

...Мы смотрим на жизнь с разных сторон, потому что когда вы смотрите на жизнь с разных сторон, вы видите все так, как бывает на самом деле.

(Из пролога к фильму «Айболит-66»)

Информация и данные.

Информатика, 1 курс.
Лекция 3.

Санкт-Петербургский государственный Политехнический университет
Кафедра «Компьютерные интеллектуальные технологии»

В ЭТОЙ ЛЕКЦИИ:

- Формы представления информации
- Свойства информации
- Различные подходы к изучению информации
- Способы измерения информации

Формы представления информации

(из учебного пособия для 5 класса)



Словесная информация – информация, в которой информационной единицей является слово – символ или набор символов, имеющий смысл в данном языке.

Графическая информация – информация, ассоциирующаяся с некоторым зрительным образом.

Числовая информация – информация, отражающая количественные взаимоотношения.

Звуковая информация – информация, представляющая собой звуковой сигнал, ассоциирующаяся со звуковым образом

Это – формы представления информации с точки зрения восприятия человеком. В процессе передачи, хранения информации она принимает иные формы.



Сделайте паузу в листании лекции
и постарайтесь осмыслить плюсы и минусы
каждой формы представления информации

...

Высокая скорость восприятия

Форма представления информации	Достоинства	Недостатки
Словесная		
Графическая		
Числовая		
Звуковая		

Чёткие критерии
сравнения

Традиционна в
человеческом
общении

Необходимость
пояснений в
другой форме

Трудоёмкость
представления

Трудно формализуема

Ограниченный набор
воспринимаемых
сигналов

Лёгкость
машинной
обработки

...

...

Данные

Данные – информация, представленная в конкретной форме, находящаяся на каком-либо носителе или передаваемая по каналу связи.



Свойства информации

Объективность –
независимость от
субъекта,
действующего с
информацией

На улице холодно

На улице +3

Достоверность –
уверенность в
истинности
информации

У меня сегодня двоек нет!

Да, это МарьяИванна. Действительно, у
Вовы сегодня нет двоек.

Полнота –
достаточность
информации для
принятия решения

Задуманное натуральное число больше 3
и меньше 5

Задуманное натуральное число больше 8
и меньше 11.

Свойства информации

Релевантность –
значимость для
получателя
информации

У сикорах семейства
голопузиковых голубые глаза

Завтра не будет первой пары.

Актуальность –
своевременность
получения
информации

Завтра – последняя попытка сдачи
зачёта по матанализу

Вчера была последняя попытка
сдачи зачёта по матанализу

Ясность – соответствие
формы представления
возможностям
получателя

Пошёл вон!

Мун тотыш!

Свойства
взаимосвязаны, для их
названий есть много
синонимов.



Различные подходы к изучению информации

- **аксиологический** (с точки зрения содержания);
- **прагматический** (с точки зрения необходимости и достаточности информации для принятия решения);
- **семантический** (связь информации с ранее накопленной системой знаний);
- **структурный** (представление, организация, способ хранения);
- **статистический** (количественные меры).

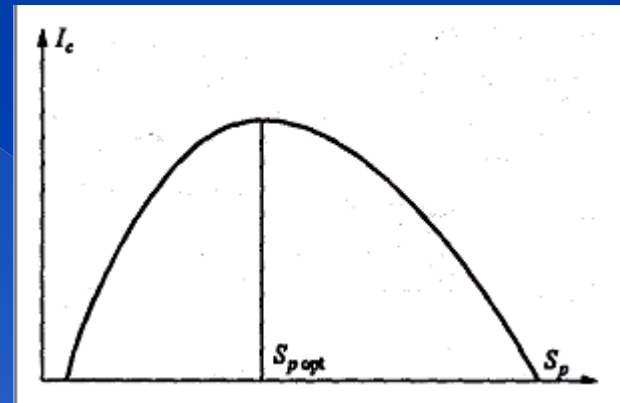
Семантический подход к изучению информации

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

S – смысловое содержание информации

S_p – тезаурус пользователя

I_c - количество семантической информации, воспринимаемой пользователем и включаемой им в дальнейшем в свой тезаурус.



Статистический подход к изучению информации

N – количество возможных исходов какого-то события (количество возможных значений чего-либо).

H – **энтропия** – величина неопределённости, связанной с исходом события.

$H=f(N)$ – энтропия зависит от числа возможных исходов.

Пусть получена некоторая информация об исходе события.

H_1 – неопределённость до получения информации.

H_2 – неопределённость после получения информации.

$H_1 \geq H_2$ (с получением информации неопределённость не возрастает, она может остаться неизменной, если информация бесполезна).

I – количество информации в сообщении.

$$I = H_1 - H_2$$

Осталось вычислить неопределённость.

$$H = \log_2 N$$

– формула Хартли

Пример

Бросили кубик с 6 гранями.
Имеются три высказывания о
результате броска:

- 1) выпало чётное число;
- 2) выпало число, меньшее 3;
- 3) выпало не число 5.

Определить количество
информации в каждом
высказывании.

До получения информации:

$$N=6. \quad H = \log_2 6 \approx 2.59$$

После получения высказывания 1:

$$N=3 \quad H_1 = \log_2 3 \approx 1.59 \quad I_1 = 2.59 - 1.59 = 1$$

После получения высказывания 2:

$$N=2 \quad H_2 = \log_2 2 = 1 \quad I_1 = 2.59 - 1 = 1.59$$

После получения высказывания 3:

$$N=5 \quad H_3 = \log_2 5 \approx 2.33 \quad I_1 = 2.59 - 2.33 = 0.26$$

Формула Хартли верна для равновероятных исходов. В противном случае действует **формула Шеннона**:

$$H = \sum_{i=1}^N P_i \cdot \log_2 \left(\frac{1}{P_i} \right).$$

Пример

Когда экзамен принимает профессор А, все оценки (от 2 до 5) равновероятны. А вот профессор Б на любом экзамене ставит 25% пятёрок, 50% четвёрок и 25% троек. Подсчитать количество информации в сообщениях об оценке, полученной студентом на экзамене у профессора А и профессора Б.

При равновероятных исходах (профессор А) – формула Хартли:

$$H = \log_2 4 \approx 2 \text{ бита}$$

Если исходы не равновероятны (профессор Б) – формула Шеннона:

$$H = \frac{1}{4} \log_2 4 + \frac{1}{2} \log_2 2 + \frac{1}{4} \log_2 4 = 1.5 \text{ бита}$$

Объёмный анализ информации

Сколько разрядов понадобится
для кодирования чего-либо,
имеющего N возможных значений?

$$\log_2 N,$$

округлённый в БОльшую сторону.

Итоги:

- Информация представляется в различных формах в зависимости от её источника и действий, которые с ней предполагается выполнять.
- Для тех, кто работает с информацией важны её свойства, поскольку от них зависит возможности её использования.
- Специалисты разных областей совершенно по-разному подходят к изучению информации.
- Измерение информации основано на понятии неопределённости и на том, насколько информация уменьшает эту неопределённость.
- Важнейшие формулы для измерения информации – формула Хартли (при равновероятных значениях) и формула Шеннона (когда значения имеют разные вероятности)