

Информатика

I семестр: 20ч. лекций, 28 ч. практ. занятий

II семестр: 16ч. лекций, 28 ч. практ. занятий
каф. ИТ-8

доц. Карулина Валентина Алексеевна
МГУПИ, 2013



Наше понятие информатики



Информатика – наука об общих свойствах информации, закономерностях и методах её поиска и получения, хранения, передачи, переработке и распространения, а также науку о способах её использования для решения различных прикладных задач.

Информатика как наука



- **Предметом** Информатики как науки являются общие закономерности управления данными, принципы организации и поиска данных и т.д..
- **Объектом** приложений Информатики стали различные сферы деятельности. Наиболее впечатляют реализации Информационных Технологий.



Происхождение термина информатика



В СССР в 60-е годы – **кибернетика**.



Кибернетика - наука об общих закономерностях в управлении и связи в различных системах: искусственных, биологических, социальных.

1948 г. Н. Винер: "Кибернетика или управление и связь в животном и машине".

В США и англоязычных странах: **computer science** – компьютерная наука.

Термин **информатика** предложен Карлом Штейнбухом в 1957.

Фр. Ф. Дрейфус : Informatique:

информатика = информация + автоматика.

1978г. Международный конгресс по информатике



Понятие Информатики - охватывает области, связанные с разработкой, созданием, использованием, и материально-техническим обслуживанием систем обработки информации, включая машины, оборудование, математическое обеспечение, организационные аспекты, а также комплекс промышленного, коммерческого, административного и социального воздействия.



Структура информатики как научной и прикладной дисциплины

И
Н
Ф
О
Р
М
А
Т
И
К
А



Место информатики в системе наук

Естественные науки

Общественные науки

Информатика

Математические науки

Технические науки

Гуманитарные науки



- **Естественные** науки - имеют дело с объективным сущностями мира, существующими независимо от нашего сознания.
- **Гуманитарные** науки - связаны с вкладом в развитие социальной сферы.
- **Технические** науки - позволяют создать машинные системы.
- **Общественные** науки – социальные
- **Математические** науки – изучают величины, их количественные отношения, а также пространственные формы.
- **Технические** науки - позволяют создать машинные системы

Интуитивное понятие информации



Информация – сведения и данные, отражающие свойства объектов в природных, социальных и технических системах и передаваемые звуковым, графическим (в т.ч. письменным) или иным способом без применения или с применением технических средств.

Интуитивное понятие информации



Информация – сведения об объектах и явлениях, их параметрах, свойствах и состоянии, которые уменьшают степень неопределенности о них или неполноты знаний.

Интуитивное понятие информации



Цель информатики - поиск нового знания.

Информатика - технология обработки накопленного знания и построения нового знания. Информатика изучает методы анализа знания о методах построения нового знания как своего собственного, так и знания других наук.



Знание - форма существования и систематизации результатов познавательной деятельности человека.

Виды знания:

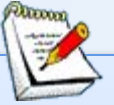
- научное, обыденное (здравый смысл), интуитивное, религиозное и др.
- научное и ненаучное (пара-, анти-, псевдо-,
- личностные,
- «народные знания»

Информация



- **Физическая информация:** третий компонент материального мира (наряду с веществом и энергией), управляющий порядком и хаосом в микромире, в космосе, в термодинамике, статистической физике, биосистемах и т. п. Не было человека на Земле, информация была, не будет его – информация будет. Физическая информация имеет два подвида: термодинамическую и статистическую. Именно на примере физической информации мы попытаемся установить связь между информацией и энтропией (информация, как отрицательная энтропия).

Информация



- **Смысловая (семантическая) информация:** информация, которую воспринимает человек (и которая, к примеру, передаётся средствами массовой информации). Это то, что можно осмыслить, оценить, купить-продать-подарить, накапливать, хранить, охранять, терять; она способна исчезать и появляться, может быть полезной и вредной, истинной и ложной, переходя в дезинформацию.

Информация



- **Техническая информация** (она же кибернетическая или компьютерная) – информация, передаваемая азбукой Морзе, по радио- или телеканалу, информация в компьютерах и прочих технических машинах. К ней относятся вопросы передачи информации по линии связи, вопросы кодирования-декодирования информации, и способы переработки информации компьютерами

Свойства смысловой информации

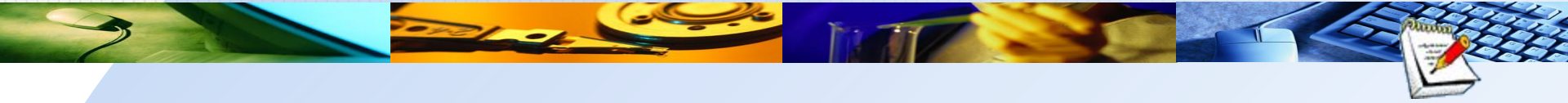


запоминаемость,
преобразуемость,
стираемость,
достоверность,
точность,
ценность,
понятность,
краткость и т. д.

передаваемость,
воспроизводимость,
объективность,
полнота,
полезность,
актуальность,
доступность,

*Пример: «На улице тепло» субъективная информация
«На улице +22 °C» объективная, но с
точностью погрешности измерения*

Хранение информации



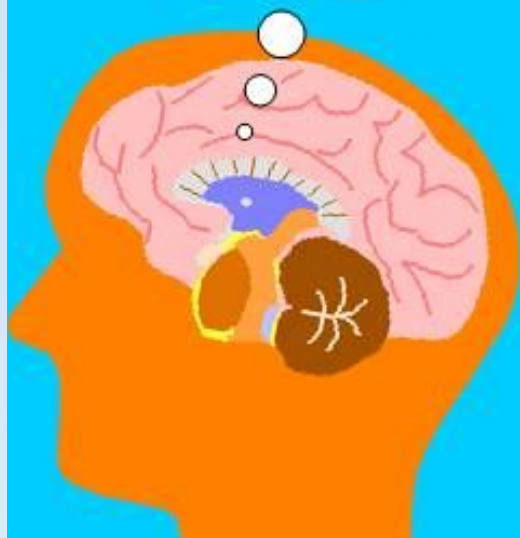
- **Письменность** — графическое изображение символов на камне, глине, папирусе, бумаге. Огромное значение в развитии этого направления имело изобретение книгопечатания.
- Хранение информации осуществляется с помощью её переноса на некоторые материальные носители. Семантическая информация, зафиксированная на материальном носителе, называется **документом**.
- Хранение информации в виде последовательностей двоичных символов. Для реализации этих методов используются разнообразные **запоминающие устройства**.

Хранение информации



Внутренняя
память
человека

$$2 \times 2 = 4$$



Внешняя память
(на внешних носителях)

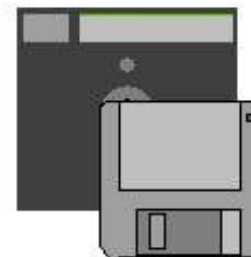
Записные книжки

Справочники

Энциклопедии

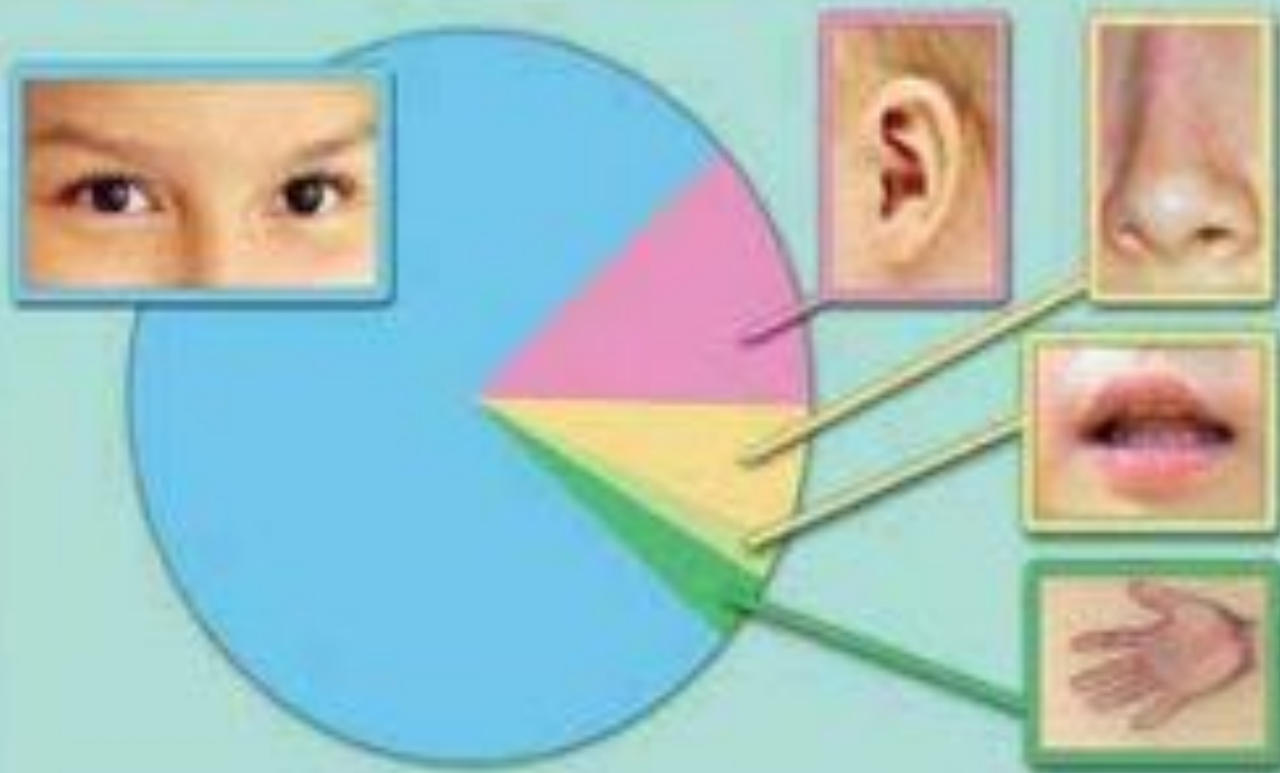
Магнитные
записи

Оптические
записи

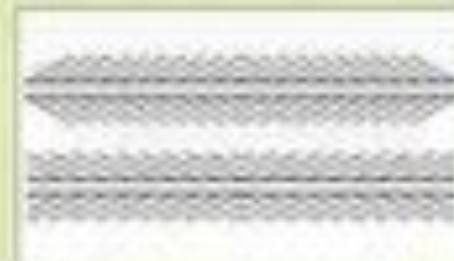


КАК МЫ ВОСПРИНИМАЕМ ИНФОРМАЦИЮ

Информация:



ЗРИТЕЛЬНЫЕ ИЛЛЮЗИИ



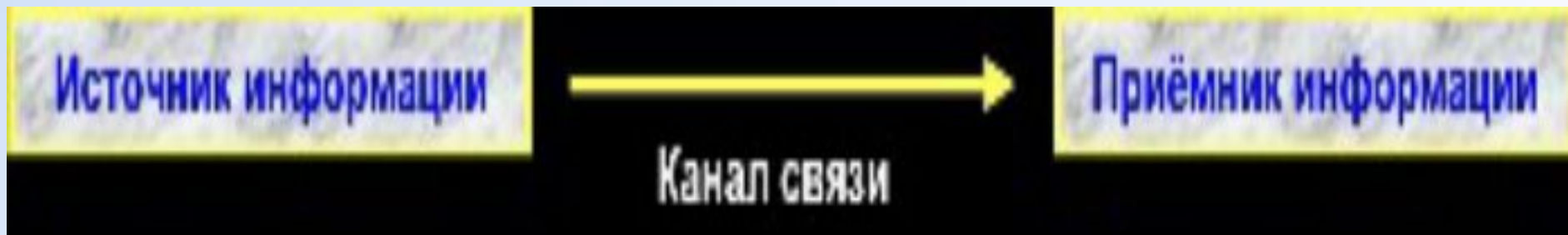
Обработка информации

Информация, предназначенная для обработки, называется данными. Для достижения цели обработки данных должен быть известен порядок действий над информацией, приводящий к заданной цели - алгоритм. Кроме самого алгоритма необходимо также некоторое устройство, реализующее этот алгоритм. В научных теориях такое устройство называется автоматом.



Передача информации

Передачей информации называется процесс её пространственного переноса от источника к получателю. (Речь, письмо и т.д.) Для осуществления передачи информация должна быть некоторым образом оформлена. Для представления информации используются различные **знаковые системы**. Представленная с их помощью семантическая информация о каком-либо объекте, явлении или процессе называется **сообщением**.



Передача информации



Канал связи

Совокупность технических устройств, обеспечивающих передачу от передатчика к приемнику



Кодирующее устройство



Декодирующие устройство



Количество информации

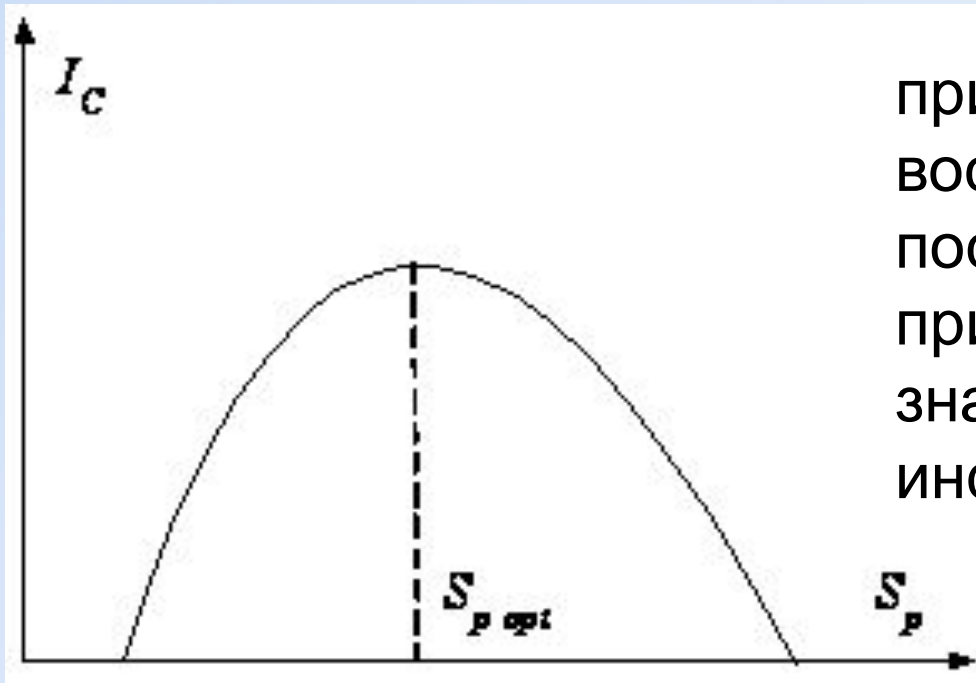
Меры информации

Мера информации	Единицы измерения	Примеры (для компьютерной области)
Синтаксическая	Степень уменьшения неопределенности (шенноновский подход)	Вероятность события
	Единицы представления информации (компьютерный подход)	Бит, байт, Кбайт и т. д.
Семантическая	Тезаурус	Пакет прикладных программ, персональный компьютер, компьютерные сети и т. д.
	Экономические показатели	Рентабельность, производительность, коэффициент амортизации и т. д.
Прагматическая	Ценность использования	Емкость памяти, производительность компьютера, скорость передачи данных и т. д. Денежное выражение Время обработки информации и принятия решений

Меры информации семантического уровня



Тезаурус — совокупность сведений, которыми располагает пользователь или система



при $S_p = 0$ пользователь не воспринимает (не понимает поступающую информацию; при $S_p \rightarrow \infty$ пользователь «все знает», и поступающая информация ему не нужна.

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

Меры информации семантического уровня



Коэффициент содержательности C -
относительная мера семантической информации:

$$C = I_c / V_d,$$

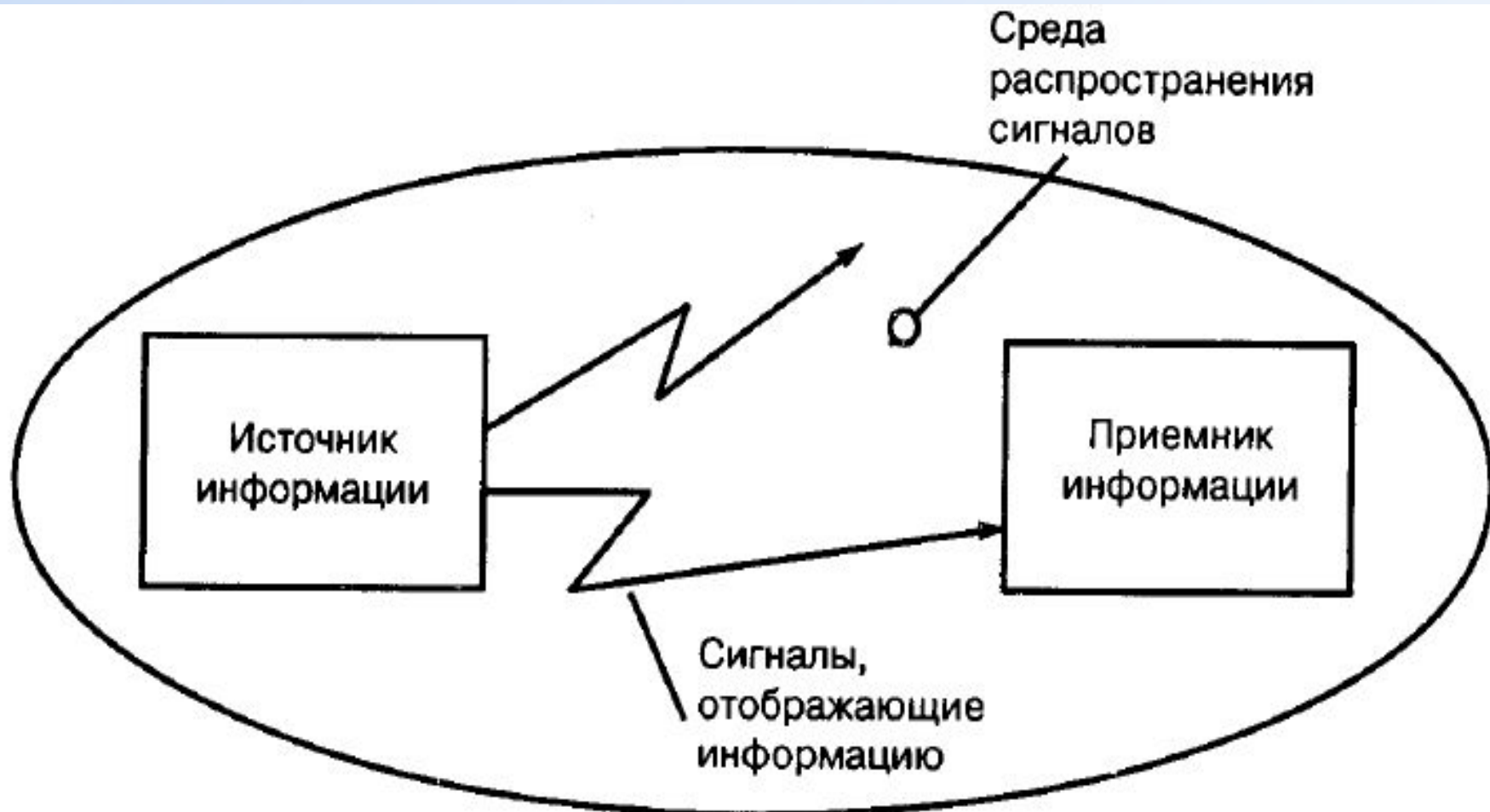
I_c - количество семантической информации,

V_d - объем семантической информации.

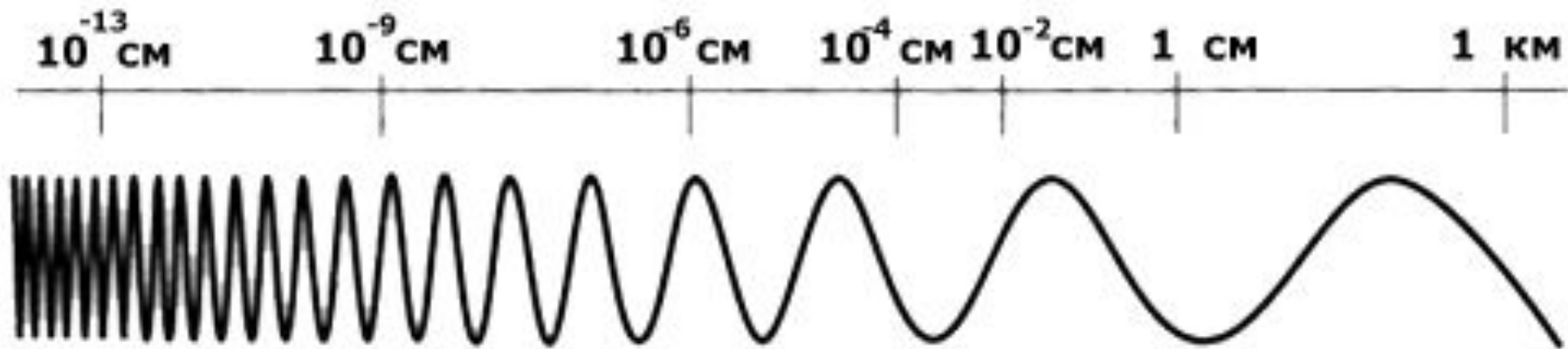
Конкретные показатели формируются на основе статистической обработки количества ссылок в различных выборках.

Компоненты обмена информацией

Сигнал - это физический процесс, несущий информацию о событии или состоянии объекта наблюдения



Спектр электромагнитных волн



Видимый спектр



Гамма лучи

Ультрафиолет

Микроволны

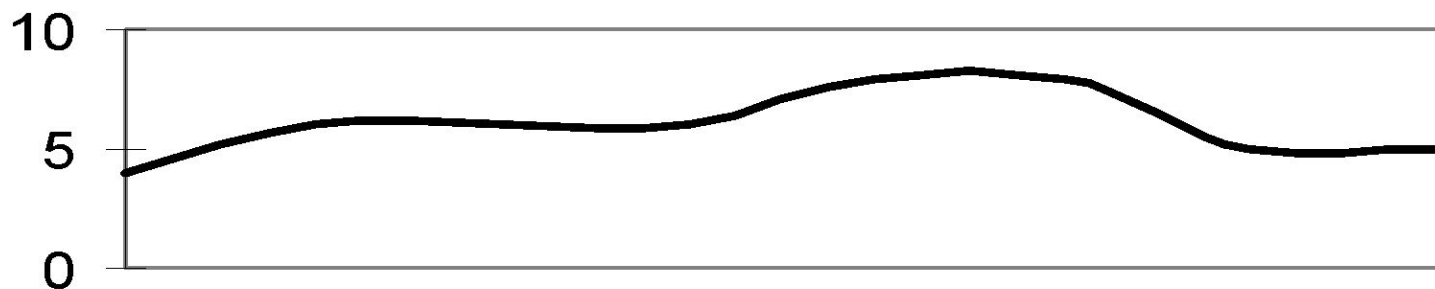
Радиоволны

Рентгеновские лучи Инфракрасные лучи

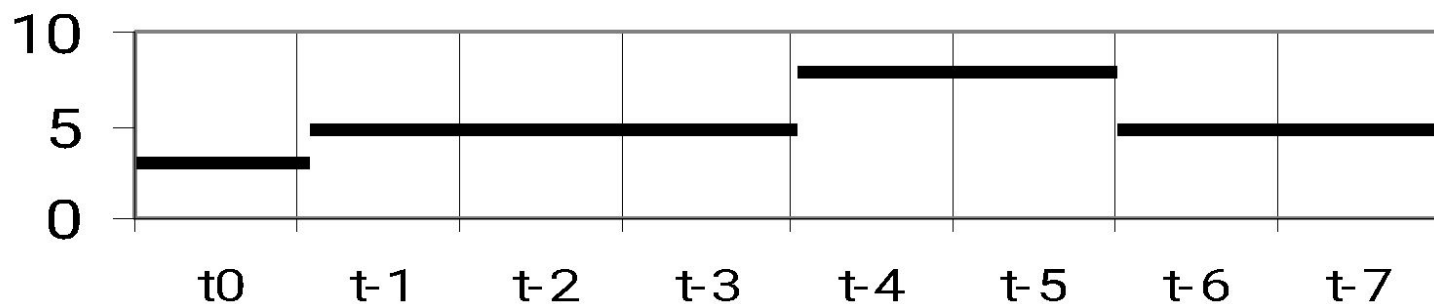
Сигналы непрерывный и дискретный



Непрерывный сигнал



Дискретный сигнал





Количество информации, воспринятой приемником, характеризуется не длиной сообщения, а разнообразием сигнала порожденного в приемнике этим сообщением.

Память носителя информации имеет некоторую физическую ёмкость, в которой она способна накапливать образы, и количество накопленной в памяти информации, характеризуется в итоге именно разнообразием заполнения этой ёмкости.

Количество информации



Минимальное разнообразие обеспечивается наличием *двух* состояний. *Информационная ёмкость одной ячейки памяти, способной находиться в двух различных состояниях, принята за единицу измерения количества информации - 1 бит.*

Количество получаемой информации может быть и меньше, чем 1 бит. Это происходит тогда, когда варианты ответов «да» и «нет» *не равновероятны.*



Клод Шеннон: «Информация - снятая неопределенность».

Неопределенность возникает в ситуации выбора. Задача снятия неопределенности – уменьшение количества рассматриваемых вариантов (уменьшение разнообразия), выбор одного соответствующего ситуации варианта из числа возможных. Снятие неопределенности дает возможность принимать обоснованные решения и действовать. В этом управляющая роль информации.

Минимальное разнообразие обеспечивается наличием **двух** состояний.



Информационная ёмкость одной ячейки памяти, способной находиться в двух различных состояниях, принята за единицу измерения количества информации - 1 бит.

1 бит (*bit* - сокращение от англ. *binary digit* - двоичное число) - единица измерения **информационной емкости и количества информации**, а также и еще одной величины – **информационной энтропии**.

**** Это справедливо только для равновероятного случая**



1928 г. Р. Хартли: процесс получения информации - выбор одного сообщения из конечного наперёд заданного множества из N равновероятных сообщений.

Формула Хартли: $I = \log_2 N$.

Примеры равновероятных сообщений:

- при бросании монеты: "выпала решка", "выпал орел";
- на странице книги: "количество букв чётное", "количество букв нечётное".
- угадать одно число из набора чисел от единицы до ста: $I = \log_2 100 > 6,644$.



Равновероятны ли сообщения "первой выйдем из дверей здания женщина" и "первым выйдем из дверей здания мужчина".

Однозначно ответить на этот вопрос нельзя. Все зависит от того, о каком именно здании идет речь. Если это, например, станция метро, то вероятность выйти из дверей первым одинакова для мужчины и женщины, а если это военная казарма, то для мужчины эта вероятность значительно выше, чем для женщины.



1948 г. Клод Шеннон. Формула определения количества информации, учитывающую возможную неодинаковую вероятность сообщений в наборе.

Формула Шеннона:

$$I = - (p_1 \log_2 p_1 + p_2 \log_2 p_2 + \dots + p_N \log_2 p_N),$$
где p_i — вероятность того, что именно i -е сообщение выделено в наборе из N сообщений.

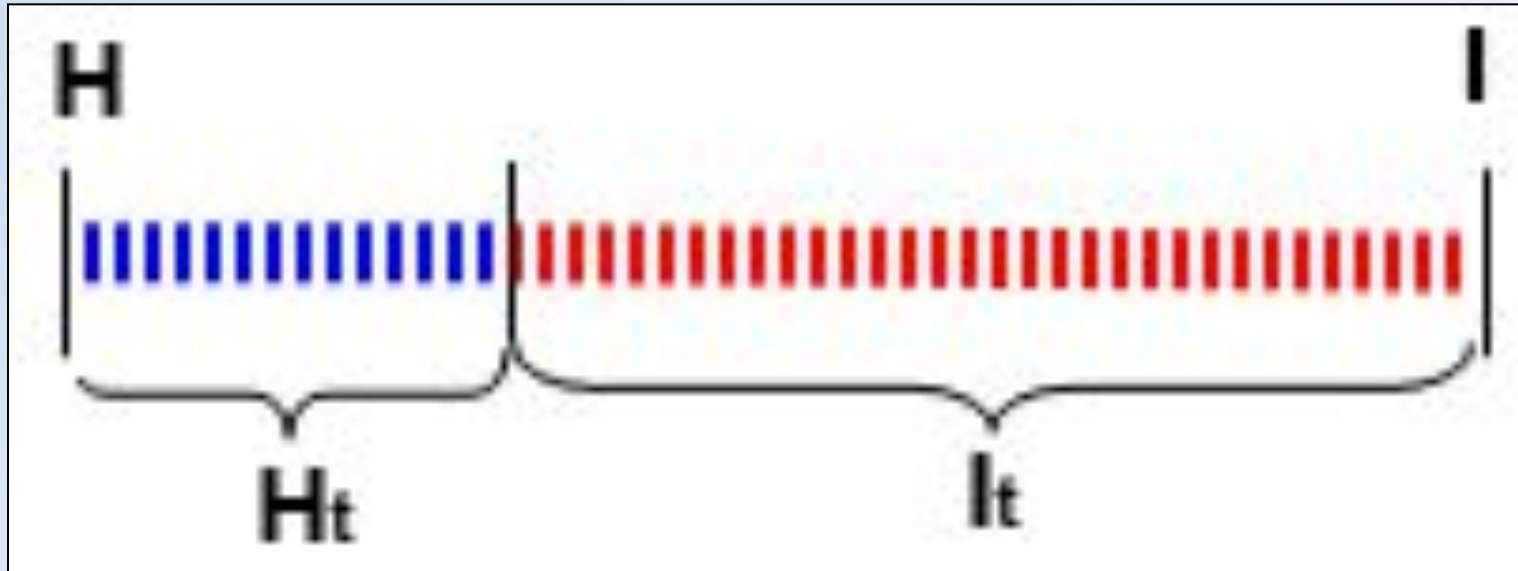
Клод Шеннон определил **информацию**, как **снятую неопределенность**.

Величина, характеризующая количество неопределенности обозначается символом H и имеет название **энтропия**, точнее **информационная энтропия**.

Энтропия (H) – мера неопределенности, или мера равномерности распределения случайной величины.



Связь между энтропией и количеством информации



I – это количество информации, которое требуется для снятия неопределенности H .

Количество информации в сообщении



В общем случае, энтропия H и количество получаемой в результате снятия неопределенности информации зависят от исходного количества рассматриваемых вариантов N и априорных вероятностей реализации каждого из них $P: \{p_0, p_1, \dots, p_{N-1}\}$, т.е. $H=F(N, P)$.

Формуле **Шеннона**:

$$H = -\sum_{i=0}^{N-1} p_i \log_2(p_i) = \sum_{i=0}^{N-1} p_i \log_2\left(\frac{1}{p_i}\right)$$



Пример. Пусть в некотором учреждении состав работников распределяется так: 3/4 - женщины, 1/4 - мужчины. Тогда неопределенность относительно того, кого вы встретите первым, зайдя в учреждение, будет рассчитана рядом действий, показанных в таблице.

	p_i	$1/p_i$	$I_i = \log_2(1/p_i), \text{ бит}$	$p_i * \log_2(1/p_i), \text{ бит}$
Ж	3/4	4/3	$\log_2(4/3) = 0,42$	$3/4 * 0,42 = 0,31$
М	1/4	4/1	$\log_2(4) = 2$	$1/4 * 2 = 0,5$
Σ	1			$H = 0,81 \text{ бит}$

Количество информации в сообщении

Формула Хартли – частный случай формулы Шеннона для равновероятных альтернатив. Подставив в формулу Шеннона вместо p_i его (в равновероятном случае не зависящее от i) значение $p_i = \frac{1}{N}$ получим:

$$H = \sum_{i=0}^{N-1} \frac{1}{N} \log_2 \left(\frac{N}{1} \right) = \frac{1}{N} * N * \log_2(N) = \log_2(N)$$

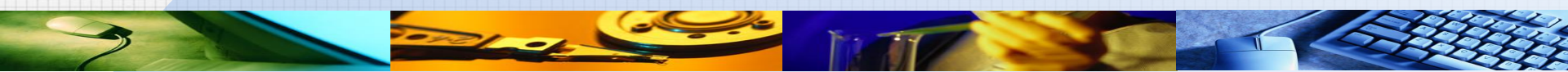
Формула Хартли

$$H = \log_2(N)$$

или

$$N = 2^H$$

Количество информации в сообщении



Неопределенность может быть снята только частично, поэтому количество информации I , получаемой из некоторого сообщения, вычисляется как уменьшение энтропии, произошедшее в результате получения данного сообщения.

$$I = H_{до} - H_{после}$$

Для равновероятного случая, используя для расчета энтропии формулу Хартли, получим:

$$I = \log_2(N_{до}) - \log_2(N_{после}) = \log_2\left(\frac{N_{до}}{N_{после}}\right)$$

$$\text{Если } N_{после} = 1, \text{ то } I = \log_2\left(\frac{N_{до}}{1}\right) = \log_2(N_{до}) = H_{до}$$

$$\text{Если } N_{после} = N_{до}, \text{ то } I = \log_2(1) = 0$$

$$\text{Если } N_{после} < N_{до}, \text{ то } \frac{N_{до}}{N_{после}} > 1 \Rightarrow I > 0,$$

$$\text{если } N_{после} > N_{до}, \text{ то } \frac{N_{до}}{N_{после}} < 1 \Rightarrow I < 0.$$

Некто вынимает одну карту из колоды (36 карт).

Изначальная неопределенность $H = \log_2(36) \approx 5,17$ бит.



Вытянувший карту сообщает нам часть информации. Какое количество информации мы получаем из этих сообщений?

А. “Это карта красной масти”. $I = \log_2(36/18) = \log_2(2) = 1$ бит (красных карт в колоде половина, неопределенность уменьшилась в 2 раза).

В. “Это карта пиковой масти”. $I = \log_2(36/9) = \log_2(4) = 2$ бита (пиковые карты составляют четверть колоды, неопределенность уменьшилась в 4 раза).

С. “Это одна из старших карт: валет, дама, король или туз”. $I = \log_2(36) - \log_2(16) = 5,17 - 4 = 1,17$ бита (неопределенность уменьшилась больше чем в два раза, поэтому полученное количество информации больше одного бита).

Д. “Это одна карта из колоды”. $I = \log_2(36/36) = \log_2(1) = 0$ бит (неопределенность не уменьшилась).

Е. “Это дама пик”. $I = \log_2(36/1) = \log_2(36) = 5,17$ бит (неопределенность полностью снята).



Задача 1. Какое количество информации будет содержать зрительное сообщение о цвете вынутого шарика, если в непрозрачном мешочке находится 50 белых, 25 красных, 25 синих шариков?

Решение.

1) всего шаров $50+25+25=100$

2) вероятности шаров $50/100=1/2$, $25/100=1/4$, $25/100=1/4$

3) $I = -(1/2 \log_2 1/2 + 1/4 \log_2 1/4 + 1/4 \log_2 1/4) =$
 $= -(1/2(0-1) + 1/4(0-2) + 1/4(0-2)) = 1,5$ бит

Задача 2. В корзине лежит 16 шаров разного цвета. Сколько информации несет сообщение, что достали белый шар?

Решение. Т.к. $N = 16$ шаров, то $I = \log_2 N = \log_2 16 = 4$ бит.



Задача 3. В корзине лежат черные и белые шары. Среди них 18 черных шаров. Сообщение о том, что достали белый шар, несет 2 бита информации. Сколько всего шаров в корзине?

Решение. Найдем по формуле Шеннона вероятность получения белого шара: $\log_2 N = 2$, $N = 4$, следовательно, вероятность получения белого шара равна $1/4$ (25%), а вероятность получения черного шара соответственно $3/4$ (75%). Если 75% всех шариков черные, их количество 18, тогда 25% всех шариков белые, их количество $(18 \cdot 25) / 75 = 6$. Осталось найти количество всех шариков в корзине $18 + 6 = 24$.
Ответ: 24 шарика.



Решение. Количество символов для кодирования номера составляет: 30 букв + 10 цифр = 40 символов. Количество информации несущий один символ равен 6 бит ($2^6=64$, но количество информации не может быть дробным числом, поэтому берем ближайшую степень двойки большую количества символов $2^6=64$).

Мы нашли количество информации, заложенное в каждом символе, количество символов в номере равно 5, следовательно, $5*6=30$ бит. Каждый номер равен 30 битам информации, но по условию задачи каждый номер кодируется одинаковым и минимально возможным количеством байт. Если разделить 30 на 8 получится дробное число, а нам необходимо найти целое количество байт на каждый номер. Находим ближайший множитель 8-ки, который превысит количество бит, это 4 ($8*4=32$). Каждый номер кодируется 4 байтами.

Для хранения 50 автомобильных номеров потребуется:
 $4*50=200$ байт.

Выбор оптимальной стратегии в игре "Угадай число".

Первый участник загадывает целое число (например, 3) из заданного интервала (например, от 1 до 16), а второй - должен "угадать" задуманное число.



Начальная неопределенность знаний для второго участника составляет 16 возможных событий.

При оптимальной стратегии интервал чисел всегда должен делиться пополам, тогда количество возможных событий (чисел) в каждом из полученных интервалов будет одинаково и отгадывание интервалов равновероятно. В этом случае на каждом шаге ответ первого игрока ("Да" или "Нет") будет нести максимальное количество информации (1 бит).

Вопрос второго участника	Ответ первого участника	Неопределенность знаний (количество возможных событий)	Полученное количество информации
Число больше 8?	Нет	8	1 бит
Число больше 4?	Нет	4	1 бит
Число больше 2?	Да	2	1 бит
Число 3?	Да	1	1 бит