

ИЗМЕРЕНИЕ И ПРЕДСТАВЛЕНИЕ ИНФОРМАЦИИ



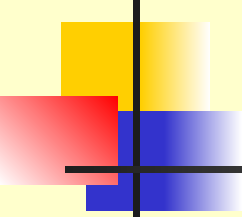
- 1.ИНФОРМАЦИЯ И ЕЕ СВОЙСТВА**
- 2.МЕРЫ ИНФОРМАЦИИ**
- 3.КАЧЕСТВО ИНФОРМАЦИИ**

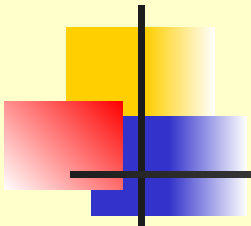


1. ИНФОРМАЦИЯ И ЕЕ СВОЙСТВА

Существование области и предмета информатики немислимо без ее основного ресурса— информации. Информацию как один из основных стратегических ресурсов общества, необходимо уметь оценивать с качественной и с количественной стороны.

Существуют проблемы из-за нематериальной природы этого ресурса и субъективности восприятия конкретной информации.

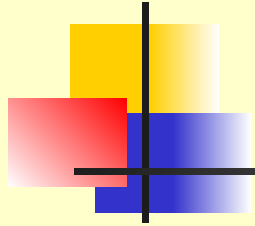
- 
-
- Термин *информация* от латинского *informatio*, означает разъяснение, осведомление, изложение. С позиции материалистической философии информация есть отражение реального мира с помощью сведений (сообщений).
 - Сообщение — это форма представления информации в виде речи, текста, изображения, цифровых данных, графиков, таблиц и т.п.



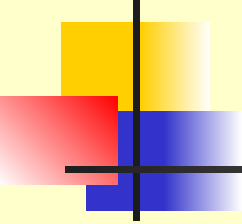
Информатика рассматривает информацию как концептуально связанные между собой сведения, данные, понятия, изменяющие наши представления о явлении или объекте окружающего мира.

Наряду с информацией в информатике часто употребляется понятие *данные*. Покажем, в чем их отличие.

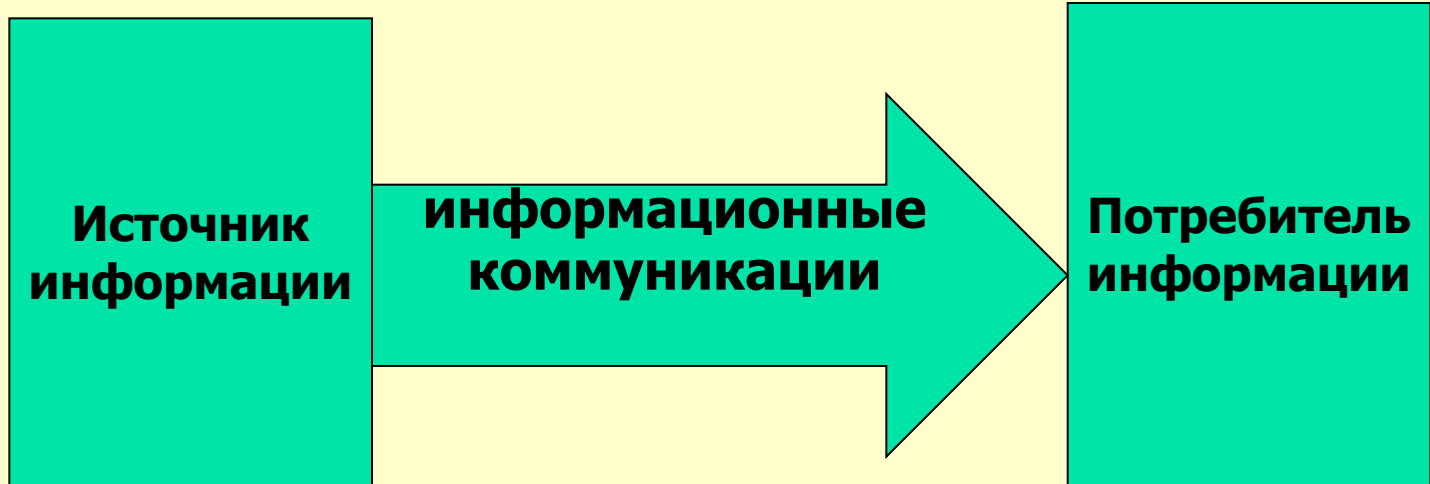
Данные могут рассматриваться как признаки или записанные наблюдения, которые по каким-то причинам не используются, а только хранятся. В том случае, если появляется возможность использовать эти данные для уменьшения неопределенности о чем-либо, данные превращаются в информацию.

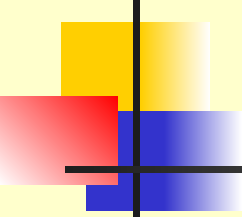


При работе с информацией всегда имеется ее источник и потребитель (получатель). Пути и процессы, обеспечивающие передачу сообщений от источника информации к ее потребителю, называются информационными коммуникациями.



Источник информации

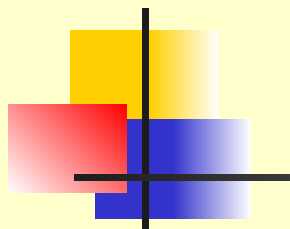




Для потребителя информации очень важной характеристикой является ее адекватность.

Адекватность информации — это определенный уровень соответствия создаваемого с помощью полученной информации образа реальному объекту, процессу, явлению и т.п.

От степени адекватности информации реальному состоянию объекта или процесса зависит правильность принятия решений человеком.

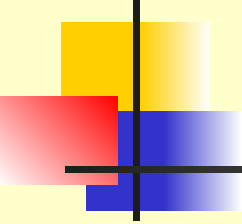


Формы адекватности информации:

семантическая

синтаксическая

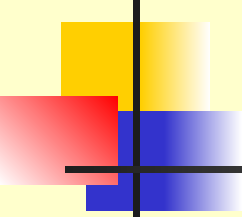
прагматическая



Синтаксическая адекватность. Отображает формально-структурные характеристики информации и не затрагивает ее смыслового содержания.

На синтаксическом уровне учитываются тип носителя и способ представления информации, скорость передачи и обработки, размеры кодов представления информации, надежность и точность преобразования этих кодов и т.п.

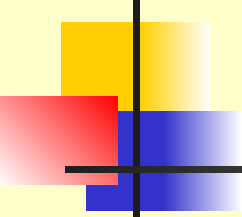
Информацию, рассматриваемую только с синтаксических позиций, называют данными, так как при этом не имеет значения смысловая сторона.



Семантическая (смысловая) адекватность. Эта форма определяет степень соответствия образа объекта и самого объекта. Предполагает учет смыслового содержания информации.

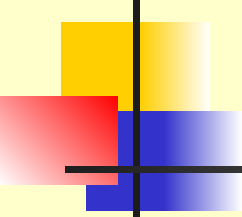
На этом уровне анализируются те сведения, которые отражает информация, рассматриваются смысловые связи.

В информатике устанавливаются смысловые связи между кодами представления информации. Эта форма служит для формирования понятий и представлений, выявления смысла, содержания информации и ее обобщения.



Прагматическая (потребительская) адекватность.
Отражает отношение информации и ее потребителя, соответствие информации цели управления, которая на ее основе реализуется.

Проявляются прагматические свойства информации при наличии единства информации (объекта), пользователя и цели управления.



Прагматический аспект связан с ценностью, полезностью использования информации при выработке потребителем решения для достижения своей цели. С этой точки зрения анализируются потребительские свойства информации. Эта форма адекватности непосредственно связана с практическим использованием информации, с соответствием ее целевой функции деятельности системы.



2. МЕРЫ ИНФОРМАЦИИ

Классификация мер информации.

Для измерения информации вводятся два параметра: количество информации I и объем данных V_d .

Эти параметры имеют разные выражения и интерпретацию в зависимости от рассматриваемой формы адекватности. Каждой форме адекватности соответствует своя мера количества информации и объема данных (см.рис.).

Меры информации

Синтаксическая
мера

Объем данных V_D

Количество информации
 $I_\beta(\alpha) = H(\beta) - H(\alpha)$,
где $H(\alpha)$ — энтропия

Семантическая
мера

Количество информации
 $I_c = C V_D$,
где C — коэффициент
содержательности

Прагматическая
мера



Синтаксическая мера информации

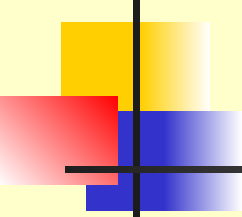
Эта мера количества информации оперирует с обезличенной информацией, не выражающей смыслового отношения к объекту.

Объем данных I_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 (двоичный разряд);

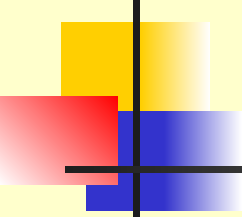
В ЭВМ наряду с минимальной единицей измерения данных "бит" используется укрупненная единица измерения "байт", равная 8 бит.

в десятичной системе счисления единица измерения — дит (десятичный разряд).



Количество информации I на синтаксическом уровне невозможно определить без рассмотрения понятия неопределенности состояния системы (энтропии системы). Рассмотрим это понятие.

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

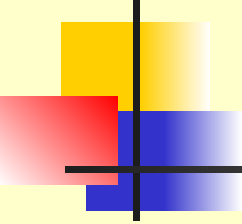


После получения некоторого сообщения β получатель приобрел некоторую дополнительную информацию $I_{\beta}(\alpha)$, уменьшившую его априорную неосведомленность так, что апостериорная (после получения сообщения β) неопределенность состояния системы стала $H_{\beta}(\alpha)$.

Тогда количество информации $I_{\beta}(\alpha)$, о системе, полученной в сообщении β , 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 неполное знание заменится полным знанием и количество информации $I_{\beta}(\alpha) = H(\alpha)$. Иными словами, *энтропия системы $H(\alpha)$ может рассматриваться как мера недостающей информации.*

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

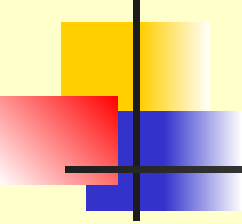
$$H(a) = - \sum_{i=1}^N P_i \log P_i$$

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

Для случая, когда все состояния системы равновероятны, т.е. их вероятности равны

$P_i = \frac{1}{N}$ ее энтропия определяется соотношением

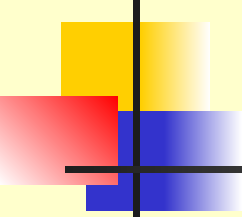
$$H(a) = - \sum_{i=1}^N \frac{1}{N} \log \frac{1}{N}$$



Информация кодируется числовыми кодами в той или иной системе счисления. Одно и то же количество разрядов в разных системах счисления может передать разное число состояний отображаемого объекта, что можно представить в виде соотношения

$$N=m^n,$$

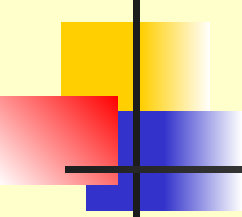
где N — число всевозможных отображаемых состояний;
 m — основание системы счисления (разнообразие символов, применяемых в алфавите);
 n — число разрядов (символов) в сообщении



Пример. По каналу связи передается n -разрядное сообщение, использующее m различных символов. Так как количество всевозможных кодовых комбинаций будет $N=m^n$, то при равновероятности появления любой из них количество информации, приобретенной абонентом в результате получения сообщения, будет

$I = \log N = n \log m$ — формула Хартли.

Если в качестве основания логарифма принять m , то $I = n$. В данном случае количество информации (при условии полного априорного незнания абонентом содержания сообщения) будет равно объему данных $I = V_d$, полученных по каналу связи. Для неравновероятных состояний системы всегда $I < V_d = n$.



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.

$$Y = I / V_d$$

причем $0 < Y < 1$.

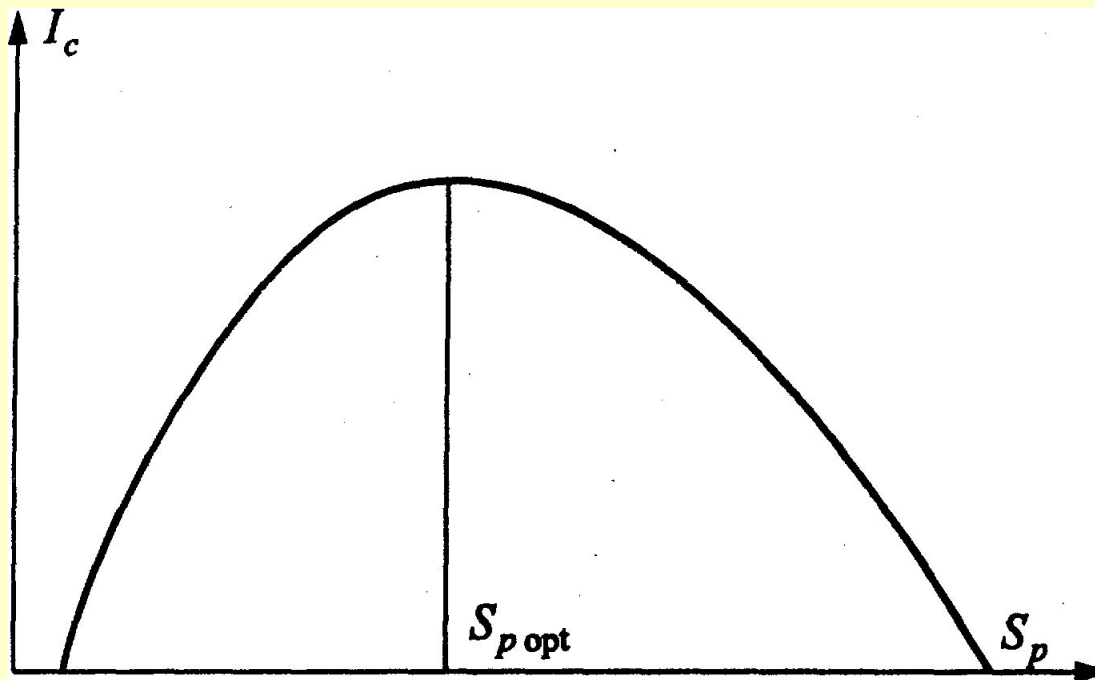
С увеличением Y уменьшаются объемы работы по преобразованию информации (данных) в системе. Поэтому стремятся к повышению информативности, для чего разрабатываются специальные методы оптимального кодирования информации.

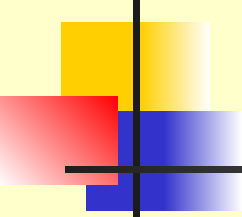


Семантическая мера информации

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 пользователя**.

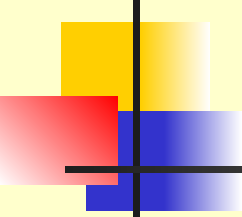
Зависимость количества семантической информации, воспринимаемой потребителем, от его тезауруса $I_c=f(S_p)$





Максимальное количество семантической информации I_s 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 тоже сообщение может иметь смысловое содержание для компетентного пользователя и быть бессмысленным (семантический шум) для пользователя некомпетентного.



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

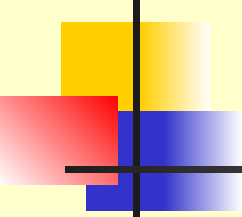
Относительной мерой количества семантической информации может служить коэффициент содержательности C , который определяется как отношение количества семантической информации к ее объему:

$$C = I_c/V_d$$



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 самых единицах (или близких к ним), в которых измеряется целевая функция.



Пример: В экономической системе прагматические свойства (ценность) информации можно определить приростом экономического эффекта функционирования, достигнутым благодаря использованию этой информации для управления системой:

$$\text{In}\beta(\gamma) = \Pi(\gamma/\beta) - \Pi(\gamma),$$

где $\text{In}\beta(\gamma)$ — ценность информационного сообщения β для системы управления γ ,

$\Pi(\gamma)$ — априорный ожидаемый экономический эффект функционирования системы управления γ ,

$\Pi(\gamma/\beta)$ — ожидаемый эффект функционирования системы γ при условии, что для управления будет использована информация, содержащаяся в сообщении β .



3. КАЧЕСТВО ИНФОРМАЦИИ

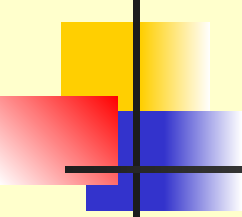
Потребительские показатели качества:

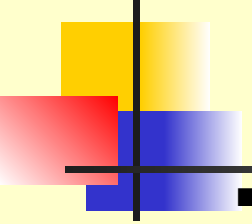
- 1. Репрезентативность**
- 2. содержательность,**
- 3. достаточность,**
- 4. доступность,**
- 5. актуальность,**
- 6. своевременность,**
- 7. точность,**
- 8. достоверность,**
- 9. устойчивость**

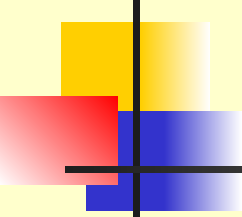


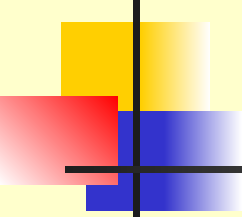
Показатели качества информации:

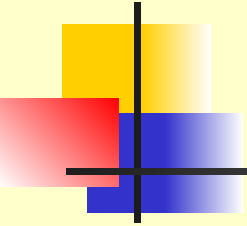
- **Репрезентативность**-правильность отбора и формирования информации для адекватного отображения свойств объекта. (правильность концепции, на базе которой сформулировано исходное понятие; обоснованность отбора существенных признаков и связей отображаемого явления.)

- 
-
- **Содержательность** – семантическая емкость - отношение количества семантической информации в сообщении к объему обрабатываемых данных, т.е. $C = I_s/V_d$. Можно использовать и коэффициент информативности, характеризующийся отношением количества синтаксической информации (по Шеннону) к объему данных
 $Y = I/V_d$.

- 
- **Достаточность** – минимальный, но достаточный объем для принятия решения.
 - **Доступность** – выполнение процедур получения и преобразования информации. Это достигается путем согласования ее семантической формы с тезаурусом пользователя.
 - **Актуальность** – ценность информации в момент ее использования, зависит от динамики изменения ее характеристик и от интервала времени, прошедшего с момента возникновения данной информации.
 - **Своевременность** – поступление информации не позже назначенного времени, согласованного с временем решения поставленной задачи.

- 
-
- **Точность** – степень близости информации реальному состоянию объекта. 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** – необходимая точность отражения существующих объектов. 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** – способность реагировать на изменение исходных данных



Такие параметры качества информации, как репрезентативность, содержательность, достаточность, доступность, устойчивость, целиком определяются на методическом уровне разработки информационных систем. Параметры актуальности, своевременности, точности и достоверности обуславливаются в большей степени также на методическом уровне, однако на их величину существенно влияет надежность системы.