

Предмет информатики
Информация, измерение,
единицы измерения

Информатика – это дисциплина, изучающая структуру и общие свойства информации, закономерности и методы её создания, хранения, поиска, преобразования, передачи и применения в различных сферах человеческой деятельности.

Возникновение термина

Термин информатика возник в 60-х гг. во Франции для названия области, занимающейся автоматизированной обработкой информации с помощью электронных вычислительных машин. Французский термин *informatique* (информатика) образован путем слияния слов *information* (информация) и *automatique* (автоматика) и означает "информационная автоматика или автоматизированная переработка информации". В англоязычных странах этому термину соответствует синоним *computer science* (наука о компьютерной технике).

В 1963 г. Ф.Е. Темнников (1906-1993), предложил термин ИНФОРМАТИКА в широком смысле - применительно к системам сбора, передачи, хранения и обработки данных: «Информатика – это “научная дисциплина, связывающая вопросы сбора, передачи, обращения, переработки и использования информации”»

Определил информатику как науку об информации вообще, состоящую из трех частей: 1) теории информационных элементов; 2) теории информационных процессов и 3) теории информационных систем

Ф.Е. Темников (1906-1993), д.т.н., профессор – профессор МЭИ, выдающийся советский и российский учёный-энциклопедист и педагог.

Широко известен как основоположник отечественной телемеханики – создатель теории развёртывающихся систем (центротехники) – концептуальной, алгоритмической и схемотехнической основы импульсных и цифровых АСУТП – систем централизованного измерения, контроля и управления сложными технологическими объектами.

Менее известен как один из основоположников информатики и системотехники.

Автор книг «Теория развёртывающихся систем» и «Теоретические основы информационной техники».

Выделение информатики как самостоятельной области человеческой деятельности в первую очередь связано с развитием компьютерной техники.

Причем основная заслуга в этом принадлежит микропроцессорной технике, появление которой в середине 70-х гг. послужило началом второй электронной революции. С этого времени элементной базой вычислительной машины становятся интегральные схемы и микропроцессоры, а область, связанная с созданием и использованием компьютеров, получила мощный импульс в своем развитии.

Термин "информатика" приобретает новое дыхание и используется не только для отображения достижений компьютерной техники, но и связывается с процессами передачи и обработки информации.

В нашей стране в 1983 г. на сессии годовичного собрания Академии наук СССР об организации нового отделения информатики, вычислительной техники и автоматизации.

Информатика трактовалась как "комплексная научная и инженерная дисциплина, изучающая все аспекты разработки, проектирования, создания, оценки, функционирования основанных на ЭВМ систем переработки информации, их применения и воздействия на различные области социальной практики".

Информатика в таком понимании нацелена на разработку общих методологических принципов построения информационных моделей. Поэтому методы информатики применимы всюду, где существует возможность описания объекта, явления, процесса и т.п. с помощью информационных моделей. Существует множество определений информатики, что связано с многогранностью ее функций, возможностей, средств и методов.

Информатика - это область человеческой деятельности, связанная с процессами преобразования информации с помощью компьютеров и их взаимодействием со средой применения.

Информация

Понятие *информация* точно и однозначно не определяется, хотя используется повсеместно. Оно вводится путём объяснения, которое опирается на интуицию, здравый смысл или бытовое применение этого термина.

Термин *информация* происходит от латинского *informatio* , что означает разъяснение, осведомление, изложение.

В Федеральном законе Российской Федерации от 27 июля 2006г. №149-ФЗ «Об информации, информационных технологиях и о защите информации»

(<http://www.rg.ru/2006/07/29/informacia-dok.html>) дается следующее определение этого термина:

«информация — сведения (сообщения, данные) независимо от формы их представления».

Толковый словарь русского языка Ожегова

1. Сведения об окружающем мире и протекающих в нем процессах, воспринимаемые человеком или специальным устройством.
2. Сообщения, осведомляющие о положении дел, о состоянии чего-нибудь. (Научно-техническая и газетная информация, средства массовой информации — печать, радио, телевидение, кино).

Различные трактовки в разных предметных областях

Например, информация может пониматься как:

- 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.

Сигнал (сообщение) – информационный поток, который в процессе передачи информации поступает к приёмнику.

Данные – это зарегистрированные на материальном носителе сигналы.

Одна и та же информация может передаваться с помощью разных сообщений. Например, сведения о выпуске книги могут быть переданы с помощью телевидения, устного разговора, рекламных щитов и т.д.

И, наоборот, одно и то же сообщение может нести различную информацию. Пример: сообщение на китайском языке несёт какую-то информацию только для тех, кто этот язык знает.

Энтропия и информация

Энтропия – это мера неопределённости наших знаний об объекте или явлении.

Энтропию иногда называют антиинформацией.

Например, если мы интересуемся сведениями о полностью засекреченном объекте или явлении, или получили зашифрованное сообщение, а ключа к расшифровке не знаем, то для нас информация о нём равна нулю, а энтропия – максимальна.

Количество информации как мера уменьшения

неопределенности знаний

Подход к информации как к мере уменьшения неопределённости наших знаний позволяет количественно измерять информацию, полученную через некоторое сообщение.

Например, после сдачи зачета Вы получаете одно из двух информационных сообщений: "зачет" или "незачет", а после сдачи эк-замена одно из четырех информационных сообщений: "2", "3", "4" или "5".

Информационное сообщение об оценке за зачет приводит к уменьшению неопределенности вашего знания в два раза, так как реализуется один из двух возможных вариантов. Информационное сообщение об оценке за экзамен приводит к уменьшению неопределенности вашего знания в четыре раза, так как получено одно из четырех возможных информационных сообщений.

Формула определения количества информации

Клод Шеннон предложил в 1948 году формулу для определения количества информации, которую мы получаем после получения одного из N возможных сообщений:

$$I = - (p_1 \log_2 p_1 + p_2 \log_2 p_2 + \dots + p_i \log_2 p_i + \dots + p_N \log_2 p_N)$$

Здесь p_i – вероятность того, что будет получено именно i -е сообщение.

формула Хартли

Если все сообщения равновероятны, то
все $p_i = 1/N$,

и из этой формулы получается формула
Хартли:

$$I = \log_2 N$$

- Из формулы Хартли следует: если $I=1$, то $N=2$, то есть в качестве единицы измерения информации можно взять тот объём информации, который мы получаем при принятии сигнала о том, что же произошло в ситуации с двумя возможными исходами. Такая единица названа ***битом***.

Наряду с единицей бит иногда используют в качестве единиц информации количества, взятые по логарифмам с другими основаниями:

дит – по десятичному логарифму (за единицу информации выбирается количество информации, необходимой для различения десяти равновероятных сообщений),

трип – по основанию три,

нат – по натуральному основанию.

Используя формулу Хартли можно, также, зная количество информации, пришедшее с одним из равновероятных сообщений, определить, сколько сообщений вообще можно было ожидать в данной ситуации. Решив это уравнение относительно N , получим при равновероятных исходах:

$$I = \log_2 N$$

$$N = 2^I$$

Например, на экзамене вы берете экзаменационный билет, и учитель сообщает, что зрительное информационное сообщение о его номере несет 5 бит информации.

Если вы хотите определить количество экзаменационных билетов, то достаточно определить количество возможных информационных сообщений об их номерах из формулы Хартли:

$$5 = \log_2 N$$

$$N = 2^5 = 32.$$

Алфавитный подход к определению количества информации

Представим себе, что необходимо передать информационное сообщение по каналу передачи информации от отправителя к получателю. Пусть сообщение кодируется с помощью знаковой системы, алфавит которой состоит из N знаков $\{1, \dots, N\}$ и вероятности появления каждого знака в сообщении равны.

В простейшем случае, когда длина кода сообщения составляет один знак, отправитель может послать N разных сообщений. Количество информации I , которое несет каждое сообщение, то есть один знак, можно рассчитать по формуле Хартли

$$I = \log_2 N$$

Эта величина называется ***информационной емкостью знака***.

С помощью этой формулы можно, например, определить информационную емкость знака двоичной знаковой системы:

$$I = \log_2 2 = 1 \text{ бит}$$

Единица измерения количества информации "бит" (bit) получила свое название от английского словосочетания "binary digit" – "двоичная цифра"

Чем большее количество знаков содержит алфавит знаковой системы, тем большее количество информации несет один знак.

В качестве примера определим количество информации, которое несет буква русского алфавита. В русский алфавит входят 33 буквы, однако на практике часто для передачи сообщений используются только 32 буквы (исключается буква "ё").

С помощью формулы Хартли определим количество информации, которое несет буква русского алфавита:

$$N = 32 \quad I = \log_2 32$$

$$I = \log_2 2^5 \quad I = 5 \text{ бит.}$$

Таким образом, информационная емкость буквы русского алфавита равна 5 битам (если считать, что все буквы используются в сообщении с равной вероятностью).

Количество информации в сообщении

Сообщение состоит из последовательности знаков, каждый из которых несет определенное количество информации.

Если знаки несут одинаковое количество информации, то количество информации I_c в сообщении можно подсчитать, умножив количество информации I_z , которое несет один знак, на длину кода K (количество знаков в сообщении):

$$I_c = I_z * K$$

Например, каждая цифра двоичного компьютерного кода несет информацию в 1 бит.

Следовательно, две цифры несут информацию в 2 бита, три цифры - в 3 бита и т. д.

Количество информации в битах равно количеству цифр двоичного компьютерного кода

В русской письменной речи частота использования букв в тексте различна, так в среднем на 1000 знаков осмысленного текста приходится 200 букв "а" и в сто раз меньшее количество буквы "ф" (всего 2).

Таким образом, с точки зрения теории информации, информационная емкость знаков русского алфавита различна (у буквы "а" она наименьшая, а у буквы "ф" - наибольшая) и информацию, которое несёт текстовое сообщение, надо рассчитывать с учетом вероятности появления букв, входящих в него.

Бит – это минимальная единица измерения количества информации.

Более крупные единицы формируются в информатике способом, который несколько отличается от принятых в большинстве наук. Первой более крупной, чем бит, единицей измерения информации, выбран ***байт***:

$$1 \text{ байт} (1\text{В}) = 8 \text{ бит} = 2^3 \text{ бит}$$

Почему реальный объем жесткого диска меньше, чем указан на этикетке диска?

У меня два жестких диска 160 и 320 Gb. Система определяет 149 и 298Gb соответственно. Где болтаются остальные Gb не понятно. Как мне найти потерявшиеся Gb?

Десятичные приставки (Система СИ)

1 килобайт (1 kB) = 1000 В (1000 байт);

1 мегабайт (1 MB) = 1000 kB ;

1 гигабайт (1 GB) = 1000 MB;

1 терабайт (1 TB) = 1000 GB;

1 петабайт (1 PB) = 1000 TB;

1 эксабайт (1 EB) = 1000 PB;

1 зеттабайт (1 ZB) = 1000 EB;

1 йоттабайт(1 YB) = 1000 ZB.

Двоичные приставки

- 1 кибибайт (1 Кбайт, 1 KiB) = 2^{10} байт = 1024 байт;
- 1 мебибайт (1 Мбайт, 1 MiB) = 2^{20} байт = 1024 Кбайт;
- 1 гигабайт (1 Гбайт, 1 GiB) = 2^{30} байт = 1024 Мбайт;
- 1 тебибайт (1 Тбайт, 1 TiB) = 2^{40} байт = 1024 Гбайт;
- 1 пебибайт (1 Пбайт, 1 PiB) = 2^{50} байт = 1024 Тбайт;
- 1 эксбибайт (1 Эбайт, 1 EiB) = 2^{60} байт = 1024 Пбайт;
- 1 зебибайт (1 Збайт, 1 ZiB) = 2^{70} байт = 1024 Эбайт;
- 1 йобибайт (1 Йбайт, 1 YiB) = 2^{80} байт = 1024 Збайт.

В десятичной системы приставка кило- означает 10^3 , мега- 10^6 , гига- 10^9 , а в двоичной приставки кило- 2^{10} , мега- 2^{20} , и гига- 2^{30} таким образом в одном случае 1 килобайт равен 1000, а в другом 1024 байта.

Если у вас жесткий диск размером 160 Гбайт (указано на самом диске), то его объем равен 160×10^9 в байтах, теперь если это число поделить 2^{30} , мы получим размер в гигабайтах "настоящих" компьютерных - 149 гигабайт

Количество информации I на синтаксическом уровне(1)

Количество информации I на синтаксическом уровне невозможно определить без рассмотрения понятия неопределенности состояния системы (энтропии системы). Получение информации о какой-либо системе всегда связано с изменением степени неосведомленности получателя о состоянии этой системы.

Пусть до получения информации потребитель имеет некоторые предварительные (априорные) сведения о системе α . Мерой его неосведомленности о системе является функция $H(\alpha)$, которая в то же время служит и мерой неопределенности состояния системы.

Количество информации I на синтаксическом уровне(2)

После получения некоторого сообщения β получатель приобрел некоторую дополнительную информацию $I_{\beta}(\alpha)$, уменьшившую его априорную неосведомленность так, что апостериорная (после получения сообщения β) неопределенность состояния системы стала $H_{\beta}(\alpha)$.

Тогда количество информации $I_{\beta}(\alpha)$ о системе, полученной в сообщении β , определится как

$$I_{\beta}(\alpha) = H(\alpha) - H_{\beta}(\alpha),$$

т.е. количество информации измеряется изменением (уменьшением) неопределенности состояния системы.

Если конечная неопределенность $H_\beta(\alpha)$ обратится в нуль, то первоначальное неполное знание заменится полным знанием и количество информации $I_\beta(\alpha) = H(\alpha)$.

Иными словами, энтропия системы $H(\alpha)$ может рассматриваться как мера недостающей информации.

Энтропия системы $H(\alpha)$, имеющая N возможных состояний, согласно формуле Шеннона, равна:

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

где p_i — вероятность того, что система находится в i -м состоянии.

Для случая, когда все состояния системы равновероятны, т. е. их вероятности равны $p_i = 1/N$, ее энтропия определяется соотношением:

$$I = \log_2 N$$

Коэффициент информативности (лаконичность) сообщения

Коэффициент (степень) информативности (лаконичность) сообщения определяется отношением количества информации к объему данных, т.е.

$$Y = I / V_d, \text{ причем } 0 < Y < 1$$

С увеличением Y уменьшаются объемы работы по преобразованию информации в системе.

Поэтому стремятся к повышению информативности, для чего разрабатываются специальные методы оптимального кодирования информации.

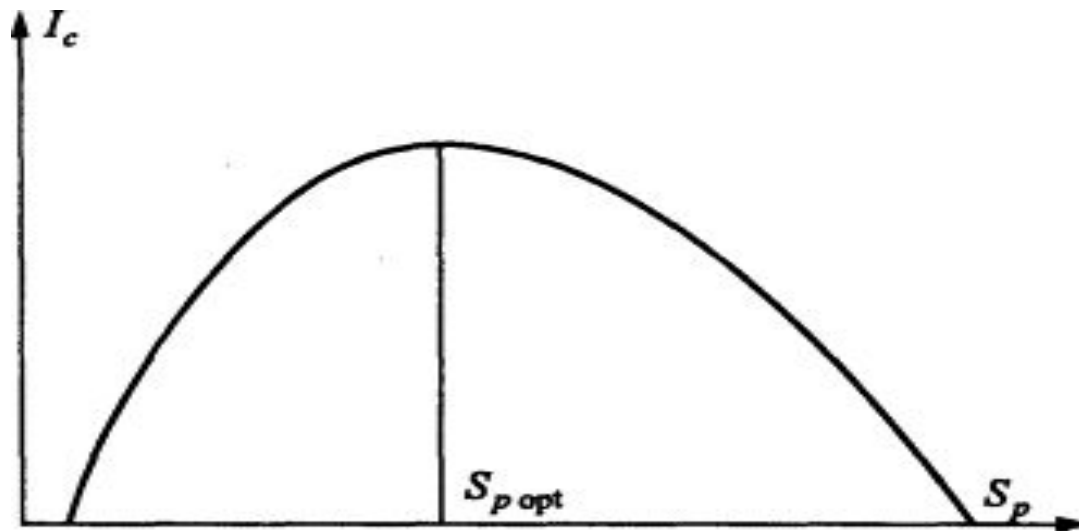
Семантическая мера информации

Для измерения смыслового содержания информации, т.е. ее количества на семантическом уровне, наибольшее признание получила тезаурусная мера, которая связывает семантические свойства информации со способностью пользователя принимать поступившее сообщение.

Для этого используется понятие тезаурус пользователя.

Тезаурус — это совокупность сведений, которыми располагает пользователь или система.

В зависимости от соотношений между смысловым содержанием информации S и тезаурусом пользователя S_p изменяется количество семантической информации I_s , воспринимаемой пользователем и включаемой им в дальнейшем в свой тезаурус.



Характер такой зависимости показан на рисунке. Рассмотрим два предельных случая, когда количество семантической информации I_c равно 0:

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.

- Максимальное количество семантической информации I_s потребитель приобретает при согласовании ее смыслового содержания S со своим тезаурусом S_p ($S_p = S_p \text{ opt}$), когда поступающая информация понятна пользователю и несет ему ранее не известные (отсутствующие в его тезаурусе) сведения.
- При оценке семантического (содержательного) аспекта информации необходимо стремиться к согласованию величин S и S_p .

Следовательно, количество семантической информации в сообщении, количество новых знаний, получаемых пользователем, является величиной относительной.

Одно и то же сообщение может иметь смысловое содержание для компетентного пользователя и быть бессмысленным (семантический шум) для пользователя некомпетентного.

Относительной мерой количества семантической информации может служить коэффициент содержательности C , который определяется как отношение количества семантической информации к ее объему:

$$C = I_c / V$$

Прагматическая мера информации

Эта мера определяет полезность информации (ценность) для достижения пользователем поставленной цели.

Эта мера также величина относительная, обусловленная особенностями использования этой информации в той или иной системе. Ценность информации целесообразно измерять в тех же самых единицах (или близких к ним), в которых измеряется целевая функция, например в экономической системе – экономический эффект

Мера информации

Единицы измерения

Примеры(для компьютерной области)

Синтаксическая:
шенноновский подход
Компьютерный(содержательный,
алфавитный) подход

Степень уменьшения
неопределенности
Единицы представления
информации

Вероятность события Бит, байт.
Кбайт и т.д.

Семантическая

Тезаурус
Экономические показатели

Пакет прикладных программ,
персональный компьютер,
компьютерные сети и т.д.
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

Спасибо за внимание!

