

ИНФОРМАЦИЯ. ИНФОРМАТИКА

1.1. Общие сведения об информации и информатике

Информацией называют сведения о предметах, процессах и явлениях окружающего нас мира.

Информацию человек воспринимает с помощью органов чувств. Зрение, слух, обоняние, вкус и осязание позволяют нам, например, получить информацию о размере, форме, цвете, твердости, аромате и вкусе яблока.

Информацию человек не только воспринимает, но и сам является ее источником с помощью речи, жестов, мимики, графики. Сам образ человека несет информацию, по которой его узнают, отличают от других людей.

Источниками информации для человека служат окружающие его предметы и явления. Много информации он получает по телевидению, радио, печати, читая книги, просматривая фильмы, в разговорах с другими людьми.

Со второй половины XX века наблюдается «информационный взрыв». Развитие науки и техники, литературы и искусства привели к значительному увеличению количества информации. Одновременно существенно увеличились темп жизни и цена времени. Человек не в силах одновременно слушать и смотреть сотни и тысячи теле- и радиопередач, читать сотни и тысячи издаваемых ежедневно книг, газет и журналов. Остро встала пробле-

ма сортировки и выбора информации, разработки средств передачи и переработки информации. На помощь пришли новейшие средства телекоммуникаций, важными элементами которых служат компьютеры.

В соответствии с увеличившимся значением информации и средств телекоммуникаций, а также углубленным пониманием роли информации и информационных процессов не только в жизни человечества, но и во вселенском масштабе, меняется взгляд на научную картину мира.

Известны два поля дальнего взаимодействия — электромагнитное, осуществляющее взаимодействие между электрически заряженными частицами, и гравитационное (поле тяготения), осуществляющее взаимодействие между массами. В настоящее время многие ученые к этим полям добавляют **информационное поле**. Современная научная картина мира строится на трех основополагающих понятиях: **вещества, энергии и информации**.

Информацию выдают, воспринимают, используют, перерабатывают все представители животного и растительного мира, все автоматически работающие механизмы, устройства. Важно отметить, что информационные процессы в живой природе, обществе, в различных механизмах и устройствах имеют много общего и подчиняются одним и тем же законам. Это позволяет рассматривать и изучать их с единых позиций, строить единую для них теорию информации.

Информатикой называют отрасль науки, изучающую структуру и общие свойства информации, а также вопросы, связанные с ее сбором, хранением, поиском, преобразованием, распространением и использованием в различных сферах деятельности человека.

Информационными технологиями называют раздел информатики, занимающийся разработкой, внедрением и использованием технических и программных средств, с помощью которых наиболее эффективно выполняются разнообразные операции по обработке информации во всех сферах человеческой деятельности. Среди областей применения информационных технологий следует выделить экономику и финансы, науку и производство, просвещение и культуру, социальную сферу и оборону, развлечения и средства связи.

Информация не может существовать без объектов, выдающих ее, передаваться без соответствующих объектов, способных ее переносить, и восприниматься без объектов, способных ее принимать. Простейшая **информационная система** состоит из источника информации, канала связи и приемника информации (рис. 1.1).

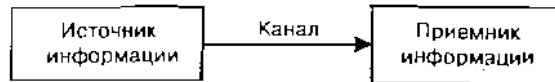


Рис. 1.1. Простейшая информационная система

Информатика тесно связана с кибернетикой.

Кибернетика — это наука о процессах управления в машинах, живых организмах, обществе. Она также занимается изучением общих законов получения, хранения, передачи и переработки информации в управляющих системах.

В кибернетике рассматриваются более сложные **информационные управляющие системы**, включающие управляющий объект, канал связи и управляемый объект (рис. 1.2). При этом управляющий и управляемый объекты могут представляться в виде «черного ящика», что предполагает изучение работы системы независимо от внутреннего содержания и структуры этих объектов. Они могут быть любой природы: механическими, электрическими, биологическими, общественными структурами и т.д.

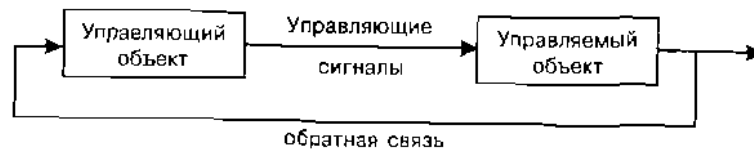


Рис. 1.2. Информационная управляющая система

Кроме понятия черного ящика в кибернетике используют важное понятие **обратной связи**. Функционирование управляющей системы рассматривается с разомкнутой и замкнутой обратной связью, которая может быть положительной или отрицательной.

Положительная обратная связь усиливает результаты работы управляемого объекта. Как правило, положительная обратная связь при-

водит к усилению управляющих сигналов, что ведет к перевозбуждению и неустойчивой работе системы. **Отрицательная обратная связь** ослабляет управляющие воздействия на управляемый объект.

Для стабильной, устойчивой, без излишних колебаний работы системы необходимо, чтобы управляющие сигналы в каждый момент времени имели соответствующую величину и знак. Это обеспечивается работой **устройства сравнения**. Оно находится на входе управляющего объекта, сравнивает передаваемые по обратной связи реальные результаты работы управляемого объекта с требуемыми результатами и вырабатывает сигнал **ошибки** соответствующего знака и величины. Это корректирует в нужном направлении управляющие сигналы и работу управляемого объекта.

Пусть управляющий объект — шофер автомашины, управляемый объект — автомашина, управляющие сигналы — это воздействия шофера на рулевое управление, обратная связь — зрительная. Качественная работа такой системы может определяться плавным, без колебаний движением автомобиля в виражи дороги. Без обратной связи такая система не работоспособна.

Примером управляющей системы может служить система учитель-ученик. Управляющие сигналы в такой системе — это информация, передаваемая на учебных занятиях. Качество функционирования системы оценивается уровнем знаний и умений ученика, которые существенно зависят от работы обратной связи — различных видов контроля усвоения учебного материала.

1.2. Свойства информации

К свойствам информации можно отнести ее важность (ценность, полезность), достоверность (истинность, правильность), полноту, оперативность (своевременность), понятность (доступность) и некоторые другие ее качества.

Важность

Полученная в разговоре по телефону информация может быть важной для собеседников или представлять известные им све-

депия. Информация о возможных заморозках имеет различную важность для огородника, шофера и для водителя трамвая. В этом проявляется субъективность в оценке информации.

Можно сказать, что цена информации определяется связанными с ней материальными и духовными приобретениями или потерями. Более ценной информацией соответствует и больший выигрыш или меньший проигрыш.

Достоверность

Это свойство связано с истинностью и ложностью информации. Известно, что истина объективна по содержанию (не зависит от человека), но субъективна по форме, так как является результатом деятельности человека.

Истина относительна. Она отражает предмет не полностью ввиду ограниченности человеческих знаний. В процессе познания человека она стремится к истине абсолютной, полностью исчерпывающей предмет познания.

Например, великий Пифагор считал, что молнии мечет на Землю разгневанный Зевс. Теперь же каждый школьник знаком с электричеством. Долгое время за истину принималась геоцентрическая система Птолемея. В настоящее время истинность, достоверность информации о гелиоцентричности Солнечной системы бесспорна.

Высокую достоверность информации получаемой по каналам связи и в результате преобразования в компьютере связывают с их высокой надежностью, безошибочностью работы. Однако следует понимать, что отправлено (в смысле истинности) на передающем конце безошибочно работающей линии связи, то и будет получено на приемном конце.

Полнота

Полнота информации о предмете, процессе, явлении зависит от ее количества, подробности, всесторонности. Понятие полноты информации о предмете так же субъективно и относительно, как и понятие истины. Информацию даже о простейшем предмете невозможно исчерпать полностью. Всегда можно что-то добавить, уточнить.

Оперативность, своевременность информации

С этой характеристикой информации мы встречаемся практически ежедневно. Все средства массовой информации стремятся оперативно получать информацию и передавать ее зрителям, слушателям, читателям. Устаревшая информация нас, как правило, мало интересует.

Доступность информации

Связана с возможностью ее воспринимать, понимать и использовать. Сообщение, передаваемое вам, например, азбукой Морзе, если вы ее не знаете, не несет никакой информации.

Можно заметить, что все перечисленные характеристики информации взаимосвязаны, взаимозависимы. Например, ценность информации определяется ее достоверностью, полнотой и оперативностью.

1.3. Кодирование информации. Единицы информации

Кодом называют совокупность знаков (символов) предназначенных для представления информации в соответствии с определенными правилами. Такое представление называют кодированием. Кодирование информации с целью ее передачи, хранения, преобразования. Совокупность предназначенных для кодирования знаков называют алфавитом языка кодирования.

Различают естественные и искусственные (формальные) языки и их алфавиты. Естественные языки развивались веками и служат для общения людей между собой. Формальные языки разрабатываются для специальных применений. Примером формальных языков могут служить языки программирования, языки кодирования информации для ее передачи, хранения и т.п.

Одно и то же понятие на различных языках может кодироваться различными способами. Например, звук а, издаваемый человеком, кодируется в некоторых языках буквой А. Буква А в азбуке Морзе кодируется так: • — (точка, тире). В компьютере буква А латинского алфавита в привычной для нас десятичной системе

кодируется числом 65, которое в «привычной» для компьютера двоичной системе (цифры только 0 и 1) кодируется так: 01000001.

Различны и алфавиты языков кодирования. Это буквы А, В, С, D, ... и буквы А, Б, В, Г, Д, ... в латинском и русском алфавитах соответственно; точка и тире в азбуке Морзе, арабские цифры 0, 1, 2, 3, ..., 9, с помощью которых записываются числа, кодирующие различные количества; красный, желтый и зеленый цвет в светофоре; цифры 0 и 1 в компьютере и т.п.

Количество и графическое изображение символов в алфавитах естественных языков определяется характерными особенностями языка, историей его развития, традициями. Например, русский алфавит имеет 33 буквы, латинский — 26, итальянский — 21, армянский — 39, арабский — 28, китайский — несколько тысяч иероглифов. Количество и изображение знаков в формальных языках могут также существенно различаться.

Наименьший по числу знаков алфавит имеет только один знак. Пусть этот знак 1 (единица). Тогда три цвета светофора можно закодировать, например, так: красный — 1, желтый — 11, зеленый — 111. Такой алфавит самый неэкономичный по записи кодов. В этом легко убедиться, если попытаться записать в этом алфавите, например, число десять: 1111111111.

В информатике и вычислительной технике широко используется алфавит, имеющий два знака, две альтернативы, например, 1 и 0, да и нет, истина и ложь, включено и выключено. Такой алфавит называют двоичным. В соответствии с этим введена и наименьшая единица информации — бит (англ. *bit*, от *binary* — двоичный и *digit* — знак).

Одного бита информации достаточно, чтобы передать слово да или нет, закодировать, например, состояние электролампочки. Кстати, на некоторых выключателях пишут 1 — включено и 0 — выключено. Сообщение о том, что выключатель включен, содержит количество информации, равное одному биту.

Если требуется закодировать в двоичном алфавите красный, желтый и зеленый цвет светофора, то требуется уже два бита. Закодировать три цвета можно, например, так: 00, 01 и 10. Сообщение о том, что включен, например, красный цвет светофора, содержит информации больше одного бита.

Для кодирования четырех сторон света (север, восток, юг и запад) требуется также два бита ($2^2 = 4$, $\log_2 4 = 2$): 00, 01, 10, 11.

Поэтому сообщение о том, что выбрана одна из сторон света, содержит ровно два бита информации.

При кодировании восьми углов куба потребуется три ($\log_2 8 = 3$, $2^3 = 8$) бита: 000, 001, 010, 011, 100, 101, 110, 111. При кодировании от 9 до 16 объектов потребуется $\log_2 16 = 4$ бита и т. д.

Рассмотренные примеры позволяют сделать вывод, что для кодирования в двоичном алфавите N альтернатив (событий, сообщений, состояний и т.п.) требуется $I = \log_2 N$ бит информации. Последовательность символов называют словом. Можно сказать, что для кодирования N объектов в двоичном алфавите требуется I -битовое двоичное слово, где

$$I = \log_2 N.$$

Эту формулу определения количества информации, которое имеем в случае получения одного из N равновероятных сообщений, называют по имени предложившего ее американского инженера формулой Хартли. Напомним, что логарифм I числа N по основанию 2 — это показатель степени, в которую нужно возвести число 2, чтобы результат был равен N , т.е. $2^I = N$.

Пример 1.

Пусть требуется отгадать задуманное число из набора чисел от 0 до 31. В соответствии с формулой Хартли количество информации в сообщении о том, какое число загадано, равно $I = \log_2 32 = 5$ бит. Поэтому, чтобы отгадать число, требуется получить 5 бит информации.

Известен оптимальный по числу попыток прием, который заключается в делении сначала всего множества чисел, затем выбранной части пополам. После каждого деления задается, например, такой вопрос: «Задуманное число осталось в первой половине?». Ответ (да или нет) несет информацию объемом в один бит. Неопределенность при этом уменьшается в два раза.

Количество чисел, из которых нужно найти заданное, уменьшается при каждом делении вдвое: 16, 8, 4, 2, 1. Поэтому достаточно сделать всего пять попыток, получить пять ответов да или нет, то есть получить пять бит информации: $\log_2 32 = 5$. Эти 5 бит составляют пятибитовое слово в двоичном алфавите, которым можно закодировать все 32 числа: 00000, 00001, 00010, ..., 11111.

Если требуется отгадать одно число из 50, то на это потребуется $I = \log_2 50 \approx 5,64$ бита информации. Если чисел 100, то требуемое количество информации увеличивается на один бит: $I = \log_2 100 = 6,64$ бита.

Сообщение о том, что горит один из трех сигналов светофора, например, зеленый, несет информацию в количестве $I = \log_2 3 = 1,585$ бит. Сообщение о том, что ваш знакомый живет в 21 квартире дома, имеющего 64 квартиры, несет информацию в количестве $I = \log_2 64 = 6$ бит. Можно сделать вывод, что с увеличением числа возможных событий количество информации в сообщении, что одно из этих событий произошло, также увеличивается.

Для пояснения вероятностного подхода к определению количества информации формулу Хартли запишем в следующем виде:

$$I = \log_2 N = \log_2(1/p) = -\log_2 p,$$

где N — число равновероятных событий. Так как вероятности появления событий равны, то можно записать $p = 1/N$, откуда $N = 1/p$. Проверим равенство для простейшего случая двух равновероятных событий: $I = -\log_2(1/2) = -(-1) = 1$ бит информации.

Пример 2.

Пусть ставится задача отгадать задуманное число из набора чисел от 1 до 32. В предыдущей аналогичной задаче все числа делились на две равные части. Поэтому информация о том, в какой части находится задуманное число, была равна одному биту — $I = \log_2 2 = 1$ (при вероятностном подходе — $I = -\log_2(1/2) = 1$).

Заметим, что вероятность немедленно, с первой попытки угадать задуманное число, равна $1/32$. Разделим все числа на две неравные части, например, 8 и 24. Вероятности нахождения задуманного числа в группах различны. Чем больше чисел в группе, тем больше эта вероятность. Для первой группы она равна $p_1 = 8/32 = 1/4$. Для второй — $p_2 = 24/32 = 3/4$.

Можно ожидать, что и количество информации, содержащейся в ответе на вопрос «в какой группе находится загаданное число» будет различным для первой и второй группы. Можно также предположить, что, чем меньше вероятность нахождения числа в

группе, тем большее количество информации содержится в сообщении, что число находится в этой группе.

Вычислим количества информации. Сообщение о нахождении числа в первой группе содержит $I_1 = -\log_2(1/4) = 2$ бита. Сообщение о нахождении числа во второй группе содержит $I_2 = -\log_2(3/4) = 0,415$ бита. Предположения подтвердились.

Пример 3.

В корзине находится 16 фруктов: 8 яблок, 6 груш и 2 лимона. Какое количество информации содержит сообщение о том, что из корзины случайным образом были взяты яблоко (I_1), груша (I_2), лимон (I_3)?

Эти события имеют разную вероятность: взять яблоко — $p_1 = 8/16 = 1/2$, взять грушу — $p_2 = 6/16 = 3/8$, взять лимон — $p_3 = 2/16 = 1/8$.

Для решения задачи воспользуемся формулой $I = \log_2(1/p)$. Подставляя в формулу вероятности событий, получаем: $I_1 = -\log_2(1/2) = 1$ бит, $I_2 = -\log_2(3/8) = 1,415$ бит, $I_3 = -\log_2(1/8) = 3$ бита информации. Снова подтверждается предположение: чем меньше вероятность события, тем больше получаем информации, когда это событие происходит.

Решим обратную задачу: задано количество полученной информации и требуется определить количество благоприятных событий. Например, в отношении предыдущей задачи обратную задачу можно было бы сформулировать так: взяли из корзины лимон, получили при этом 3 бита информации и требуется определить, сколько было в корзине лимонов.

Пример 4.

В доме 32 квартиры. Среди них имеются квартиры трехкомнатные, двухкомнатные и однокомнатные. Сообщение о том, что ваш знакомый живет в двухкомнатной квартире, содержит 2 бита информации. Сколько в доме двухкомнатных квартир?

Для решения применим формулу $I = \log_2(1/p)$. Напомним, что I — это показатель степени, в которую нужно возвести основание логарифма 2, чтобы получить число $1/p$. В нашем примере $I = 2$. Подставим значения в формулу: $2^2 = 32/x$, откуда получаем $x = 8$, т.е. в доме 8 двухкомнатных квартир.

При **неравновероятных** сообщениях для определения количества информации в более сложных задачах применяют формулу, предложенную американским ученым Шенноном:

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

где p_i — вероятность появления i -го сообщения из N возможных.

На практике чаще применяются производные, более объемные единицы информации, чем бит. Восьмибитовое двоичное слово называют байтом. Байт позволяет кодировать до 256 альтернатив ($2^8 = 256$). Это коды 00000000, 00000001, 00000010, 00000011, ..., 11111111. Такими кодами в компьютерах кодируют все буквы, цифры, знаки препинания, пробелы в тексте, специальные символы. В битах, байтах измеряют объем памяти запоминающих устройств компьютеров, количество информации, содержащееся в сообщениях.

Байт до недавнего времени был наименьшей, имеющей адрес, единицей памяти компьютера. В настоящее время осуществляется активный переход на двухбайтовое кодирование символов и команд, позволяющее получать 65536 различных кодов.

Количество информации (требуемый для хранения информации объем памяти) $2^{10} = 1024$ байта в обиходе называют килобайтом (кбайтом), 2^{10} кбайт называют мегабайт (Мбайт), 2^{10} Мбайт — гигабайт (Гбайт), 2^{10} Гбайт — терабайт.

Пример 5.

Подсчитаем объем памяти, требуемый для записи и хранения в памяти компьютера книги «Программируем на языке Quick BASIC 4.5» Г.А. Вельднера. В книге 420 страниц. На каждой странице в среднем по 40 строк. В каждой строке по 60 символов. Итого в книге $60 \times 40 \times 420 \approx 1\,000\,000$ знаков. Каждый знак требует для записи 1 байтв памяти. Поэтому для записи всей книги нужно около 1 Мбайта. А это всего одна гибкая дискета с объемом памяти 1,44 Мбайта.

Часто отождествляют понятия количества и объема информации. Поясним на примере их различие. Пусть имеется большой энциклопедический словарь в 1400 страниц и такого же объема книга, на каждой странице которой все строки заполнены вопросительными знаками. Объем словаря и книги одинаков — 1400

страниц текста. Поэтому для хранения словаря и книги требуется одинаковый объем памяти. Количество же информации в словаре и книге существенно различается. Всю информацию книги можно записать одной фразой: «1400 страниц с вопросительными знаками». Но даже эта фраза имеет сомнительную ценность и для многих людей не несет никакой полезной информации.

В целях экономии памяти при записи информации на хранение (архивации) широко используют процедуру ее сжатия. Так, рассматриваемый выше энциклопедический словарь можно сжать примерно в десять раз. При этом следует понимать, что путем сжатия уменьшается не количество информации, а только ее объем.

Задания для самостоятельной работы

1. Придумать примеры: а) информации, имеющей различную ценность для различных людей, что подтверждает субъективность в оценках информации; б) информации, ценность которой существенно зависит от ее оперативности; в) информации, ценность которой зависит от ее полноты.
2. Определить число бит, необходимое для кодирования в двоичном шрифите: а) всех пальцев на руках и ногах; б) всех квартир пятидесятиквартирного дома.
3. Сколько бит информации нужно получить, чтобы отгадать одно задуманное число из 8 чисел, из 64 чисел?
4. В корзине с грибами 8 белых грибов, 8 подосиновиков и 16 подберезовиков. Сколько информации содержат события, состоящие в том, что из корзины случайным образом взяли один белый гриб, один подберезовик?
5. Из кошелька с восемью монетами наугад взяли монету 1 рубль, и это событие содержит 2 бита информации. Сколько в кошельке монет достоинством в один рубль?
6. Рассмотреть работу кибернетических систем с обратной связью учитель — ученик, избиратель — депутат. Описать процесс работы таких систем, определить, что в них выступает в качестве устройств сравнения, сигналов ошибки. Выбрать и описать примеры подобных систем.
7. Выбрать любую книгу и подсчитать потребный для ее записи объем памяти.