

Федеральное государственное бюджетное образовательное
учреждение высшего профессионального образования
«Вятский государственный гуманитарный университет»

**Дополнительная подготовка школьников
по дисциплине
«Информатика и информационные технологии»**

**Учебный модуль
Представление информации в компьютере**

Е. В. Котельников

Киров
2011

СОДЕРЖАНИЕ

1. Представление текстовой информации	3
1.1. ASCII.....	3
1.2. Unicode.....	5
2. Представление графической информации.....	6
2.1. Способы представления графической информации	6
2.2. Цветовые модели	10
3. Представление звуковой информации.....	14
4. Задания для самостоятельного выполнения	16
Литература.....	17

1. Представление текстовой информации

1.1. ASCII

В ячейки компьютерной памяти можно записывать только двоичные числа, следовательно, для хранения текстовой информации (как и любой другой) нужен способ, позволяющий преобразовывать (кодировать) символы в двоичное представление и обратно. Такой способ называется *кодировкой*.

Обычный подход в этом случае – составить таблицу кодировки, в которой назначить каждому символу определенный код. Тогда компьютер при необходимости преобразования символ → код и обратно, обращается к этой таблице.

Самая известная таблица кодировки, реализованная в большинстве компьютеров, называется *ASCII (American Standard Code for Information Interchange)* – американский стандартный код для обмена информацией).

Это 7-битовый двоичный код, в котором представляются алфавитные, цифровые и специальные символы. Общее количество представимых символов:

$$2^7 = 128.$$

Символы в коде ASCII делятся на *управляющие* (десятичные коды 0..31, 127) и *печатаемые* (десятичные коды 32..126). Управляющие символы не отображаются на экране, а управляют выводом текста, например, символ с кодом 9 обозначает табуляцию, а символ с кодом 10 – перевод строки. Печатаемые символы (табл. 1) отображаются при выводе на экран или принтер.

Таблица 1. Печатаемые символы в кодировке ASCII (коды 32–127)
(Dec – десятичный код, Hex – шестнадцатеричный код)

Dec	Hex	Символ	Dec	Hex	Символ	Dec	Hex	Символ
32	20	<i>пробел</i>	64	40	@	96	60	'
33	21	!	65	41	A	97	61	a
34	22	“	66	42	B	98	62	b
35	23	#	67	43	C	99	63	c
36	24	\$	68	44	D	100	64	d
37	25	%	69	45	E	101	65	e
38	26	&	70	46	F	102	66	f
39	27	`	71	47	G	103	67	g

40	28	(72	48	H	104	68	h
41	29)	73	49	I	105	69	i
42	2A	*	74	4A	J	106	6A	j
43	2B	+	75	4B	K	107	6B	к
44	2C	,	76	4C	L	108	6C	l
45	2D	–	77	4D	M	109	6D	m
46	2E	.	78	4E	N	110	6E	n
47	2F	/	79	4F	O	111	6F	o
48	30	0	80	50	P	112	70	p
49	31	1	81	51	Q	113	71	q
50	32	2	82	52	R	114	72	r
51	33	3	83	53	S	115	73	s
52	34	4	84	54	T	116	74	t
53	35	5	85	55	U	117	75	u
54	36	6	86	56	V	118	76	v
55	37	7	87	57	W	119	77	w
56	38	8	88	58	X	120	78	x
57	39	9	89	59	Y	121	79	y
58	3A	:	90	5A	Z	122	7A	z
59	3B	;	91	5B	[123	7B	{
60	3C	<	92	5C	\	124	7C	
61	3D	=	93	5D]	125	7D	}
62	3E	>	94	5E	^	126	7E	~
63	3F	?	95	5F	_	127	7F	

Поскольку таблица ASCII 7-битовая, а минимальная адресуемая ячейка памяти имеет размер 1 байт, остается незадействованным 1 бит (или 128 символов из 256 доступных в байте). Поэтому коды 128..255 были выделены для национальных символов и алфавитов, а также для символов псевдографики. В этом случае таблица ASCII называется *расширенной*.

Например, для кириллицы существуют несколько вариантов расширенных таблиц – KOI-8R, Windows-1251, CP-866, ISO 8859-5. Во всех них первые 128 символов совпадают и представлены в таблице 1.

1.2. Unicode

Таблицы кодировки, использующие один байт, не вмещают все необходимые символы (например, иероглифы, арабский алфавит, математические символы и др.). В 1991 году был опубликован Unicode – стандарт кодировки, использующий для представления одного символа более одного байта и позволяющий закодировать все широко используемые алфавиты в мире. В 2011 году вышла 6-ая версия стандарта¹.

В настоящее время символ в Unicode может быть представлен ячейкой памяти от 1 до 4 байт. Например, если это буква латинского алфавита или цифра, то на её представление отводится 1 байт, как и в ASCII-кодировке. Для кириллической буквы будет выделено два байта, для китайского иероглифа – 3 байта, а музыкальные символы будут представлены 4-мя байтами. Всего в действующем стандарте Unicode может быть представлено 1 112 064 символов².

Обозначаются коды Unicode следующим образом:

U+*<шестнадцатеричный код символа>*.

Например, код U+0041 обозначает заглавную латинскую букву А (десятичный код 65, совпадает с ASCII), а код U+0410 отвечает за кириллическую букву А. По адресу <http://www.unicode.org/charts/> можно найти код любого символа, представленного в Unicode и, наоборот, по коду определить символ. Также в операционной системе Windows имеется стандартное приложение «Таблица символов», которое позволяет просмотреть и скопировать широко используемые Unicode-символы.

¹ См. <http://unicode.org/>

² Это количество ($2^{20} + 2^{16} - 2048$) можно представить в одном из форматов Unicode – UTF-16.

2. Представление графической информации

2.1. Способы представления графической информации

Существует три основных способа представления графических изображений: растровый, векторный и фрактальный. Растровые форматы хорошо подходят для изображений со сложными гаммами цветов, оттенков и форм (фотографии, рисунки). Векторные форматы более приспособлены для схем, чертежей и простых изображений. Фрактальная графика применяется для создания сложных изображений, структура которых описывается при помощи математических выражений.

В *растровых форматах* изображение представляется прямоугольной матрицей точек – пикселей¹ (pixel – picture element), положение которых в матрице соответствует координатам точек на экране. Помимо координат каждый пиксель характеризуется своим цветом, цветом фона или градацией яркости. Количество битов, выделяемых для указания цвета пикселя, изменяется в зависимости от формата.

Пример. В высококачественных изображениях (т. н. *TrueColor*) цвет пикселя описывают 24 битами, что дает $2^{24} = 16\,777\,216$ цветов.



Рис. 1. Пример исходного (32×32 пикселя) и увеличенного в 7 раз растрового изображения

¹ Существует вариант произношения без мягкого знака «пиксел». По поводу того, как правильнее, см. обзор <http://yuriy-apostol.livejournal.com/10097.html>.

Таким образом, растровое изображение характеризуется следующими параметрами:

- размер матрицы пикселей – *разрешение*;
- количество битов для хранения цвета или количество представимых цветов – *глубина цвета*;
- используемая *цветовая модель*, например, RGB или CMYK (см. параграф 2.2).

Основным достоинством растровой графики является высокая точность и возможность отображения малейших нюансов любого изображения – фото, художественной картины, чертежа, графика и т. д.

Недостатки растровой графики заключаются, во-первых, в большой емкости памяти, требуемой для хранения изображения, во-вторых, при масштабировании (увеличении) может проявляться ухудшение четкости, размытие изображения.

Чтобы преодолеть первый недостаток, часто применяют сжатие растровой графики, которое бывает двух видов – без потерь и с потерями. При сжатии без потерь применяются специальные алгоритмы, которые позволяют при хранении уменьшать размер графического файла, а для представления на экране – восстанавливать изображение в исходном виде без искажений (без потерь). В случае сжатия с потерями возможно гораздо значительнее уменьшить размер файла, но при отображении часть информации будет потеряна и качество изображения несколько ухудшится.

В настоящее время существует множество форматов графических файлов, различающихся алгоритмами сжатия и способами представления растровых изображений, а также сферой применения. Некоторые из распространенных форматов растровых графических файлов перечислены в таблице 2.

Таблица 2. Форматы файлов растровой графики

Формат	Полное название	Глубина цвета	Сжатие	Применение
BMP	Bitmap Picture	1–6 байт	Без потерь	ОС Windows
GIF	Graphics Interchange Format	1 байт (256 цветов)	Без потерь	Интернет, анимация
PNG	Portable Network Graphics	1–6 байт	Без потерь	Интернет
JPEG	Joint Photographic Experts Group	1–3 байта	С потерями	Фотографии

Формат	Полное название	Глубина цвета	Сжатие	Применение
TIFF	Tagged Image File Format	1–8 байт	Без потерь и с потерями	Фотографии, сканеры, издательские системы
PSD	Photoshop Document	1–6 байт	Без потерь	Adobe Photoshop

Для создания и редактирования растровых изображений используются специальные программы – растровые графические редакторы. В качестве примеров можно назвать: Microsoft Paint (входящий в стандартную поставку Windows), Adobe Photoshop, GIMP, Paint.NET, Corel PaintShop Photo и др.

Векторное представление, в отличие от растровой графики, определяет описание изображения не пикселями, а т. н. *графическими примитивами* – прямыми, окружностями (эллипсами), многоугольниками, сплайнами.

Сплайн – это гладкая кривая, которая проходит через две или более опорные точки, управляющие формой сплайна. В векторной графике наиболее распространены сплайны на основе кривых Безье. Суть сплайна: любую элементарную кривую можно построить, зная четыре коэффициента P_0 , P_1 , P_2 и P_3 , соответствующие четырем точкам на плоскости (рис. 2). Перемещение этих точек влечет за собой изменение формы кривой.

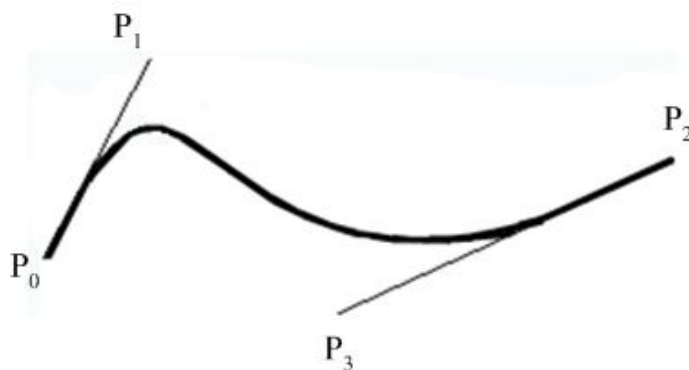


Рис. 2. Вариант сплайна

Пример. Для представления в растровом формате отрезка прямой из 100 пикселей с глубиной цвета 3 байта понадобится $100 \cdot 3 = 300$ байт; та же прямая в векторном формате будет описана координатами начала и конца

отрезка, параметрами его толщины и цвета. Если каждая координата занимает 2 байта, толщина – 1 байт и цвет – 3 байта, то вся прямая будет храниться в $4 \cdot 2 + 1 + 3 = 12$ байтах.

Достоинства векторной графики:

- описание объекта является простым и занимает мало памяти;
- простота масштабирования изображения без ухудшения его качества.

Недостатком векторного формата является высокая сложность представления сложных изображений, например, фотографий, полутоновых рисунков и т. д.

Как и для растровой графики, существует несколько форматов графических векторных файлов (таблица 3).

Таблица 3. Форматы файлов векторной графики

Формат	Полное название	Применение
AI	Adobe Illustrator	Adobe Illustrator
CDR	Corel Drawing	CorelDRAW
DXF	Drawing Exchange Format	Обмен информацией между САПР ¹
SVG	Scalable Vector Graphics	Открытый формат
VSD	Microsoft Visio format	Microsoft Office Visio

Работа с векторной графикой осуществляется при помощи векторных редакторов: CorelDRAW, Adobe Illustrator, AutoCAD, Microsoft Office Visio, OpenOffice.org Draw, Inkscape и др.

Фрактал – это бесконечно самоподобная геометрическая фигура, каждый фрагмент которой повторяется при уменьшении масштаба. Изображения во фрактальной графике создаются путем математических расчетов, но в отличие от векторной графики в памяти компьютера геометрические объекты не хранятся, а изображения строятся исключительно на основе уравнений. Поэтому ничего, кроме формулы, хранить не надо. Изменив коэффициенты в уравнении, можно получить совершенно другую картину.

¹ САПР – системы автоматизированного проектирования (AutoCAD, T-FLEX CAD, OrCAD, КОМПАС).



Рис. 3. Пример фрактала – множество Жюлиа

Фракталы применяются в различных областях компьютерной графики: для изображения облаков, снега, береговых линий, деревьев и кустов, для создания текстур и фоновых изображений, в компьютерных играх.

Построенные фрактальные изображения обычно сохраняются в одном из форматов растровой графики – BMP, JPG или PNG.

2.2. Цветовые модели

Цветовая модель – это способ представления информации о цвете. Цветовые модели основаны на законах *Грассмана*¹.

Закон трехмерности: любой цвет может быть выражен посредством трех линейно независимых цветов. Линейная независимость цвета означает, что его нельзя представить с помощью двух других цветов.

Закон непрерывности: при непрерывном изменении составляющих цвет меняется также непрерывно. Для любого цвета существует бесконечно близкий к нему цвет.












По закону трехмерности для построения цветовой модели нужно выбрать три независимых или базовых цвета. Рассмотрим две наиболее часто применяющиеся в информатике цветовые модели, в которых выбор базовых цветов осуществлен по-разному.

Модель RGB (Red, Green, Blue) – в качестве базовых выбраны красный, зеленый и синий цвета. В этой модели любой цвет получается смешением трех базовых цветов. Используется модель RGB при выводе изображений на экран монитора, а также в большинстве форматов графических файлов.

¹ Грассман, Герман Гюнтер (1809–1877) – немецкий физик, математик и филолог.

Для хранения в компьютерной памяти каждого цвета может быть использовано разное количество бит (глубина цвета). Рассмотрим глубину цвета TrueColor, в которой каждый основной цвет представлен одним байтом (диапазон 0–255) (табл. 4). В модели RGB цвет обозначается набором шестнадцатеричных цифр, например, если оранжевый цвет имеет красную составляющую 255, зеленую – 128 и синюю – 0, то его представление в шестнадцатеричном виде будет следующим: FF 80 00.

Таблица 4. Представление разных цветов в модели RGB

Цвет	Образец	Основные цвета			Обозначение (HEX)
		Red	Green	Blue	
Черный (Black)		0	0	0	00 00 00
Красный (Red)		255	0	0	FF 00 00
Оранжевый (Orange)		255	128	0	FF 80 00
Желтый (Yellow)		255	255	0	FF FF 00
Зеленый (Green)		0	255	0	00 FF 00
Голубой (Sky Blue)		0	191	255	00 BF FF
Циан или сине-зеленый (Cyan)		0	255	255	00 FF FF
Синий (Blue)		0	0	255	00 00 FF
Фиолетовый (Violet)		128	0	255	80 00 FF
Пурпурный (Magenta)		255	0	255	FF 00 FF
Белый (White)		255	255	255	FF FF FF

В модели CMYK (Cyan, Magenta, Yellow, black) в качестве основных используются цвета циан (сине-зеленый), пурпурный, желтый и черный. Применяется модель CMYK в цветных принтерах, а также в издательской деятельности.










Вывод цветного изображения на экран монитора и печать на принтере принципиально отличаются. Дело в том, что в мониторе за вывод каждого



пикселя отвечают три элемента красного, зеленого и синего цветов, которые излучают собственный свет, таким образом, чтобы получить некоторый цвет, нужно включить эти элементы с соответствующей интенсивностью. Поэтому модель RGB называют *аддитивной* (англ. *add* – добавлять).

С другой стороны, при печати на бумаге краска не излучает сама, а поглощает часть света (обычно белого, т.е. включающего все остальные цвета) и мы видим отраженную часть света. Таким образом, чтобы напечатать красным цветом, краска должна поглотить зеленую и синюю составляющие, поэтому модель CMYK называется *субтрактивной* (англ. *subtract* – вычитать). Базовые цвета в модели CMYK выбраны так, чтобы поглощать свет определенных цветов – циан поглощает красный свет, пурпурная краска поглощает зеленый, а желтая – синий. Чтобы поглотить свет всех цветов, т.е. получить черный цвет, теоретически достаточно смешать все три краски – циановую, пурпурную и желтую. Однако реальные краски неидеальны и вместо черного получается темно-серый цвет. Поэтому в модель добавлена черная краска.

Конкретный цвет в модели CMYK описывается 4-мя числами, от 0 до 100, которые означают степень яркости соответствующей краски (табл. 5).

Таблица 5. Представление разных цветов в модели CMYK

Цвет	Образец	Основные цвета			
		Cyan	Yellow	Magenta	Black
Черный (Black)		0	0	0	100
Красный (Red)		0	100	100	0
Оранжевый (Orange)		0	65	95	0
Желтый (Yellow)		0	0	100	0
Зеленый (Green)		100	50	100	0
Голубой (Sky Blue)		70	0	0	0
Циан или сине-зеленый (Cyan)		100	0	0	0
Синий (Blue)		100	100	0	0
Фиолетовый (Violet)		93	94	0	0

Пурпурный (Magenta)		0	0	100	0
Белый (White)		0	0	0	0

Кроме рассмотренных моделей RGB и CMYK, существуют ещё несколько моделей: XYZ, HSB, YUV и др.

3. Представление звуковой информации

Понятие аудио связано со звуками, которые способно воспринимать человеческое ухо. Частоты аудиосигналов лежат в диапазоне от 20 Гц до 20 кГц, а сигналы по своей природе являются непрерывными (аналоговыми). Прежде чем быть представленной в компьютере, аудиоинформация должна быть преобразована в цифровую форму (оцифрована) с использованием временной дискретизации и квантования сигнала (рис. 4). Для этого значения звуковых сигналов (*выборки, samples*), взятые через малые интервалы времени (временная дискретизация), с помощью аналого-цифровых преобразователей (АЦП) переводятся в двоичный код (квантование сигнала). Обратное действие выполняется цифро-аналоговыми преобразователями (ЦАП). Чем чаще производятся выборки, т. е. чем больше *частота дискретизации*, тем выше может быть точность последующего воспроизведения исходного сигнала, но тем большая емкость памяти требуется для хранения оцифрованного звука.

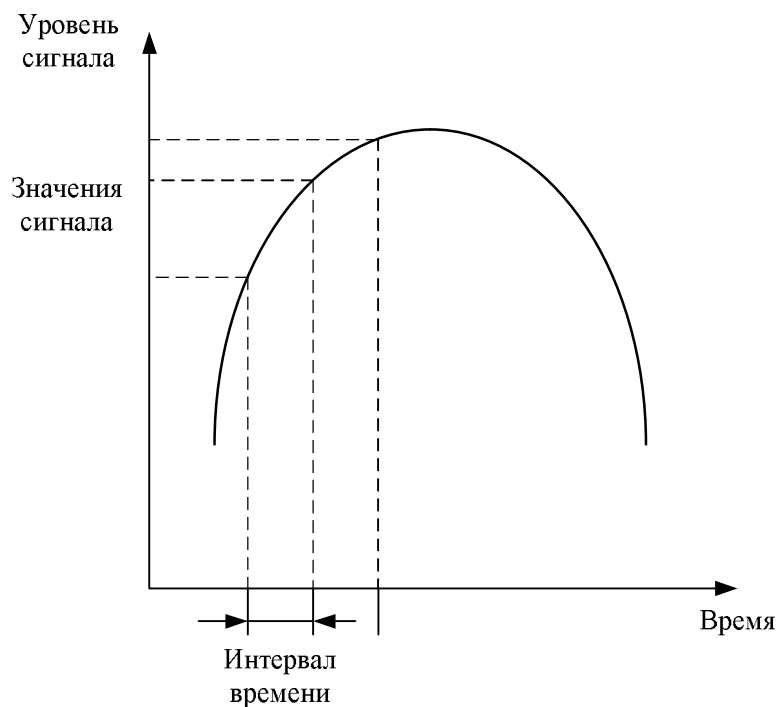


Рис. 4. Процесс оцифровки аудиосигнала

Для запоминания уровня каждого сигнала отводится ячейка памяти n бит, размер которой определяет *глубину кодирования звука* (n в разных аудиоформатах меняется от 8 до 32). Тогда количество возможных уровней будет 2^n и это число конечно, в отличие от количества значений уровня

реального звука (аудиосигнал непрерывен). Таким образом, вследствие дискретного измерения по времени и по уровню оцифрованный звук всегда имеет худшее качество, чем оригинальный аудиосигнал. Однако на практике стараются выбирать частоту дискретизации и размер ячейки памяти настолько большими, чтобы человек не в состоянии был заметить присутствующие искажения.

По *теореме Найквиста-Котельникова* частоту дискретизации следует выбирать не менее чем в два раза больше, чем максимальная частота измеряемого сигнала.

Пример. Для записи Audio CD (аудио компакт-дисков) используется частота дискретизации 44,1 кГц, поскольку она более чем в два раза превосходит максимальную частоту, воспринимаемую человеческим ухом.

Цифровой эквивалент аудиосигналов обычно хранится в виде файлов, причем широко используются различные методы сжатия такой информации. Как правило, к методам сжатия аудиоинформации предъявляется требование возможности восстановления непрерывного сигнала без заметного ухудшения его качества. В настоящее время распространен целый ряд форматов хранения аудиоинформации. Некоторые из них перечислены в таблице.

Таблица 6. Форматы аудиофайлов

Формат	Полное название	Применение
WAV	WAVE («волна»)	Простое хранение оцифрованного звука без сжатия
WMA	Windows Media Audio	Закрытый формат аудиофайлов от Microsoft; использует сжатие с потерями
AAC	Advanced Audio Coding	Формат аудиофайлов с эффективным сжатием с потерями
OGG	Ogg	Формат аудиофайлов со сжатием с потерями
MIDI	Musical Instrument Digital Interface	Формат для обмена данными между электронными музыкальными инструментами
MP3	Moving Picture Experts Group (MPEG) Layer 3	Один из самых популярных форматов аудиофайлов со сжатием с потерями

4. Задания для самостоятельного выполнения

1. Найдите информацию об управляющих символах ASCII-кода и их современном применении.
2. Какие есть отличия в кодировании кириллицы в расширенных таблицах KOI-8R, Windows-1251, CP-866, ISO 8859-5?
3. Найдите коды кириллических букв в Unicode.
4. Найдите коды алхимических символов (Alchemical Symbols) в Unicode.
5. Перечислите бесплатные редакторы растровой и векторной графики.
6. Назовите форматы растровой графики и их особенности.
7. Назовите форматы векторной графики и их особенности.
8. Найдите таблицу цветов, определяемых в спецификации HTML 4.01 и используемых в Интернете.
9. Найдите информацию о частоте дискретизации и глубине кодирования звука для форматов, перечисленных в табл. 6.
10. Перечислите программы работы со звуком; выделите среди них платные и бесплатные.

Литература

1. Андреева Е. В., Босова Л. Л., Фалина И. Н. Математические основы информатики. Элективный курс: Учебное пособие. – М.: БИНОМ. Лаборатория знаний, 2005.
2. Могилев А. В., Пак Н. И., Хеннер Е. К. Информатика. 7-ое изд., стереотип. – М.: Академия, 2009.
3. Петцольд Ч. Код. – М.: Русская редакция, 2001.