

§ 1.6

Измерение информации

Ключевые слова:

- бит
- информационный вес символа
- информационный объём сообщения
- единицы измерения информации

1.6.1. Алфавитный подход к измерению информации

Одно и то же сообщение может нести много информации для одного человека и не нести её совсем для другого человека. При таком подходе количество информации определить однозначно затруднительно.

Алфавитный подход позволяет измерить информационный объём сообщения, представленного на некотором языке (естественном или формальном), независимо от его содержания.

Для количественного выражения любой величины необходима, прежде всего, единица измерения. Измерение осуществляется путём сопоставления измеряемой величины с единицей измерения. Сколько раз единица измерения «укладывается» в измеряемой величине, таков и результат измерения.

При алфавитном подходе считается, что каждый символ некоторого сообщения имеет определённый **информационный вес** — несёт фиксированное **количество информации**. Все символы одного алфавита имеют один и тот же вес, зависящий от мощности алфавита. Информационный вес символа двоичного алфавита принят за минимальную единицу измерения информации и называется **1 бит**.



Обратите внимание, что название единицы измерения информации «бит» (*bit*) происходит от английского словосочетания *binary digit* — «двоичная цифра».



За минимальную единицу измерения информации принят **1 бит**. Считается, что таков информационный вес символа двоичного алфавита.

1.6.2. Информационный вес символа произвольного алфавита

Ранее мы выяснили, что алфавит любого естественного или формального языка можно заменить двоичным алфавитом. При этом мощность исходного алфавита N связана с разрядностью двоичного кода i , требуемой для кодирования всех символов исходного алфавита, соотношением: $N = 2^i$.

Разрядность двоичного кода принято считать информационным весом символа алфавита. Информационный вес символа алфавита выражается в битах.



Информационный вес символа алфавита i и мощность алфавита N связаны между собой соотношением: $N = 2^i$.



Задача 1. Алфавит племени Пульти содержит 8 символов. Каков информационный вес символа этого алфавита?

Решение. Составим краткую запись условия задачи.

$$N = 8$$

$$i = ?$$

Известно соотношение, связывающее величины i и N : $N = 2^i$. С учётом исходных данных: $8 = 2^i$. Отсюда: $i = 3$.

Полная запись решения в тетради может выглядеть так:

$$\begin{array}{l|l|l} N = 8 & N = 2^i & 8 = 2^i, i = 3 \text{ бита} \\ \hline i = ? & & \end{array}$$

Ответ: 3 бита.

1.6.3. Информационный объём сообщения

Информационный объём сообщения (количество информации в сообщении), представленного символами естественного или фор-

мального языка, складывается из информационных весов составляющих его символов.

Информационный объём сообщения I равен произведению количества символов в сообщении K на информационный вес символа алфавита i : $I = K \cdot i$.

Задача 2. Сообщение, записанное буквами 32-символьного алфавита, содержит 140 символов. Чему равен информационный объём этого сообщения?

Решение

$$\begin{array}{l|l} N = 32 & I = K \cdot i, N = 2^i \\ K = 140 & 32 = 2^i, i = 5, \\ I = ? & I = 140 \cdot 5 = 700 \text{ (бит)} \end{array}$$

Ответ: 700 бит.

Задача 3. Информационное сообщение объёмом 720 битов состоит из 180 символов. Какова мощность алфавита, с помощью которого записано это сообщение?

Решение

$$\begin{array}{l|l} I = 720 & N = 2^i, \\ K = 180 & I = K \cdot i, i = I/K \\ N = ? & i = 720/180 = 4 \text{ (бита)} \\ & N = 2^4 = 16 \text{ (символов)} \end{array}$$

Ответ: 16 символов.

1.6.4. Единицы измерения информации

В наше время подготовка текстов в основном осуществляется с помощью компьютеров. Можно говорить о «компьютерном алфавите», включающем следующие символы: строчные и прописные русские и латинские буквы, цифры, знаки препинания, знаки арифметических операций, скобки и др. Такой алфавит содержит 256 символов. Поскольку $256 = 2^8$, информационный вес каждого символа этого алфавита равен 8 битам. Величина, равная восьми битам, называется **байтом**. 1 байт — информационный вес символа алфавита мощностью 256.

$$1 \text{ байт} = 8 \text{ бит}$$

Бит и байт — «мелкие» единицы измерения. На практике для измерения информационных объёмов используются более крупные единицы:

1 Кбайт (килобайт) = 1024 байта = 2^{10} байт
 1 Мбайт (мегабайт) = 1024 Кбайт = 2^{10} Кбайт
 1 Гбайт (гигабайт) = 1024 Мбайт = 2^{10} Мбайт
 1 Тбайт (терабайт) = 1024 Гбайт = 2^{10} Гбайт

Задача 4. Информационное сообщение объёмом 4 килобайта (Кбайт) состоит из 4096 символов. Каков информационный вес символа используемого алфавита? Сколько символов содержит алфавит, с помощью которого записано это сообщение?

Решение

$I = 4 \text{ Кбайт}$ $K = 4096$ $i = ?$ $N = ?$	$I = K \cdot i, i = I/K$ $N = 2^i$	$I = 4 \text{ Кбайт} =$ $= 4 \cdot 1024 \cdot 8 \text{ бит}$ $i = 4 \cdot 1024 \cdot 8 / 4096 = 8 \text{ бит}$ $N = 2^8 = 256 \text{ символов}$
---	---------------------------------------	--

Ответ: 8 бит, 256 символов.

Задача 5. В велокроссе участвуют 128 спортсменов. Специальное устройство регистрирует прохождение каждым из участников промежуточного финиша, записывая номер участника цепочкой из нулей и единиц минимальной длины, одинаковой для каждого спортсмена. Каков будет информационный объём сообщения, записанного устройством после того, как промежуточный финиш пройдут 80 велосипедистов?

Решение. Номера 128 участников кодируются с помощью двоичного алфавита. Требуемая разрядность двоичного кода (длина цепочки) равна 7, так как $128 = 2^7$. Иначе говоря, зафиксированное устройством сообщение о том, что промежуточный финиш прошёл один велосипедист, несёт 7 бит информации. Когда промежуточный финиш пройдут 80 спортсменов, устройство запишет $80 \cdot 7 = 560$ бит, или 70 байт информации.

Ответ: 70 байт.

САМОЕ ГЛАВНОЕ

При алфавитном подходе считается, что каждый символ некоторого сообщения имеет определённый информационный вес — несёт фиксированное количество информации.

1 бит — минимальная единица измерения информации.

Информационный вес символа алфавита i и мощность алфавита N связаны между собой соотношением: $N = 2^i$.

Информационный объём сообщения I равен произведению количества символов в сообщении K на информационный вес символа алфавита i : $I = K \cdot i$.

1 байт = 8 бит.

Байт, килобайт, мегабайт, гигабайт, терабайт — единицы измерения информации. Каждая следующая единица больше предыдущей в 1024 (2^{10}) раза.

Вопросы и задания



1. Ознакомьтесь с материалами презентации к параграфу, содержащейся в электронном приложении к учебнику. Используйте эти материалы при подготовке ответов на вопросы и выполнении заданий.
2. В чём суть алфавитного подхода к измерению информации?
3. Что принято за минимальную единицу измерения информации?
4. Что нужно знать для определения информационного веса символа алфавита некоторого естественного или формального языка?
5. Определите информационный вес i символа алфавита мощностью N , заполняя в тетради таблицу:

N	$N = 2^i$	i (бит)
8		
32		
64		
128		
256		

6. Как определить информационный объём сообщения, представленного символами некоторого естественного или формального языка?
7. Определите информационный объём сообщения из K символов алфавита мощностью N , заполняя в тетради таблицу:

N	$N = 2^i$	i (бит)	K	$I = K \cdot i$ (бит)
8			200	
16			110	
64			120	
128			100	
256			80	

8. Племя Мульти пишет письма, пользуясь 16-символьным алфавитом. Племя Пульти пользуется 32-символьным алфавитом. Вожди племён обменялись письмами. Письмо племени Мульти содержит 120 символов, а письмо племени Пульти — 96. Сравните информационные объёмы сообщений, содержащихся в письмах.
9. Информационное сообщение объёмом 650 бит состоит из 130 символов. Каков информационный вес каждого символа этого сообщения?
10. Выразите количество информации в различных единицах, заполняя в тетради таблицу:

Бит	Байт	Кбайт
24 576		
	2048	
		1,5
2^{13}		
	2^{11}	
		4 (2^2)

11. Информационное сообщение объёмом 375 байт состоит из 500 символов. Каков информационный вес каждого символа этого сообщения? Какова мощность алфавита, с помощью которого записано это сообщение?
12. Для записи текста использовался 64-символьный алфавит. Какой объём информации в байтах содержат 3 страницы текста, если на каждой странице расположено 40 строк по 60 символов в строке?
13. Сообщение занимает 6 страниц по 40 строк, в каждой строке записано по 60 символов. Информационный объём всего сообщения равен 9000 байт. Каков информационный вес одного символа? Сколько символов в алфавите языка, на котором записано это сообщение?
14. Метеорологическая станция ведёт наблюдение за влажностью воздуха. Результатом одного измерения является целое число от 0 до 100 процентов, которое записывается цепочкой из нулей и единиц минимальной длины, одинаковой для каждого измерения. Станция сделала 8192 измерения. Определите информационный объём результатов наблюдений.
15. Племя Пульти пользуется 32-символьным алфавитом. Свод основных законов племени хранится на 512 глиняных табличках, на каждую из которых нанесено ровно 256 символов. Какой объём информации содержится на каждом носителе? Какой объём информации заключён во всём своде законов?





Тестовые задания для самоконтроля

1. Какое из следующих утверждений точнее всего раскрывает смысл понятия «информация» с обыденной точки зрения?
 - а) последовательность знаков некоторого алфавита
 - б) книжный фонд библиотеки
 - в) сведения об окружающем мире и протекающих в нём процессах, воспринимаемые человеком непосредственно или с помощью специальных устройств
 - г) сведения, содержащиеся в научных теориях
2. Непрерывным называют сигнал:
 - а) принимающий конечное число определённых значений
 - б) непрерывно изменяющийся во времени
 - в) несущий текстовую информацию
 - г) несущий какую-либо информацию
3. Дискретным называют сигнал:
 - а) принимающий конечное число определённых значений
 - б) непрерывно изменяющийся во времени
 - в) который можно декодировать
 - г) несущий какую-либо информацию
4. Информацию, не зависящую от личного мнения или суждения, называют:
 - а) понятной
 - б) актуальной
 - в) объективной
 - г) полезной

5. Информацию, существенную и важную в настоящий момент, называют:
 - а) полезной
 - б) актуальной
 - в) достоверной
 - г) объективной
6. По способу восприятия человеком различают следующие виды информации:
 - а) текстовая, числовая, графическая, табличная и пр.
 - б) научная, социальная, политическая, экономическая, религиозная и пр.
 - в) обыденная, производственная, техническая, управленческая
 - г) визуальная, аудиальная, тактильная, обонятельная, вкусовая
7. Известно, что наибольший объем информации физически здоровый человек получает при помощи:
 - а) органов слуха
 - б) органов зрения
 - в) органов осязания
 - г) органов обоняния
 - д) вкусовых рецепторов
8. Укажите «лишний» объект с точки зрения соглашения о смысле используемых знаков:
 - а) буквы
 - б) дорожные знаки
 - в) цифры
 - г) нотные знаки
9. Укажите «лишний» объект с точки зрения вида письменности:
 - а) русский язык
 - б) английский язык
 - в) китайский язык
 - г) французский язык
10. К формальным языкам в наше время можно отнести:
 - а) русский язык
 - б) латынь
 - в) китайский язык
 - г) французский язык

11. По форме представления информацию можно условно разделить на следующие виды:
- а) математическая, биологическая, медицинская, психологическая и пр.
 - б) знаковая и образная
 - в) обыденная, научная, производственная, управленческая
 - г) визуальная, аудиальная, тактильная, обонятельная, вкусовая
12. Дискретизация информации — это:
- а) физический процесс, изменяющийся во времени
 - б) количественная характеристика сигнала
 - в) процесс преобразования информации из непрерывной формы в дискретную
 - г) процесс преобразования информации из дискретной формы в непрерывную
13. Дайте самый полный ответ.
При двоичном кодировании может использоваться алфавит, состоящий из:
- а) 0 и 1
 - б) слов ДА и НЕТ
 - в) знаков + и –
 - г) любых двух символов
14. Сколько существует различных последовательностей из символов «плюс» и «минус» длиной ровно шесть символов?
- а) 64 б) 50 в) 32 г) 20
15. Для пяти букв латинского алфавита заданы их двоичные коды (для некоторых букв — из двух битов, для некоторых — из трёх битов). Эти коды представлены в таблице:

А	В	С	Д	Е
000	01	100	10	011

Определите, какой набор букв закодирован двоичной строкой 0110100011000.

- а) EVCEA б) BDDEA в) BDCEA г) EBAEA
16. Шахматная доска состоит из 8 столбцов и 8 строк. Какое минимальное количество битов потребуется для кодирования координат каждого шахматного поля?
- а) 4 б) 5 в) 6 г) 7

17. В какой строке единицы измерения информации расположены по возрастанию?
 а) гигабайт, мегабайт, килобайт, байт, бит
 б) бит, байт, мегабайт, килобайт, гигабайт
 в) байт, бит, килобайт, мегабайт, гигабайт
 г) бит, байт, килобайт, мегабайт, гигабайт
18. Объём сообщения равен 11 Кбайт. Сообщение содержит 11 264 символа. Какова мощность алфавита, с помощью которого записано сообщение?
 а) 64 б) 128 в) 256 г) 512
19. Дан текст из 600 символов. Известно, что символы берутся из таблицы размером 16×32 . Определите информационный объём текста в битах.
 а) 1000 б) 2400 в) 3600 г) 5400
20. Два текста содержат одинаковое количество символов. Первый текст составлен из символов алфавита мощностью 16, а второй текст — из символов алфавита мощностью 256. Во сколько раз количество информации во втором тексте больше, чем в первом?
 а) 12 б) 2 в) 24 г) 4
21. Информационные процессы — это:
 а) процессы строительства зданий и сооружений
 б) процессы химической и механической очистки воды
 в) процессы сбора, хранения, обработки, передачи информации
 г) процессы производства электроэнергии
22. Под носителем информации принято подразумевать:
 а) линию связи
 б) сеть Интернет
 в) компьютер
 г) материальный объект, на котором можно тем или иным способом зафиксировать информацию
23. В какой строке верно представлена схема передачи информации?
 а) источник → кодирующее устройство → декодирующее устройство → приёмник
 б) источник → кодирующее устройство → канал связи → декодирующее устройство → приёмник



- в) источник → кодирующее устройство → помехи → декодирующее устройство → приёмник
- г) источник → декодирующее устройство → канал связи → кодирующее устройство → приёмник

24. Гипертекст — это:

- а) очень большой текст
- б) текст, в котором могут осуществляться переходы по ссылкам
- в) текст, набранный на компьютере
- г) текст, в котором используется шрифт большого размера

25. Поисковой системой НЕ является:

- а) Google
- б) Internet Explorer
- в) Bing
- г) Яндекс

26. Даны запросы к поисковой системе. По какому запросу будет найдено наибольшее количество соответствующих ему документов?

- а) разведение & содержание & меченосцы & сомики
- б) содержание & меченосцы
- в) содержание | меченосцы | сомики
- г) содержание & меченосцы & сомики

Для проверки знаний и умений по теме «Информация и информационные процессы» вы можете воспользоваться интерактивным тестом к главе 1, содержащимся в электронном приложении к учебнику.