряд чисел x_1, x_2,..., x_k. Значение аргумента t может быть любым в интервале (-T, T).
4. Дискретная функция дискретного аргумента. Значения, которые могут принимать функция $x(t)$ и аргумент t, образуют дискретные ряды чисел x_1, x_2, ..., x_k и t_1, t_2, ..., t_k, заполняющие интервалы (x_{min}, x_{max}) и $(-T,T)$ соответственно.

Преобразование информации

Операцию, переводящую информацию непрерывного вида в информацию дискретного вида, называют ***квантованием по времени***, или ***дискретизацией***. Следовательно, дискретизация состоит в преобразовании сигнала $x(t)$ непрерывного аргумента t в сигнал $x(t_i)$ дискретного аргумента t_i.

Квантование по уровню состоит в преобразовании непрерывного множества значений сигнала $x(t_i)$ в дискретное множество значений x_k, k = =0,1,..., $(m$ - 1); $x_k \in (x_{\min}, x_{\max})$ (третий вид сигнала).

Преобразование информации

Совместное применение операций дискретизации и квантования по уровню позволяет **преобразовать непрерывный сигнал** $\mathbf{x}$**(t) в дискретный** по координатам $\mathbf{x}$ и t(четвертая разновидность).

В результате дискретизации исходная функция $\mathbf{x}(t)$ заменяется совокупностью отдельных значений $\mathbf{x}(t_i)$. По значениям функции $\mathbf{x}(t_i)$ можно восстановить исходную функцию $\mathbf{x}(t)$ с некоторой погрешностью. Функцию, полученную в результате восстановления (**интерполяции**) по значениям $\mathbf{x}(t_i)$, будем называть воспроизводящей и обозначать $\mathbf{V}(t)$.

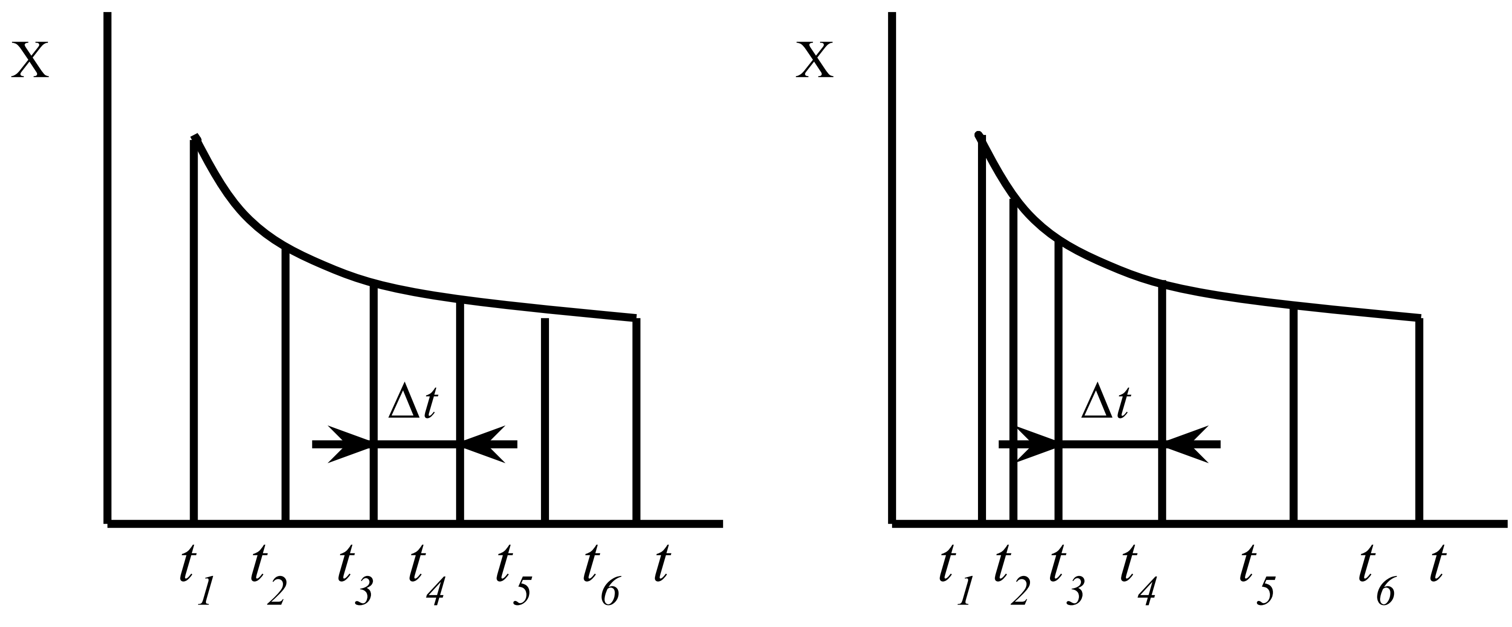
Преобразование информации

При дискретизации сигналов приходится решать вопрос о том, как часто следует проводить ***отсчеты*** функции, т. е. каков должен быть **шаг дискретизации** $\Delta t_i = t_i - t_{i-1}$. При *малых шагах* дискретизации количество отсчетов функции на отрезке обработки будет большим и точность воспроизведения — *высокой*. При *больших шагах* дискретизации количество отсчетов уменьшается, но при этом, как правило, *снижается точность восстановления*. *Оптимальной* является такая дискретизация, которая обеспечивает представление исходного сигнала с *заданной точностью при* минимальном количестве

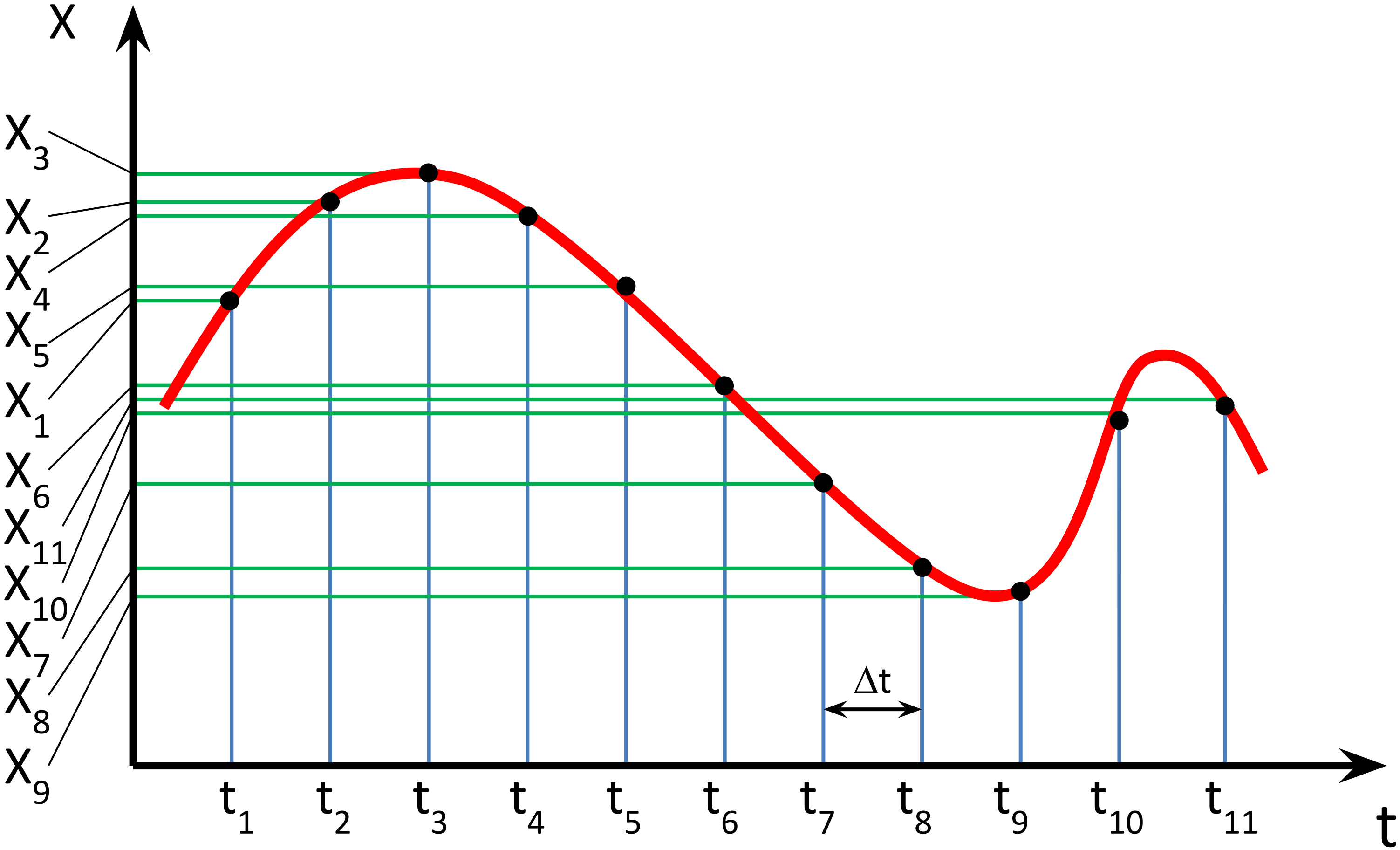
Преобразование информации

Дискретизация называется равномерной (рис. *а*), если длительность интервалов $\Delta t_i = \mathrm{const}$ на всем отрезке обработки сигнала.

Дискретизация называется неравномерной (рис. *б*), если длительность интервалов между отсчетами Δt_i , различна, т. е. $\Delta t_i = \mathrm{var}$.



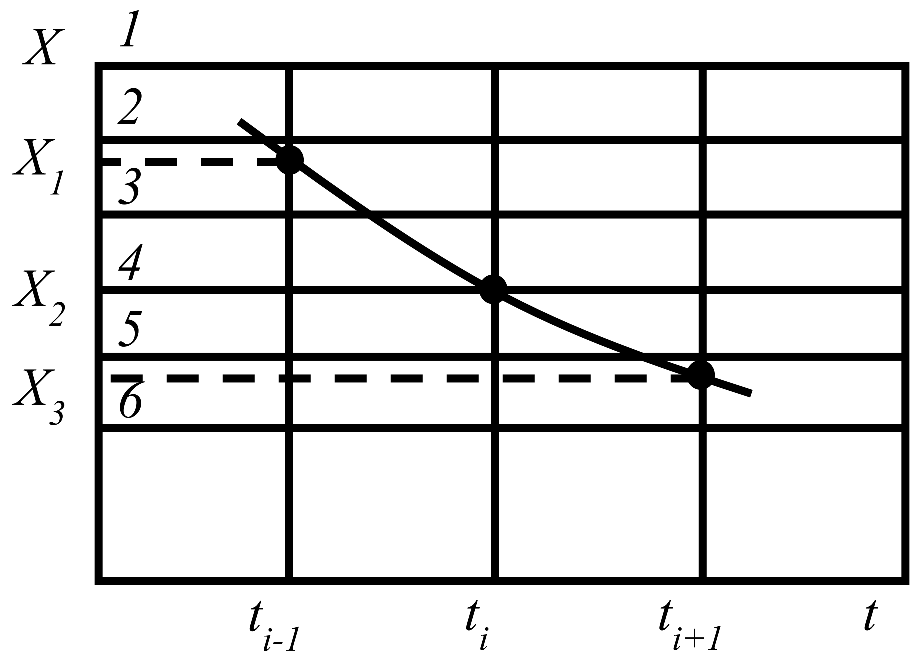
а *б*



x
X_3
X_2
X_4
X_5
X_1
X_6
X_{11}
X_{10}
X_7
X_8
X_9
Δt
t_1
t_2
t_3
t_4
t_5
t_6
t_7
t_8
t_9
t_{10}
t_{11}
t

Преобразование информации

Квантование по уровню состоит в преобразовании непрерывных значений сигнала $x(t_i)$ в моменты отсчета t_i, в дискретные значения.



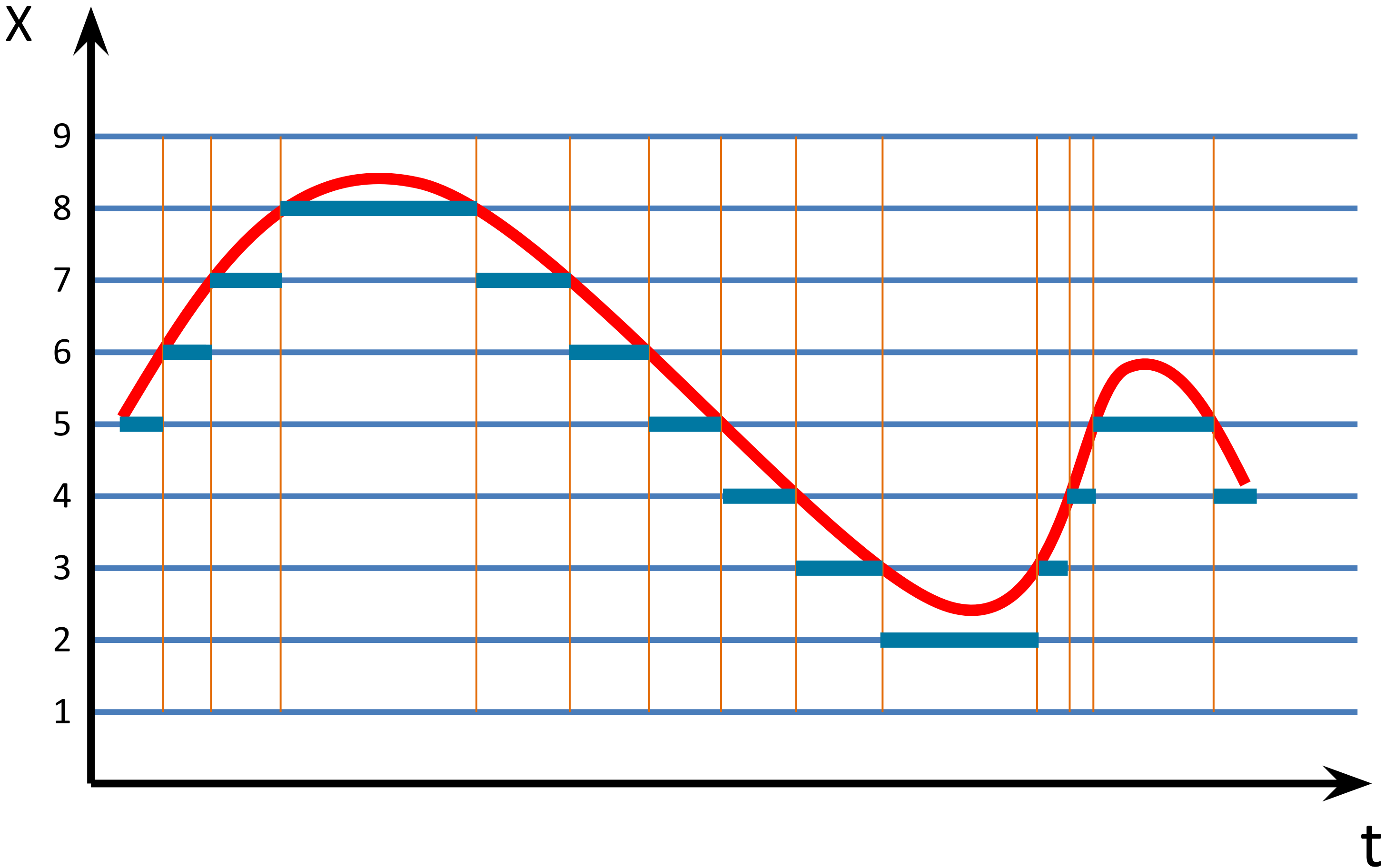
В соответствии с графиком изменения функции $x(t)$ ее истинные значения представляются в виде заранее заданных дискретных уровней *1, 2, 3, 4, 5* или *6*.

Преобразование информации

Квантование по уровню может быть равномерным и неравномерным в зависимости от величины шага квантования. Под шагом (интервалом) квантования δ_m понимается разность $\delta_m = \mathbf{x}_m - x_{m\text{-}1}$, где $\mathbf{x}_m$, $x_{m\text{-}1}$ – соседние уровни квантования.

Уровень квантования для заданного значения сигнала $x(t)$ можно выразить двумя способами:

1. сигнал $x(t_{\mathrm{i}})$ отождествляется с **ближайшим уровнем** квантования;
2. сигнал $x(t_{\mathrm{i}})$ отождествляется с **ближайшим меньшим (или большим) уровнем** квантования.


X
9
8
7
6
5
4
3
2
1
t

Преобразование информации

Так как в процессе преобразования значение сигнала $x(t)$ отображается уровнем квантования x_m , а каждому уровню m может быть поставлен в соответствие свой номер (число), то при передаче или хранении информации можно вместо истинного значения величины x_m использовать соответствующее число m.

Такое преобразование сопровождается **шумами** или **погрешностью квантования**. Погрешность квантования связана с заменой истинного значения сигнала $x(t_i)$ значением, соответствующим уровню квантования x_m.

Преобразование информации

Метод дискретизации при преобразовании непрерывной информации в дискретную влияет на количество информации, которую надо хранить или преобразовывать в ЭВМ. Важна **теорема Котельникова**, согласно которой функция, имеющая ограниченный спектр частот, полностью определяется дискретным множеством своих значений, взятых с частотой отсчетов: $F_0 = 2f_m$, где $f_m = 2\pi\omega_m$ – максимальная частота в спектре частот $S(j\omega)$ сигнала $x(t)$; ω_m –угловая скорость.

Функция $x(t)$ воспроизводится без погрешностей по точным значениям $x(t_i)$ в виде ряда Котельникова:

$$x(t) = \sum_{k=-\infty}^{0} x(k_{\Delta t}) \frac{\sin \omega_m (t - k_{\Delta t})}{(t - k_{\Delta t})}$$

где Δt – шаг дискредитации.

Преобразование информации

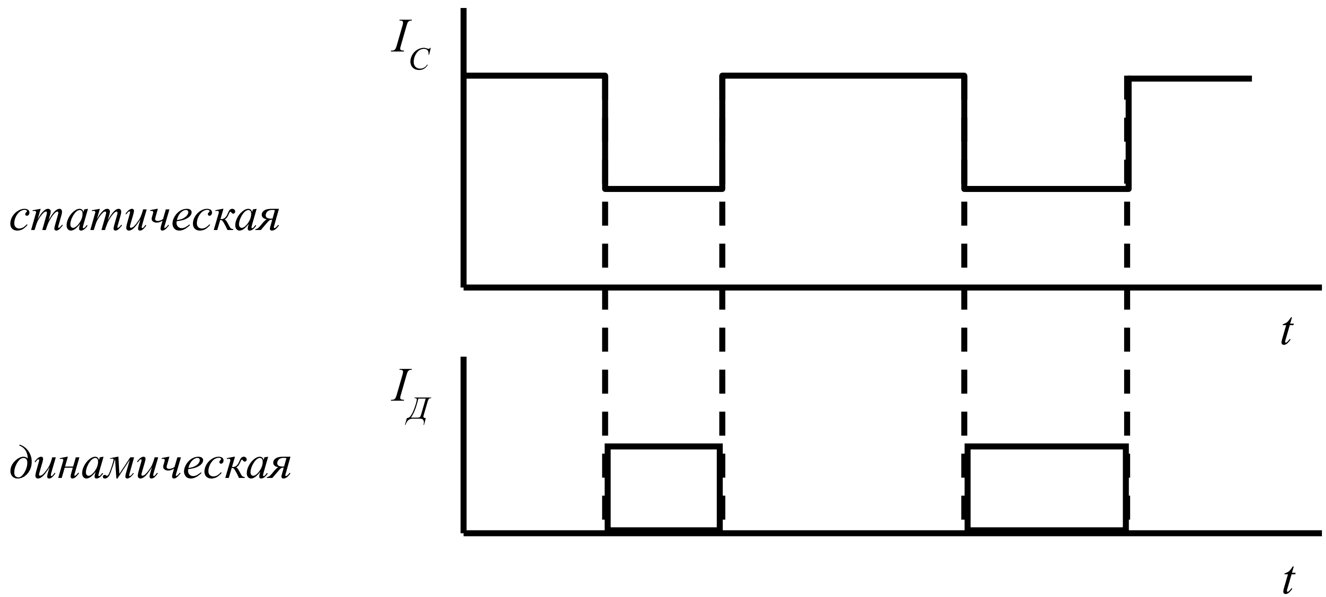
Для практических задач, однако, идеально точное восстановление функций не требуется, необходимо лишь восстановление с заданной точностью. Поэтому теорему Котельникова можно рассматривать как приближенную для функций с неограниченным спектром. На практике частоту отсчетов часто определяют по формуле

$$F_0 = 2f_{max}k_з,$$

где $k_з$ — коэффициент запаса (обычно $1{,}5<k_з< 6$); $f_{\max}$ — максимальная допустимая частота в спектре сигнала $x(t)$, например, с учетом доли полной энергии, сосредоточенной в ограниченном частотой спектре сигнала.

Формы представления информации

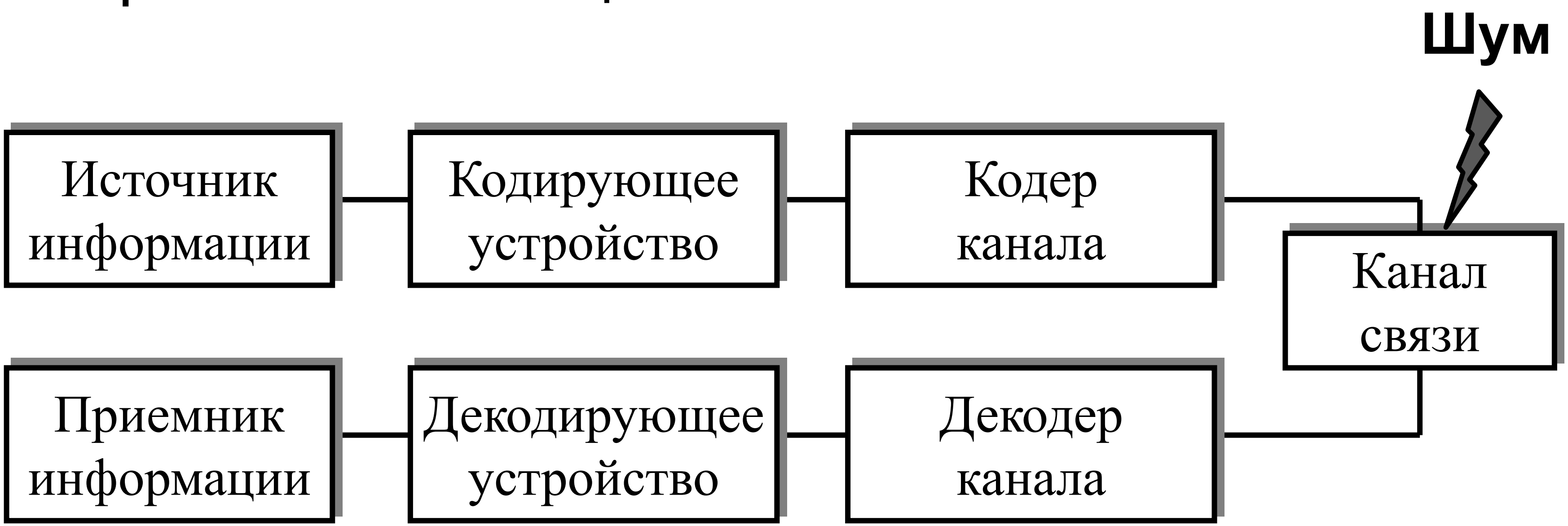
- Статическая информация
- Динамическая информация



Информационная модель канала связи

Формы представления информации

- *Кодирование* – преобразование сообщения в форму, удобную для передачи по данному каналу
- *Декодирование* – операция восстановления принятого сообщения



Информационная модель канала связи с шумами

Передача информации

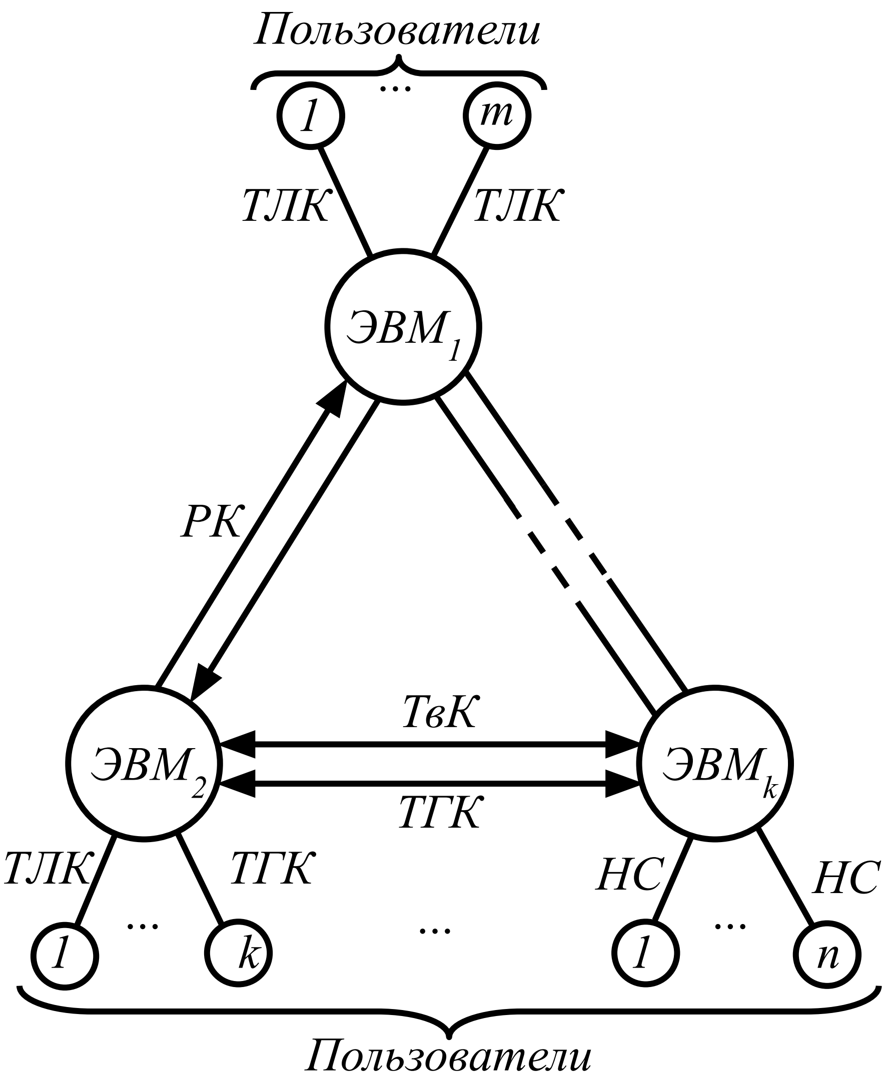
Каналы связи: непосредственная связь, телефонный канал, телеграфный канал, радиоканал, телевизионный канал и т.д.

Пропускная способность – характеристика канала связи, которая не зависит от скорости передачи информации.

Пропускная способность канала с шумами – максимальная скорость передачи информации при условии, что канал связи без помех согласован с источником информации.

Передача информации

Каналы связи в вычислительных сетях



Передача информации

Передача информации по каналу без помех

Если через канал связи без помех передается последовательность дискретных сообщений длительностью T, то скорость передачи информации по каналу связи (бит/с)

$$v = \lim_{T \to \infty} (I / T)$$

где I – количество информации, содержащейся в последовательности сообщений.

Предельное значение скорости передачи информации называется пропускной способностью канала связи без помех $c = v_{max}$.

Передача информации

Передача информации по каналу без помех

Количество информации в сообщениях максимально при равной вероятности состояний. То

$$\nu = \lim_{T \to \infty} \log_2 k / T$$

Скорость передачи информации в общем случае зависит от статистических свойств сообщений и параметров канала связи.

Передача информации

Передача информации по каналу без помех

Для наиболее эффективного использования канала связи необходимо, чтобы скорость передачи информации была как можно ближе к пропускной способности канала связи. Если скорость поступления информации на вход канала связи превышает пропускную способность канала, то по каналу будет передана не вся информация. Основное условие согласования источника информации и канала связи $v \leq c$.

Согласование осуществляется путем соответствующего кодирования сообщений.

Передача информации

Передача информации по каналу с помехами

При передаче информации через канал с помехами сообщения искажаются, и на приемной стороне нет уверенности в том, что принято именно то сообщение, которое передавалось. Следовательно, сообщение недостоверно, вероятность правильности его после приема не равна единице. В этом случае количество получаемой информации уменьшается на величину неопределенности, вносимой помехами, т. е. вычисляется как разность энтропии сообщения до и после приема: $I' = H(i) - H_i(i)$, где $H(i)$ – энтропия источника сообщений; $H_i(i)$ – энтропия сообщений на приемной стороне.

Таким образом скорость передачи по каналу связи с помехами:

$$\nu' = \lim_{T \to \infty} \frac{H(i) - H_i(i)}{T}$$

Передача информации

Передача информации по каналу с помехами

Пропускной способностью канала с шумами называется максимальная скорость передачи информации при условии, что канал связи без помех согласован с источником информации:

$$c = \lim_{T \to \infty} \frac{I_{max}}{T}$$

Если энтропия источника информации не превышает пропускной способности канала ($H \le c$), то существует код, обеспечивающий передачу информации через канал с помехами со сколь угодно малой частотой ошибок или сколь угодно малой недостоверностью.

Передача информации

Передача информации по каналу с помехами

Пропускная способность канала связи при ограниченной средней мощности аналогового сигнала:

$$c = F_m \log_2 (1 + W_c / W_m)$$

где F_m – полоса частот канала (Гц); W_c – средняя мощность сигнала; W_m — средняя мощность помех (равномерный спектр) с нормальным законом распределения амплитуд в полосе частот канала связи.

Передача информации

Передача информации по каналу с помехами

Следовательно, можно передавать информацию по каналу с помехами без ошибок, если скорость передачи информации меньше пропускной способности канала. Для скорости $v > c$ при любой системе кодирования частота ошибок принимает конечное значение, причем оно растет с увеличением значения v. Для канала с весьма высоким уровнем шумов ($W_m >> W_c$) максимальная скорость передачи близка к нулю.

Фазы преобразования информации

1. Подготовка информации
2. Регистрация информации
3. Сбор и передача
4. Обработка
5. Вывод и воспроизведение

Наряду с крупными этапами или фазами преобразования информации существуют более мелкие операции, связанные с отдельными воздействиями на информацию для получения каких-то данных по заранее известным алгоритмам: классификация, синтез.

Фазы преобразования информации

Независимо от фазы преобразования информации каждый вид ее обладает определенными характеристиками, среди которых полезно выделить связанные с функционированием ИС следующие характеристики:

- Цель информации
- Формат
- Избыточность
- Периодичность появления
- Верность