Меры и единицы количества и объема информации

ПОДХОДЫ К ИЗМЕРЕНИЮ ИНФОРМАЦИИ

ТЕОРЕТИЧЕСКИЕ ПОЛОЖЕНИЯ

Объемный (алфавитный)
подход к измерению информации

Кодирование текстовой информации

- Текст – набор символов (буквы, цифры, знаки препинания, спецсимволы, пробелы между словами).
- Формула определения количества информации: **$N = 2^I$**
- Достаточный алфавит - мощностью 256 символов. 1 символ – 1 байт
- Каждому символу ставится в соответствие восьмиразрядный двоичный код.

Таблицы кодировки

- Таблица кодировки - таблица, в которой всем символам компьютерного алфавита поставлены в соответствие порядковый номер и восьмизначный двоичный код.
- Существуют различные таблицы кодировок кириллицы.

Двоичный код	Десятичный код	КОИ8	CP1251	CP866	Mac	ISO
11000010	194	б	В	-	-	Т

Таблица кодировки **ASCII**

8-рядная

255 символов

□ Структура таблицы кодировки ASCII

Порядковый номер		Символ
0-31	00000000 - 00011111	Управляющими символами Их функция – управление процессом вывода текста на экран или печать, подача звукового сигнала, разметка текста и т.п.
32-127	00100000 - 01111111	Стандартная часть таблицы (английский). Строчные и прописные буквы латинского алфавита, десятичные цифры, знаки препинания, всевозможные скобки, коммерческие и другие символы.
128-255	10000000 - 11111111	Альтернативная часть таблицы (русская). Вторая половина кодовой таблицы называется кодовой страницей. В первую очередь используется для размещения национальных алфавитов, отличных от латинского.

Таблица ASCII

символ	10-й код	2-й код									
	32	00100000	8	56	00111000	P	80	01010000	h	104	01101000
!	33	00100001	9	57	00111001	Q	81	01010001	i	105	01101001
"	34	00100010	:	58	00111010	R	82	01010010	j	106	01101010
#	35	00100011	;	59	00111011	S	83	01010011	k	107	01101011
\$	36	00100100	<	60	00111100	T	84	01010100	l	108	01101100
%	37	00100101	=	61	00111101	U	85	01010101	m	109	01101101
&	38	00100110	>	62	00111110	V	86	01010110	n	110	01101110
'	39	00100111	?	63	00111111	W	87	01010111	o	111	01101111
(	40	00101000	@	64	01000000	X	88	01011000	p	112	01110000
)	41	00101001	A	65	01000001	Y	89	01011001	q	113	01110001
*	42	00101010	B	66	01000010	Z	90	01011010	r	114	01110010
+	43	00101011	C	67	01000011	[	91	01011011	s	115	01110011
,	44	00101100	D	68	01000100	\	92	01011100	t	116	01110100
-	45	00101101	E	69	01000101	]	93	01011101	u	117	01110101
.	46	00101110	F	70	01000110	^	94	01011110	v	118	01110110
/	47	00101111	G	71	01000111	_	95	01011111	w	119	01110111
0	48	00110000	H	72	01001000	`	96	01100000	x	120	01111000
1	49	00110001	I	73	01001001	a	97	01100001	y	121	01111001
2	50	00110010	J	74	01001010	b	98	01100010	z	122	01111010
3	51	00110011	K	75	01001011	c	99	01100011	{	123	01111011
4	52	00110100	L	76	01001100	d	100	01100100		124	01111100
5	53	00110101	M	77	01001101	e	101	01100101	}	125	01111101
6	54	00110110	N	78	01001110	f	102	01100110	~	126	01111110
7	55	00110111	O	79	01001111	g	103	01100111	□	127	01111111

символ	10-й код	2-й код									
Ъ	128	10000000		160	10100000	А	192	11000000	а	224	11100000
Г	129	10000001	Ѹ	161	10100001	Б	193	11000001	б	225	11100001
,	130	10000010	ѹ	162	10100010	В	194	11000010	в	226	11100010
г	131	10000011	Ј	163	10100011	Г	195	11000011	г	227	11100011
..	132	10000100	Ѱ	164	10100100	Д	196	11000100	д	228	11100100
...	133	10000101	Ѓ	165	10100101	Е	197	11000101	е	229	11100101
†	134	10000110	Ѕ	166	10100110	Ж	198	11000110	ж	230	11100110
‡	135	10000111	Ї	167	10100111	З	199	11000111	з	231	11100111
€	136	10001000	Е	168	10101000	И	200	11001000	и	232	11101000
‰	137	10001001	©	169	10101001	Й	201	11001001	й	233	11101001
Љ	138	10001010	€	170	10101010	К	202	11001010	к	234	11101010
<	139	10001011	«	171	10101011	Л	203	11001011	л	235	11101011
Њ	140	10001100	–	172	10101100	М	204	11001100	м	236	11101100
Ќ	141	10001101	-	173	10101101	Н	205	11001101	н	237	11101101
Ђ	142	10001110	®	174	10101110	О	206	11001110	о	238	11101110
Џ	143	10001111	Ї	175	10101111	П	207	11001111	п	239	11101111
ђ	144	10010000	°	176	10110000	Р	208	11010000	р	240	11110000
‘	145	10010001	±	177	10110001	С	209	11010001	с	241	11110001
’	146	10010010	І	178	10110010	Т	210	11010010	т	242	11110010
“	147	10010011	і	179	10110011	У	211	11010011	у	243	11110011
”	148	10010100	г	180	10110100	Ф	212	11010100	ф	244	11110100
•	149	10010101	и	181	10110101	Х	213	11010101	х	245	11110101
–	150	10010110	¶	182	10110110	Ц	214	11010110	ц	246	11110110
—	151	10010111	·	183	10110111	Ч	215	11010111	ч	247	11110111
□	152	10011000	ë	184	10111000	Ш	216	11011000	ш	248	11111000
™	153	10011001	№	185	10111001	Щ	217	11011001	щ	249	11111001
љ	154	10011010	€	186	10111010	Ъ	218	11011010	ъ	250	11111010
›	155	10011011	»	187	10111011	Ы	219	11011011	ы	251	11111011
њ	156	10011100	ј	188	10111100	Ь	220	11011100	ь	252	11111100
ќ	157	10011101	ѕ	189	10111101	Э	221	11011101	э	253	11111101
ћ	158	10011110	ѕ	190	10111110	Ю	222	11011110	ю	254	11111110
џ	159	10011111	ї	191	10111111	Я	223	11011111	я	255	11111111

Международный стандарт **Unicode**

- **16-разрядная** кодировка.
- Допускает включение **до 65536** символов.
- Включает в себя все существующие, вымершие и искусственно созданные алфавиты мира, а также множество математических, музыкальных, химических и прочих символов.

ТЕОРЕТИЧЕСКИЕ ПОЛОЖЕНИЯ

Содержательный подход
к измерению информации

Основные понятия

- **Информация** – это снятая неопределенность.
- Мера неопределенности в теории информации – **энтропия**.
- Один бит – это такое количество информации, которое уменьшает неопределенность в два раза.
- ○ **Формула Хартли** (когда все исходы события равновероятны) **$i = \log_2 N$**
(i – количество информации)
- **Формула Шеннона** **$I = \sum_{i=1}^N P_i \log_2 \left(\frac{1}{P_i} \right)$**
- p_i - вероятность (частота) наступления i -го события

ПРАКТИЧЕСКИЕ ЗАДАНИЯ

Задача 1

- Информационный объем одного символа некоторого сообщения равен 6 битам. Сколько символов входит в алфавит, с помощью которого было составлено это сообщение?

Задача 2

- Два текста содержат одинаковое количество символов. Первый текст составлен в алфавите мощностью 32 символа, второй – мощностью 64 символа. Во сколько раз отличается количество информации в этих текстах?

Задача 3

- При кодировании (Unicode) информационный объем фразы **«Ученье - свет, а неученье - тьма.»** составляет ...

Задача 4

- Автоматическое устройство осуществило перекодировку информационного сообщения на русском языке, первоначально записанного в 8-битном коде, в 16-битную кодировку Unicode. При этом информационное сообщение увеличилось на 2048 байт. Каков был информационный объем сообщения до перекодировки?

Задача 5

- Десятичный код (номер) буквы «e» в таблице кодировки символов ASCII равен 101. Какая последовательность десятичных кодов будет соответствовать слову: **file**

Задача 6

- Информационный объем текста, подготовленного с помощью компьютера, равен 3,5 Кб. Сколько символов содержит этот текст?

Задача 7

- В корзине лежат 32 клубка шерсти, из них 4 красных. Сколько бит информации несет сообщение о том, что достали клубок красной шерсти?

Задача 8

- Для регистрации на сайте некоторой страны пользователю требуется придумать пароль. Длина пароля – ровно 11 символов. В качестве символов используются десятичные цифры и 12 различных букв местного алфавита, причём все буквы используются в двух начертаниях: как строчные, так и заглавные (регистр буквы имеет значение!). Под хранение каждого такого пароля на компьютере отводится минимально возможное и одинаковое целое количество байтов, при этом используется посимвольное кодирование и все символы кодируются одинаковым и минимально возможным количеством битов. Определите объём памяти, который занимает хранение 60 паролей.

Задача 9

- В конкурсе участвовали 20 студентов, 8 школьников и 4 учащихся колледжа. Количество информации в сообщении о том, что победил школьник, считая, что победа любого из участников равновероятна, составит ____ бит(-а).

Задача 10

- Каждая ячейка памяти компьютера, работающего в троичной системе счисления, может принимать три различных значения $(-1, 0, 1)$. Для хранения некоторой величины отвели 4 ячейки памяти. Сколько различных значений может принимать эта величина?

Задача 11

- В некоторой стране автомобильный номер состоит из 8 символов. Первый символ – одна из 26 латинских букв, остальные семь – десятичные цифры. Пример номера – A1234567. Каждый символ кодируется одинаковым и минимально возможным количеством бит, а каждый номер – одинаковым и минимально возможным количеством байт. Определите объем памяти, необходимый для хранения 30 автомобильных номеров.

Задача 12

- В базе данных хранятся записи, содержащие информацию о датах. Каждая запись содержит три поля: год (число от 1 до 2100), номер месяца (число от 1 до 12) и номер дня в месяце (число от 1 до 31). Каждое поле записывается отдельно от других полей с помощью минимально возможного числа бит. Определите минимальное количество бит, необходимых для кодирования одной записи.