

КОЛИЧЕСТВО И КАЧЕСТВО ИНФОРМАЦИИ

1. Уровни проблем передачи информации
2. Меры информации
 - 2.1. синтаксического уровня
 - 2.2. семантического уровня
 - 2.3. прагматического уровня
3. Качество информации (с/р)

ОСНОВНЫЕ ВОПРОСЫ ЛЕКЦИИ

1. Сообщение
2. Уровни изучения сообщений. Синтаксический уровень. Семантический уровень. Прагматический уровень
3. Меры информации
4. Объемный подход
5. Энтропийный подход
6. Виды неопределенности
7. Количество информации
8. Формула Хартли
9. Формула Шеннона
10. Тезаурус
11. Зависимость количества информации от тезауруса получателя
12. Коэффициент содержательности
13. Исследования Харкевича А.А.
14. Качество информации. Составляющие качества информации.

1. УРОВНИ ПРОБЛЕМ ПЕРЕДАЧИ ИНФОРМАЦИИ

При реализации информационных процессов всегда происходит перенос информации в пространстве и времени от источника к приемнику. При этом для передачи информации используют различные знаки или символы, позволяющие выразить ее в некоторой форме, называемой сообщением.

Сообщение – форма представления информации в виде совокупности знаков (символов), используемая для передачи.

СООБЩЕНИЕ КАК СОВОКУПНОСТЬ ЗНАКОВ МОЖЕТ ИЗУЧАТЬСЯ НА ТРЕХ УРОВНЯХ:

- **Синтаксическом**, где рассматриваются внутренние свойства сообщений, т.е. отношения между знаками, отражающие структуру данной знаковой системы;
- **Семантическом**, где анализируются отношения между знаками и обозначаемыми ими предметами, действиями, качествами, т.е. смысловое содержание сообщения, его отношение к источнику информации;
- **Прагматическом**, где рассматриваются отношения между сообщением и получателем, т.е. потребительское содержание сообщения, его отношение к получателю.

СИНТАКСИЧЕСКИЙ УРОВЕНЬ

- Проблемы *синтаксического уровня* касаются создания теоретических основ построения ИС и повышения эффективности их использования.
- Это чисто технические проблемы совершенствования методов передачи сообщений и их материальных носителей – сигналов.
- На этом уровне информацию называют **данными**, так как смысловая сторона при этом не имеет значения.

Современная теория информации исследует в основном проблемы этого уровня. Она опирается на понятие «количество информации», являющееся мерой частоты употребления знаков.

В связи с этим часто говорят, что современная теория информации находится на синтаксическом уровне.

СЕМАНТИЧЕСКИЙ УРОВЕНЬ

- Проблемы **семантического уровня** связаны с формализацией и учетом смысла передаваемой информации, определения степени соответствия образа объекта и самого объекта.
- На данном уровне анализируются те сведения, которые отражает информация, рассматриваются смысловые связи, выявляется содержание информации, осуществляется ее обобщение.
- Проблемы этого уровня чрезвычайно сложны, так как смысловое содержание информации больше зависит от получателя, чем от семантики сообщения, представленного на каком-либо языке.

ПРАГМАТИЧЕСКИЙ УРОВЕНЬ

- На **прагматическом уровне** интересуют последствия от получения и использования данной информации потребителем.
- Проблемы этого уровня связаны с определением ценности и полезности использования информации при выработке потребителем решения для достижения своей цели.
- Основная сложность здесь состоит в том, что полезность информации может быть совершенно различной для различных получателей, кроме того она зависит от ряда факторов, таких например как своевременность ее доставки и использования.

2. МЕРЫ ИНФОРМАЦИИ

2.1. МЕРЫ ИНФОРМАЦИИ СИНТАКСИЧЕСКОГО УРОВНЯ

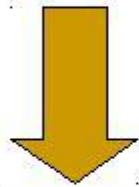
Количественная оценка информации этого уровня не связана с содержательной стороной информации.

Для измерения информации на синтаксическом уровне вводятся два параметра:

- *Объем информации (данных)* - V_d объемный подход
- *Количество информации* – I энтропийный подход.

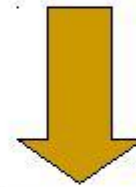
Синтаксическая мера информации

На синтаксическом уровне для измерения информации вводятся два параметра



V_D

**Объем информации
(объемный подход)**



I

**Количество информации
(вероятностный подход)**

ОБЪЕМ ИНФОРМАЦИИ- V_D

При реализации информационных процессов информация передается в виде сообщения, представляющего собой совокупность символов какого-либо алфавита.

При этом каждый новый символ в сообщении увеличивает количество информации, представленной последовательностью символов данного алфавита.

Если теперь количество информации, содержащейся в сообщении из одного символа, принять за единицу, то объем информации (данных) в любом другом сообщении будет равен количеству символов (разрядов) в этом сообщении.

Так как одна и та же информация может быть представлена разными способами (с использованием разных алфавитов), то и единица измерения информации (данных) соответственно будет меняться.

ОБЪЕМ ИНФОРМАЦИИ- V_D

Так в десятичной системе счисления один разряд имеет вес равный 10 и соответственно единицей измерения информации будет дит (десятичный разряд).

В этом случае сообщение в виде n -разрядного числа имеет объем данных $V_D = n$ дит.

- Например, в $(CC)_{10}$ четырехразрядное число **2019** имеет объем данных 4 дит.

В двоичной системе счисления один разряд имеет вес, равный 2, и соответственной единицей измерения информации будет бит (binary digit- двоичный разряд).

В этом случае сообщение в виде n -разрядного числа имеет объем данных $V_D = n$ бит.

- Например, восьмиразрядный двоичный код **11001011** имеет объем данных 8 бит.

ПРИ РАБОТЕ С БОЛЬШИМИ ОБЪЕМАМИ ИНФОРМАЦИИ ПРИМЕНЯЮТ БОЛЕЕ КРУПНЫЕ ЕДИНИЦЫ ИЗМЕРЕНИЯ, ТАКИЕ КАК, **КИЛОБАЙТ, МЕГАБАЙТ, ГИГАБАЙТ, ТЕРАБАЙТ.**

- $1 \text{ кбайт} = 1024 = 2^{10} \text{ байт}$
- $1 \text{ Мбайт} = 1024 \text{ кбайт} = 2^{20} \text{ байт} = 1\,048\,576 \text{ байт}$
- $1 \text{ Гбайт} = 1024 \text{ Мбайт} = 2^{30} \text{ байт} = 1\,073\,741\,824 \text{ байт}$
- $1 \text{ Тбайт} = 1024 \text{ Гбайт} = 2^{40} \text{ байт} = 1099\,511\,627\,776 \text{ байт}$

КОЛИЧЕСТВО ИНФОРМАЦИИ – I (ЭНТРОПИЙНЫЙ ПОДХОД)

Этот подход основан на том, что факт получения информации всегда связан с уменьшением разнообразия или неопределенности (**энтропии**) системы.



Количество информации I

- Количество информации о системе, полученное в сообщении, измеряется уменьшением неопределенности о состоянии системы.

- Мету неопределенности в теории информации называют **“энтропия”**.

- Неопределенность не отделима от понятия вероятности.

Неопределенность может иметь разное происхождение.

Один из ее видов – **неизвестность** – рассматривается теорией познания и философией; «*Есть ли жизнь на других планетах?*» или «*Существуют ли другие цивилизации?*» и т.п.

Другой вид неопределенности – **расплывчатость, размытость**, – например, «*Сколько надо взять песчинок, чтобы образовать небольшую кучу?*»

Размытость – характерное свойство языка: «*в комнату (какую?) вошел высокий (какого роста?) молодой (какого конкретно возраста?) человек (кто он?) и т.п.*

Третий вид неопределенности – **случайность**. В ее основе лежат статистические закономерности, устанавливаемые теорией вероятности.

Отличительная особенность статистической неопределенности заключается в том, что для нее можно установить количественную меру, о которой пойдет речь далее.

Рассмотрим несколько простейших ситуаций, которые будем именовать *опытами* А, В и С.

-
- **Опыт А** заключается в бросании монеты. В этом опыте возможны два *исхода* ($n=2$): “орел или решка”. Очевидно, вероятность каждого исхода ($p_i=1/2$).
 - **Опыт В** – бросание игральной шестигранной кости. В этом опыте возможны уже шесть исходов ($n=6$). Вероятность каждого исхода $p_i=1/6$.
 - **Опыт С** предполагает одновременное бросание двух костей. Для этого опыта $n=36$ и $p_i=1/36$.

Оценка неопределённости результатов опытов есть оценка трудности предугадывания исхода опыта. Интуитивно ясно, что из всех описанных ситуаций опыт **С** имеет максимальную неопределённость, поскольку число исходов здесь самое большое и заранее предвидеть исход этого опыта труднее всего.

Таким образом, при энтропийном подходе под информацией понимается количественная величина исчезнувшей в ходе какого-либо процесса (испытания, измерения) неопределенности.

При этом в качестве меры неопределенности вводится **энтропия (H)**, а количество информации равно:

КОЛИЧЕСТВО ИНФОРМАЦИИ:

$$I = H_{apr} - H_{aps}, \text{ где}$$

H_{apr} - априорная энтропия о состоянии исследуемой системы
или процесса

H_{aps} – апостериорная энтропия

- *Апостериори* (из последующего) – происходящее из опыта.
- *Априори* (из предшествующего) – понятие характеризующее знание, предшествующее опыту (испытанию).

В случае, когда в ходе испытания имевшаяся неопределенность снята (получен конкретный результат, т.е. $H_{aps} = 0$), количество полученной информации совпадает с первоначальной энтропией

$$I = H_{apr}$$

Формула для вычисления меры неопределенности была предложена американским ученым **Р.Хартли в 1928г.**

$$H = \log N$$

в информатике в качестве меры неопределенности обычно используют формулу:

$$H = \log_2 N$$

При этом единица неопределенности называется двоичной единицей или битом и представляет собой неопределенность выбора из двух равновероятных событий.

ФОРМУЛУ ХАРТЛИ МОЖНО СЛОВЕСНО ВЫРАЗИТЬ ИНАЧЕ:

количество информации равно степени, в которую необходимо возвести число 2, чтобы получить число равновероятных вариантов выбора, т.е.

$$2^I = 16, \text{ где } I = 4 \text{ (бита)}$$

Формула Хартли

$$2^i = N$$

N – число равновероятных событий

i – количество информации в сообщении

$$i = \log_2 N$$

Предложенная мера позволяет решать определенные практические задачи, когда все возможные состояния источника информации имеют одинаковую вероятность.

Американский ученый К.Шеннон обобщил понятие меры неопределенности выбора **H** на случай, когда **H** зависит не только от числа состояний, то и от вероятностей этих состояний.

Эту меру называют *энтропией дискретного источника информации*

ФОРМУЛА ШЕННОНА

$$I = -\sum_{i=1}^n p(i) * \log_i(p(i))$$

n - количество возможных событий

I - количество информации

p_i - вероятность отдельных событий

При равновероятных выборах все $p_i = 1/N$ и формула Шеннона преобразуется в формулу Хартли:

$$H = - \sum_{i=1}^N p_i \log_2 p_i = -N \frac{1}{N} \log_2 \frac{1}{N} = -N \frac{1}{N} (\log_2 1 - \log_2 N) = \log_2 N$$

2. МЕРЫ ИНФОРМАЦИИ

2.2. МЕРЫ ИНФОРМАЦИИ СЕМАНТИЧЕСКОГО УРОВНЯ

Под семантической информацией понимают смысловое содержание, извлекаемое получателем из сообщения.

Общепринятой количественной меры для его измерения в настоящее время не существует.

Среди известных подходов к семантической оценке информации наибольшее распространение получил **тезаурусный подход**, когда содержащийся в принятом сообщении смысл оценивается путем соотнесения с тезаурусом получателя(приемника), его способностью понимать и принимать поступившее сообщение.

Для понимания и использования информации ее получатель должен обладать определенным запасом знаний. Полное незнание предмета не позволяет извлечь полезной информации из принятого сообщения об этом предмете.

Чем больше знаний о предмете, тем больше и количество полезной информации, извлекаемой из сообщения.

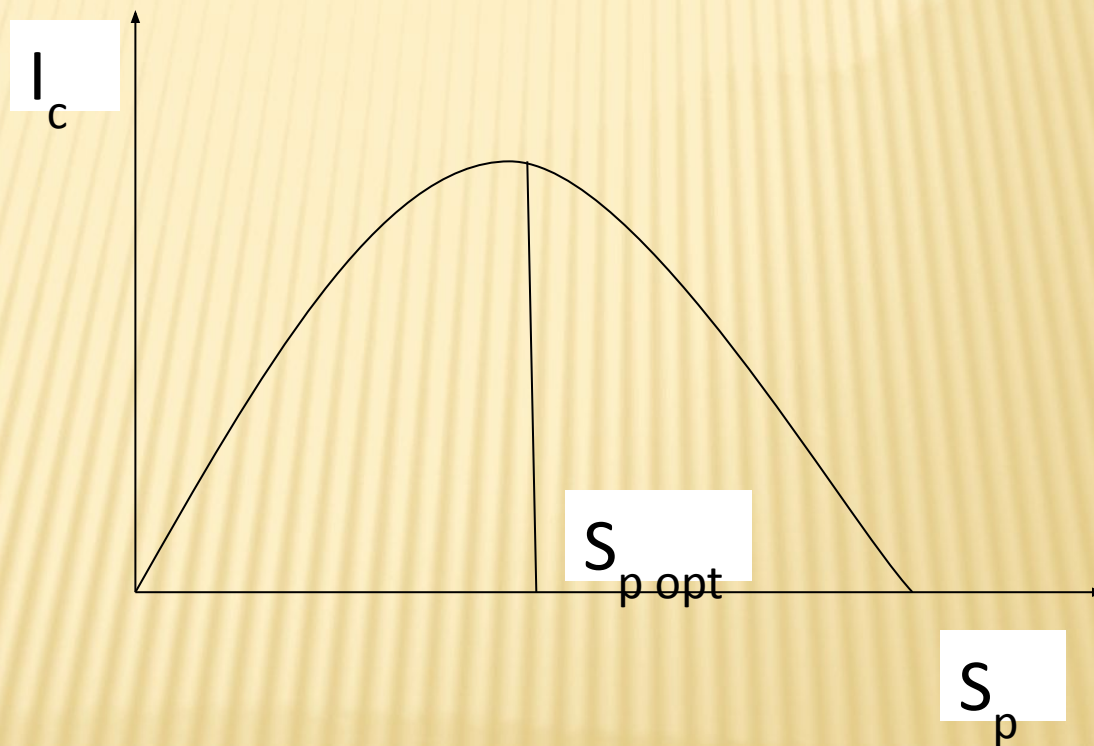
Семантическая мера информации

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*.

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

Тезаурус – в широком смысле – совокупность сведений, которыми располагает пользователь (система), в более узком – запас знаний в виде словаря, отражающего семантические связи между словами и другими смысловыми элементами данного языка.

ЗАВИСИМОСТЬ КОЛИЧЕСТВА СЕМАНТИЧЕСКОЙ ИНФОРМАЦИИ,
ВОСПРИНИМАЕМОЙ ПОТРЕБИТЕЛЕМ, ОТ ЕГО ТЕЗАУРУСА.



Количество семантической информации I_c , извлекаемой получателем из поступающих сообщений и включаемой им в дальнейший свой тезаурус, зависит от степени подготовленности (полноты) тезауруса S_p для восприятия такой информации.

При недостаточно развитом тезаурусе получатель извлекает из сообщения малое количество семантической информации, а в предельном случае, когда $S_p = 0$ (отсутствие у получателя исходных знаний) он вообще не воспринимает (не может понять) поступающее сообщение.

При достаточно развитом тезаурусе получателя также воспринимается малое количество семантической информации (тезаурус незначительно изменяется под воздействием данного сообщения), при этом в предельном случае, когда $S_p \rightarrow \infty$ получатель все знает, и извлекаемая из сообщения информация ему не нужна.

Максимальное количество семантической информации, содержащейся в сообщении, является величиной относительной.

Однако одно и то же сообщение может иметь смысловое содержание для получателя с более развитым тезаурусом (компетентного пользователя) и быть бессмысленным для получателя с недостаточно развитым тезаурусом (некомпетентного пользователя)

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

$$C = I_c / V_d$$

2. МЕРЫ ИНФОРМАЦИИ

2.3. МЕРЫ ИНФОРМАЦИИ ПРАГМАТИЧЕСКОГО УРОВНЯ

Эта мера определяет полезность информации (ценность) для достижения пользователем заданной цели.

Она также величина относительная, обусловленная особенностями использования информации в той или иной системе.

Одним из первых российских ученых к проблеме оценки информации прагматического уровня обратился Харкевич А.А., который предложил принять за меру ценности информации количество информации, необходимое для достижения поставленной цели.

Если до получения информации вероятность достижения цели равнялась p_0 , а после ее получения – p_1 , то ценность информации определяется как логарифм отношения p_1/p_0 :

$$I = \log_2 p_1 - \log_2 p_0 = \log_2 \frac{p_1}{p_0}$$

Максимальной ценностью обладает то количество информации, которое уменьшает потери до нуля при достижении поставленной цели.

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.

3. КАЧЕСТВО ИНФОРМАЦИИ. СОСТАВЛЯЮЩИЕ КАЧЕСТВА ИНФОРМАЦИИ.

Качество информации – совокупность свойств информации, характеризующих степень ее соответствия потребностям (целям, ценностям) пользователей (средств автоматизации, персонала и др.).

1. **Репрезентативность**
2. **Содержательность**
3. **Достаточность**
4. **Доступность**
5. **Устойчивость**
6. **Своевременность**
7. **Точность**
8. **Достоверность**
9. **Актуальность**