

Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
«Тихоокеанский государственный университет»

Измерение информации

Методические указания
к выполнению практической работы по информатике
для студентов всех направлений дневной формы обучения

Хабаровск
Издательство ТОГУ
2013

УДК 004(076.5)

Измерение информации : методические указания к выполнению практической работы по информатике для студентов всех направлений дневной формы обучения / сост. Н. Д. Берман, Н. И. Шадрина. – Хабаровск : Изд-во Тихоокеан. гос. ун-та, 2013. – 27 с.

Методические указания составлены на кафедре информатики. Включают общие сведения об информации, свойствах информации, представлении информации в компьютере, примеры решения задач и задания для самостоятельного выполнения.

Печатается в соответствии с решениями кафедры информатики и методического совета факультета компьютерных и фундаментальных наук.

Понятие информации

Термин «информация» происходит от латинского *information*, что означает разъяснение, осведомление, изложение.

С позиции материалистической философии информация есть отражение реального мира с помощью сведений (сообщений). 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.

Информация, предназначенная для передачи, называется сообщением. Сообщение может быть представлено в виде знаков и символов, преобразовано и закодировано с помощью электрических сигналов.

Информация, представленная в виде, пригодном для обработки (человеком, компьютером), называется данными. Данные могут быть, например, числовыми, текстовыми, графическими.

Чтобы происходил обмен информацией, должны быть источник информации, передатчик, канал связи, приемник и получатель. Обычно в качестве получателя выступает человек, который оценивает информацию с точки зрения ее применимости для решения поставленной задачи. Процедура оценки информации проходит в три этапа, определяющие ее синтаксический, семантический и прагматический аспекты.

Определенный набор данных вне зависимости от смысловых и потребительских качеств характеризует синтаксический аспект информации. Сопоставление данных с тезаурусом (тезаурус – полный систематизированный набор данных и знаний в какой-либо области) формирует знание о наблюдаемом факте, это является семантическим аспектом информации (отражает смысловое содержание информации). Оценка практической полезности информации отражает ее прагматический аспект.

Свойства информации

Информация характеризуется определенными свойствами, зависящими как от данных (содержательной части информации), так и от методов работы с ними. Свойства информации делятся на две группы: атрибутивные и потребительские.

Атрибутивные свойства - это свойства, которые отображают внутреннюю природу информации и особенности ее использования. Наиболее важными из этих свойств являются следующие:

- информация представляет новые сведения об окружающем мире, отсутствовавшие до ее получения;
- информация не материальна, несмотря на то, что она проявляется в форме знаков и сигналов на материальных носителях;
- знаки и сигналы могут предоставить информацию только для получателя, способного их воспринять и распознать;
- информация неотрывна от физического носителя, но в то же время не связана ни с конкретным носителем, ни с конкретным языком;
- информация дискретна – она состоит из отдельных фактических данных, передающихся в виде сообщений;
- информация непрерывна – она накапливается и развивается поступательно.

Качество информации определяется ее свойствами, отвечающими потребностям пользователя.

Рассмотрим наиболее важные потребительские свойства информации:

- полнота (достаточность);
- достоверность;
- адекватность;
- доступность;
- актуальность.

Полнота (достаточность) информации. Под полнотой информации понимают ее достаточность для принятия решений.

Достоверность информации. Под достоверностью информации понимают ее соответствие объективной реальности окружающего мира. Свойство достоверности информации имеет важное значение в тех случаях, когда ее используют для принятия решений.

Адекватность информации – это степень соответствия информации, полученной потребителем, тому, что автор вложил в ее содержание. Адекватность информации иногда путают с ее достоверностью. Это разные свойства. Можно привести пример адекватной, но недостоверной информации. Так, если 1 апреля в газете появится заведомо ложное сообщение, то его можно считать адекватным. Адекватно толковать его не как информационное, а как развлекательное. То же сообщение, опубликованное 2 апреля, будет и недостоверным, и неадекватным.

Доступность информации – это мера возможности получить ту или иную информацию.

Актуальность информации – это степень соответствия информации текущему моменту времени. Нередко с актуальностью, как и с полнотой, связывают коммерческую ценность информации. Поскольку информационные процессы растянуты во времени, то достоверная и адекватная, но устаревшая информация может приводить к ошибочным решениям.

Единицы измерения количества информации

За единицу количества информации принимается такое ее количество, которое содержит сообщение, уменьшающее неопределенность в два раза.

Единица измерения информации называется бит (bit) – сокращение от английских слов *binarydigit*, что означает «двоичная цифра». Если положить в мешок два шара разного цвета, то, вытащив вслепую один шар, получим информацию о цвете шара в 1 бит.

В компьютерной технике бит соответствует физическому состоянию носителя информации: намагничено – не намагничено, есть сигнал – нет сигнала. При этом одно состояние принято обозначать цифрой 1, а другое - цифрой 0.

В информатике часто используется величина, называемая байтом (byte) и равная 8 битам. И если бит позволяет выбрать один вариант из двух возможных, то байт, соответственно, 1 из 256 (2^8).

Наряду с байтами для измерения количества информации используются более крупные единицы:

1 Килобайт (Кбайт) = 1024 байт = 2^{10} байт = 2^{13} бит;

1 Мегабайт (Мбайт) = 1024 Кбайт = 2^{10} Кбайт = 2^{20} байт = 2^{23} бит;

1 Гигабайт (Гбайт) = 1024 Мбайт = 2^{10} Мбайт = 2^{30} байт = 2^{33} бит;

1 Терабайт (Тбайт) = 1024 Гбайт = 2^{10} Гбайт = 2^{40} байт = 2^{43} бит;

1 Петабайт (Пбайт) = 1024 Тбайт = 2^{10} Тбайт = 2^{50} байт = 2^{53} бит;

1 Эксабайт (Эбайт) = 1024 Пбайт = 2^{10} Пбайт = 2^{60} байт = 2^{63} бит.

Представление информации в компьютере

Любой компьютер предназначен для обработки, хранения, преобразования данных. Для выполнения этих функций компьютер должен обладать некоторыми свойствами представления этих данных. Представление этих данных заключается в их преобразовании в вид, удобный для последующей обработки либо пользователем, либо компьютером. В зависимости от этого данные имеют внешнее и внутреннее представление. Во внешнем представлении (для пользователя) все данные хранятся в виде файлов. Простейшими способами внешнего представления данных являются:

- числовые данные (вещественные и целые);
- текст (последовательность символов);
- изображение (графика, фотографии, рисунки, схемы);
- звук.

Внутреннее представление данных определяется физическими принципами, по которым происходит обмен сигналами между аппаратными средствами компьютера, принципами организации памяти, логикой работы компьютера.

Для автоматизации работы с данными, относящимися к различным типам, важно унифицировать их форму представления. Для этого обычно используется прием кодирования, т. е. выражение данных одного типа через данные другого типа.

Любые данные для обработки компьютером представляются последовательностью двух целых чисел – единицы и нуля. Такая форма представления данных получила название двоичного кодирования.

Кодирование числовых данных

Числовая информация в памяти компьютера хранится и обрабатывается в двоичном коде.

Применяется две формы кодирования двоичных чисел: с фиксированной и плавающей запятой.

В форме с фиксированной запятой хранятся и обрабатываются целые числа. В ячейках памяти из n разрядов можно закодировать (записать) целые числа от $X = 0$ до $X = 2^n - 1$. Например, с помощью 8 бит можно закодировать целые числа от 0 до 255.

В форме с плавающей запятой хранятся и обрабатываются вещественные числа. При этом предполагается запись вещественного числа в экспоненциальном виде:

$$X = \pm m \cdot q^p,$$

где m – мантисса числа ($|M| < 1$); q – основание системы счисления; p – порядок числа (p – целое число).

Кодирование логических данных

Логические данные принимают два значения: «Истина» или «Ложь» (1 или 0). В компьютере для логического значения отводится 2 байта, или 16 разрядов (бит), которые заполняются единицами, если значение «Ложь», и нулями, если значение «Истина».

Кодирование текстовых (символьных) данных

Правило кодирования символьных данных (букв алфавита и других символов) заключается в том, что каждому символу ставится в соответствие двоичный код – совокупность нулей и единиц.

Технически это выглядит просто, но существуют организационные сложности. В первые годы развития вычислительной техники эти сложности были связаны с отсутствием необходимых стандартов, а в настоящее время вызваны, наоборот, избытком одновременно действующих и противоречивых стандартов. Для того чтобы весь мир одинаково кодировал текстовые данные, нужны единые таблицы кодирования.

Наиболее распространенный стандарт кодировки символов ASCII-код (American Standard Code for Information Interchange – американский стандартный код для обмена информацией) был введен институтом стандартизации США в 1963 г. и после модификации в 1977 г. был принят в качестве всемирного стандарта. Каждому символу в этой таблице поставлено в соответствие двоичное число от 0 до 255 (8-битовый двоичный код), например, A – 01000001, B – 01000010, C – 01000011, D – 01000100 и т. д.

В системе ASCII закреплены две таблицы кодирования – базовая и расширенная. Базовая таблица закрепляет значения кодов от 0 до 127, а расширенная относится к символам с номерами от 128 до 255. Первые 32 кода отданы производителям аппаратных средств. В этой области размещаются так называемые управляющие коды, которым не соответствуют никакие символы языков, и, соответственно, эти коды не выводятся ни на экран, ни на устройства печати, но ими можно управлять, например, тем, как производится вывод прочих данных.

Начиная с 32-го кода по 127-й размещены коды символов английского алфавита, знаков препинания, цифр, знаков арифметических действий, некоторые вспомогательные символы (число 127 представляет команду DELETE).

Базовая таблица кодировки ASCII приведена в таблице.

Число	Символ	Число	Символ	Число	Символ	Число	Символ	Число	Символ	Число	Символ
32	пробел	48	0	64	@	80	P	96	`	112	p
33	!	49	1	65	A	81	Q	97	a	113	q
34	“	50	2	66	B	82	R	98	b	114	r
35	#	51	3	67	C	83	S	99	c	115	s
36	\$	52	4	68	D	84	T	100	d	116	t
37	%	53	5	69	E	85	U	101	e	117	u
38	&	54	6	70	F	86	V	102	f	118	v
39	‘	55	7	71	G	87	W	103	g	119	w
40	(	56	8	72	H	88	X	104	h	120	x
41	)	57	9	73	I	89	Y	105	i	121	y
42	*	58	:	74	J	90	Z	106	j	122	z
43	+	59	;	75	K	91	[	107	k	123	{
44	,	60	<	76	L	92	\	108	l	124	
45	-	61	=	77	M	93	]	109	m	125	}
46	.	62	>	78	N	94	^	110	n	126	~
47	/	63	?	79	O	95	_	111	o	127	DEL

Национальные системы кодирования занимают расширенную часть, определяющую значения кодов с 128 до 255.

В России наиболее широкое применение нашли кодировки Windows 1251 (была введена компанией Microsoft), КОИ-8 (код обмена информацией восьмизначный), ISO (International Standard Organization – Международный институт стандартизации) – международная кодировка, в которой предусмотрено кодирование символов русского алфавита.

Организационные трудности, связанные с созданием единой системы кодирования текстовых данных, вызваны ограниченным набором кодов (256). Если, например, кодировать символы не восьмиразрядными двоичными числами, а числами с большим количеством разрядов, то и диапазон возможных значений кодов станет намного больше. Такая система, основанная на 16-разрядном кодировании символов, получила название универсальной UNICODE. Шестнадцать разрядов позволяют обеспечить уникальные коды для 65 536 (2^{16}) различных символов – этого поля достаточно для размещения в одной таблице символов большинства языков планеты.

Несмотря на тривиальную очевидность такого подхода, простой механический переход на данную систему долгое время сдерживался из-за недостаточных ресурсов средств вычислительной техники (в системе кодирования UNICODE все текстовые документы автоматически становятся вдвое длиннее).

Во второй половине 90-х г. технические средства достигли необходимого уровня обеспеченности ресурсами, и сегодня переход документов и программных средств на универсальную систему кодирования в основном осуществлен.

Кодирование графических данных

Графические данные хранятся и обрабатываются в двоичном коде.

Существуют два принципиально разных подхода к кодированию (представлению) графических данных: растровый и векторный.

При растровом представлении вся область данных разбивается на множество точечных элементов – пикселей, каждый из которых имеет свой цвет. Число пикселей по горизонтали и вертикали определяет *разрешение* изображения.

При растровом способе представления графических данных под каждый пиксель отводится определенное число битов, называемое *битовой глубиной* или *информационной емкостью одного пикселя* и используемое для кодирования цвета пикселя. Каждому цвету соответствует двоичный код. Например, если битовая глубина равна 1, то под каждый пиксель отводится 1 бит. В этом случае 0 соответствует черному цвету, 1 – белому, а изображение может быть только черно-белым. Если битовая глубина равна 2, то каждый пиксель может быть закодирован цветовой гаммой из 4 цветов (2^2) и т. д. Для качественного представления графических данных в современных компьютерах используются цветовые схемы с битовой глубиной 8, 24, 32, 40, т. е. каждый пиксель может иметь 2^8 , 2^{24} , 2^{32} , 2^{40} оттенков. Количество цветов N , отображаемых на экране монитора, может быть вычислено по формуле

$$N = 2^i, \quad (1)$$

где i – битовая глубина.

Если известны размеры (в пикселях) рисунка по высоте X и ширине Y , а также битовая глубина i , то занимаемый объем V будет равен

$$V = X \cdot Y \cdot i. \quad (2)$$

Основным недостатком растровой графики является большой объем памяти, необходимый для хранения изображения. Это объясняется тем, что запоминается цвет каждого пикселя, общее число которых задается разрешением.

При векторном представлении графических данных задается и впоследствии сохраняется математическое описание графического примитива – геометрического объекта (отрезка, окружности, прямоугольника и т. п.), из которых формируется изображение. Например, для воспроизведения окружности достаточно запомнить положение ее центра, радиус, толщину и цвет линии. Благодаря этому для хранения векторных графических данных требуется значительно меньше памяти.

Программы для работы с графическими данными делятся на растровые графические редакторы (Paint, Photoshop) и векторные графические редакторы (CorelDraw, Adobe Illustrator, Visio).

Приведем краткие характеристики наиболее популярных графических форматов.

BMP (Bit Map image) – растровый формат, используемый в системе Windows. Поддерживается большинством графических редакторов (в частности, Paint и Photoshop). Применяется для хранения отсканированных изображений и обмена данными между различными приложениями.

TIFF (Tagged Image File Format) – растровый формат, поддерживаемый различными операционными системами. Включает алгоритм сжатия без потери качества изображения. Используется в сканерах, а также для хранения и обмена данными.

GIF (Graphics Interchange Format) – растровый формат, включающий в себя алгоритм сжатия, значительно уменьшающий объем файла без потери информации. Поддерживается приложениями для различных операционных систем. Применяется в изображениях, содержащих до 256 цветов, а также для создания анимации. Используется для размещения графики в Интернете.

JPEG (Joint Photographic Expert Group) – растровый формат, содержащий алгоритм сжатия, который уменьшает объем файла в десятки раз, но приводит к необратимой потере части информации. Поддерживается большинством операционных систем. Используется для размещения графических изображений на web-страницах в Интернете.

PNG (Portable Network Graphic) – растровый формат, аналогичный GIF. Используется для размещения графики в Интернете.

WMF (Windows Meta File) – векторный формат для Windows-приложений.

EPS (Encapsulated Post Script) – векторный формат, поддерживаемый большинством операционных систем.

CDR – векторный формат, поддерживаемый графической системой Corel-Draw.

Для представления цвета используются цветовые модели.

Цветовая модель – это правило, по которому может быть вычислен цвет. Самая простая цветовая модель – битовая. В ней для описания цвета каждого пикселя (черного или белого) используется всего один бит. Для представления полноцветных изображений используются более сложные модели, среди которых самые известные – модели RGB и CMYK.

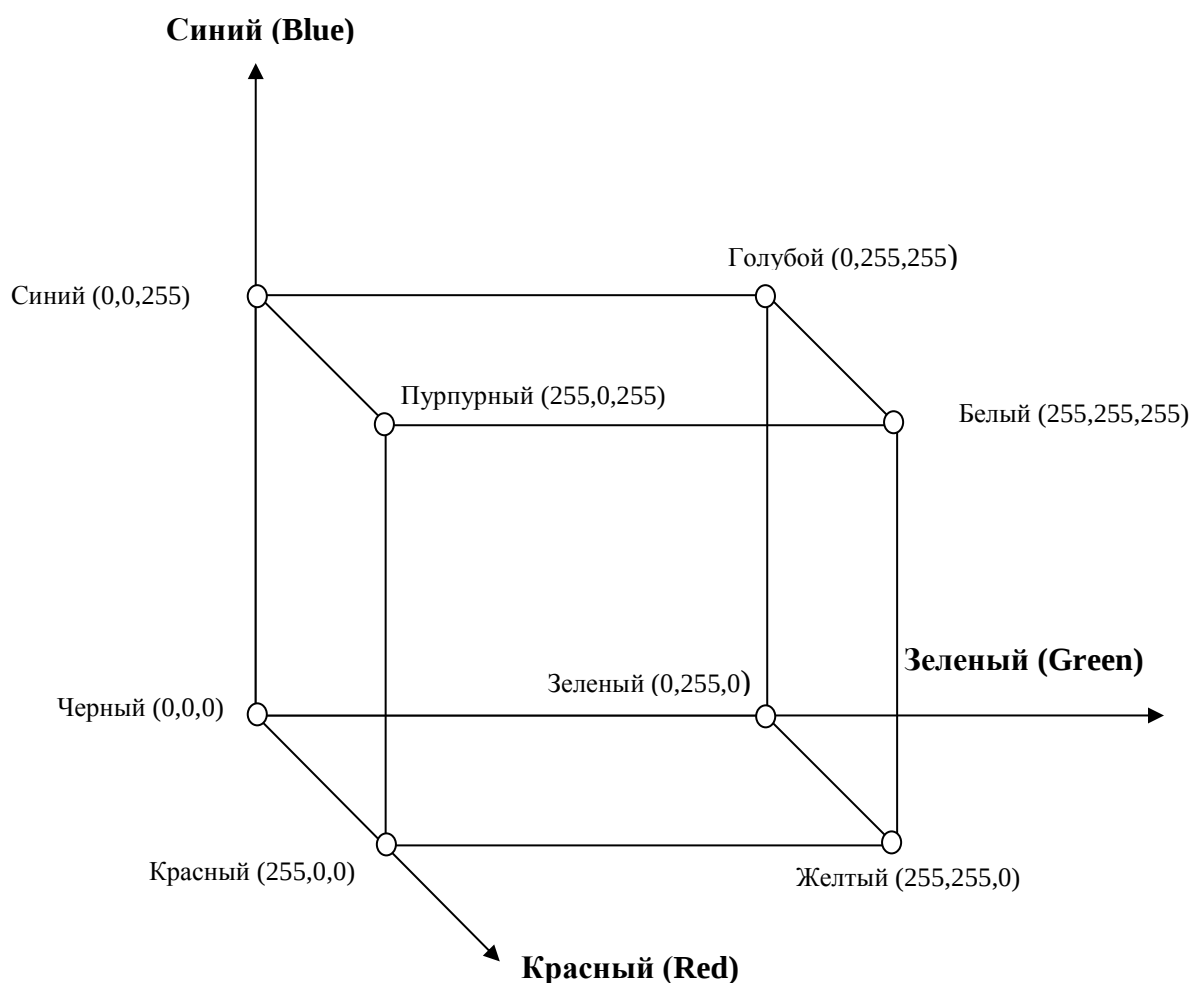
Цветовая модель RGB используется в таких устройствах, как телевизионные кинескопы, компьютерные мониторы.

Цветовая модель RGB (Red-Green-Blue, красный-зеленый-синий) основана на том, что любой цвет может быть представлен как сумма трех основных цветов: красного, зеленого и синего.

В основе цветовой модели лежит декартова система координат. Цветовое пространство представляет собой куб сочетаний трех базовых цветов (рисунок). Любой оттенок цвета при этом выражается набором из трех чисел. На каждое число отводится один байт, поэтому интенсивность одного цвета имеет 256 значений (0–255), общее количество оттенков цвета – $1\,677\,7216(2^{24})$. Белый цвет в RGB представляется как (255,255,255), черный – (0,0,0,0), красный – (255,0,0), зеленый – (0,255,0), синий – (0,0,255).

Цветовая модель CMYK используется в полиграфии.

Цветовая модель CMY является производной модели RGB и также построена на базе трех цветов: C – Cyan (голубого), M – Magenta (пурпурного), Y – Yellow (желтого), которые образуются следующим образом.



Голубой цвет $C(0,255,255)$ является комбинацией синего и зеленого, желтый цвет $Y(255,255,0)$ – зеленого и красного, а пурпурный цвет $M(255,0,255)$ – красного и синего, в противном случае каждому из основных цветов ставится в соответствие дополнительный цвет (дополняющий основной до белого).

Дополнительными цветами для красного является голубой, для зеленого – пурпурный, для синего – желтый.

Смешение голубого, пурпурного и желтого цветов должно давать черный цвет, который, однако, выглядит осветленным по сравнению с оригиналом. Поэтому для получения чистого черного цвета при печати цветовая модель CMY расширяется до модели CMYK, содержащей четвертый основной цвет – черный (K – black).

Кодирование звуковой информации

Звук представляет собой звуковую волну с непрерывно меняющейся амплитудой и частотой. В процессе кодирования непрерывного сигнала производится его временная дискретизация и квантование.

Дискретизация заключается в замерах величины аналогового сигнала огромное множество раз в секунду. Полученной величине аналогового сигнала сопоставляется определенное значение из заранее выделенного диапазона: 256 (8 бит) или 65 536 (16 бит). Приведение в соответствие уровня сигнала определенной величине диапазона называется квантованием.

Как бы часто ни проводились измерения, все равно часть информации будет теряться. Но чем чаще проводятся замеры, тем точнее будет соответствовать цифровой звук своему аналоговому оригиналу.

С одной стороны, чем больше бит отведено под кодирование уровня сигнала (квантование), тем точнее соответствие. С другой стороны, звук хорошего качества будет содержать больше данных и, следовательно, больше занимать места на цифровом носителе информации.

Определить информационный объем V цифрового аудиофайла, длительность звучания которого составляет t секунд при частоте дискретизации N и разрешении i битов (квантуют i битами), можно по формуле

$$V = N \cdot i \cdot t. \quad (3)$$

Если требуется определить информационный объем стереоаудиофайла, то полученные вычисления умножаются на 2:

$$V = N \cdot i \cdot t \cdot 2. \quad (4)$$

Измерение количества информации

Рассмотрим два подхода к измерению информации – содержательный (вероятностный) и символьный (алфавитный).

В содержательном подходе возможна качественная оценка информации: новая, срочная, важная и т. д. Согласно К. Шеннону, информативность сообщения характеризуется содержащейся в нем полезной информацией – той частью сообщения, которая снимает полностью или уменьшает неопределенность ка-

кой-либо ситуации. Неопределенность некоторого события – это количество возможных исходов данного события. Например, неопределенность погоды на завтра обычно заключается в диапазоне температуры воздуха и возможности выпадения осадков.

Содержательный подход часто называют субъективным, так как разные люди (субъекты) информацию об одном и том же предмете оценивают по-разному. Но если число исходов не зависит от суждений людей (например, случай бросания кубика или монеты), то информация о наступлении одного из возможных исходов является объективной.

Формулу для вычисления количества информации, учитывающую неодинаковую вероятность событий, предложил К. Шеннон в 1948 г. Количественная зависимость между вероятностью события p и количеством информации I в сообщении о нем выражается формулой Шеннона

$$I = -\log_2 p. \quad (5)$$

Качественную связь между вероятностью события и количеством информации в сообщении об этом событии можно выразить следующим образом: чем меньше вероятность некоторого события, тем больше информации содержит сообщение об этом событии.

Количество информации для событий с различными вероятностями определяется по формуле (эту формулу также называют формулой Шеннона)

$$I = - \sum_{i=1}^n p_i \log_2 p_i. \quad (6)$$

Рассмотрим пример. В коробке имеется 100 шаров. Из них 80 белых и 20 черных. Очевидно, вероятность того, что при вытаскивании случайным образом попадет белый шар, больше, чем вероятность попадания черного. Проведем количественную оценку вероятности для каждой ситуации. Обозначим $p_{\text{ч}}$ – вероятность, что вытащили черный шар, $p_{\text{б}}$ – вероятность, что вытащили белый шар. Тогда: $p_{\text{ч}} = 20/100 = 0,2$, $p_{\text{б}} = 80/100 = 0,8$. Заметим, что вероятность попадания белого шара в 4 раза больше, чем черного.

Количество информации в сообщении, что вынутый случайным образом шар является черным, вычисляется по формуле

$$I_ч = -\log_2(0,2) = 2,321928 \text{ бит.}$$

Количество информации в сообщении, что вынутый случайным образом шар является белым, вычисляется по формуле

$$I_б = -\log_2(0,8) = 0,321928 \text{ бит.}$$

Количество информации в сообщении о цвете вынутого случайным образом шара вычисляется по формуле

$$-0,2 \log_2(0,2) - 0,8 \log_2(0,8) = 0,2 \cdot 2,321928 + 0,8 \cdot 0,321928 = 0,721928 \text{ бит.}$$

Если события равновероятны ($p_i = 1/N$, где N – число возможных событий), то величина количества информации I вычисляется по формуле Р. Хартли:

$$I = \log_2 N. \quad (7)$$

Используя формулу (7), можно записать и формулу, которая связывает количество возможных событий N и количество информации I

$$N = 2^I. \quad (8)$$

Алфавитный подход основан на том, что всякое сообщение можно закодировать с помощью конечной последовательности символов некоторого алфавита. С позиций информатики носителями информации являются любые последовательности символов, которые хранятся, передаются и обрабатываются с помощью компьютера. Информативность последовательности символов зависит не от содержания сообщения, а определяется минимально необходимым количеством символов для кодирования этой последовательности символов. Алфавитный подход является объективным, т. е. он не зависит от субъекта, воспринимающего сообщение. Смысл сообщения либо учитывается на этапе выбора алфавита кодирования, либо не учитывается вообще.

При алфавитном подходе, если допустить, что все символы алфавита встречаются в тексте с одинаковой частотой (равновероятно), то количество информации, которое несет каждый символ (информационный вес одного символа), вычисляется по формуле

$$I = \log_2 N, \quad (9)$$

где N – мощность алфавита (полное количество символов, составляющих алфавит выбранного кодирования). Тогда мощность алфавита можно вычислить по формуле

$$N = 2^I. \quad (10)$$

В алфавите, который состоит из двух символов (двоичное кодирование), каждый символ несет 1 бит ($\log_2 2 = 1$) информации; из четырех символов – каждый символ несет 2 бита информации ($\log_2 4 = 2$); из восьми символов – 3 бита ($\log_2 8 = 3$) и т. д. Один символ из алфавита мощностью 256 несет в тексте 8 битов ($\log_2 256 = 8$) информации.

Если весь текст состоит из k символов, то при алфавитном подходе размер содержащейся в нем информации H определяется по формуле

$$H = k \cdot I, \quad (11)$$

где I – информационный вес одного символа в используемом алфавите.

Максимальное количество слов L из m букв, которое можно составить из алфавита мощностью N , определяется по формуле

$$L = N^m. \quad (12)$$

Примеры решения задач

Задача 1. Два игрока играют в «крестики нолики» на поле размером 4×4 . Определить, какое количество информации I получит второй игрок после первого хода первого игрока.

Решение. Первый игрок может для первого хода выбрать любое поле из 16 возможных ($N = 4 \cdot 4 = 16$). Тогда по формуле (7)

$$I = \log_2 16 = \log_2 2^4 = 4 \text{ бита.}$$

Количество информации I можно также найти из соотношения (8)

$$16 = 2^I \quad 2^4 = 2^I \quad I = 4 \text{ бита.}$$

Ответ: $I = 4$ бита.

Задача 2. В группе 24 студента. За экзамен были получены следующие оценки: 3 пятерки, 12 четверок, 6 троек, 3 двойки.

1) Определить, какое количество информации I содержит сообщение, что студент Романов получил оценку «четыре».

2) Определить, какое количество информации I содержит сообщение об оценке любого студента группы.

Решение.1) Вероятность события, что случайным образом выбранный студент получил оценку «четыре», равна $p = \frac{12}{24} = \frac{1}{2}$. Используя формулу (5), получим

$$I = -\log_2 p = -\log_2 \frac{1}{2} = -\log_2 2^{-1} = \log_2 2 = 1 \text{ бит.}$$

Ответ: $I = 1$ бит.

2) Для решения задачи воспользуемся формулой Шеннона (6). Вероятности событий, что случайным образом выбранный студент получил оценку «пять», «четыре», «три», «два», соответственно равны: $p_1 = \frac{3}{24} = \frac{1}{8}$, $p_2 = \frac{12}{24} = \frac{1}{2}$, $p_3 = \frac{6}{24} = \frac{1}{4}$, $p_4 = \frac{3}{24} = \frac{1}{8}$.

$$I = -(p_1 \cdot \log_2 p_1 + p_2 \cdot \log_2 p_2 + p_3 \cdot \log_2 p_3 + p_4 \cdot \log_2 p_4) =$$
$$-(\frac{1}{8} \log_2 \frac{1}{8} + \frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{4} \log_2 \frac{1}{4} + \frac{1}{8} \log_2 \frac{1}{8}) = \frac{3}{8} + \frac{1}{2} + \frac{2}{4} + \frac{3}{8} = 1,75.$$

Ответ: $I = 1,75$ бита.

Задача 3. В коробке лежат красные и синие карандаши, всего в коробке 24 карандаша. Информация о том, что из коробки случайным образом достали синий карандаш, равна 2 битам. Определить, сколько в коробке красных и синих карандашей.

Решение. Обозначим за x число синих карандашей в коробке. Для решения задачи воспользуемся формулой (5): $2 = -\log_2 p$.

Из этого соотношения найдем p – вероятность того, что случайным образом вынутый шар является синим: $p = \frac{1}{4}$. Теперь определим x из соотношения

$$\frac{x}{24} = \frac{1}{4} \quad x = 6.$$

Ответ: В коробке 6 синих и 18 красных карандашей.

Задача 4. Растровое графическое изображение 20x20 точек содержит не более 256 цветов. Сколько памяти потребуется для хранения изображения?

Решение. Для решения воспользуемся формулой (1). Одна точка может иметь один из 256 цветов ($N = 256$). Найдем сколько бит i , требуется для ее хранения (битовая глубина) из соотношения:

$$256 = 2^i \quad i = 8 \text{ бит.}$$

Для хранения изображения 20x20 точек требуется $20 \cdot 20 \cdot 8 = 3200$ бит или 400 байт ($3200/8 = 400$).

Ответ: Для хранения изображения потребуется 400 байт.

Задача 5. Сообщение из 30 символов было записано в 8-битной кодировке Windows-1251. После вставки в текстовый редактор сообщение было перекодировано в 16-битный код Unicode. На сколько байт увеличилось при этом количество памяти?

Решение. При перекодировке из Windows-1251 в Unicode объем памяти увеличивается в два раза, т. е. если в кодировке Windows-1251 сообщение занимало $30 \cdot 8 = 240$ бит, то в кодировке Unicode сообщение займет $30 \cdot 16 = 480$ бит, т. е. количество памяти увеличилось на $480 - 240 = 240$ бит, или $240/8 = 30$ байт.

Ответ: Сообщение увеличилось на 30 байт.

Задача 6. Отправлено SMS-сообщение:

А не могу без тебя жить!

Мне и в дожди без тебя – сушь,

Мне и в жару без тебя – стыть,

Мне без тебя и Москва – глушь.

В мобильном телефоне адресата установлено ограничение размера входящего SMS-сообщения 64 байтами (при превышении этого размера сообщение автоматически делится на части). Каждый символ кодируется 16 битами. На сколько частей будет разбито сообщение?

Решение. Всего символов в сообщении 114. Так как каждый символ кодируется 16 битами (2 байтами), то сообщение занимает $114 \cdot 2 = 228$ байт. Теперь вычислим, на сколько частей будет разбито сообщение: $228/64 = 3,56$.

Ответ: Сообщение будет разбито на 4 части.

Задача 7. Сообщение содержит 4096 символов. Объем сообщения при использовании равномерного кода составил 1/512 Мбайт. Какова мощность алфавита, с помощью которого записано сообщение?

Решение. Определим, какой объем памяти (в битах) занимает один символ. Для этого переведем 1/512 Мбайт в биты и полученный результат разделим на число символов, содержащееся в сообщении:

$$\frac{1024 \cdot 1024 \cdot 8}{512 \cdot 4096} = \frac{2^{10} \cdot 2^{10} \cdot 2^3}{2^9 \cdot 2^{12}} = \frac{2^{23}}{2^{21}} = 4 \text{ бита.}$$

Для определения мощности алфавита используем формулу (10).

$$N = 2^4 = 16 \text{ символов.}$$

Ответ: Мощность алфавита 16 символов.

Задача 8. Скорость передачи данных через ADSL соединения равна 256 000 бит/сек. Передача файла заняла 4 минуты. Определить размер файла в Кбайтах.

Решение. Определим размер файла как произведение скорости передачи на время:

$$256\,000 \cdot 4 \cdot 60 \text{ бит} = 256\,000 \cdot 4 \cdot 60 / 8 / 1024 \text{ Кбайт} = 7500 \text{ Кбайт.}$$

Ответ. Размер файла составляет 7500 Кбайт.

Задача 9. Определить информационный объем цифрового стереоаудиофайла, длительность звучания которого составляет 10 секунд при частоте дискретизации 22,05 кГц и разрешении 8 битов (квантуется 8 битами).

Решение. Для определения информационного объема цифрового стерео аудио файла воспользуемся формулой (4).

$$V = 22\,050 \cdot 8 \cdot 10 \cdot 2 = 3528000 (\text{бит}) = 3528000 / 8 / 1024 / 1024 (\text{Мбайт}) = 0,42 (\text{Мбайт}).$$

Ответ: Информационный объем цифрового стерео аудио файла составляет 0,42 Мбайт.

Задача 10. Световое табло состоит из лампочек. Каждая лампочка может находиться в одном из трех состояний («включено», «выключено», «мигает»).

Какое наименьшее количество лампочек должно быть на табло, чтобы с его помощью можно было передать 18 различных сигналов.

Решение. Воспользуемся формулой (12). Мощность алфавита $N = 3$. Требуется найти m (наименьшее количество лампочек). Так как в формуле (12) определяется максимальное количество слов, а необходимо передать только 18 сигналов (слов), то m будем находить из соотношения $18 \leq 3^m$. Следовательно, $m = 3, 4, 5, \dots$. Поскольку нужно найти наименьшее количество лампочек, то $m = 3$.

Ответ: На табло должно быть 3 лампочки.

Задача 11. В велокроссе участвуют 720 спортсменов. Устройство регистрирует прохождение промежуточного финиша каждым из участников, записывая его номер с использованием минимально возможного количества бит, одинакового для всех номеров. Каков информационный объем сообщения, записанного устройством после того, как промежуточный финиш прошли 100 велосипедистов?

Решение. Для регистрации одного любого номера необходимо 10 бит, поскольку с помощью 10 бит можно закодировать $2^{10} = 1024$ различных номеров (9 бит будет недостаточно). Для регистрации 100 номеров потребуется $100 \cdot 10 = 1000$ бит = $1000/8$ байт = 125 байт.

Ответ: Информационный объем сообщения равен 125 байт.

Задачи для самостоятельного решения

1. В корзине лежат 16 шаров. Все шары разного цвета. Сколько информации несет сообщение о том, что из корзины выкатился красный шар?
2. Сколько бит информации несет сообщение о том, что из колоды в 32 карты достали даму пик?
3. Какое сообщение содержит большее количество информации?
 - 1) Монета упала «решкой» вверх.
 - 2) В библиотеке книга нашлась в 5-м шкафу из 8.

- 3) Роман получил за экзамен 3 балла (единицы не ставятся) по 5-балльной системе.
- 4) Из колоды карт (32 шт.) выпала семерка пик.
4. В корзине лежат шары. Все разного цвета. Сообщение о том, что достали синий шар, несет 5 бит информации. Сколько всего шаров в корзине?
5. В соревновании участвуют 4 команды. Сколько информации в сообщении, что выиграла 3-я команда?
6. В коробке 5 синих и 15 красных шариков. Какое количество информации несет сообщение, что из коробки достали синий шарик?
7. В коробке находятся кубики трех цветов: красного, желтого и зеленого, причем желтых в два раза больше красных, а зеленых на 6 больше, чем желтых. Сообщение о том, что из коробки случайно вытащили желтый кубик, содержало 2 бита информации. Сколько было зеленых кубиков?
8. Студенты группы изучают один из трех языков: английский, немецкий или французский, причем 12 студентов не учат английский. Сообщение, что случайно выбранный студент Петров изучает английский, несет $\log_2 3$ бит информации, а что Иванов изучает французский – 1 бит. Сколько студентов изучают немецкий язык?
9. В составе 16 вагонов, среди которых К – купейные, П – плацкартные и СВ – спальные. Сообщение о том, что ваш друг приезжает в СВ, несет 3 бита информации. Сколько в поезде вагонов СВ?
10. Студенческая группа состоит из 21 человека, которые изучают немецкий или французский языки. Сообщение о том, что студент А изучает немецкий язык, несет $\log_2 3$ бит информации. Сколько человек изучают французский язык?
11. Сколько информации несет сообщение о том, что было угадано число в диапазоне целых чисел от 684 до 811?
12. При угадывании целого числа в некотором диапазоне было получено 8 бит информации. Сколько чисел содержал этот диапазон?
13. Сообщение о том, что ваш друг живет на 10-м этаже, несет 4 бита информации. Сколько этажей в доме?

14. На железнодорожном вокзале 8 путей отправления поездов. Вам сообщили, что ваш поезд прибывает на четвертый путь. Сколько информации вы получили?
15. Какое количество информации содержит сообщение, уменьшающее неопределенность знаний в 8 раз?
16. Сколько бит информации содержит сообщение объемом 4 мегабайта?
17. Сколько мегабайт информации содержит сообщение объемом 223 бит?
18. Сколько Гбайт содержится в 219 Кбайтах информации?
19. Книга состоит из 64 страниц. На каждой странице 256 символов. Какой объем информации содержится в книге, если используемый алфавит состоит из 32 символов?
20. Световое табло состоит из лампочек, каждая из которых может находиться в двух состояниях («включено» или «выключено»). Какое наименьшее количество лампочек должно находиться на табло, чтобы с его помощью можно было передать 50 различных сигналов?
21. Метеорологическая станция ведет наблюдение за влажностью воздуха. Результатом одного измерения является целое число от 0 до 100 %, которое записывается при помощи минимально возможного количества бит. Станция сделала 80 измерений. Каков информационный объем результатов наблюдений?
22. Для передачи секретного сообщения используется код, состоящий из десятичных цифр. При этом все цифры кодируются одним и тем же (минимально возможным) количеством бит. Определите информационный объем сообщения длиной в 150 символов.
23. Два текста содержат одинаковое количество символов. Первый текст составлен в алфавите мощностью 16 символов. Второй текст в алфавите мощностью 256 символов. Во сколько раз количество информации во втором тексте больше, чем в первом?
24. В некоторой кодировке слово из 15 букв занимает информационный объем на 39 байт больше, чем слово из двух букв. Каким количеством бит кодиру-

ется одна буква, если учесть, что под все символы этой кодировки выделяется равный объем памяти?

25. Одна ячейка памяти «троичной ЭВМ» может принимать одно из трех возможных состояний. Для хранения некоторой величины отвели 5 ячеек памяти. Сколько значений может принимать эта величина?
26. Какое наименьшее количество символов должно быть в алфавите, чтобы при помощи всевозможных слов, состоящих из четырех символов данного алфавита, можно было передать не менее 100 различных сообщений?
27. Для передачи сигналов на флоте используется «флажковая азбука». Какое количество различных сигналов можно передать при помощи двух сигнальных флажков, если всего имеются флаги шести различных видов?
28. Азбука Морзе позволяет кодировать символы для радиосвязи, задавая комбинацию из точек и тире. Сколько различных символов (цифр, букв, знаков пунктуации и т.д.) можно закодировать, используя код Морзе длиной не менее пяти и не более шести сигналов (точек и тире)?
29. За 45 секунд был распечатан текст. Подсчитать количество страниц в тексте, если известно, что в среднем на странице 50 строк по 75 символов в каждой, скорость печати лазерного принтера 8 Кбит/с, 1 символ – 1 байт. Ответ округлить до целой части.
30. Лазерный принтер печатает со скоростью в среднем 7 Кбит в секунду. Сколько времени понадобится для распечатки 12-страничного документа, если известно, что на одной странице в среднем по 45 строк, в строке 60 символов (1 символ – 1 байт). Результат округлить до целой части.
31. Можно ли уместить на одну дискету (1,44 Мбайт) книгу, имеющую 432 страницы, причем на каждой странице этой книги 46 строк, а в каждой строке 62 символа?
32. Имеется 2 текста на разных языках. Первый текст использует 32-символьный алфавит и содержит 200 символов, второй – 16-символьный алфавит и содержит 250 символов. Какой из текстов содержит большее количество информации и на сколько бит?

33. Автоматическое устройство осуществило перекодировку информационного сообщения на русском языке, первоначально записанного в 16-битном коде Unicode, в 8-битную кодировку КОИ-8. При этом информационное сообщение уменьшилось на 480 бит. Какова длина сообщения в символах?
34. Сообщение занимает 3 страницы по 25 строк. В каждой строке записано по 60 символов. Сколько символов в использованном алфавите, если все сообщение содержит 1125 байт?
35. Сколько символов в тексте, если мощность алфавита – 64 символа, а объем информации, содержащейся в нем, 1,5 Кбайт?
36. Для записи сообщения использовался 64-символьный алфавит. Каждая страница содержит 30 строк. Все сообщение содержит 8775 байтов информации и занимает 6 страниц. Сколько символов в строке?
37. Скорость передачи данных через модемное соединение 28 Кбит/с. Передача текстового файла заняла 10 с. Сколько символов содержал переданный текст, если известно, что он был представлен в кодировке Unicode?
38. Для хранения растрового изображения размером 128x128 пикселей отвели 4 Кбайта памяти. Каково максимально возможное число цветов в палитре изображения?
39. Укажите минимальный объем памяти (в килобайтах), достаточный для хранения любого растрового изображения размером 64×64 пикселя, если известно, что в изображении используется палитра из 256 цветов. Саму палитру хранить не нужно.
40. Скорость передачи данных через ADSL-соединение равна 128 000 бит/с. Через данное соединение передают файл размером 625 Кбайт. Определите время передачи файла в секундах.
41. Системный администратор ограничил длительность непрерывного подключения компьютеров сотрудников организации к сети Интернет 10 минутами. Сотруднику требуется переслать файл размером 100 Мбайт. Скорость передачи информации с рабочего места (компьютера) сотрудника в среднем со-

ставляет 512 Кбит/с. На сколько частей необходимо разделить файл для пересылки?

42. Два шифровальщика обменялись сообщениями по 200 закодированных символов. Кодовая таблица первого содержит N символов, второго – в 4 раза больше. На сколько больше бит информации передал второй шифровальщик?
43. На магнитном диске объемом 30 Мбайт записана книга. В книге 1552 страницы. Из них страниц с текстом на 752 больше, чем страниц с рисунками. Страница с текстом содержит 640 символов. Все рисунки восьми цветные и имеют единый формат. Каков размер рисунков?
44. Для регистрации на сайте пользователю необходимо придумать пароль длиной ровно 11 символов. В пароле можно использовать десятичные цифры и 32 различных символа местного алфавита, причем все буквы используются в двух начертаниях – строчные и прописные. Каждый символ кодируется одинаковым и минимально возможным количеством бит, а каждый пароль – одинаковым и минимально возможным количеством байт. Каков объем памяти, необходимый для хранения 50 паролей?
45. Каков объем памяти для хранения цифрового аудиофайла, время звучания которого составляет две минуты при частоте дискретизации 44,1 кГц и разрешении 16 бит.

Библиографический список

1. Акулов, О. А. Информатика: базовый курс : учеб. пособие для студентов вузов / О. А. Акулов, Н. В. Медведев. – М. : Омега-Л, 2005. – 552 с.
2. Могилев, А. В. Информатика : учеб. пособие для студ. высш. учеб. заведений / А. В. Могилев, Н. И. Пак, Е. К. Хеннер. – М. : Академия, 2008. – 848 с.
3. Могилев, А. В. Практикум по информатике : учеб. пособие для студ. высш. учеб. заведений / А. В. Могилев, Н. И. Пак, Е. К. Хеннер. – М. : Академия, 2008. – 608 с.
4. Макарова, Н. В. Информатика : учеб. для вузов / Н. В. Макарова, В. Б. Волков. – СПб. : Питер, 2011. – 576 с.
5. Дергачева, Л. М. Решение типовых экзаменационных задач по информатике : учеб. пособие / Л. М. Дергачева. – М. : БИНОМ. Лаборатория знаний, 2013. – 360 с.
6. Степанов, А. Н. Информатика : учеб. для вузов / А. Н. Степанов. – СПб. : Питер, 2008. – 684 с.

Оглавление

Понятие информации	3
Свойства информации.....	4
Единицы измерения количества информации	5
Представление информации в компьютере.....	6
Кодирование числовых данных.....	7
Кодирование логических данных.....	7
Кодирование текстовых (символьных) данных.....	8
Кодирование графических данных	10
Кодирование звуковой информации.....	14
Измерение количества информации.....	14
Примеры решения задач	17
Задачи для самостоятельного решения	21
Библиографический список.....	27

Измерение информации

Методические указания
к выполнению практической работы по информатике
для студентов всех направлений дневной формы обучения

Нина Демидовна Берман
Нина Ивановна Шадрина

Главный редактор *Л. А. Суевалова*
Редактор *Л. С. Бакаева*

Подписано в печать . Формат 60×84 ¹/₁₆. Бумага писчая. Гарнитура «Таймс».
Печать цифровая. Усл. печ. л. Тираж 200 экз. Заказ

Издательство Тихоокеанского государственного университета.
680035, Хабаровск, ул. Тихоокеанская, 136.

Отдел оперативной полиграфии издательства
Тихоокеанского государственного университета.
680035, Хабаровск, ул. Тихоокеанская, 136.