

## Тема 3. Понятие информации. Мера информации.

### 1. Определение информации и её виды.

#### 1.1. Определение информации.

Термин информация происходит от латинского *informatio*, что означает разъяснение, осведомление, изложение. С позиции материалистической философии информация есть отражение реального мира с помощью сведений (сообщений). Сообщение - это форма представления информации в виде речи, текста, изображения, цифровых данных, графиков, таблиц и т.п. В широком смысле информация - это общенаучное понятие, включающее в себя обмен сведениями между людьми, обмен сигналами между живой и неживой природой, людьми и устройствами.

**Информация** - это содержание сообщения, сигнала, памяти.

#### 1.2. Определение информационных процессов.

Люди обмениваются устными сообщениями, записками, используют для этого электронную и голосовую почту. Они передают друг другу просьбы, приказы, объявления и научные статьи; хранят старые письма и документы; размышляют над полученными известиями, реагируют на них. Все это – информационные процессы.

**Информационные процессы** - это процессы передачи, хранения и переработки информации.

#### 1.3. Виды информации.

Виды информации:

- Числовая представляется цифрами и отражает результаты некоторого вычисления.
- Текстовая представлена в виде слов, состоящих из символов.
- Кодовая представлена кодами (напр., азбука Морзе, машинные коды и т.п.).
- Графическая представлена графическими объектами с учетом их геометрических и оптических свойств.
- Акустическая представлена звуками.
- Телевизионная - это изображения в ТВ-формате.
- Видео-информация - это видео- и кинофильмы в специальных форматах.

#### 1.4. Представление информации.

Информация всегда связана с материальным носителем, а ее передача - с затратами энергии. Одну и ту же информацию можно хранить в различном материальном виде: на бумаге, в виде негатива фотографии, на магнитной ленте и т.д. и передавать с различными затратами энергии - по почте, по телефону, с курьером и т.д.

При любых видах работы с информацией всегда идет речь о ее представлении в виде определенных символических структур. Наиболее распространены одномерные представления информации.

Одномерное представление информации имеет вид последовательности символов,

например, - в письменных текстах (набор букв), при передаче по каналам связи (уровень электрического сигнала), в памяти ЭВМ (наличие или отсутствие сигнала).

Многомерное представление информации - это рисунки, схемы, объемные макеты и плюс к этому множественность признаков используемых символов. Например, информацию могут нести не только значения букв и цифр, но и их цвет, размер, вид шрифта.

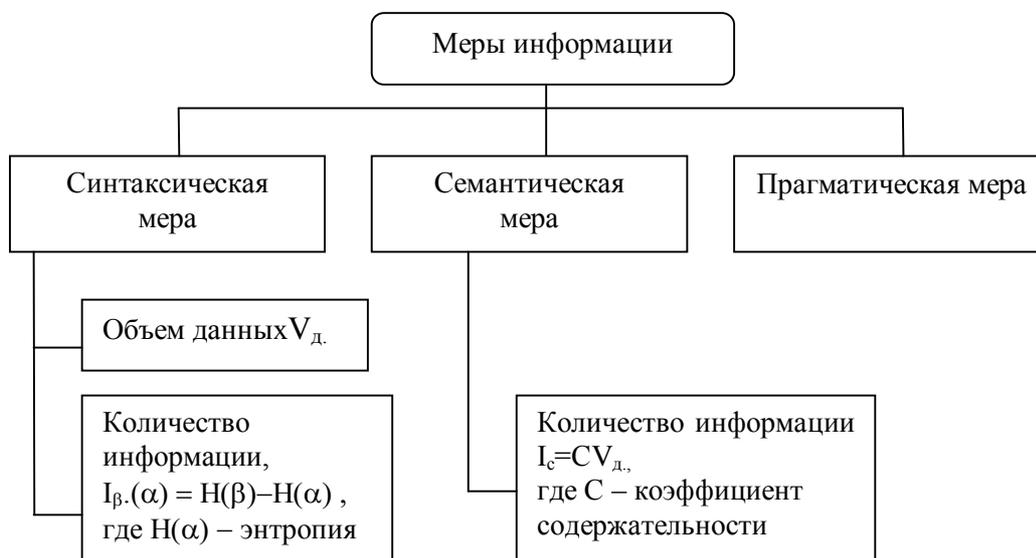
## 2. Теории информации.

Структурная теория информации рассматривает структуру построения отдельных информационных сообщений. Единица количества информации - элементарная структурная единица квант.

Статистическая теория оценивает информацию с точки зрения меры неопределенности. Основное внимание уделяется распределению вероятностей, либо появлению сигналов, либо изменению характеристик этих сигналов и построению на его основе некоторых обобщенных характеристик, позволяющих оценить количество информации.

Семантическая теория занимается изучением именно смысловых характеристик информации: ценности, содержательности, полезности. Помогает связать количество и ценность информации с такими обобщенными характеристиками системы, как эффективность, информационная пропускная способность, информационная помехоустойчивость.

## 3. Меры информации.



### Единицы измерения информации и примеры

Мера информации	Единицы измерения	Примеры
<u>Синтаксическая</u> (шенноновский подход)	Степень уменьшения неопределенности	Вероятность события

компьютерный подход)	Единицы представления информации	Бит, байт, Кбайт и т.д.
<u>Семантическая</u>	Тезаурус Экономический показатель	Пакет прикладных программ, ПК, компьютерные сети Рентабельность, производительность и т.д.
<u>Прагматическая</u>	Ценность использования	Емкость памяти, производительность ПК, скорость передачи данных и т.д. Денежное выражение
<u>Алгоритмическая</u>	Минимальное число внутренних состояний машины	Машина Тьюринга

### 3.1. Определение энтропии системы.

**Энтропия системы** - это степень (мера) неопределенности (неосведомленности) о состоянии системы.

### 3.2. Синтаксическая мера информации

Объем данных  $V_d$  в сообщении измеряется количеством символов (разрядов) в этом сообщении. В различных системах счисления один разряд имеет различный вес и соответственно меняется единица измерения данных:

- в двоичной системе счисления единица измерения - бит (bit-binary digit-двоичный разряд);
- в десятичной системе счисления единица измерения – дигит (десятичный разряд).

Количество информации  $I$  на синтаксическом уровне невозможно определить без рассмотрения понятия неопределенности состояния системы (энтропии системы). Получение информации о какой-либо системе всегда связано с изменением степени неосведомленности получателя о состоянии этой системы (теория Шеннона).

### 3.2. Мера информации по Шеннону

Пусть до получения информации потребитель имеет некоторые предварительные (априорные) сведения о системе  $\alpha$ . Мерой его неосведомленности о системе является функция  $H(\alpha)$ , которая в то же время служит и мерой неопределенности состояния системы.

После получения некоторого сообщения  $\beta$  получатель приобрел некоторую дополнительную информацию  $I_\beta(\alpha)$ , уменьшившую его априорную неосведомленность так, что апостериорная (после получения сообщения  $\beta$ ) неопределенность состояния системы стала  $H_\beta(\alpha)$ .

Тогда количество информации  $I_\beta(\alpha)$  о системе, полученной в сообщении  $\beta$ , определится как:

$$I_\beta(\alpha) = H(\alpha) - H_\beta(\alpha),$$

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

Если конечная неопределенность  $H_\beta(\alpha)$  обратится в нуль, то первоначальное непол-

ное знание заменится полным знанием и количество информации  $I_B(\alpha) = H(\alpha)$ . Иными словами, энтропия системы  $H(\alpha)$  может рассматриваться как мера недостающей информации.

Энтропия системы  $H(\alpha)$ , имеющая  $N$  возможных состояний, согласно формуле Шеннона, равна:

$$H(\alpha) = -\sum_{i=1}^N P_i \log P_i$$

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

Для случая, когда все состояния системы равновероятны, т.е. их вероятности равны  $P_i = (1/N)$ , ее энтропия определяется соотношением:

$$H(\alpha) = -\sum_{i=1}^N \frac{1}{N} \log \frac{1}{N}$$

Часто информация кодируется числовыми кодами в той или иной системе счисления, особенно это актуально при представлении информации в компьютере. Естественно, что одно и то же количество разрядов в разных системах счисления может передать разное число состояний отображаемого объекта, что можно представить в виде соотношения

$$N = m^n,$$

где  $N$  — число всевозможных отображаемых состояний;

$m$  - основание системы счисления (разнообразие символов, применяемых в алфавите);

$n$  — число разрядов (символов) в сообщении.

Наиболее часто используются двоичные и десятичные логарифмы. Единицами измерения в этих случаях будут соответственно бит и дит.

Коэффициент (степень) информативности (лаконичность) сообщения определяется отношением количества информации к объему данных, т.е.:

$$Y = \frac{1}{V_d}, \text{ причем } 0 < Y < 1.$$

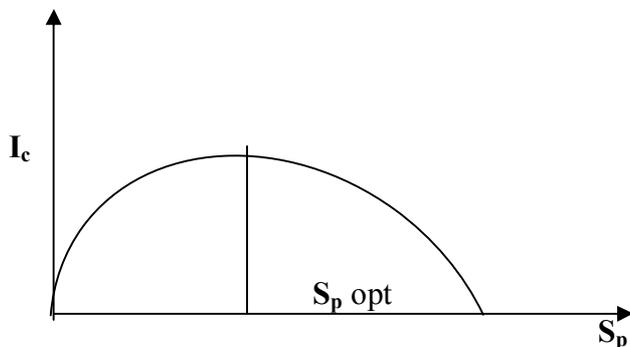
С увеличением  $Y$  уменьшаются объемы работы по преобразованию информации (данных) в системе. Поэтому стремятся к повышению информативности, для чего разрабатываются специальные методы оптимального кодирования информации.

#### **4. Семантическая мера информации.**

Для измерения смыслового содержания информации, т.е. ее количества на семантическом уровне, наибольшее признание получила тезаурусная мера, которая связывает семантические свойства информации со способностью пользователя принимать поступившее сообщение. Для этого используется понятие тезаурус.

**Тезаурус** - это совокупность сведений, которыми располагает пользователь или система.

В зависимости от соотношений между смысловым содержанием информации  $S$  и тезаурусом пользователя  $S_p$  изменяется количество семантической информации  $I_c$ , воспринимаемой пользователем и включаемой им в дальнейшем в свой тезаурус.



- при  $S_p \approx 0$  пользователь не воспринимает, не понимает поступающую информацию;

- при  $S_p \rightarrow \infty$  пользователь все знает, и информация ему не нужна.

Рис. 2. Зависимость количества семантической информации, воспринимаемой потребителем, от его тезауруса  $I_c = F(S_p)$

Максимальное количество информации  $I_c$  потребитель приобретает при согласовании ее смыслового содержания  $S$  со своим тезаурусом  $S_p$  ( $S_p = S_p \text{ opt}$ ), когда поступающая информация понятна пользователю и несет ему ранее не известные (отсутствующие в его тезаурусе) сведения.

Следовательно, количество семантической информации в сообщении, количество новых знаний, получаемых пользователем, является величиной относительной. Одно и то же сообщение может иметь смысловое содержание для компетентного пользователя и быть бессмысленным для пользователя некомпетентного.

При оценке семантического (содержательного) аспекта информации необходимо стремиться к согласованию величин  $S$  и  $S_p$ .

Относительной мерой количества семантической информации может служить коэффициент содержательности  $C$ , который определяется как отношение количества семантической информации к ее объему:

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

## **5. Прагматическая мера информации.**

Эта мера определяет полезность информации (ценность) для достижения пользователем поставленной цели. Эта мера также величина относительная, обусловленная особенностями использования этой информации в той или иной системе. Ценность информации целесообразно измерять в тех же единицах (или близких к ним), в которых измеряется целевая функция.

## **6. Алгоритмическая мера информации.**

Каждый согласится, что слово 0101...01 сложнее слова 00...0, а слово, где 0 и 1 выбираются из эксперимента - бросания монеты (где 0-герб, 1-решка), сложнее обоих предыдущих.

Любому сообщению можно приписать количественную характеристику, отражающую сложность (размер) программы, которая позволяет ее произвести.

Так как имеется много разных вычислительных машин и разных языков программирования (разных способов задания алгоритма), то для определенности задаются некоторой конкретной вычислительной машиной, например машиной Тьюринга.

Сложность слова (сообщения) определяется как минимальное число внутренних состояний машины Тьюринга, требующиеся для его воспроизведения.

## 7. Качество информации

Потребительские показатели качества:

- репрезентативность, содержательность, достаточность
- актуальность, своевременность, точность
- достоверность, устойчивость

Репрезентативность связана с правильностью ее отбора и формирования в целях адекватного отражения свойств объекта. Важнейшее значение здесь имеют:

- правильность концепции, на базе которой сформулировано исходное понятие;
- обоснованность отбора существенных признаков и связей отображаемого явления.

Содержательность отражает семантическую емкость, равную отношению количества семантической информации в сообщении к объему обрабатываемых данных, т.е.

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

С увеличением содержательности информации растет семантическая пропускная способность информационной системы (для получения одних и тех же сведений требуется преобразовать меньший объем данных).

Наряду с коэффициентом содержательности  $C$ , отражающим семантический аспект, можно использовать и коэффициент информативности, характеризующийся отношением количества синтаксической информации (по Шеннону) к объему данных:

$$Y = \frac{1}{V_d}$$

Достаточность (полнота) означает, что она содержит минимальный, но достаточный для принятия правильного решения состав (набор показателей). Понятие полноты информации связано с ее смысловым содержанием (семантикой) и прагматикой. Как неполная, т.е. недостаточная для принятия правильного решения, так и избыточная информация снижает эффективность принимаемых пользователем решений.

Доступность восприятию обеспечивается выполнением соответствующих процедур ее получения и преобразования. Например, в информационной системе информация преобразовывается к доступной и удобной для восприятия пользователя форме (в частности, и путем согласования ее семантической формы с тезаурусом пользователя).

Актуальность информации определяется степенью сохранения ценности информации для управления в момент ее использования, зависит от динамики изменения ее характеристик и от интервала времени, прошедшего с момента возникновения данной информации.

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

Точность информации определяется степенью близости получаемой информации к

реальному состоянию объекта, процесса, явления и т.п. Для информации, отображаемой цифровым кодом, известны четыре классификационных понятия точности:

- формальная точность, измеряемая значением единицы младшего разряда числа;
- реальная точность, определяемая значением единицы последнего разряда числа, верность которого гарантируется;
- максимальная точность, которую можно получить в конкретных условиях функционирования системы;
- необходимая точность, определяемая функциональным назначением показателя.

Достоверность информации определяется ее свойством отражать реально существующие объекты с необходимой точностью. Измеряется достоверность информации доверительной вероятностью необходимой точности, т.е. вероятностью того, что отображаемое информацией значение параметра отличается от истинного значения этого параметра в пределах необходимой точности.

Устойчивость информации отражает ее способность реагировать на изменения исходных данных без нарушения необходимой точности. Устойчивость информации, как и репрезентативность, обусловлена выбранной методикой ее отбора и формирования.

Репрезентативность, содержательность, достаточность, доступность, устойчивость определяются на методическом уровне разработки информационных систем.

Актуальность, своевременность, точность и достоверность обуславливаются на методическом уровне, однако на их величину существенно влияет и характер функционирования системы (надежность).

Параметры актуальности и точности жестко связаны соответственно с параметрами своевременности и достоверности.