

Лекция 2-ср

Информация

Цель лекции:

Рассмотреть систематизированные основы знаний по основным понятиям информации

Учебные вопросы:

1. Определение понятия **Информация**
2. Представление и измерение информации
3. Классификация информации
4. Методы получения информации

1. Определение понятия **Информация**

Информация - это настолько общее и глубокое понятие, что его нельзя объяснить одной фразой. В это слово вкладывается различный смысл в технике, науке и в житейских ситуациях.

Термин "*информация*" происходит от латинского слова "*informatio*", что означает сведения, разъяснения, изложение.

Понятие «информация» имеет различные трактовки в разных предметных областях.

Понятие информации – одно из фундаментальных понятий информатики:

- **абстракция, абстрактная модель** рассматриваемой системы (в математике);
- **сигналы для управления**, приспособления рассматриваемой системы (в кибернетике);
- **мера хаоса** в рассматриваемой системе (в термодинамике);
- **вероятность выбора** в рассматриваемой системе (в теории вероятностей);
- **мера разнообразия** в рассматриваемой системе (в биологии) и др.

В обиходе **информацией** называют **любые данные или сведения**, которые кого-либо интересуют.

Информация - сведения об объектах и явлениях окружающей среды, их параметрах, свойствах и состоянии, которые воспринимают информационные системы (живые организмы, управляющие машины и др.) в процессе жизнедеятельности и работы.

Информация - это общенаучное понятие включающие, обмен сведениями между людьми, человеком и автоматом, обмен сигналами в животном и растительном мире (передача признаков от клетки к клетке, от организма к организму).

Более узкое определение дается в технике, где это понятие **информация** включает в себя все **сведения**, являющиеся **объектом хранения, передачи и преобразования**.

С понятием информации связаны такие понятия как **сигнал, сообщение и данные**.

Сигнал представляет собой любой процесс, несущий информацию.

Сообщение - это информация, представленная в определенной форме и предназначенная для передачи.

Данные - это информация, представленная в формализованном виде и предназначенная для обработки ее техническими средствами, например ЭВМ.

Одно и то же информационное сообщение (статья в газете, объявление, письмо, телеграмма, справка, рассказ, чертёж, радиопередача и т.п.) может содержать разное количество информации для разных людей - в зависимости от их предшествующих знаний, от уровня понимания этого сообщения и интереса к нему.

Так, сообщение, составленное на японском языке, не несёт никакой новой информации человеку, не знающему этого языка, но может быть высоко информативным для человека, владеющего японским.

Никакой новой информации не содержит и сообщение, изложенное на знакомом языке, если его содержание непонятно или уже известно.

Связь информации и сообщения

Информация – содержание *сообщения*, *сообщение* – форма *информации*.

Информация – это некоторая упорядоченная последовательность сообщений, отражающих, передающих и увеличивающих наши знания.

Информация есть характеристика не сообщения, а *соотношения между сообщением и его потребителем*.

Без наличия потребителя, хотя бы потенциального, говорить об информации бессмысленно.

Представление информации в компьютере.

Единицы измерения информации

Компьютер может обрабатывать только информацию, представленную в числовой форме.

Вся другая информация для обработки на компьютере должна быть преобразована в числовую форму.



Единицы измерения информации

Бит - единица информации, представляющая собой двоичный разряд, который может принимать значение 0 или 1.

Байт - восемь последовательных битов. В одном байте можно кодировать значение одного символа из 256 возможных ($256 = 2^8$).

Более крупными единицами информации являются следующие: 1 Кбайт = 2^{10} = 1024 байта;
1 Мбайт = 2^{20} байт = 1024 Кбайта; 1 Гбайт = 2^{30} байта и т. д.

Количество информации

- Содержательный подход $2^i = N$

Сообщение, уменьшающее неопределенность знаний в два раза, несет один **бит** информации

Шарик под первыми двумя стаканчиками? - Нет
Шарик под третьим стаканчиком? - Да



Клод Шеннон
(1916 - 2001)

Количество информации

- Алфавитный подход

Каждый символ некоторого сообщения имеет определенный информационный вес – несет фиксированное количество информации

В двоичном коде каждая двоичная цифра несет одну единицу информации, которая называется один **бит**

$$2^{i=1} = 2 \Rightarrow i = 1 \text{ бит}$$

i – информационный вес символа

$$N \Rightarrow 2^i;$$

N – мощность алфавита (количество символов в алфавите)

Информационный вес символа

- Алфавитный подход $2^i = N$
N – количество символов в алфавите
i – информационный вес символа

В русском алфавите (не считая букву ё) 32 символа. Определить информационный вес одного символа.

Информационный вес СИМВОЛА

- Алфавитный подход $2^i = N$ N – количество символов в алфавите
i – информационный вес символа

В русском алфавите (не считая букву ё) 32 символа. Определить информационный вес одного символа.

$$\begin{array}{l|l|l} N = 32 & 2^i = N & 2^i = 32 \Rightarrow 2^i = 2^5 \Rightarrow i = 5(\text{бит}) \\ \hline i - ? & & \end{array}$$

Ответ: 5
бит.

Информационный объем сообщения

- Алфавитный подход $2^i = N$ $I = K \times i$

В русском алфавите (не считая букву ё) 32 символа. Определить информационный объем сообщения «*информатика – наука об информации*»

Информационный объем сообщения

- Алфавитный подход $2^i = N$ $I = K \times i$

В русском алфавите (не считая букву ё) 32 символа. Определить информационный объем сообщения «информатика – наука об информации» (учесть все символы, включая пробел)

$$\begin{array}{l} N = 32 + 2 = 34 \\ K = 34 \\ I = ? \end{array} \left| \begin{array}{l} 2^i = N \\ I = K \times i \end{array} \right| \begin{array}{l} 2^i = 34 \Rightarrow 2^i = 2^6 \Rightarrow \text{бит} (\quad) \\ \text{бит} 34 \times 6 = 204 (\quad) \end{array}$$

Ответ: 204

бит.

Основные соотношения между единицами измерения:

1 *бит* (**binary digit** – двоичное число) = 0 или 1,

1 *байт* 8 битов,

1 *килобайт* (1К) = 2^{10} бит,

1 *мегабайт* (1М) = 2^{20} бит,

1 *гигабайт* (1Г) = 2^{30} бит,

1 *терабайт* (1Т) = 2^{40} бит,

1 *петабайт* (1П) = 2^{50} бит,

1 *эксабайт* (1Э) = 2^{60} бит.

Вопрос

Ы

Какое количество информации содержит один разряд двоичного числа?

Как записать число $(17)_{10}$ в двоичной системе счисления?

Для чего необходима процедура актуализации данных?

2. Представление и измерение информации

Формальное определение алфавита:

Алфавит - конечное множество различных знаков, символов, для которых определена операция **конкатенации** (приписывания, присоединения символа к символу или цепочке символов).

С помощью **конкатенации** по определенным правилам соединения символов и слов можно получать слова (цепочки знаков) и словосочетания (цепочки *слов*) в этом *алфавите* (над этим *алфавитом*).

Определение буквы - знака

Буквой или *знаком* называется любой элемент x алфавита X ,

где

$$x \in X$$

Понятие знака x неразрывно связано с тем, что им обозначается y («со смыслом»).

Знак x и его смысл y вместе могут рассматриваться как пара элементов (x, y) .

Примеры алфавитов: множество из десяти цифр, множество из знаков русского языка, точка и тире в азбуке Морзе и др.

В алфавите цифр знак 5 связан с понятием «быть в количестве пяти элементов».

Понятие слова

Конечная последовательность **букв алфавита**

называется **словом** в алфавите (или над алфавитом).

Длиной $|p|$ некоторого слова p над алфавитом **X** называется число составляющих его букв.

Слово (обозначаемое символом \emptyset) имеющее **нулевую длину**, называется пустым словом: $|\emptyset| = 0$.

Множество различных слов над алфавитом **X** обозначим через **$S(X)$** и назовем **словарным запасом (словарем) алфавита** (над алфавитом) **X** .

В отличие от конечного алфавита, словарный запас может быть и бесконечным.

Слова над некоторым заданным алфавитом определяют сообщения

Пример.

Слова над алфавитом кириллицы:

"Информатика", "инто", "иини", "и".

Слова над алфавитом десятичных цифр и знаков арифметических операций:

"1256", "23+78", "35-6+89", "4".

Слова над алфавитом азбуки Морзе:

".", ". . _", "_ _ _".

В алфавите должен быть определен порядок следования букв (порядок типа "предыдущий элемент – последующий элемент"), то есть любой алфавит имеет упорядоченный вид

Вывод

Таким образом, *алфавит* должен позволять решать задачу лексикографического (алфавитного) упорядочивания, или задачу расположения слов над этим алфавитом, в соответствии с порядком, определенным в алфавите (то есть по символам алфавита).

3. Классификация информации

Информация по отношению к источнику или приемнику бывает трех типов: входная, выходная и внутренняя.

Информация по отношению к конечному результату бывает исходная, промежуточная и результирующая.

Информация по ее изменчивости бывает постоянная, переменная и смешанная.

Информация по стадии ее использования бывает первичная и вторичная.

Информация по ее полноте бывает избыточная, достаточная и недостаточная.

Информация по доступу к ней бывает открытая и закрытая.

Есть и другие типы классификации информации.

Основные свойства информации:

объективность

достоверность;

полнота;

точность;

актуальность;

адекватность;

полезность;

массовость;

устойчивость;

ценность и др.

Объективность информации. Объективный –

существующий вне и независимо от человеческого сознания. **Информация** – это отражение внешнего объективного мира. **Информация объективна**, если она **не зависит от методов ее фиксации, чьего-либо мнения, суждения.**

Пример. Сообщение «На улице тепло» несет субъективную информацию, а сообщение «На улице 22° С» – объективную, но с точностью, зависящей от погрешности средства измерения.

Объективную информацию можно получить с помощью исправных датчиков, измерительных приборов.

Отражаясь в сознании конкретного человека, информация перестает быть объективной, так как, преобразовывается (в большей или меньшей степени) в зависимости от мнения, суждения, опыта, знаний конкретного субъекта.

Достоверность информации.

Информация достоверна, если она отражает истинное положение дел. Объективная информация всегда достоверна, но достоверная информация может быть как объективной, так и субъективной. Достоверная информация помогает принять нам правильное решение.

Недостоверной информация может быть по следующим причинам:

- преднамеренное искажение (дезинформация) или непреднамеренное искажение субъективного свойства;
- искажение в результате воздействия помех («испорченный телефон») и недостаточно точных средств ее фиксации.

Полнота информации. Информацию можно назвать полной, если ее достаточно для понимания и принятия решений. Неполная информация может привести к ошибочному выводу или решению.

Точность информации определяется степенью ее близости к реальному состоянию объекта, процесса, явления и т. п.

Актуальность информации – важность для настоящего времени, злободневность, насущность. Только вовремя полученная информация может быть полезна.

Полезность (ценность) информации. Полезность может быть оценена применительно к нуждам конкретных ее потребителей и оценивается по тем задачам, которые можно решить с ее помощью.

Количество информации – число, адекватно характеризующее разнообразие (структурированность, определенность, выбор состояний и т.д.) в оцениваемой системе.

Количество информации часто оценивается в битах, причем такая оценка может выражаться и в долях битов (так речь идет не об измерении или кодировании сообщений).

Мера информации – критерий оценки количества информации.

Обычно мера информации задана некоторой неотрицательной функцией, определенной на множестве событий и являющейся аддитивной, то есть мера конечного объединения событий (множеств) равна сумме мер каждого события. При оценке информации различают три аспекта:

- синтаксический,
- семантический,
- прагматический.

Мера информации		Единицы измерения	Примеры (для компьютерной области)
Синтаксическая	Вероятностный подход	Степень уменьшения неопределенности	Вероятность события
	Компьютерный подход	Единицы представления информации	Бит, байт, килобайт, ...
Семантическая		Тезаурус (совокупность сведений, которыми располагает пользователь или система)	Пакеты прикладных программ, компьютерные сети, система искусственного интеллекта
Прагматическая		Ценность использования	Емкость памяти, производительность компьютера, скорость передачи данных и т. Д.

Синтаксическая мера (вероятностный подход)

Мера Р. Хартли

Пусть известны N состояний системы S . Если каждое состояние системы закодировать двоичными кодами, то длину кода d необходимо выбрать так, чтобы число всех различных комбинаций было бы не меньше, чем N :

$$2^d \geq N.$$

Логарифмируя это неравенство, можно записать:

$$D \geq \log_2 N.$$

Наименьшее решение этого неравенства или мера разнообразия множества состояний системы задается формулой Р. Хартли:

Пример

Чтобы определить состояние системы из четырех возможных состояний, то есть получить некоторую *информацию* о системе, необходимо задать 2 вопроса.

Первый вопрос, например: "Номер состояния больше 2?".

Узнав ответ ("да", "нет"), мы увеличиваем суммарную *информацию* о системе на 1 *бит* ($I = \log_2 2$).

Далее необходим еще один уточняющий вопрос, например, при ответе "да": "Состояние – номер 3?". Итак, количество *информации* равно 2 *битам* ($I = \log_2 4$).

Если система имеет **n** различных состояний, то максимальное количество информации равно

Если во множестве $X = \{x_1, x_2, \dots, x_n\}$ искать произвольный элемент, то для его нахождения (по Хартли) необходимо иметь не менее $\log_2 n$ (единиц) информации.

Уменьшение H говорит об уменьшении разнообразия состояний N системы.

Увеличение H говорит об увеличении разнообразия состояний N системы.

Мера Хартли подходит лишь для идеальных, абстрактных систем, так как в реальных системах состояния системы не одинаково осуществимы (не равновероятны).

Мера Шеннона

Для реальных систем используют более подходящую меру К. Шеннона. Мера

Шеннона

оценивает *информацию* отвлеченно от ее смысла:

$$I = - \sum_{i=1}^n p_i \log_2 p_i$$

где n – число состояний системы;

p_i – вероятность (относительная частота)

перехода

Обозначим величину:

$$f_i = -n \cdot \log_2 p_i.$$

Тогда из формулы К. Шеннона следует, что количество информации I можно понимать как среднеарифметическое величин f_i , то есть величину f_i можно интерпретировать как информационное содержание символа алфавита с индексом i и величиной вероятности появления p_i этого символа в любом сообщении (слове), передающем информацию.

Пример

Если положение точки в системе из 10 клеток известно, например если точка находится во второй клетке, то есть

$$p_i = 0; \quad i = 1, 3, 4, \dots, 10; \quad p_2 = 1,$$

то тогда получаем количество информации, равное нулю:

$$I = \log_2 1 = 0.$$

Связь энтропии (меры хаоса) и информации

Увеличение меры Шеннона свидетельствует об **уменьшении энтропии** (увеличении порядка) системы;
уменьшение меры Шеннона свидетельствует об **увеличении энтропии** (увеличении беспорядка) системы.

Положительная сторона формулы Шеннона – ее **отвлеченность от смысла информации.**

В отличие от формулы Хартли она учитывает различность состояний, что делает ее пригодной для практических вычислений.

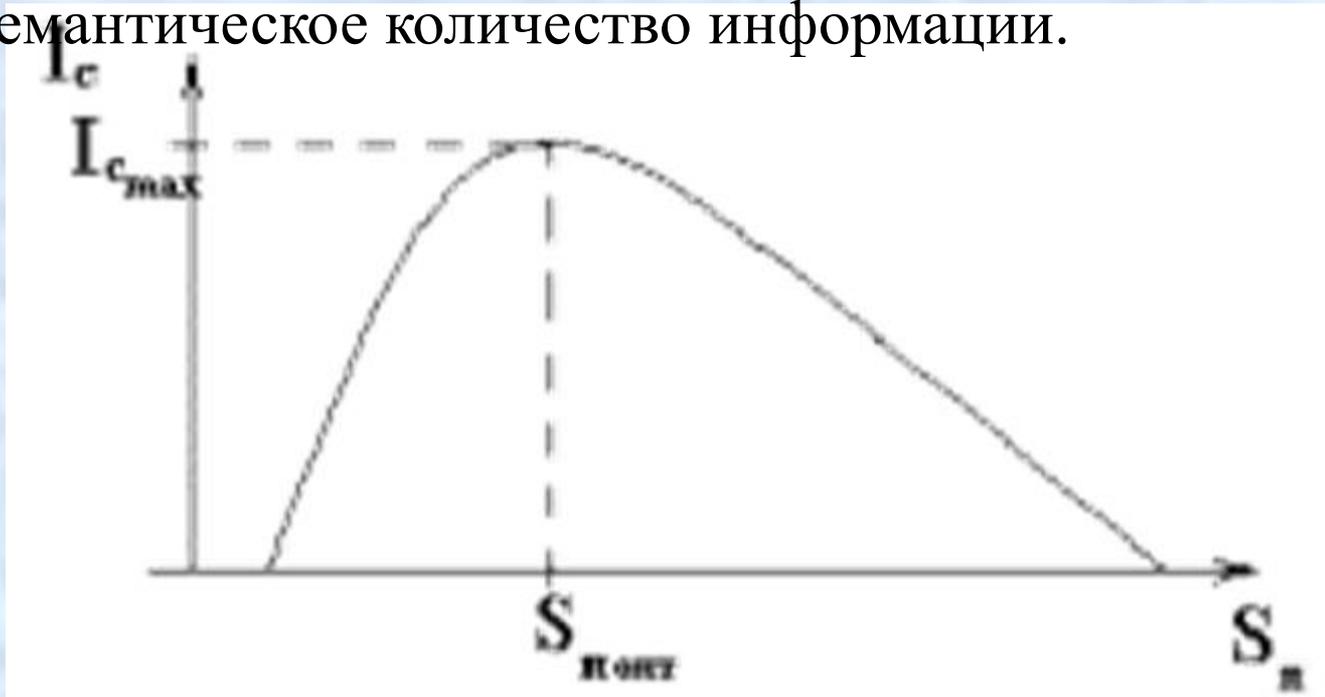
Основная отрицательная сторона формулы Шеннона – она не распознает различные состояния системы с одинаковой вероятностью.

Семантическая мера

Характеризует смысловое содержание информации

S_n - тезаурусная мера (совокупность сведений) получателя;

I_c - семантическое количество информации.



при $S_n \rightarrow 0$ пользователь не воспринимает, не понимает поступающую информацию;

при $S_n \rightarrow \infty$ пользователь все знает, и поступающая информация ему не нужна.

Прагматическая мера

$$I_n = \log_a \frac{P_1}{P_0},$$

I_n - прагматическое количество информации;
 $a > 1$;

P_0 - вероятность достижения цели до получения информации;

P_1 - вероятность достижения цели после

4. Методы получения информации

Методы получения информации можно разбить на:

- Эмпирические методы или методы получения эмпирических данных.
- Теоретические методы или методы построения различных теорий.
- Эмпирико-теоретические методы (смешанные) или методы построения теорий на основе полученных эмпирических данных об объекте, процессе, явлении.

Эмпирические методы

Наблюдение – сбор первичной информации об объекте, процессе, явлении.

Сравнение – обнаружение и соотнесение общего и различного.

Измерение – поиск с помощью измерительных приборов эмпирических фактов.

Эксперимент – преобразование, рассмотрение объекта, процесса, явления с целью выявления каких-то новых свойств.

Кроме классических форм их реализации, в последнее время используются **опрос, интервью, тестирование** и другие.

Эмпирико-теоретические методы

Моделирование (простое моделирование) – получение знания о целом или о его частях с помощью модели или приборов.

Исторический метод – поиск знаний с использованием предыстории, реально существовавшей или же мыслимой.

Логический метод – поиск знаний путем воспроизведения частей, связей или элементов в мышлении.

Макетирование – получение информации по макету, представлению частей в упрощенном, но целостном виде.

Актуализация – получение информации с помощью перевода целого или его частей (а следовательно, и целого) из статического состояния в динамическое состояние.

Визуализация – получение информации с помощью наглядного или визуального представления состояний

Эмпирико-теоретические методы

Абстрагирование – выделение наиболее важных для исследования свойств, сторон исследуемого объекта, процесса, явления и игнорирование несущественных и второстепенных.

Анализ – разъединение целого на части с целью выявления их связей.

Декомпозиция – разъединение целого на части с сохранением их связей с окружением.

Синтез – соединение частей в целое с целью выявления их взаимосвязей.

Композиция — соединение частей целого с сохранением их взаимосвязей с окружением.

Индукция – получение знания о целом по знаниям о частях.

Дедукция – получение знания о частях по знаниям о целом.

Эвристики, использование эвристических процедур – получение знания о целом по знаниям о частях и по наблюдениям, опыту, интуиции, предвидению.

Теоретические методы

Восхождение от абстрактного к конкретному – получение знаний о целом или о его частях на основе знаний об абстрактных проявлениях в сознании, в мышлении.

Идеализация – получение знаний о целом или его частях путем представления в мышлении целого или частей, не существующих в действительности.

Формализация – получение знаний о целом или его частях с помощью языков искусственного происхождения (формальное описание, представление).

Аксиоматизация – получение знаний о целом или его частях с помощью некоторых аксиом (не доказываемых в данной теории утверждений) и правил получения из них (и из ранее полученных утверждений) новых верных утверждений.

Виртуализация – получение знаний о целом или его частях

Пример. Для построения модели планирования и управления производством в рамках страны, региона или крупной отрасли нужно решить следующие проблемы:

- определить структурные связи, уровни управления и принятия решений, ресурсы; (наблюдения, сравнения, измерения, эксперимента, анализа и синтеза, дедукции и индукции, эвристический, исторический и логический методы, макетирование и др.);
- определить гипотезы, цели, возможные проблемы планирования; (наблюдение, сравнение, эксперимент, абстрагирование, анализ, синтез, дедукция, индукция, эвристический, исторический, логический и др.);
- конструирование эмпирических моделей; (абстрагирование, анализ, синтез, индукция, дедукция, формализация, идеализация и др.);
- поиск решения проблемы планирования и просчет различных вариантов, директив планирования, поиск оптимального решения; (измерение, сравнение, эксперимент, анализ, синтез, индукция, дедукция, актуализация, макетирование, визуализация,

Управление системой

Суть задачи управления системой – отделение ценной *информации* от "шумов" (бесполезного, иногда даже вредного для системы возмущения *информации*) и выделение *информации*, которая позволяет этой системе существовать и развиваться.

Информационная система – это система, в которой элементы, структура, цель, ресурсы рассматриваются на информационном уровне (хотя, естественно, имеются и другие уровни рассмотрения).

Информационная среда – это среда (система и ее окружение) из взаимодействующих информационных систем, включая и информацию, актуализируемую в этих системах.

Выво

А Установление отношений и связей, описание их формальными средствами, языками, разработка соответствующих описаниям моделей, методов, алгоритмов, создание и актуализация технологий, поддерживающих эти модели и методы, и составляет **основную задачу информатики как науки, образовательной области, сферы человеческой деятельности.**

Таким образом, информатику можно определить как науку, изучающую неизменные сущности (инварианты) информационных процессов, которые протекают в различных предметных областях, в обществе, в познании, в природе.