

# Распределенные информационные системы

Информация и данные

# Вопросы

- Информация и данные
  - Адекватность
  - Мера
  - Качество

# ТЕОРИЯ ИНФОРМАЦИИ

# Теория информации

- Теория информации — раздел прикладной математики. Она **аксиоматически** определяет:
  - понятие информации
  - ее свойства
- и устанавливает предельные соотношения для систем передачи данных
- Используется математический аппарат теории вероятностей и математической статистики
- Основные разделы теории информации:
  - Кодирование источника (сжимающее)
  - Канальное (помехоустойчивое) кодирование

# Помехи – определения

- Помеха:
  - Вообще: то, что мешает кому-либо, чему-либо; затруднение, препятствие
  - В линиях связи: электромагнитные колебания, не связанные с принимаемым сигналом, искажающие его
- В РИС будем понимать под помехой искажение данных с некоторой вероятностью. Возникает из-за:
  - Воздействия окружающей среды (радиопомехи, космические лучи, радиация)
  - Ошибок обработки ( в т.ч. при передаче)

# Помехи – единицы измерения

- Уровень помех можно определить через интенсивность битовых ошибок (Bit Error Rate, BER):
  - Для каналов связи без дополнительных средств защиты от ошибок составляет,  $10^{-4}$  —  $10^{-6}$
  - Оптоволокно —  $10^{-9}$

# Помехоустойчивость в целом

- Помехоустойчивость системы – способность выполнять функции при наличии помех
  - Оценивается интенсивностью (уровнем) помех, при которой нарушение функций не превышает допустимых пределов
- Помехоустойчивость линии — способность игнорировать помехи внешней среды. Зависит от:
  - Свойств физической среды передачи данных
  - Средств для экранирования и подавления помех (против электромагнитных помех: экранирование, скручивание проводников)
    - Наименьшая: у радиолиний
    - Лучше: кабельные линии
    - Совсем хорошо: волоконно-оптические линии

# Кодирование: помехоустойчивость

- Помехоустойчивое кодирование – способ представления информации, который:
  - Позволяет обнаружить факт искажения информации
  - Позволяет исправить одну или более ошибок (неверно переданных/поврежденных бит)
- Примеры:
  - Бит четности – позволяет обнаружить однократную (нечетное число) ошибок
  - Коды Рида – Соломона – исправляют  $n$ , обнаруживают  $n+1$
  - Свертки (хеш, контрольные суммы) – md5 , ГОСТ Р 34.11-2011)



# Базовые понятия теории информации

- Информация – *нематериальная* сущность, при помощи которой с **любой точностью** можно описывать сущности разной природы:
  - Реальные (материальные объекты и явления окружающей среды)
  - Виртуальные (возможные, моделируемые)
  - Понятийные (терминологические)
- Иначе: информация – сведения о сущностях, их **параметрах, свойствах и состоянии**, которые *уменьшают* имеющуюся степень неопределенности, неполноты знаний о них

**ИНФОРМАЦИЯ: АДЕКВАТНОСТЬ**

# Адекватность информации – 1/3

- Уровень соответствия образа, создаваемого с помощью полученной информации, реальному объекту, процессу, явлению
- Синтаксическая – отображает формально-структурные характеристики информации, не затрагивая ее смыслового содержания
  - Структура
  - Размер
  - Тип **данных**

# Адекватность информации – 2/3

- Семантическая – степень соответствия образа объекта самому объекту.  
Учитывается смысловое содержание информации
  - Сведения, отражаемые информацией
  - Смысловые связи
  - Проявляется при наличии единства информации и пользователя

# Адекватность информации – 3/3

- Прагматическая – соответствие информации цели управления, реализуемой на ее основе
  - Проявляются при наличии единства информации, пользователя и цели управления
  - Свойства информации, связанные с ее практическим использованием, с соответствием целевой функции деятельности системы:
    - Полезность
    - Своевременность
    - Полнота

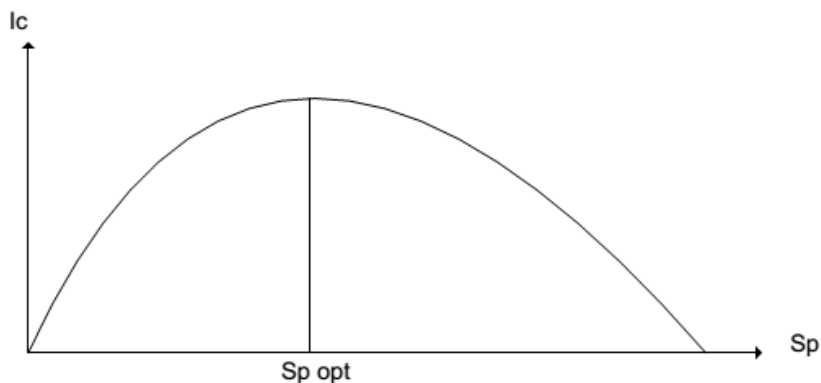
Синтаксическая  
Семантическая  
Прагматическая

**ИНФОРМАЦИЯ: МЕРЫ**

# Меры информации – 1/3

- Синтаксическая мера – оперирует с данными – обезличенной информацией, не выражающей смыслового отношения
  - Объем данных  $V_D$  в битах, байтах и т.д.
- Семантическая мера - измеряет смысловое содержание информации
  - Количество информации  $I$
  - Коэффициент содержательности  $C=I/V_D$
  - Тезаурусная мера (см. далее)

# Семантическая мера информации: тезаурус – 2/3

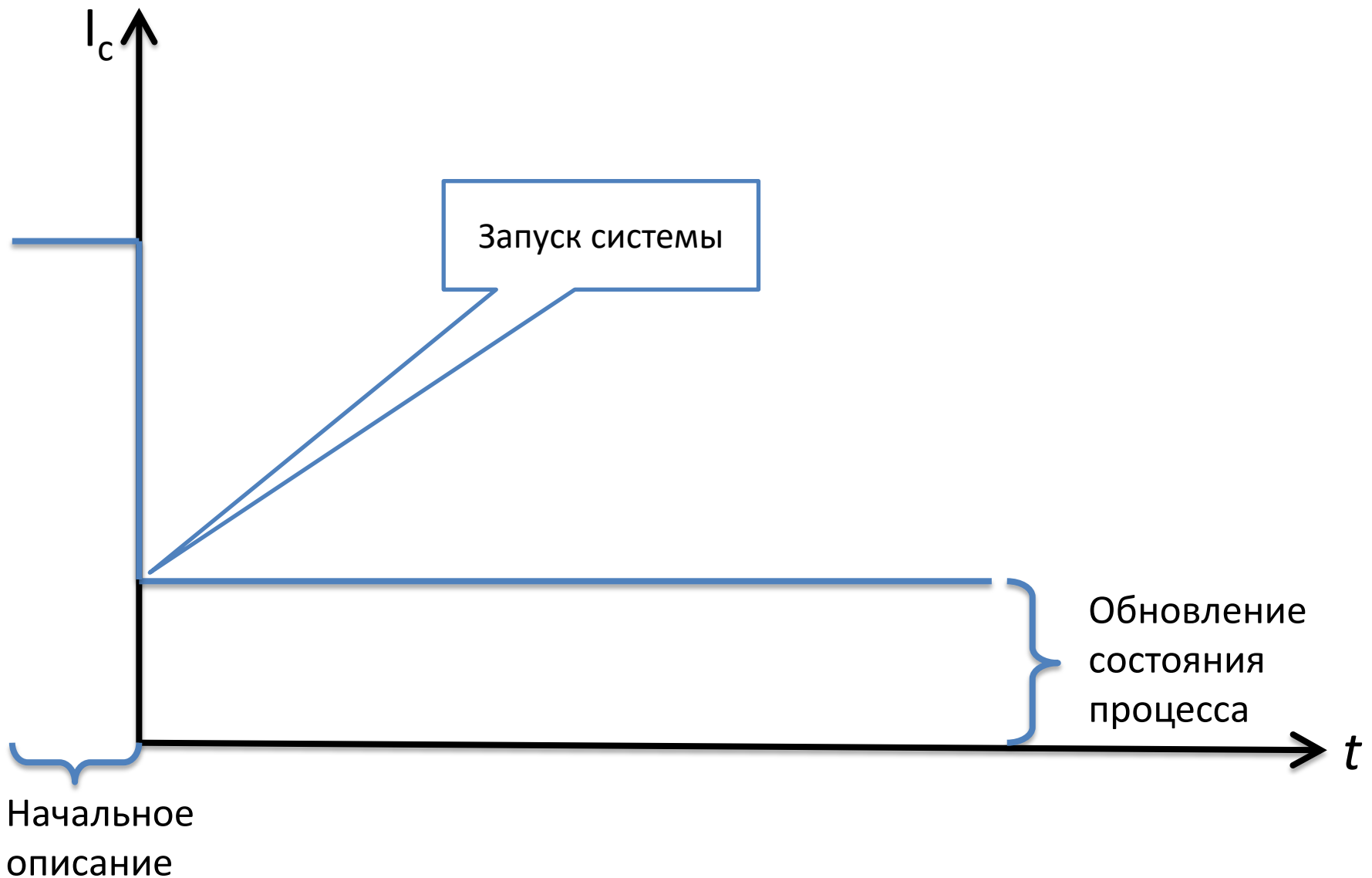


Максимальное количество семантической информации потребитель получает при согласовании ее смыслового содержания со своим тезаурусом, когда поступающая информация понятна пользователю и несет ему ранее не известные сведения

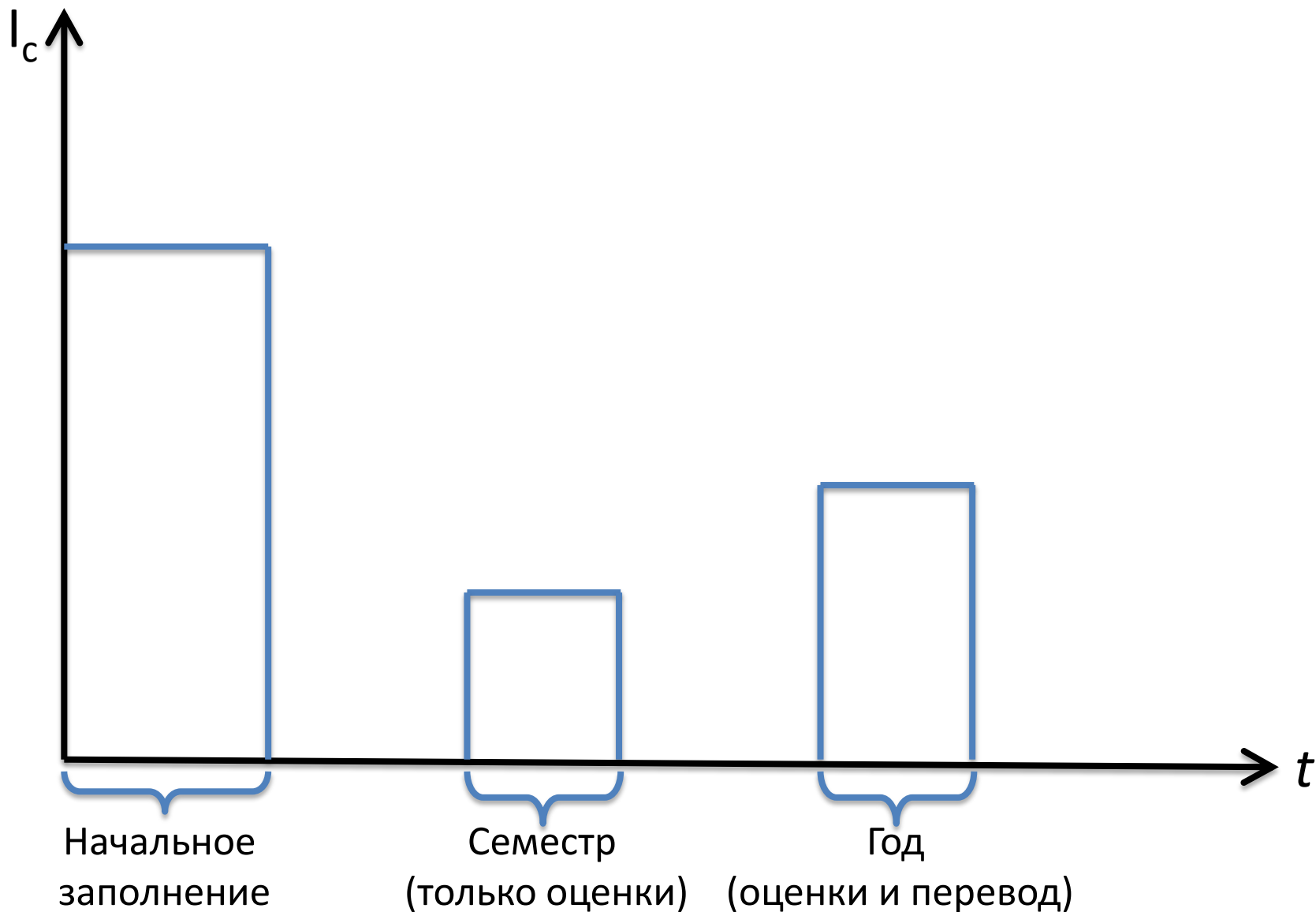
- Тезаурус это совокупность сведений, которыми располагает пользователь или система
- График зависимости меры информации, извлекаемой из одного сообщения от начальной осведомленности
- $S_p=0$  – информация не может быть усвоена
- $S_p=\infty$  – вся информация известна, новая не нужна



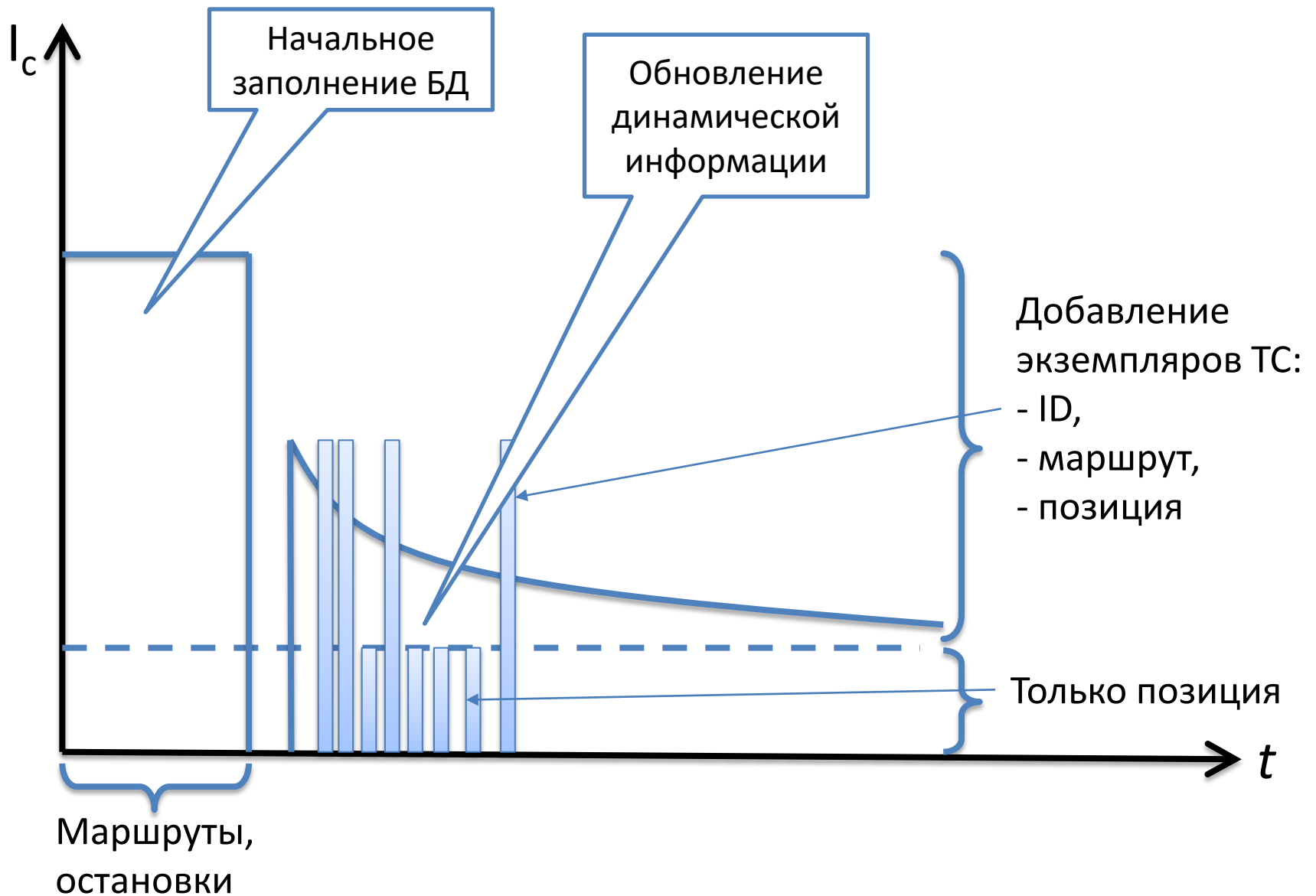
# Тезаурус пример: Химзавод



# Тезаурус пример: БД студентов



# Тезаурус пример: Общественный транспорт



# Меры информации – 3/3

- Прагматическая мера – описывает полезность информации, ее ценность для процесса управления
  - Обычно ценность информации измеряется в тех же единицах, что и целевая функция управления системой
  - Она зависит от конкретной системы и ее текущих целей, обычно является объектом рассмотрения теории искусственного интеллекта

**ИНФОРМАЦИЯ: КАЧЕСТВО**

# Качество информации – 1/3

- Репрезентативность
  - Правильностью отбора и формирования в целях адекватного отражения свойств объекта
- Содержательность (семантическая емкость)
- Достаточность (полнота)
  - Набора показателей для правильного функционирования РИС (принятия правильных решений)
- Пертинентность
  - Соответствие полученной информации информационной потребности пользователя

# Качество информации – 2/3

- Доступность
  - РИС: наличие источников информации
  - Пользователю: представление ее в удобном для восприятия формате
- Актуальность
  - Степень сохранения ценности информации в момент ее использования
- Своевременность
  - Поступление не позже заранее назначенного момента времени, согласованного с работой РИС (например, временем начала выполнения расчетов)

# Качество информации – 3/3

- Точность
  - Степень близости получаемой информации к реальному состоянию объекта, процесса, явления
- Достоверность
  - Доверительная вероятность необходимой точности
- Устойчивость
  - Способность реагировать на изменения исходных данных без нарушения необходимой точности