

Лекция 2

**Информация:
формы, измерение,
количество и качество**

Понятие информации

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

Несмотря на то, что с термином "информация" мы сталкиваемся ежедневно, строгого и общепризнанного ее определения до сих пор не существует. Поэтому вместо определения используют понятие об информации.

Понятие информации

- Информация есть информация, а не материя или энергия (Норберт Винер)
- Информация – мера снижения неопределенности системы (Клод Шеннон)
- Информация есть отражение реального мира с помощью сведений (сообщений)
- **Информация – это сведения об объектах и явлениях окружающей среды, их параметрах, свойствах и состоянии, которые уменьшают степень неопределенности и неполноты имеющихся о них знаний.**

Свойства информации

(с философской точки зрения)

- информация приносит сведения, об окружающем мире, которых в рассматриваемой точке не было до ее получения;
- информация нематериальна, но она проявляется в форме материальных носителей дискретных знаков или первичных сигналов;
- знаки и первичные сигналы несут информацию только для получателя, способного ее распознать.

Данные, сообщения, сигналы

Данные – это информация, закодированная определенным образом с целью передачи, обработки, хранения, поиска или извлечения.

Данные становятся информацией при наличии **методов** ее извлечения.

Информация — это продукт взаимодействия данных и адекватных им методов

- **Сообщение** – форма представления информации.
- **Сигнал** – материальный носитель, который фиксирует информацию для переноса ее от источника к потребителю.

Виды сообщений

Сообщения могут быть представлены в виде:

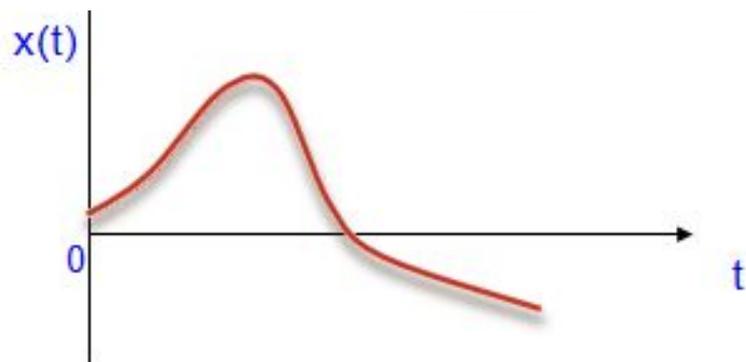
- первичных сигналов (музыка, речь, зрительный образ и т.д.);
- знаков (текст, числа, коды и т.д.).

Сообщения могут быть:

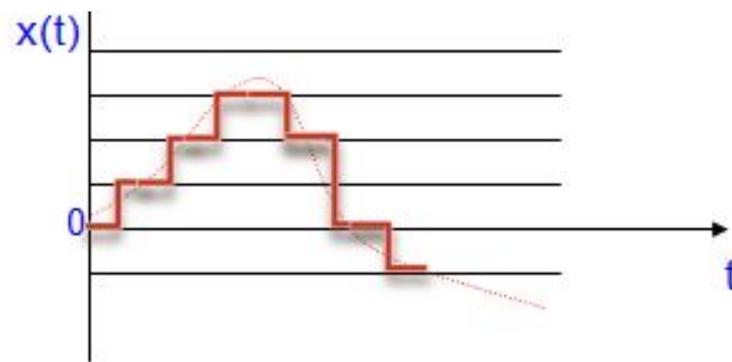
- динамичными;
- статичными.

Виды сигналов

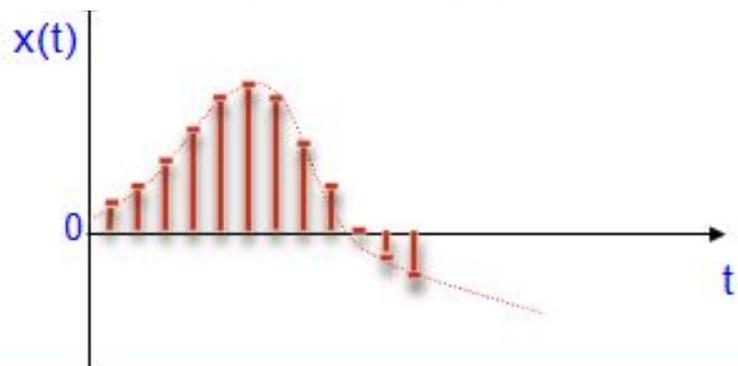
Непрерывный (аналоговый)



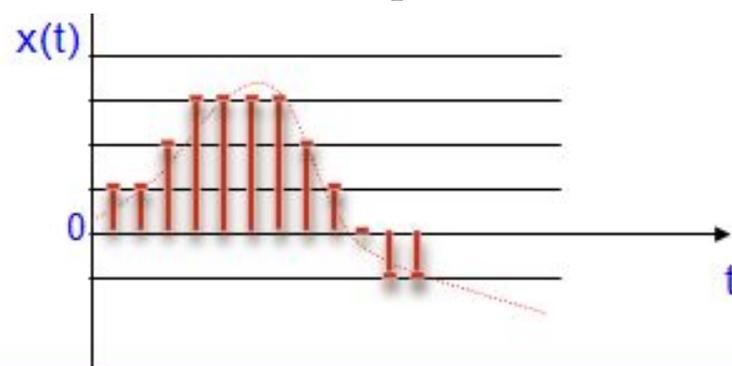
Непрерывно-дискретный



Дискретно-непрерывный

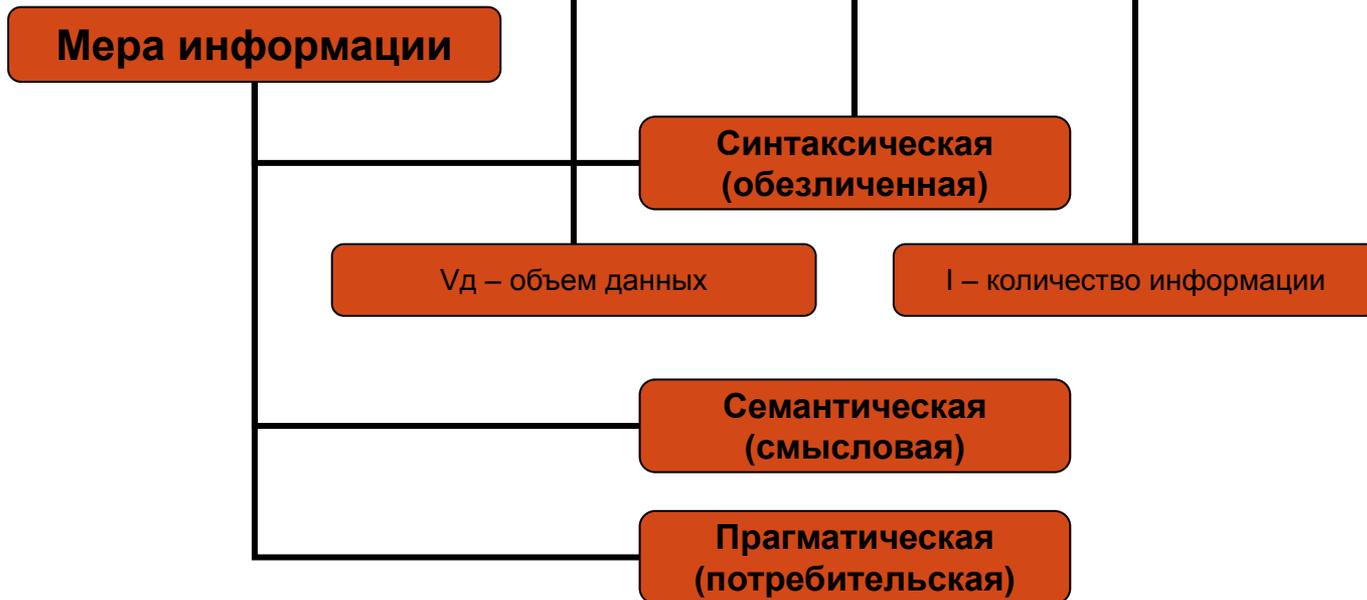


Дискретный



Меры информации

- **Синтаксическая мера** отражает формально-структурные характеристики информации, не затрагивая его смысловое содержание.
- **Семантическая мера** отражает смысловое содержание информации.
- **Прагматическая мера** отражает отношение информации и ее потребителя, соответствие информации поставленной цели.



Синтаксическая мера. Объем данных

оперирует с обезличенной информацией, не выражающей смыслового отношения к объекту.

Объём данных (V_D) в сообщении измеряется количеством символов (разрядов) в этом сообщении.

Единицы измерения:

- В десятичной системе счисления – **дит** (dit - **d**ecimal **d**igit)
- В двоичной системе счисления – **бит** (bit - **b**inary **d**igit)
- Производные единицы измерения объема данных:

1 Кбайт (один килобайт) = 2^{10} байт = 1024 байта (1 kB);

1 Мбайт (один мегабайт) = 2^{10} Кбайт = 1024 Кбайта (1 MB);

1 Гбайт (один гигабайт) = 2^{10} Мбайт = 1024 Мбайта (1 GB);

1 Тбайт (один терабайт) = 2^{10} Гбайт = 1024 Гбайта (1 TB);

1 Пбайт (один петабайт) = 2^{10} Тбайт = 1024 Тбайта (1 PB);

1 Эбайт (один эксабайт) = 2^{10} Пбайт = 1024 Пбайта (1 EB);

1 Збайт (один зеттабайт) = 2^{10} Эбайт = 1024 Эбайта (1 ZB);

1 Йбайт (один йоттабайт) = 2^{10} Збайт = 1024 Збайта (1 YB).

Измерение информации (Алфавитный подход)

Алфавитный подход -

это способ измерения
информационного объема текста,
не связанного с его содержанием.

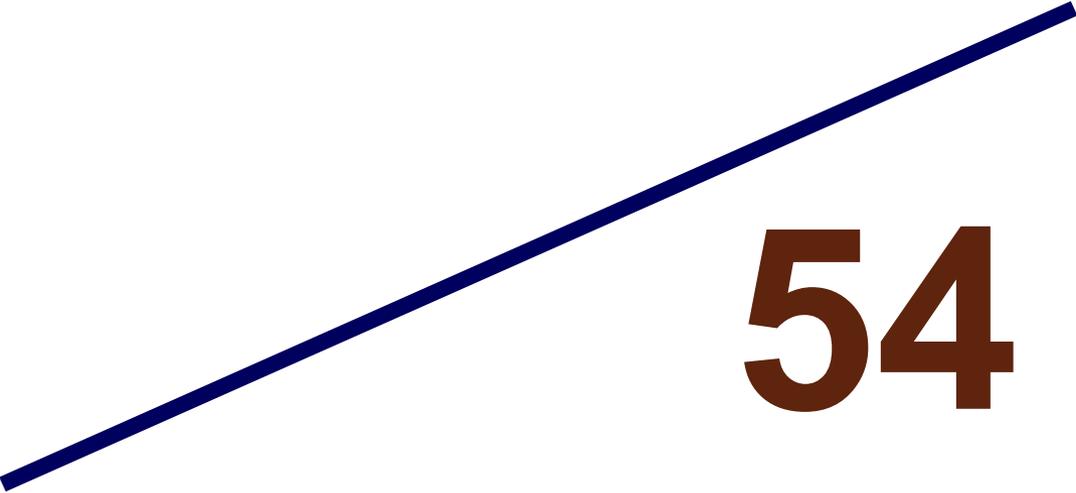
Алфавит :

- набор букв;
- знаков препинания;
- цифр;
- скобок и других символов, используемых в тексте;
- пробел между словами.



МОЩНОСТЬ РУССКОГО АЛФАВИТА:

- 33 буквы
- 10 цифр
- 11 знаков препинания
- скобки
- пробел



54

$$N = 2^i$$

где N – количество символов знаковой системы (**мощностью алфавита**);
 i – информационный вес символа.

Самое наименьшее число символов в алфавите: 2 (0 и 1)-
двоичный алфавит.

- Информационный вес символа двоичного алфавита принят за единицу информации и называется
1 БИТ.

- Алфавит, из которого составляется «компьютерный текст», содержит 256 СИМВОЛОВ.

$$N = 2^i$$

$$256 = 2^8$$

следовательно

- *1 символ компьютерного алфавита «весит» 8 битов.*

Задачи

- Сообщение записанное буквами из 128 – символьного алфавита, содержит 30 символов. Какой объем информации оно несет?

$$N = 2^i$$

$$N = 128 = 2^7,$$

$i = 7$ бит (вес одного символа).

Сообщение содержит 30 символов,
следовательно

$$7 \times 30 = 210 \text{ бит}$$

Сколько байтов составляет
сообщение, содержащее 1000
бит?

ОТВЕТ:

1 байт = 8 бит

1000: 8 = 125 байт

Информационное сообщение объёмом 5 Кбайт содержит 8192 символа. Сколько символов содержит алфавит, при помощи которого было записано это сообщение?

РЕШЕНИЕ:

$$N = 2^i$$

5 Кб = 5120 байт = 40960 бит

Сообщение содержит 8192 символа,
следовательно

$i = 40960 : 8192 = 5$ бит (вес одного
символа).

$$N = 2^5 = 32$$

Для записи текста
использовался 32 – символный
алфавит. Каждая страница
содержит 40 строк по 50
символов в строке. Какой объем
информации содержит 7
страниц текста?

РЕШЕНИЕ: $N = 2^i$

- $N = 32 = 2^5$
- $i = 5$ бит (вес одного символа)
- Вес одной страницы = $40 \times 50 \times 5 = 10000$ бит = 1250 байт
- Объём 7 страниц = $7 \times 1250 = 8750$ байт

- Сообщение занимает 5 страниц по 35 строк. В каждой строке записано по 60 символов. Сколько символов в алфавите, если все сообщение содержит 3750 байтов?

РЕШЕНИЕ: $N = 2^i$

- 3750 байт = 30000 бит
- $5 \times 25 \times 60 = 7500$ СИМВОЛОВ
- Один символ «весит» 30000:
 $7500 = 4$ бита
- $N = 2^4 = 16$

Синтаксическая мера. Количество информации

Количество информации I на синтаксическом уровне определяется через понятие энтропии системы.

Энтропия $H(\alpha)$ – мера неосведомленности о системе α .

$I_{\beta}(\alpha)$ - количество информации, полученное из сообщения β о системе.

$H_{\beta}(\alpha)$ - апостериорная (после получения) неопределенность состояния системы α после получения сообщения β .

$$I_{\beta}(\alpha) = H(\alpha) - H_{\beta}(\alpha).$$

т.е. количество информации измеряется изменением (уменьшением) неопределенности состояния системы.

Если конечная неопределенность $H_{\beta}(\alpha) = 0$, то $I_{\beta}(\alpha) = H(\alpha)$.

Другими словами, энтропия $H(\alpha)$ может рассматриваться как мера недостающей информации.

Коэффициент (степень) информативности (лаконичность)

$$Y = \frac{I}{V_d}$$

Полная информация о системе

Энтропия системы, имеющей N возможных состояний :

$$H = -\sum_{i=1}^N p_i \log_2 p_i, \quad (\text{формула Шеннона})$$

где p_i — вероятность того, что система находится в i -м состоянии.

Если все N состояний системы равновероятны

$$H = -\sum_{i=1}^N \frac{1}{N} \log_2 \frac{1}{N} = \log_2 N. \quad (\text{формула Хартли})$$

Для числового символа равно основанию системы счисления m , а для сообщения из n цифр

$$N = m^n$$

тогда количество информации в таком сообщении

$$I = \log_2 m^n = n \log_2 m$$

Содержательный подход к измерению информации



- Какова вероятность того, что камень сам по себе будет лететь вверх?
 - Какова вероятность того, что вы будете читать книгу с конца?
 - Какова вероятность того, что сегодня будет дождь?
- События могут быть равновероятными и не быть таковыми.

Содержательный подход к измерению информации для равновероятных событий

- Количество информации определяется содержанием информации (событиями).
- Знания можно увеличивать. Знания увеличиваются – неопределенность знания уменьшается.

Единицы измерения информации

- Сообщение, уменьшающее неопределенность знания в 2 раза, несет 1 бит информации.
- Мы получаем 1 бит информации, получая ответ ДА или НЕТ на заданный вопрос.



Информационный вес события

- Количество информации (i), содержащееся в сообщении о том, что произошло одно из N определяется из решения показательного уравнения

$$N = 2^i$$

Решение задач

Сколько информации несет сообщение, что Иванов живет в 3-ей квартире 64-х квартирному дома?

1) 3 бита; 2) 3 байта; 3) 2 бита; 4) 6 бит.

Число событий $N=64$.

По формуле Хартли рассчитаем b .

$$64 = 2^6$$

$$64 = 2^b$$

ОВ

$$N = 2^b$$

Семантическая мера

служит для измерения смыслового содержания информации.

Тезаурусная мера связывает семантические свойства информации со способностью пользователя принимать поступившие сообщения.

Тезаурус — это совокупность сведений, которыми располагает пользователь или система.

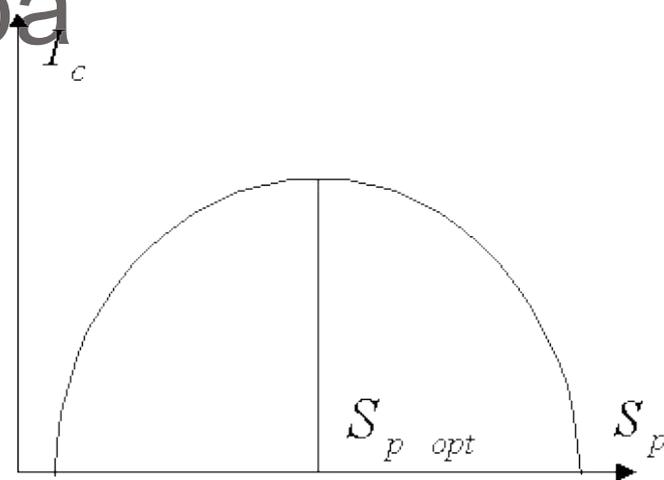
$$S_p \rightarrow 0 \Rightarrow I_c \rightarrow 0$$

$$S_{p \text{ opt}} \Rightarrow I_c \rightarrow \max$$

$$S_p \rightarrow \infty \Rightarrow I_c \rightarrow 0$$

Коэффициент содержательности

$$C = \frac{I_c}{V_d}$$



I_c — количество семантической информации

S_p — тезаурус пользователя

Прагматическая мера

определяет полезность информации (ценность) для достижения пользователем поставленной цели.

Ценность информации целесообразно измерять в тех же самых единицах (или близких к ним), в которых измеряется целевая функция.

Качество информации

Совокупность свойств, обуславливающих способность информации удовлетворять потребности людей.

Репрезентативность - правильность отбора и формирования информации в целях адекватного отражения свойств объекта.

Достаточность (полнота) информации означает, что она содержит минимальный, но достаточный для принятия правильного решения состав (набор показателей).

Доступность информации восприятию пользователя обеспечивается выполнением процедур ее получения и преобразования.

Актуальность информации – это степень сохранения ценности информации на момент использования.

Качество информации

Своевременность информации означает ее поступление не позже заранее назначенного момента времени, согласованного с временем решения поставленной задачи.

Точность информации определяется степенью близости получаемой информации к реальному состоянию объекта, процесса, явления и т.п.

Достоверность информации определяется ее свойством отражать реально существующие объекты с необходимой точностью.

Устойчивость информации отражает ее способность реагировать на изменения исходных данных без нарушения необходимой точности.

Классификация информации

По восприятию органами

чувств:

- *визуальная* (символы, рисунки, жесты, зрительные образы);
- *звуковая* (аудиальная);
- *вкусовая*;
- *обонятельная* (запахи);
- *тактильная* (ощущения);
- *органолептическая*.

По общественному назначению:

- *массовая*;
- *личная*;
- *обыденная*;
- *общественно – политическая*;
- *эстетическая*.

По области возникновения:

- *элементарная* (процессы и явления неодушевленной природы);
- *биологическая* (процессы животного и растительного мира);
- *социальная* (человеческое общество).

Классификация информации, циркулирующей в организации

Информация				
По месту возникновения	По стабильности	По стадии обработки	По способу отображения	По функции управления
Входная	Переменная	Первичная	Текстовая	Плановая
Выходная	Постоянная	Вторичная	Графическая	Нормативно-справочная
Внутренняя		Промежуточная		Учетная
Внешняя		Результатная		Оперативная