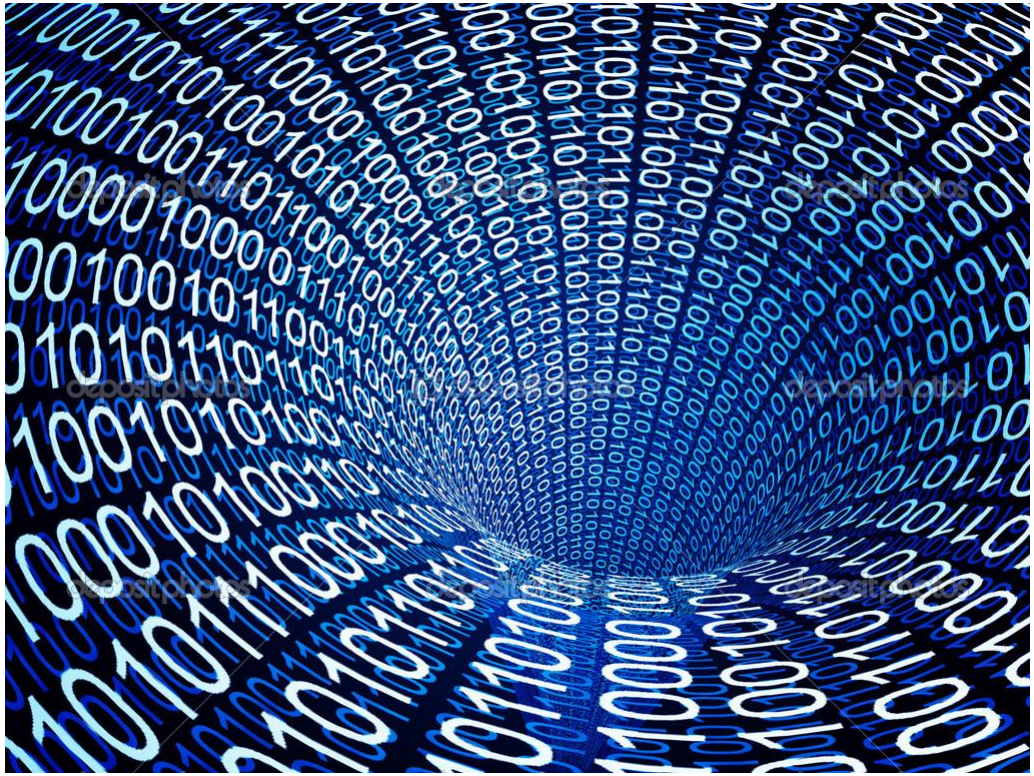
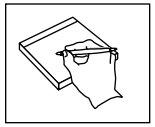
Представление текстов в памяти ПК

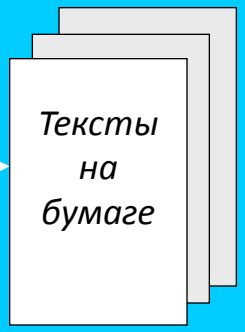


СПОСОБЫ ОБРАБОТКИ И ХРАНЕНИЯ ТЕКСТОВ

Ручная запись текста на бумаге



Печать на пишущей машинке

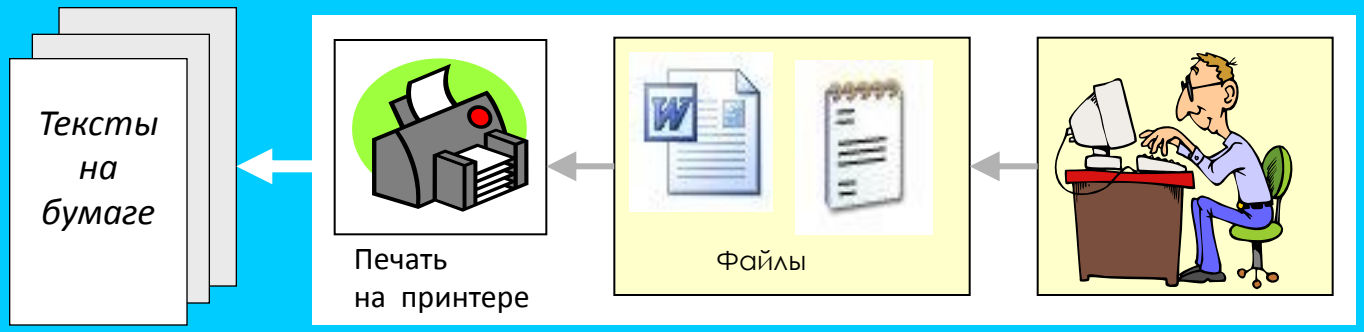


Бумажная технология

Недостатки бумажной технологии

- проблемы редактирования бумажных документов (зачеркивание, стирание, заклеивание, переписывание заново)
- книги, документы на бумажном носителе занимают очень много места
- одноразовое использование, перерасход бумаги
- трудности копирования бумажных текстовых документов
- трудности передачи текстовой бумажной информации на большие расстояния

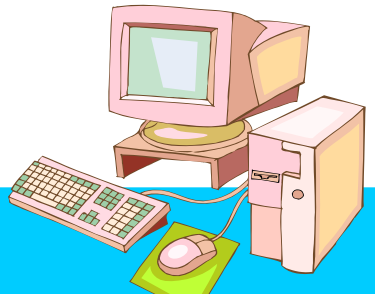
СПОСОБЫ ОБРАБОТКИ И ХРАНЕНИЯ ТЕКСТОВ



Компьютерная технология

Достоинства компьютерной технологии

- компактное хранение в файлах на носителях
- возможность многократного использования носителя
- простота копирования и размножения
- простота редактирования – внесения изменений
- быстрая передача по сетевым каналам



КОДИРОВАНИЕ ТЕКСТА. ТАБЛИЦА КОДИРОВКИ

Структура кодовой таблицы

№	Символ	Двоичный код
0	NUL	00000000
1	ЗОН	00000001
...	.....	.....
31	US	00011111
...	.....	.....
82	R	01010010
83	S	01010011
84	T	01010100
...	.....	.....
127	□	01111111
...	.....	.....
195	Г	11000011
196	Д	11000100
197	Е	11000101
...	.....	.....
254	ю	11111110
255	я	11111111

Кодировка в ОС Windows – код ANSI

American National Standard Institute (американский институт стандартизации)

Стандартная часть кода ANSI/ASCII

0 (00000000) **. . .** **127** (01111111)

Символы с кодами **0 . . . 31** выполняют управляющие функции. При выводе текста они не отображаются графическими знаками.

Кодовая страница CP-1251 (разная для разных стран – национальные алфавиты)

128 (10000000) **. . .** **255** (11111111)

Содержит русские буквы, упорядоченные в алфавитной последовательности (это правило соблюдается не во всех кодовых страницах)

ТЕКСТОВЫЕ ФАЙЛЫ

Двоичный код каждого символа состоит из 8 цифр (бит) и занимает 1 байт памяти ПК

В текстовом файле вся информация представлена символами кодовой таблицы, используемой в операционной системе. Для национальных алфавитов существуют специальные кодовые страницы. Например: CP1251 – “русская” кодовая страница

ИНФОРМАТИКА → 200 205 212 206 208 204 192 210 200 202 192

Последовательность кодов CP1251

11001000 11001101 11010100 11001110
11010000 11001100 11000000 11010010 11001000 11001010 11000000

Кодовая таблица Windows (ANSI, CP-1251)

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0				©	Ё	§	Є	·		°						
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1																
	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
2		!	"	#	$	%	&		(	)	*	+				/
	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47
3	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63
4	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79
5	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95
6	‘	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111
7	p	q	r	s	t	u	v	w	x	y	z	{	|	}]	~	□
	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127
8	Ъ	Ѓ	‚	ѓ	„	…	†	‡		‰	Љ	‹	Њ	Ќ	Ћ	Џ
	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
9	ђ	‘	’	"	"		–	—		™	љ	›	ѣ	ќ	ћ	џ
	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
A		Ў	ў	J	¤	Ґ	¦	§	Ё	©	Є	«		-	®	Ї
	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
B	°	±	І	і	ґ	µ		·	ё	№	є	»	j	S	s	ї
	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191
C	А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П
	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207
D	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я
	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	▪ 223
E	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п
	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
F	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я
	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255

А	Б	В	Г	Д	Е	Ж	З
192	193	194	195	196	197	198	199
Р	С	Т	У	Ф	Х	Ц	Ч
208	209	210	211	212	213	214	215
а	б	в	г	д	е	ж	з
224	225	226	227	228	229	230	231
р	с	т	у	ф	х	ц	ч
240	241	242	243	244	245	246	247

И	Й	К	Л	М	Н	О	П
200	201	202	203	204	205	206	207
Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я
216	217	218	219	220	221	222	▪ 223
и	й	к	л	м	н	о	п
232	233	234	235	236	237	238	239
ш	щ	ъ	ы	ь	э	ю	я
248	249	250	251	252	253	254	255

Задание

Как в текстовом файле будет хранится запись Вашего имени (в кодах CP1251)

Пример: Р о м а

десятичный код 208 238 236 224

двоичный код (дома по таблице учебника)

ГИПЕРТЕКСТ

1
Страница 5
Страница 8

2
Страница 12

3
Страница 10

4
Страница 2

5
Страница 8

6
Страница 4
Страница 7

7
Страница 3

8
Страница 6

9
Страница 8

10
Страница 2

11
Страница 3

12
Страница 1
Страница 5

ГИПЕРТЕКСТ – это текст, организованный так, что его можно просматривать в последовательности смысловых связей между его отдельными фрагментами. Такие связи называются **гиперсвязями** (гиперссылками)

ГИПЕРТЕКСТ

1
Страница 5
Страница 8

2
Страница 12

3
Страница 10

4
Страница 2

5
Страница 8

6
Страница 4
Страница 7

7
Страница 3

8
Страница 6

9
Страница 8

10
Страница 2

11
Страница 3

12
Страница 1
Страница 5

ГИПЕРТЕКСТ – это текст, организованный так, что его можно просматривать в последовательности смысловых связей между его отдельными фрагментами. Такие связи называются **гиперсвязями** (гиперссылками)

ГИПЕРТЕКСТ

1 Страница 5

Страница 8

2 Страница 12

3 Страница 10

4 Страница 2

5 Страница 8

6 Страница 4

Страница 7

7 Страница 3

8 Страница 6

9 Страница 8

10 Страница 2

11 Страница 3

12 Страница 1

Страница 5

ГИПЕРТЕКСТ – это текст, организованный так, что его можно просматривать в последовательности смысловых связей между его отдельными фрагментами. Такие связи называются **гиперсвязями** (гиперссылками)

ГИПЕРТЕКСТ

1 Страница 5

Страница 8

2 Страница 12

3 Страница 10

4 Страница 2

5 Страница 8

6 Страница 4

Страница 7

7 Страница 3

8 Страница 6

9 Страница 8

10 Страница 2

11 Страница 3

12 Страница 1

Страница 5

ГИПЕРТЕКСТ – это текст, организованный так, что его можно просматривать в последовательности смысловых связей между его отдельными фрагментами. Такие связи называются **гиперсвязями** (гиперссылками)

ГИПЕРТЕКСТ

1 Страница 5

Страница 8

2 Страница 12

3 Страница 10

4 Страница 2

5 Страница 8

6 Страница 4

Страница 7

7 Страница 3

8 Страница 6

9 Страница 8

10 Страница 2

11 Страница 3

12 Страница 1

Страница 5

ГИПЕРТЕКСТ – это текст, организованный так, что его можно просматривать в последовательности смысловых связей между его отдельными фрагментами. Такие связи называются **гиперсвязями** (гиперссылками)

ГИПЕРТЕКСТ

1
Страница 5

Страница 8

2
Страница 12

3
Страница 10

4
Страница 2

5
Страница 8

6
Страница 4

Страница 7

7
Страница 3

8
Страница 6

9
Страница 8

10
Страница 2

11
Страница 3

12
Страница 1

Страница 5

ГИПЕРТЕКСТ – это текст, организованный так, что его можно просматривать в последовательности смысловых связей между его отдельными фрагментами. Такие связи называются **гиперсвязями** (гиперссылками)

ГИПЕРТЕКСТ

1 Страница 5

Страница 8

2 Страница 12

3 Страница 10

4 Страница 2

5 Страница 8

6 Страница 4

Страница 7

7 Страница 3

8 Страница 6

9 Страница 8

10 Страница 2

11 Страница 3

12 Страница 1

Страница 5

ГИПЕРТЕКСТ – это текст, организованный так, что его можно просматривать в последовательности смысловых связей между его отдельными фрагментами. Такие связи называются **гиперсвязями** (гиперссылками)

ГИПЕРТЕКСТ

1 Страница 5

Страница 8

2 Страница 12

3 Страница 10

4 Страница 2

5 Страница 8

6 Страница 4

Страница 7

7 Страница 3

8 Страница 6

9 Страница 8

10 Страница 2

11 Страница 3

12 Страница 1

Страница 5

ГИПЕРТЕКСТ – это текст, организованный так, что его можно просматривать в последовательности смысловых связей между его отдельными фрагментами. Такие связи называются **гиперсвязями** (гиперссылками)

ГИПЕРТЕКСТ

1
Страница 5
Страница 8

2
Страница 12

3
Страница 10

4
Страница 2

5
Страница 8

6
Страница 4
Страница 7

7
Страница 3

8
Страница 6

9
Страница 8

10
Страница 2

11
Страница 3

12
Страница 1
Страница 5

ГИПЕРТЕКСТ – это текст, организованный так, что его можно просматривать в последовательности смысловых связей между его отдельными фрагментами. Такие связи называются **гиперсвязями** (гиперссылками)

ГИПЕРТЕКСТ

1
Страница 5
Страница 8

2
Страница 12

3
Страница 10

4
Страница 2

5
Страница 8

6
Страница 4
Страница 7

7
Страница 3

8
Страница 6

9
Страница 8

10
Страница 2

11
Страница 3

12
Страница 1
Страница 5

ГИПЕРТЕКСТ – это текст, организованный так, что его можно просматривать в последовательности смысловых связей между его отдельными фрагментами. Такие связи называются **гиперсвязями** (гиперссылками)

Решение задач

Сколько бит памяти компьютера займет слово **клавиатура** ?

Ответ:

Решение задач

Сколько бит памяти компьютера займет слово **клавиатура** ?

Ответ: 80

Каждая буква занимает 8 бит (1 байт) памяти.

$10 \times 8 = 80$ бит.

Решение задач

Текст занимает полных 7 страниц. На каждой странице размещается 30 строк по 50 символов в строке. Какой объем оперативной памяти займет этот текст?

Ответ: _____ байт.

Решение задач

Текст занимает полных 7 страниц. На каждой странице размещается 30 строк по 50 символов в строке. Какой объем оперативной памяти займет этот текст?

Ответ: 10500 байт.

Весь текст содержит

7 × 30 × 50 = 10500 символов.

Каждый символ текста занимает 1 байт памяти.

Следовательно, текст займет 10500 байт.

Задание на дом
п. 13

Сколько **символов** содержится в тексте, использующем таблицу **ASCII**, если известно, что он занимает 24 576 бит памяти?