

Тема

Информация и ее свойства

Лекция

Понятие информации и ее свойства. Неопределенность и количество информации

Вопросы лекции:

1. Определение информации. Сообщение
2. Виды информации
3. Свойства информации
4. Неопределенность, количество информации и энтропия
5. Количество информации, получаемой в процессе сообщения
6. Единицы измерения информации

Определение информации

Понятие информации включает три подхода:

- **недетерминированный** - бытовой подход, трактующий информацию, как понятие, используемое применительно к фактам и суждениям, получаемым в повседневной жизни от других живых существ, из средств массовой информации, а также путем наблюдения явлений окружающей среды.

- **техноцентрический** - отождествление информации с данными, например, фразы: "информация хранится на жестком диске, информация обрабатывается компьютером", однако станут ли эти данные информацией, целиком зависит от аппаратных, программных методов и непосредственного участия человека;

- **антропоцентрический** - отождествление информации со сведениями и фактами, которые *теоретически* могут быть получены и усвоены, т. е. преобразованы в знания. Недостаток этого понимания в том, что нет объяснения *абстрактной* информации, не имеющей адекватного отображения в природе и обществе (сфера теологии, идеализма, некоторых разделов математики, команды, в том числе команды компьютерных программ и т. п.).

Определение информации

Ни один из подходов не дает достаточно полного понимания, что же такое информация.

В современном представлении **информация** - это динамический объект, образующийся в ходе диалектического взаимодействия объективных данных и субъективных методов.

Все процессы в неживой природе протекают в виде непрерывного **энергетического обмена**, а в живой природе, кроме этого протекает **направленный обмен веществ**.

Информационный обмен - это взаимосвязь, возникающая между процессами энергетического обмена и обмена веществ.

Информационный обмен протекает в виде **информационных процессов**.

Все процессы в природе сопровождаются **сигналами**.

Зарегистрированные сигналы отражают **данные**, которые преобразуются, транспортируются и потребляются с помощью методов.

Данные имеют объективную природу - это результат регистрации объективно существовавших сигналов, вызванных изменениями в материальных телах или полях.

Определение информации

Современное научное представление об информации сформулировал Норберт Винер, "отец" кибернетики:

Информация — это обозначение содержания, полученного из внешнего мира в процессе нашего приспособления к нему и приспособления к нему наших чувств.

Люди обмениваются информацией в форме сообщений.

Сообщение — это форма представления информации в виде речи, текстов, жестов, взглядов, изображений, цифровых данных, графиков, таблиц и т.п.

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

Определение информации

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

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

В случаях, когда говорят об автоматизированной работе с информацией посредством каких-либо технических устройств, обычно в первую очередь интересуются не содержанием сообщения, а тем, *сколько символов это сообщение содержит*.

Применительно к компьютерной обработке данных под информацией понимают некоторую последовательность символических обозначений (букв, цифр, закодированных графических образов и звуков и т.п.), несущую смысловую нагрузку и представленную в понятном компьютеру виде. Каждый новый символ в такой последовательности символов увеличивает информационный объём сообщения.

Определение информации

Методы регистрации и обработки сигналов **субъективны**, т. к. они осуществляются людьми или устройствами ими созданными. Поэтому для получения достоверной информации необходимо использовать **контекстные** методы.

Контекстные методы - это методы, являющиеся **общепринятыми** для работы с данными определенного типа:

- графические данные - наблюдение;
- текстовые данные - чтение;
- данные физики - аппаратные методы;
- данные, вводимые в компьютер - программные методы и т. п.

Информационный процесс - это всегда цикл образования информации из данных и **немедленное сохранение** её в виде **новых** данных (т. е., данные – информация – новые данные и т. д.)

Само возникновение информации **очень непродолжительно**, а информационный процесс длится столько, **сколько существуют носители информации**.

Определение информации

Особенность информационного процесса в вычислительной технике в том, что большинство его этапов происходит без участия человека, **автоматически**. В ходе этих этапов данные, представленные электрическими сигналами, взаимодействуют как с аппаратными, так и с программными методами.

В контексте автоматизации в информатике присутствует **двойственность данных и методов**.

Примером являются **компьютерные программы**: с одной стороны они проявляют себя как методы, а с другой - как данные.

В активной фазе программа работает с оборудованием (аппаратом), ее команды управляют процессами компьютера, который обрабатывает данные и взаимодействует с внешним оборудованием. В пассивной фазе программа ничем не отличается от данных: ее точно также можно хранить, передавать, воспроизводить в виде текста и редактировать.

Информация возникает и существует во время диалектического взаимодействия объективных данных и субъективных методов.

Определение информации

Информация может существовать в виде:

- текстов, рисунков, чертежей, фотографий;
- световых или звуковых сигналов;
- радиоволн;
- электрических и нервных импульсов;
- магнитных записей;
- жестов и мимики;
- запахов и вкусовых ощущений;
- хромосом, посредством которых передаются по наследству признаки и свойства организмов и т.д.

Предметы, процессы, явления материального или нематериального свойства, рассматриваемые с точки зрения их информационных свойств, называются **информационными объектами**.

Виды информации

Видов информации, как и видов ее классификации существует достаточно много. Если не учитывать предметную ориентацию информации, то в основу классификации видов информации можно положить **пять наиболее основных признаков: место возникновения, стадия обработки, способ отображения, стабильность, функция управления.**

Входная (выходная) информация - это информация, поступающая в(из) объект(а). Одна и та же информация может являться одновременно входной для одного объекта и выходной для другого.

Внутренняя (внешняя) информация возникает внутри (за пределами) объекта.

Первичная информация - это та, которая возникает непосредственно в процессе деятельности объекта и регистрируется на начальной стадии. **Вторичная** информация получается в результате обработки первичной и может быть промежуточной или результативной.

Виды информации

Промежуточная используется в качестве исходных данных для последующей обработки. **Результативная** получается в процессе обработки первичной и промежуточной информации и используется для выработки решений.

Текстовая информация - это совокупность алфавитных, цифровых и специальных символов, с помощью которых информация отображается на физическом носителе. **Графическая** информация - это различного рода графики, чертежи, схемы, рисунки и т.д.

Переменная информация меняется для каждого случая как по назначению, так и количественно. **Постоянная** - это многократно используемая в течение длительного времени информация. Она может быть справочной, нормативной.

Понятно, что эта классификация весьма условна, в конкретной ситуации она может иметь другое представление. Например, по предметной ориентации информация бывает экономическая, историческая, техническая и т.д.

Передача информации. Сообщение

В информационном процессе один и тот же объект (или система) может последовательно выступать в роли приёмника, носителя и источника информации для последующих приемников.

Для того чтобы приёмник информации мог воспринять ее от источника, они должны встретиться в одной точке физического пространства, и провести вместе некоторое время.

Информация передаётся в форме сообщений от некоторого источника информации к её приёмнику посредством канала связи между ними. Источник посылает передаваемое сообщение, которое кодируется в передаваемый сигнал. Этот сигнал посылается по каналу связи. В результате в приёмнике появляется принимаемый сигнал, который декодируется и становится принимаемым сообщением.

Сообщение (в пространстве) – совместная, общая часть, непосредственный контакт (место совмещения, соединения) источника и приёмника информации. Именно при непосредственном контакте происходит изменение свойств элементов приёмника информации под воздействием источника. Если сообщение в пространстве не установлено, восприятие информации не возможно.

Передача информации. Сообщение

Воспроизведение информации с целью ее передачи, будет результативно только тогда, когда есть непосредственное сообщение с объектом, способным выступить в роли **приемника информации и дальнейшего ее носителя**.

Система сообщений (коммуникация, система связи) — последовательная цепь сообщений, обеспечивающая опосредованную передачу информации. При этом установление отдельных сообщений в системе может быть разнесено, как в пространстве, так и во времени.

В информатике под словом “**сообщение**” чаще всего понимают получаемое или предназначенное для передачи информационное послание: устное сообщение, письмо, записку, message.

Форму процесса сообщения удобно описывать в виде сигнала.

Сигнал (от лат. signum - знак) – представление процесса сообщения в виде изменяющихся во времени значений свойств элементов памяти приёмника информации.

Воспринятый сигнал есть отражение реакции приемника на поведение источника в процессе их сообщения.

Одно из самых распространенных определений **сигнала** – **процесс, несущий информацию**.

Свойства информации

1. Объективность и субъективность информации. Эти понятия всегда взаимосвязаны: то и другое относительно. В информационном процессе из-за субъективности используемых методов степень объективности, всегда понижается. Для большей объективности получаемых данных необходимо применять как можно менее субъективные методы.

2. Достоверность информации - соответствие получаемой информации объективной реальности окружающего мира. На достоверность информации влияет как объективность данных, так и адекватность методов, применяемых для ее получения.

Иногда **недостоверные** данные могут давать **достоверную** информацию, в частности, когда заранее известна степень их недостоверности. Эти методы основываются на **сортировке, фильтрации и статистическом анализе** данных.

Свойства информации

3. Адекватность информации — степень соответствия информации, полученной потребителем, тому, что автор вложил в ее содержание, т. е. в данные.

Адекватность не следует путать с достоверностью.

Например, заведомо ложное (недостоверное) сообщение в газете 1 апреля тоже адекватно, т. е. **подходит к данному времени**, т. к. трактуется не как информационное, а как развлекательное. Но это же сообщение, опубликованное 2 апреля, уже неадекватно.

Вымысел писателя-фантаста также адекватен, т. к. соответствует стилю жанра и выполняет свою функцию.

4. Актуальность - это степень соответствия информации текущему времени. Часто с актуальностью связывают коммерческую ценность информации.

Достоверная и адекватная информация может приводить к ошибочным выводам (недостоверности новой информации) из-за старения информации в течение информационного процесса. Отсюда значимость методов шифрования, аппаратных методов защиты информации, т. к. из-за продолжительности поиска алгоритмов получения информации она теряет актуальность, или практическую ценность.

Свойства информации

5. Доступность - это мера возможности получения информации. При этом имеет значение как доступность данных, так и адекватных методов для их интерпретации.

6. Полнота информации - ее достаточность для принятия управляющего решения. Она зависит как от полноты данных, так и от наличия необходимых методов.

Например, облагая всеми необходимыми данными, мы не можем принять верное решение, т. к. не владеем методом работы с данными.

7. Избыточность информации — в основном воспринимаемое получателем как положительное качество, т. к. позволяет ему меньше напрягать свое внимание и меньше утомляться.

Чем **выше** избыточность, тем **шире** диапазон методов, с помощью которых можно получить более достоверную информацию.

Свойства информации

Избыточность позволяет **повышать достоверность информации**, путем применения **специальных методов**, основанных на теории вероятностей и математической статистики. Общий принцип - в результате **фильтрации объем данных сокращается, а их достоверность увеличивается**.

Особо важное значение избыточность приобретает в информатике, где используется **автоматическая обработка данных**.

Но при этом избыточность имеет и **негативную сторону**, т. к. при увеличении объема растут затраты на хранение, обработку и транспортировку. Для уменьшения этих затрат существуют **методы сжатия данных (архивация)**.

Эффективный подбор аппаратных и программных средств и методов, необходимых для оптимального решения вопросов - одна из основных задач информатики.

Неопределенность и количество информации

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

Представьте, что вы зашли в магазин и попросили продать вам жевательную резинку. Продавщица, у которой, скажем, 16 сортов жевательной резинки, находится в состоянии неопределенности. Она не может выполнить вашу просьбу без получения дополнительной информации. Если вы уточнили, скажем, - «Orbit», и из 16 первоначальных вариантов продавщица рассматривает теперь только 8, вы уменьшили ее неопределенность в два раза (забегая вперед, скажем, что *уменьшение неопределенности вдвое соответствует получению 1 бита информации*). Если вы, просто указали пальцем на витрине, - «вот эту!», то неопределенность была снята полностью. Опять же, забегая вперед, скажем, что этим жестом в данном примере вы сообщили продавщице **4 бита информации**.

Неопределенность и количество информации

Ситуация *максимальной неопределенности* предполагает наличие нескольких *равновероятных альтернатив (вариантов)*, т.е. ни один из вариантов не является более предпочтительным.

Причем, *чем больше равновероятных вариантов* наблюдается, тем больше неопределенность, тем сложнее сделать однозначный выбор и *тем больше информации требуется* для этого получить.

Для N вариантов эта ситуация описывается следующим распределением вероятностей: $\{1/N, 1/N, \dots, 1/N\}$.

Минимальная неопределенность равна 0, т.е. эта ситуация *полной определенности*, означающая что выбор сделан, и вся необходимая информация получена. Распределение вероятностей для ситуации полной определенности выглядит так: $\{1, 0, \dots, 0\}$.

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

Энтропия (H) – мера неопределенности, выраженная в битах. Так же энтропию можно рассматривать как *меру равномерности распределения* случайной величины.

Неопределенность и количество информации

Количество информации I и **энтропия H** характеризуют одну и ту же ситуацию, но с качественно противоположенных сторон. I – это количество информации, которое требуется для снятия неопределенности H .

Когда неопределенность снята полностью, количество полученной информации I равно изначально существовавшей неопределенности H .

При частичном снятии неопределенности, полученное количество информации и оставшаяся неснятой неопределенность составляют в сумме исходную неопределенность. $H_t + I_t = H$.

По этой причине, **формулы для расчета энтропии H** являются и **формулами для расчета количества информации I** , т.е. когда речь идет о **полном снятии неопределенности**, H в них может заменяться на I .

В общем случае, энтропия H и количество получаемой в результате снятия неопределенности информации I 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. $H=F(N, P)$. Расчет энтропии в этом случае производится **по формуле Шеннона**.

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

Формула Хартли

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

$$H = \sum_{i=0}^{N-1} \frac{1}{N} \log_2 \left(\frac{N}{1} \right) = \frac{1}{N} * N * \log_2 (N) = \log_2 (N) \quad (2)$$

Из нее явно следует, что чем больше количество альтернатив (N), тем больше неопределенность (H). Эти величины связаны в формуле (2) не линейно, а через двоичный логарифм. Логарифмирование по основанию 2 и приводит количество вариантов к единицам измерения информации – битам. Заметьте, что энтропия будет являться целым числом лишь в том случае, если N является степенью числа 2, т.е. если N принадлежит ряду: $\{1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048 \dots\}$.

Для решения обратных задач, когда известна неопределенность (H) или полученное в результате ее снятия количество информации (I) и нужно определить какое количество равновероятных альтернатив соответствует возникновению этой неопределенности, используют обратную формулу Хартли, которая выводится в соответствии с определением логарифма и выглядит еще проще:

$$N = 2^H \quad (3)$$

Количество информации, получаемой в процессе сообщения

До сих пор мы приводили формулы для расчета энтропии (неопределенности) H , указывая, что H в них можно заменять на I , потому что количество информации, получаемое *при полном снятии неопределенности* некоторой ситуации, количественно равно начальной энтропии этой ситуации.

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

$$I = H_{\text{до}} - H_{\text{после}} \quad (4)$$

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

$$I = \log_2(N_{\text{до}}) - \log_2(N_{\text{после}}) = \log_2\left(\frac{N_{\text{до}}}{N_{\text{после}}}\right) \quad (5)$$

Таким образом, в равновероятном случае I зависит от того, во сколько раз изменилось количество рассматриваемых вариантов выбора (рассматриваемое разнообразие).

Количество информации, получаемой в процессе сообщения

Исходя из (5) можно вывести следующее:

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

полное снятие неопределенности, количество полученной в сообщении информации равно неопределенности, которая существовала до получения сообщения.

Если $N_{\text{после}} = N_{\text{до}}$, то - $I = \log_2 (1) = 0$
неопределенность не изменилась, следовательно, информации получено не было.

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

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

Т.е. количество полученной информации будет положительной величиной, если в результате получения сообщения количество рассматриваемых альтернатив уменьшилось, и отрицательной, если увеличилось.

Единицы измерения информации

В качестве единицы информации Клод Шеннон предложил принять один **бит** (англ. bit — **b**inary **d**igit — двоичная цифра).

Бит в теории информации — количество информации, необходимое для различения двух равновероятных сообщений (типа "орел"—"решка", "чет"—"нечет" и т.п.).

В вычислительной технике битом называют наименьшую "порцию" памяти компьютера, необходимую для хранения одного из двух знаков "0" и "1", используемых для внутримашинного представления данных и команд.

На практике чаще применяется более крупная единица — **байт**, равная восьми битам. Именно восемь битов требуется для того, чтобы закодировать любой из 256 символов алфавита клавиатуры компьютера ($2^8=256$).

Широко используются также ещё более крупные производные единицы информации:

- 1 Килобайт (Кбайт) = 1024 байт = 2^{10} байт,
- 1 Мегабайт (Мбайт) = 1024 Кбайт = 2^{20} байт,
- 1 Гигабайт (Гбайт) = 1024 Мбайт = 2^{30} байт.
- 1 Терабайт (Тбайт) = 1024 Гбайт = 2^{40} байт,
- 1 Петабайт (Пбайт) = 1024 Тбайт = 2^{50} байт.

Единицы измерения информации

Единицы измерения информации, в названии которых есть приставки «кило», «мега» и т.д., с точки зрения теории измерений не является корректными, поскольку эти приставки используются в метрической системе мер, в которой в качестве множителей кратных единиц используется коэффициент 10^n , где $n=3, 6, 9\dots$ Для устранения этой некорректности международная организация *International Electrotechnical Commission*, занимающаяся созданием стандартов для отрасли электронных технологий, утвердила ряд новых приставок для единиц измерения количества информации: *киби* (*kibi*), *меби* (*mebi*), *гиби* (*gibi*), *теби* (*tebi*), *пету* (*peti*), *эксби* (*exbi*). Однако пока используются старые обозначения единиц измерения количества информации, и требуется время, чтобы новые названия начали широко применяться.

Вероятностный подход используется и при определении количества информации, представленной с помощью знаковых систем. Если рассматривать символы алфавита как множество возможных сообщений N , то количество информации, которое несет один знак алфавита, можно определить по формуле Шеннона (1).

Единицы измерения информации

При равновероятном появлении каждого знака алфавита в тексте сообщения для определения количества информации можно воспользоваться формулой Хартли (2).

Количество информации, которое несет один знак алфавита, тем больше, чем больше знаков входит в этот алфавит. Количество знаков, входящих в алфавит, называется **мощностью алфавита**.

Количество информации (информационный объем), содержащееся в сообщении, закодированном с помощью знаковой системы и содержащем определенное количество знаков (символов), определяется с помощью формулы: $V=I \cdot K$,

где V – информационный объем сообщения;

$I = \log_2 N$ - информационный объем одного символа (знака);

K – количество символов (знаков) в сообщении;

N – мощность алфавита.