

Сообщения, данные, сигнал,  
показатели качества  
информации, формы  
представления информации.

Системы передачи  
информации

Лектор Лузев В.С.

Информация и данные. Количество и  
качество информации

# Термин «информатика»

- Термин информатика происходит от французского слова *informatique* (объединение слов *information* — «информация» и *automatique*— «автоматика»). В англоязычных странах этому термину соответствует синоним *Computer Science* (компьютерная наука).
- Впервые термин появился в Германии, широко распространился во Франции.
- *Informatik* - Германия, 1957 г., инженер К. Штейнбух.
- *Informatique* - Франция, 1962 г., программист Ф.Дрейфус.
- *Informatics* – США, 1962 г., программист У. Ф. Бауэр
- Информатика - СССР, 1963 г., профессор МЭИ Ф.Е.Темников.

# Термин «информатика»

- Распространению термина «информатика» немало способствовал новосибирский учёный А.П. Ершов (с 1970 г. — член-корреспондент, с 1984 г. — академик АН СССР).
- В школах СССР учебная дисциплина «Информатика» появилась в 1985 году одновременно с первым учебником А. П. Ершова «Основы информатики и вычислительной техники» .



Академик Андрей Петрович Ершов (1931–1988) – один из зачинателей теоретического и системного программирования в СССР, создатель Сибирской школы информатики и неформальный лидер всего отечественного программистского сообщества. Его существенный вклад в становление информатики как новой отрасли науки и нового феномена общественной жизни широко признан в нашей стране и за рубежом.

# Определения информатики

- Информатика — это наука и сфера практической деятельности, связанная с различными аспектами получения, хранения, обработки, передачи и использования информации.

«Понятие информатики охватывает области, связанные с разработкой, созданием, использованием и материально-техническим обслуживанием систем обработки информации, включая машины, оборудование, организационные аспекты, а также комплекс промышленного, коммерческого, административного и социального воздействия» — такая формулировка была дана в 1978 г. Международным конгрессом по информатике.

# Информация и данные

- Не существует единственного «правильного» определения информации, но почти каждый человек имеет представление о том, что такое информация.
- Есть попытки дать научное определение информации:
- *Информация - является одной из фундаментальных сущностей окружающего нас мира (академик Поспелов).*
- *Информация ( Information )- содержание сообщения или сигнала; сведения, рассматриваемые в процессе их передачи или восприятия, позволяющие расширить знания об интересующем объекте [Терминологический словарь по основам информатики и вычислительной техники. Под ред. А.П. Ершова, Н.М. Шанского. Москва .: Просвещение, 1991.]*



# Информация и данные

- Информация (от лат. *informātiō* — «разъяснение, представление, понятие о ч.-л.», от лат. *informare* — «придавать вид, форму, обучать; мыслить, воображать») — сведения, независимо от формы их представления, воспринимаемые человеком или специальными устройствами как отражение фактов материального мира в процессе коммуникации (ГОСТ 7.0-99).

# Информация и данные

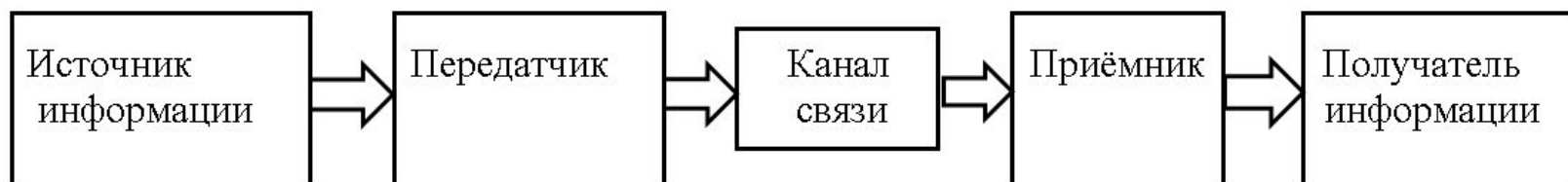
- Данные — поддающееся многократной интерпретации представление информации в формализованном виде, пригодном для передачи, связи, или обработки (ISO/IEC 2382:2015).
- Данные — формы представления информации, с которыми имеют дело информационные системы и их пользователи (ISO/IEC 10746-2:1996).

# Информация и данные

- **Данные** содержат информацию, но не являются самой информацией. По этому поводу часто **возникает путаница. В неформальном контексте эти два термина очень часто используют как синонимы.**
- Данные есть сигналы (сведения), получаемые от объектов или уже записанные на носителях в определённой знаковой системе.
- Чтобы данные воспринимались человеком как **информация**, надо применить к ним некие **методы** преобразования их в понятия, доступные сознанию человека.
- Методы могут быть настолько различными, что количество информации, извлекаемой из одних и тех же данных, может отличаться существенно.

# Передача информации

Для передачи, должны быть источник информации, передатчик, канал связи, приёмник и получатель информации.



# Качество информации

- Качество информации определяется некоторыми ее свойствами, отвечающими потребностям (целям, задачам) пользователей, например, полнота, достоверность, доступность, актуальность.
- *Полнота информации* характеризует степень её достаточности для описания объекта.
- *Достоверность информации* — это свойство, характеризующее степень соответствия информации о некотором объекте.

# Качество информации

- *Доступность информации* — это возможность получения информации при необходимости. Доступность складывается из двух составляющих: из доступности данных и доступности методов.
- *Актуальность информации* характеризует её свойство сохранить достоверность на текущий момент.

# Количество информации

Кроме качества для нас важно и *количество информации*.

Для измерения количества информации имеются два различных подхода: статистический и объёмней.

# Количество информации. Вероятностный подход

- Информационная энтропия — мера неопределённости или непредсказуемости информации, неопределённость появления какого-либо символа первичного алфавита. При отсутствии информационных потерь численно равна количеству информации на символ передаваемого сообщения.



# Количество информации. Вероятностный подход

- В 1928 г. американский инженер Р. Хартли предложил оценивать количество информации  $H$  в сообщении логарифмом числа  $N$  возможных равновероятных состояний:

$$H = \log_2 N$$

Стандартной минимальной единицей количества информации  $H = 1$  будет выбор из двух возможностей  $N = 2$  (да- нет, 0-1, истина-ложь и т. д.). Такая единица носит название **бита**.

# Количество информации.

## Вероятностный подход

- В 1948 году, исследуя проблему рациональной передачи информации через зашумлённый коммуникационный канал, Клод Шеннон (1916–2001) предложил революционный вероятностный подход к пониманию коммуникаций и создал первую, истинно математическую, теорию энтропии.
- Клод Шеннон предложил формулу определения количества информации, учитывающую возможную неодинаковую вероятность  $P$  сообщений в наборе из  $i$  элементов.
- Согласно концепции К. Шеннона, информация – это снятая неопределенность, то есть сведения, которые должны снять в той или иной степени существующую у потребителя до их получения неопределенность, расширить его понимание объекта полезными сведениями.

$$H = -\sum_{i=1}^N P_i \times \log_2 P_i.$$

# Количество информации. Вероятностный подход

- В качестве примера определим количество информации, связанное с появлением каждого символа в сообщениях, записанных на русском языке. Будем считать, что русский алфавит состоит из 33 букв и знака «пробел» для разделения слов. По формуле Хартли  $H = \log_2 34 = 5,09$  бит.
- Если воспользоваться формулой Шеннона, учитывая разную вероятность появления букв в тексте, то получим  $H = 4,72$  бит.

# Количество информации.

## Объёмный подход

- Объёмный подход предполагает количественную оценку информации в определенных единицах. За минимальную единицу информации принимается один бит (англ. bit — **b**inary **dig**it — двоичная цифра).
- Невозможно нецелое число битов (в отличие от вероятностного подхода).
- Байт является последовательностью из восьми битов. Широко используются ещё более крупные производные единицы информации:

# Количество информации. Объёмный подход

Таблица 1 – Производные единицы объёма данных(информации)

|                 |           |                             |                 |            |       |          |
|-----------------|-----------|-----------------------------|-----------------|------------|-------|----------|
| <b>байт</b>     | <b>B</b>  | <u><math>10^0</math></u>    | <b>байт</b>     | <b>B</b>   | байт  | $2^0$    |
| <u>килобайт</u> | <b>kB</b> | <u><math>10^3</math></u>    | <u>кибибайт</u> | <b>KiB</b> | Кбайт | $2^{10}$ |
| <u>мегабайт</u> | <b>MB</b> | <u><math>10^6</math></u>    | <u>мебибайт</u> | <b>MiB</b> | Мбайт | $2^{20}$ |
| <u>гигабайт</u> | <b>GB</b> | <u><math>10^9</math></u>    | <u>гибибайт</u> | <b>GiB</b> | Гбайт | $2^{30}$ |
| <u>терабайт</u> | <b>TB</b> | <u><math>10^{12}</math></u> | <u>тебибайт</u> | <b>TiB</b> | Тбайт | $2^{40}$ |