

ИНФОРМАЦИЯ И ЕЕ СВОЙСТВА

ИНФОРМАЦИЯ И ДАННЫЕ

Термин *информация* происходит от латинского *informatio*, что означает разъяснение, осведомление, изложение. С позиции материалистической философии информация есть отражение реального мира с помощью сведений (сообщений). Сообщение - это форма представления информации в виде речи, текста, изображения, цифровых данных, графиков, таблиц и т.п. В широком смысле информация - это общенаучное понятие, включающее в себя обмен сведениями между людьми, обмен сигналами между живой и неживой природой, людьми и устройствами.

Информация - сведения об объектах и явлениях окружающей среды, их параметрах, свойствах и состоянии, которые уменьшают имеющуюся о них степень неопределенности, неполноты знаний.

Информатика рассматривает информацию как концептуально связанные между собой сведения, данные, понятия, изменяющие наши представления о явлении или объекте окружающего мира. Наряду с информацией в информатике часто употребляется понятие **данные**

При работе с информацией всегда имеется ее источник и потребитель (получатель).

Пути и процессы, обеспечивающие передачу сообщений от источника информации к ее потребителю, называются **информационными коммуникациями**.

ФОРМЫ АДЕКВАТНОСТИ ИНФОРМАЦИИ

Адекватность информации - это определенный уровень соответствия создаваемого с помощью полученной информации образа реальному объекту, процессу, явлению и т.п.

Адекватность информации может выражаться в трех формах: семантической, синтаксической, прагматической.

Синтаксическая адекватность. Она отображает формально-структурные характеристики информации и не затрагивает ее смыслового содержания. На синтаксическом уровне учитываются тип носителя и способ представления информации, скорость передачи и обработки, размеры кодов представления информации, надежность и точность преобразования этих кодов и т.п. Информацию, рассматриваемую только с синтаксических позиций, обычно называют данными, так как при этом не имеет значения смысловая сторона. Эта форма способствует восприятию внешних структурных характеристик, т.е. синтаксической стороны информации.

Семантическая (смысловая) адекватность. Эта форма определяет степень соответствия образа объекта и самого объекта. Семантический аспект предполагает учет смыслового содержания информации. На этом уровне анализируются те сведения, которые отражает информация, рассматриваются смысловые связи. В информатике устанавливаются смысловые связи между кодами представления информации. Эта форма служит для формирования понятий и представлений, выявления смысла, содержания информации и ее обобщения.

Прагматическая (потребительская) адекватность. Она отражает отношение информации и ее потребителя, соответствие информации цели управления, которая на ее основе реализуется. Проявляются прагматические свойства информации только при наличии единства информации (объекта), пользователя и цели управления. Прагматический аспект рассмотрения связан с ценностью, полезностью использования информации при выработке потребителем решения для достижения своей цели. С этой точки зрения анализируются потребительские свойства информации. Эта форма адекватности непосредственно связана с практическим использованием информации, с соответствием ее целевой функции деятельности системы.

МЕРЫ ИНФОРМАЦИИ

Классификация мер

Для измерения информации вводятся два параметра: количество информации I и объем данных V_d .

Эти параметры имеют разные выражения и интерпретацию в зависимости от рассматриваемой формы адекватности. Каждой форме адекватности соответствует своя мера количества информации и объема данных (рис. 2.1).



Рис. 2.1. Меры информации

Синтаксическая мера информации

Эта мера количества информации оперирует с обезличенной информацией, не выражающей смыслового отношения к объекту.

Объем данных V_d . в сообщении измеряется количеством символов (разрядов) в этом сообщении. В различных системах счисления один разряд имеет различный вес и соответственно меняется единица измерения данных:

в двоичной системе счисления единица измерения - бит (*bit - binary digit* - двоичный разряд);

Примечание. В современных ЭВМ наряду с минимальной единицей измерения данных "бит" широко используется укрупненная единица измерения "байт", равная 8 бит.

в десятичной системе счисления единица измерения - дит (десятичный разряд).

Пример 2.3. Сообщение в двоичной системе в виде восьмиразрядного двоичного кода 10111011 имеет объем данных $V_d=8$ бит.

Сообщение в десятичной системе в виде шестиразрядного числа 275903 имеет объем данных $V_d=6$ дит.

Количество информации - на синтаксическом уровне невозможно определить без рассмотрения понятия неопределенности состояния системы (энтропии системы). Действительно, получение информации о какой-либо системе всегда связано с изменением степени неосведомленности получателя о состоянии этой системы. Рассмотрим это понятие.

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

После получения некоторого сообщения b получатель приобрел некоторую дополнительную информацию $I_b(a)$, уменьшившую его априорную неосведомленность так, что апостериорная (после получения сообщения b) неопределенность состояния системы стала $H_b(a)$.

Тогда количество информации $I_b(a)$ о системе, полученной в сообщении b , определится как

$$I_b(a) = H(a) - H_b(a),$$

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

Если конечная неопределенность $H_b(a)$ обратится в нуль, то первоначальное неполное знание заменится полным знанием и количество информации $I_b(a) = H(a)$. Иными

словами, **энтропия системы** $H(a)$ может рассматриваться как мера недостающей информации.

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

$$H(\alpha) = - \sum_{i=1}^N P_i \log P_i,$$

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

$$H(\alpha) = \sum_{i=1}^N \frac{1}{N} \log \frac{1}{N}.$$

Часто информация кодируется числовыми кодами в той или иной системе счисления, особенно это актуально при представлении информации в компьютере. Естественно, что одно и то же количество разрядов в разных системах счисления может передать разное число состояний отображаемого объекта, что можно представить в виде соотношения

$$N = m^n,$$

где N - число всевозможных отображаемых состояний;

m - основание системы счисления (разнообразие символов, применяемых в алфавите);

n - число разрядов (символов) в сообщении.

Пример 2.4. По каналу связи передается n -разрядное сообщение, использующее m различных символов. Так как количество всевозможных кодовых комбинаций будет $N = m^n$, то при равновероятности появления любой из них количество информации, приобретенной абонентом в результате получения сообщения, будет $I = \log N = \log m^n$ - формула Хартли.

Если в качестве основания логарифма принять m , то $I = n$. В данном случае количество информации (при условии полного априорного незнания абонентом содержания сообщения) будет равно объему данных $I = V_d$, полученных по каналу связи. Для неравновероятных состояний системы всегда $I < V_d = n$.

Наиболее часто используются двоичные и десятичные логарифмы. Единицами измерения в этих случаях будут соответственно бит и дит.

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

$$Y = \frac{I}{V_d},$$

причем $0 < Y < 1$

С увеличением Y уменьшаются объемы работы по преобразованию информации (данных) 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_c , воспринимаемой пользователем и включаемой им в дальнейшем в свой тезаурус.

Характер такой зависимости показан на рис.2.2. Рассмотрим два предельных случая, когда количество семантической информации I_c равно 0:
 при $S_p = 0$ пользователь не воспринимает, не понимает поступающую информацию;
 при $S_p = S_0$; пользователь все знает, и поступающая информация ему не нужна.

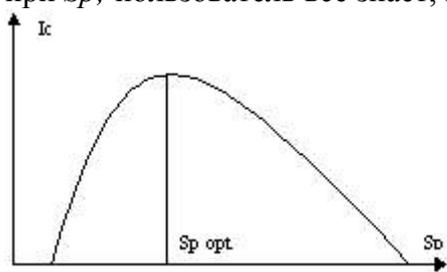


Рис. 2.2. Зависимость количества семантической информации, воспринимаемой потребителем, от его тезауруса $I_c = f(S_p)$

Максимальное количество семантической информации I_c 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.

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.

При оценке семантического (содержательного) аспекта информации необходимо стремиться к согласованию величин S и S_p .

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

$$C = \frac{I_c}{V_0}$$

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.

Пример 2.5. В экономической системе прагматические свойства (ценность) информации можно определить приростом экономического эффекта функционирования, достигнутым благодаря использованию этой информации для управления системой:

$$Inb(g) = \Pi(g/b) - \Pi(g),$$

где $Inb(g)$ - ценность информационного сообщения b для системы управления g ,

$\Pi(g)$ - априорный ожидаемый экономический эффект функционирования системы управления g ,

$\Pi(g/b)$ - ожидаемый эффект функционирования системы g при условии, что для управления будет использована информация, содержащаяся в сообщении b .

Для сопоставления введенные меры информации представим в табл.2.1.

Таблица 2.1. Единицы измерения информации и примеры

Мера информации	Единицы измерения	Примеры (для компьютерной области)
Синтаксическая: шенноновский подход компьютерный подход	Степень уменьшения неопределенности Единицы представления информации	Вероятность события Бит, байт 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. Время обработки информации и принятия решений

КАЧЕСТВО ИНФОРМАЦИИ

Возможность и эффективность использования информации обуславливаются такими основными ее потребительскими **показателями качества**, как репрезентативность, содержательность, достаточность, доступность, актуальность, своевременность, точность, достоверность, устойчивость.

- **Репрезентативность** информации связана с правильностью ее отбора и формирования в целях адекватного отражения свойств объекта. Важнейшее значение здесь имеют:
 - правильность концепции, на базе которой сформулировано исходное понятие;
 - обоснованность отбора существенных признаков и связей отображаемого явления.
- Нарушение репрезентативности информации приводит нередко к существенным ее погрешностям.
- **Содержательность** информации отражает семантическую емкость, равную отношению количества семантической информации в сообщении к объему обрабатываемых данных, т.е. $C=Ic/V\partial$.

С увеличением содержательности информации растет семантическая пропускная способность информационной системы, так как для получения одних и тех же сведений требуется преобразовать меньший объем данных.

Наряду с коэффициентом содержательности C , отражающим семантический аспект, можно использовать и коэффициент информативности, характеризующийся отношением количества синтаксической информации (по Шеннону) к объему данных $Y=I/V\partial$.

- **Достаточность (полнота)** информации означает, что она содержит минимальный, но достаточный для принятия правильного решения состав (набор показателей). Понятие полноты информации связано с ее смысловым содержанием (семантикой) и прагматикой. Как неполная, т.е. недостаточная для принятия правильного решения, так и избыточная информация снижает эффективность принимаемых пользователем решений.
- **Доступность** информации восприятию пользователя обеспечивается выполнением соответствующих процедур ее получения и преобразования. Например, в информационной системе информация преобразовывается к

доступной и удобной для восприятия пользователя форме. Это достигается, в частности, и путем согласования ее семантической формы с тезаурусом пользователя.

- **Актуальность** информации определяется степенью сохранения ценности информации для управления в момент ее использования и зависит от динамики изменения ее характеристик и от интервала времени, прошедшего с момента возникновения данной информации.
- **Своевременность** информации означает ее поступление не позже заранее назначенного момента времени, согласованного с временем решения поставленной задачи.
- **Точность** информации определяется степенью близости получаемой информации к реальному состоянию объекта, процесса, явления и т.п. Для информации, отображаемой цифровым кодом, известны четыре классификационных понятия точности:
 - формальная точность, измеряемая значением единицы младшего разряда числа;
 - реальная точность, определяемая значением единицы последнего разряда числа, верность которого гарантируется;
 - максимальная точность, которую можно получить в конкретных условиях функционирования системы;
 - необходимая точность, определяемая функциональным назначением показателя.

Достоверность информации определяется ее свойством отражать реально существующие объекты с необходимой точностью. Измеряется достоверность информации доверительной вероятностью необходимой точности, т.е. вероятностью того, что отображаемое информацией значение параметра отличается от истинного значения этого параметра в пределах необходимой точности.

Устойчивость информации отражает ее способность реагировать на изменения исходных данных без нарушения необходимой точности. Устойчивость информации, как и репрезентативность, обусловлена выбранной методикой ее отбора и формирования.

В заключение следует отметить, что такие параметры качества информации, как репрезентативность, содержательность, достаточность, доступность, устойчивость, целиком определяются на методическом уровне разработки информационных систем. Параметры актуальности, своевременности, точности и достоверности обуславливаются в большей степени также на методическом уровне, однако на их величину существенно влияет и характер функционирования системы, в первую очередь ее надежность. При этом параметры актуальности и точности жестко связаны соответственно с параметрами своевременности и достоверности.

ПРЕДСТАВЛЕНИЕ ИНФОРМАЦИИ В ЭВМ

Системы счисления и формы представления чисел

Информация в ЭВМ кодируется, как правило, в двоичной или в двоично-десятичной системе счисления.

Система счисления - это способ наименования и изображения чисел с помощью символов, имеющих определенные количественные значения.

В зависимости от способа изображения чисел системы счисления делятся на позиционные и непозиционные.

В *позиционной системе* счисления количественное значение каждой цифры зависит от ее места (позиции) в числе. В *непозиционной* системе счисления цифры не меняют своего количественного значения при изменении их расположения в числе. Количество (P) различных цифр, используемых для изображения числа в позиционной системе счисления, называется основанием системы счисления. Значения цифр лежат в пределах от 0 до $P-1$. В общем случае запись любого смешанного числа в системе счисления с основанием P будет представлять собой ряд вида:

$$a_{m-1}P^{m-1} + a_{m-2}P^{m-2} + \dots + a_1P^1 + a_0P^0 + a_{-1}P^{-1} + a_{-2}P^{-2} + \dots + a_{-s}P^{-s}, \quad (1)$$

где нижние индексы определяют местоположение цифры в числе (разряд):

- положительные значения индексов - для целой части числа (m разрядов);
- отрицательные значения - для дробной (s разрядов).

Пример 4.1. Позиционная система счисления - арабская десятичная система, в которой: основание $P=10$, для изображения чисел используются 10 цифр (от 0 до 9). Непозиционная система счисления - римская, в которой для каждого числа используется специфическое сочетание символов (XIV, CXXVII и т.п.).

Двоичная система счисления имеет основание $P=2$ и использует для представления информации всего две цифры: 0 и 1. Существуют правила перевода чисел из одной системы счисления в другую, основанные в том числе и на соотношении (1).

Пример 4.2.

$$101110,101_{(2)} = 1*2^5 + 0*2^4 + 1*2^3 + 1*2^2 + 1*2^1 + 0*2^0 + 1*2^{-1} + 0*2^{-2} + 1*2^{-3} = 46,625_{(10)},$$

т.е. двоичное число 101110,101 равно десятичному числу 46,625.

В вычислительных машинах применяются две формы представления двоичных чисел:

- естественная форма или форма с фиксированной запятой (точкой);
- нормальная форма или форма с плавающей запятой (точкой).

С **фиксированной запятой** все числа изображаются в виде последовательности цифр с постоянным для всех чисел положением запятой, отделяющей целую часть от дробной.

Пример 4.3. В десятичной системе счисления имеются 5 разрядов в целой части числа (до запятой) и 5 разрядов в дробной части числа (после запятой); числа, записанные в такую разрядную сетку, имеют вид:

+00721,35500; +00000,00328; -10301,20260.

Эта форма наиболее проста, естественна, но имеет небольшой диапазон представления чисел и поэтому не всегда приемлема при вычислениях.

С **плавающей запятой** каждое число изображается в виде двух групп цифр. Первая группа цифр называется мантиссой, вторая - порядком, причем абсолютная величина мантиссы должна быть меньше 1, а порядок - целым числом. В общем виде число в форме с плавающей запятой может быть представлено так:

$$N = \pm MP^{\pm r},$$

где M -мантисса числа ($|M| < 1$);

r - порядок числа (r - целое число);

P - основание системы счисления.

Пример 4.5. Приведенные в примере 4.3 числа в нормальной форме запишутся так:

+0,721355*10³; +0,328* 10⁻³; -0,103012026*10⁵.

Нормальная форма представления имеет огромный диапазон отображения чисел и является основной в современных ЭВМ.

Двоично-десятичная система счисления получила большое распространение в современных ЭВМ ввиду легкости перевода в десятичную систему и обратно. Она используется там, где основное внимание уделяется не простоте технического построения машины, а удобству работы пользователя. В этой системе счисления все десятичные цифры отдельно кодируются четырьмя двоичными цифрами (табл. 4.1) и в таком виде записываются последовательно друг за другом.

При программировании иногда используется **шестнадцатеричная система счисления**, перевод чисел из которой в двоичную систему счисления весьма прост - выполняется поразрядно (полностью аналогично переводу из двоично-десятичной системы).

Для изображения цифр, больших 9, в шестнадцатеричной системе счисления применяются буквы A=10, B=11, C=12, D=13, E=14, F=15.

Таблица 4.1. Таблица двоичных кодов десятичных и шестнадцатеричных цифр.

Цифра	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
Код	00 00	00 01	00 10	00 11	01 00	01 01	01 10	01 11	10 00	10 01	10 10	10 11	11 00	11 01	11 10	11 11

Пример 4.7. Десятичное число 9703 в двоично-десятичной системе выглядит так:
1001011100000011.

Пример 4.8. Шестнадцатеричное число F17B в двоичной системе выглядит так:
1111000101111011.

Варианты представления информации в ПК

Вся информация (данные) представлена в виде двоичных кодов. Для удобства работы введены следующие термины, обозначающие совокупности двоичных разрядов (табл. 4.2). Эти термины обычно используются в качестве единиц измерения объемов информации, хранимой или обрабатываемой в ЭВМ.

Таблица 4.2. Двоичные совокупности

Количество двоичных разрядов в группе	1	8	16	$8 \cdot 1024$	$8 \cdot 1024^2$	$8 \cdot 1024^3$	$8 \cdot 1024^4$
Наименование единицы измерения	Бит	Байт	Параграф	Килобайт (Кбайт)	Мегабайт (Мбайт)	Гигабайт (Гбайт)	Терабайт (Тбайт)

Последовательность нескольких битов или байтов часто называют полем данных. Биты в числе (в слове, в поле и т.п.) нумеруются справа налево, начиная с 0-го разряда. В ПК могут обрабатываться поля постоянной и переменной длины.

Поля постоянной длины:

слово - 2 байта	двойное слово - 4 байта
полуслово - 1 байт	расширенное слово - 8 байт
слово длиной 10 байт- 10 байт	

Числа с фиксированной запятой чаще всего имеют формат слова и полуслова, числа с плавающей запятой - формат двойного и расширенного слова.

Поля переменной длины могут иметь любой размер от 0 до 256 байт, но обязательно равный целому числу байтов.

КОДЫ ASCII

Распакованный формат представления двоично-десятичных чисел (иногда его называют "зонный") является следствием использования в ПК ASCII-кода для представления символьной информации.

Код ASCII (American Standard Code for Information Interchange - Американский стандартный код для обмена информацией) имеет основной стандарт и его расширение (табл. 4.3). Основной стандарт для кодирования символов использует шестнадцатеричные коды 00-7F, расширение стандарта - 80 -FF. Основной стандарт является международным и используется для кодирования управляющих символов, цифр и букв латинского алфавита; в расширении стандарта кодируются символы псевдографики и буквы национального алфавита (естественно, в разных странах разные).

Таблица 4.3. Таблица кодов ASCII

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0		▸		0	@	P	'	p	А	Р	а	▯▯▯	Л	Л	р	Ё
1	☺	◀	!	1	A	Q	а	q	Б	С	В	▯	└	т	с	ё
2	☹	↕	"	2	B	R	Ь	г	В	Т	в	▯	т	т	т	ё
3	♥	!!	#	3	C	S	с	s	Г	У	г		т	Л	у	е
4	♦	¶	\$	4	D	T	d	t	Д	Ф	д	┆	-	ь	ф	й
5	♣	§	%	5	E	U	e	u	Е	Х	е	┆	+	г	х	у
6	♠	_	&	6	F	V	f	v	Ж	Ц	ж	▯	т	г	ц	у
7	•	±	'	7	G	W	g	w	З	Ч	э	▯	т	т	ч	у
8	▯	↑	(8	H	X	h	x	И	Ш	и	▯	Л	+	ш	°
9		↓)	9	I	Y	i	y	И	Щ	й	▯	т	└	щ	°
A		→	*	:	J	Z	j	z	К	Ъ	к	▯	Л	г	ъ	•
B	♂	←	+	:	K	[k	{	Л	Ы	л	▯	т	▯	ы	└
C	♀	└	,	<	L	\	l		М	Ь	м	▯	т	▯	ь	№
D		↔	-	=	M]	m	}	Н	Э	н	▯	-	▯	э	▯
E	♫	▲	.	>	N	^	n	~	О	Ю	о	▯	т	▯	ю	▯
F	⊗	▼	/	?	O	_	o	△	П	Я	п	└	▯	▯	я	