Представление текстовой и графической информации в компьютере

1

Учебник Информатика. 10кл. Баз. Уровень Семакин и др
п.6 «Представление текста, изображений и звука в компьютере», стр 43-49

© К.Ю. Поляков, Е.А. Ерёмин, 2013 http://kpolyakov.spb.ru

Кодирование текстовой информации

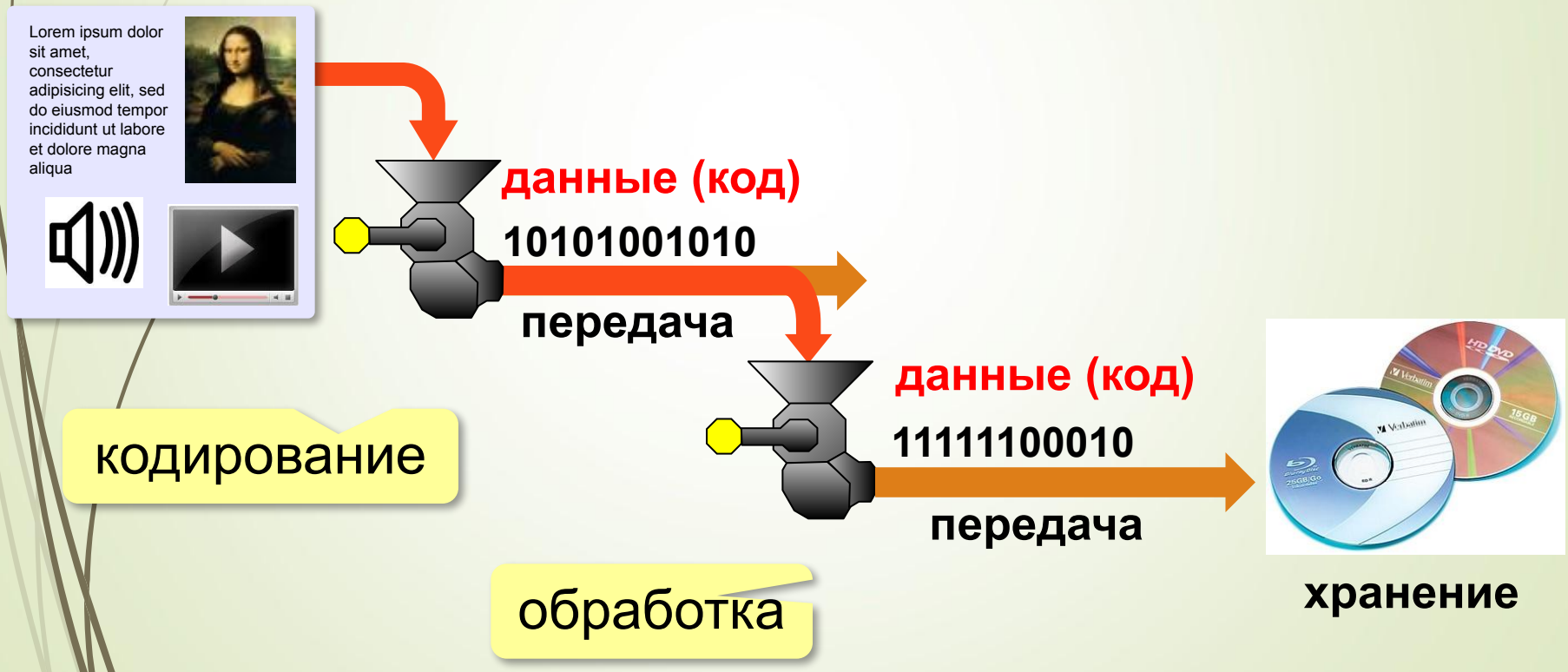
2

© К.Ю. Поляков, Е.А. Ерёмин, 2013

Зачем кодировать информацию?

Кодирование — это представление информации в форме, удобной для её хранения, передачи и обработки.

В компьютерах используется двоичный код:



Кодирование символов

Текстовый файл

- на экране (символы)
- в памяти – коды

1000001_2	1000010_2	1000011_2	1000100_2
65	66	67	68

! В файле хранятся не изображения символов, а их числовые коды!

Файлы со шрифтами: ***.fon**, ***.ttf**, ***.otf**

Кодировка ASCII (7-битная)

ASCII = *American Standard Code for Information Interchange*

Коды 0-127:

0-31 **управляющие символы**:
7 – звонок, 10 – новая строка,
13 – возврат каретки, 27 – Esc.

32 пробел

знаки препинания: . , : ; ! ?

специальные знаки: + – * / () {} []

48-57 цифры **0..9**

65-90 заглавные латинские буквы **A-Z**

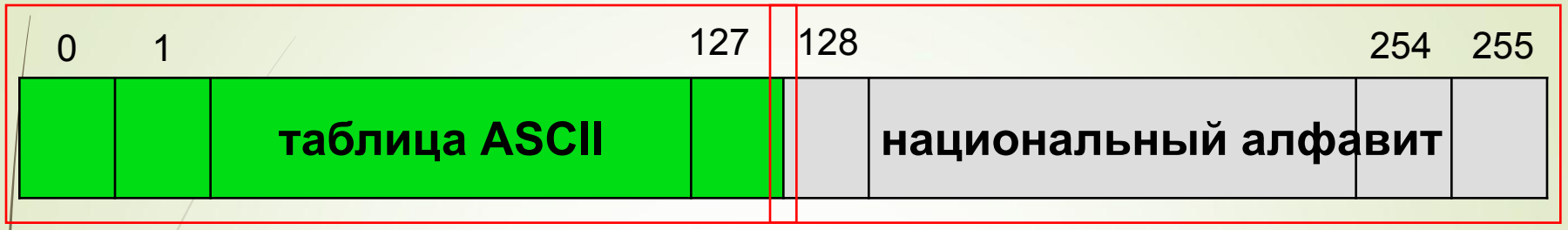
97-122 строчные латинские буквы **a-z**

? Где русские буквы?

Не предусмотрены

8-битные кодировки

Кодовые страницы (расширения ASCII):



Для русского языка:

CP-866	для *MS DOS*
CP-1251	для *Windows* (Интернет)
KOI8-R	для *UNIX* (Интернет)
MacCyrillic	для компьютеров *Apple*

Проблема:

Windows-1251	*KOI8-R*
Привет, Вася!	оПХБЕР,
	бЮЯЪ!
рТЙЧЕФ,	Привет, Вася!
чБУС!	

8-битные кодировки

- 1 байт на символ – файлы небольшого размера!
- просто обрабатывать в программах

- нельзя использовать символы разных кодовых страниц одновременно (русские и французские буквы, и т.п.)
- неясно, в какой кодировке текст (перебор вариантов!)
- для каждой кодировки нужен свой шрифт (изображения символов)

Стандарт UNICODE

1 112 064 знаков, используются около **100 000**

Windows: **UTF-16**

16 битов на распространённые символы,
32 бита на редко встречающиеся

Linux: **UTF-8**

8 битов на символ для ASCII,
от 16 до 48 бита на остальные

- совместимость с ASCII
- более экономична, чем UTF-16, если много символов ASCII

2010 г. – 50% сайтов использовали UTF-8!

Решение задач

Основные формулы:

$$2^i = N$$

$$I = i*K,$$

N – мощность алфавита (кол-во символов)
К – длина текста
i - информационный объем символа (бит)
I – информационный объем сообщения (бит)

Задача 1.

Текст длиной 32768 символов закодирован с помощью алфавита, содержащего 64 символа. Сколько килобайт занимает в памяти этот текст?

Дано:
K=32768
N=64
Найти
I -?

Решение
$2^i = 64$, i=6 (бит)
I = i*K=6*32768=6*1024*32 (бит)=6*1024*4 (байт)=6*4 (Кбайт)=24 Кбайт
Ответ: 24 Кбайт

Задача 2.

Сообщение длиной 28672 символа занимает в памяти 21 Кбайт. Найдите мощность алфавита, который использовался при кодировании.

© К.Ю. Поляков, Е.А. Ерёмин, 2013 http://kpolyakov.spb.ru

Решение задач

Основные формулы:

$$2^i = N$$

$$I = i*K,$$

N – мощность алфавита (кол-во символов)
К – длина текста
i - информационный объем символа (бит)
I – информационный объем сообщения (бит)

Задача 2.

Сообщение длиной 28672 символа занимает в памяти 21 Кбайт. Найдите мощность алфавита, который использовался при кодировании.

Дано:
K=28672
I=21 Кбайт
Найти
N -?

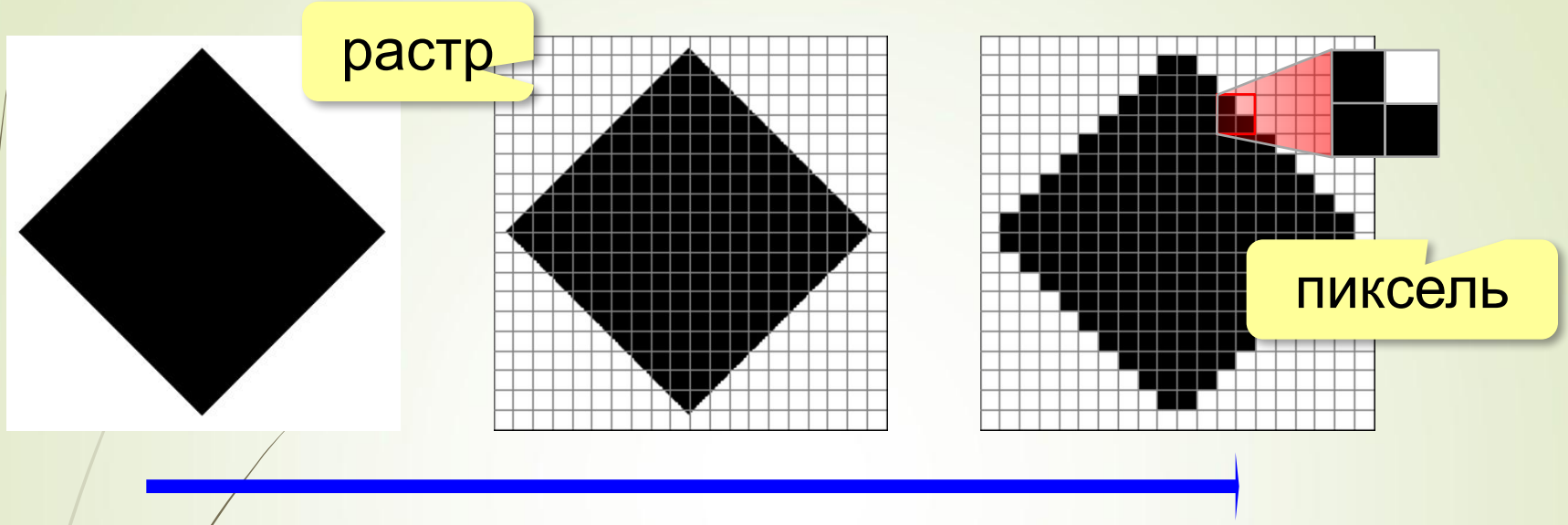
Решение
1) I = i*K,
i=I/k=21*1024*8/28672=21*1024*8/(1024*28)=6 (бит)
2) $2^i = N$, $2^6=64$ (символа)

Ответ: мощность алфавита 64 символа

Кодирование графической информации

© К.Ю. Поляков, Е.А. Ерёмин, 2013 http://kpolyakov.spb.ru

Растровое кодирование

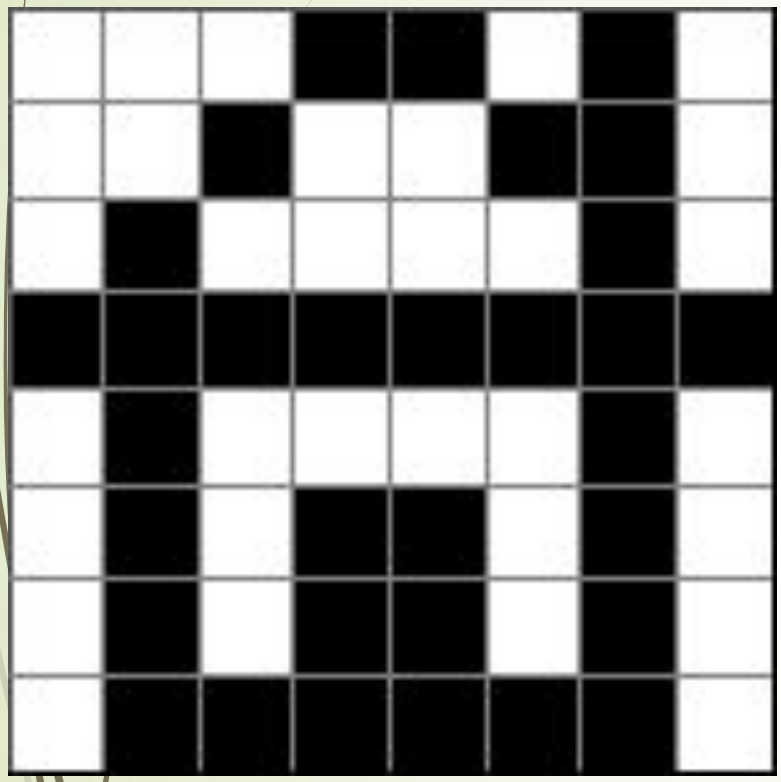


! Рисунок искажается!

Пиксель – это наименьший элемент рисунка, для которого можно задать свой цвет.

Растровое изображение – это изображение, которое кодируется как множество пикселей.

Растровое кодирование



Двоичный код

Шестнадцатиричный код

0	0	0	1	1	0	1	0	1A
0	0	1	0	0	1	1	0	26
0	1	0	0	0	0	1	0	42
1	1	1	1	1	1	1	1	FF
0	1	0	0	0	0	1	0	42
0	1	0	1	1	0	1	0	5A
0	1	0	1	1	0	1	0	5A
0	1	1	1	1	1	1	0	7E

Первый символ шестнадцатеричного кода

Второй символ

$$1A2642FF425A5A7E_{16}$$

Соответствие 10, 2 и 16 кодов

10 число	Двоичное	Шестнадцатеричное
1	0001	1
2	0010	2
3	0011	3
4	0100	4
5	0101	5
6	0110	6
7	0111	7
8	1000	8
9	1001	9
10	1010	A
11	1011	B
12	1100	C
13	1101	D
14	1110	E
15	1111	F

© К.Ю. Поляков, Е.А. Ерёмин, 2013 http://kpolyakov.spb.ru

Разрешение

Разрешение – это количество пикселей, приходящихся на дюйм размера изображения.

ppi = pixels per inch, пикселей на дюйм

1 дюйм = 2,54 см

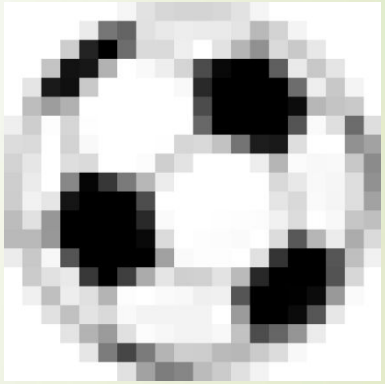
300 ppi

печать

96 ppi

экран

48 ppi



24 ppi

© К.Ю. Поляков, Е.А. Ерёмин, 2013 http://kpolyakov.spb.ru

Разрешение

Задача 1. Какой размер в пикселях должен иметь закодированный рисунок с разрешением 300 ppi, чтобы с него можно было сделать отпечаток размером 10×15 см?

высота $\dfrac{10\text{ см} \times 300\text{ пикселей}}{2{,}54\text{ см}} \approx \mathbf{1181\text{ пиксель}}$

ширина $\dfrac{15\text{ см} \times 300\text{ пикселей}}{2{,}54\text{ см}} \approx \mathbf{1771\text{ пиксель}}$

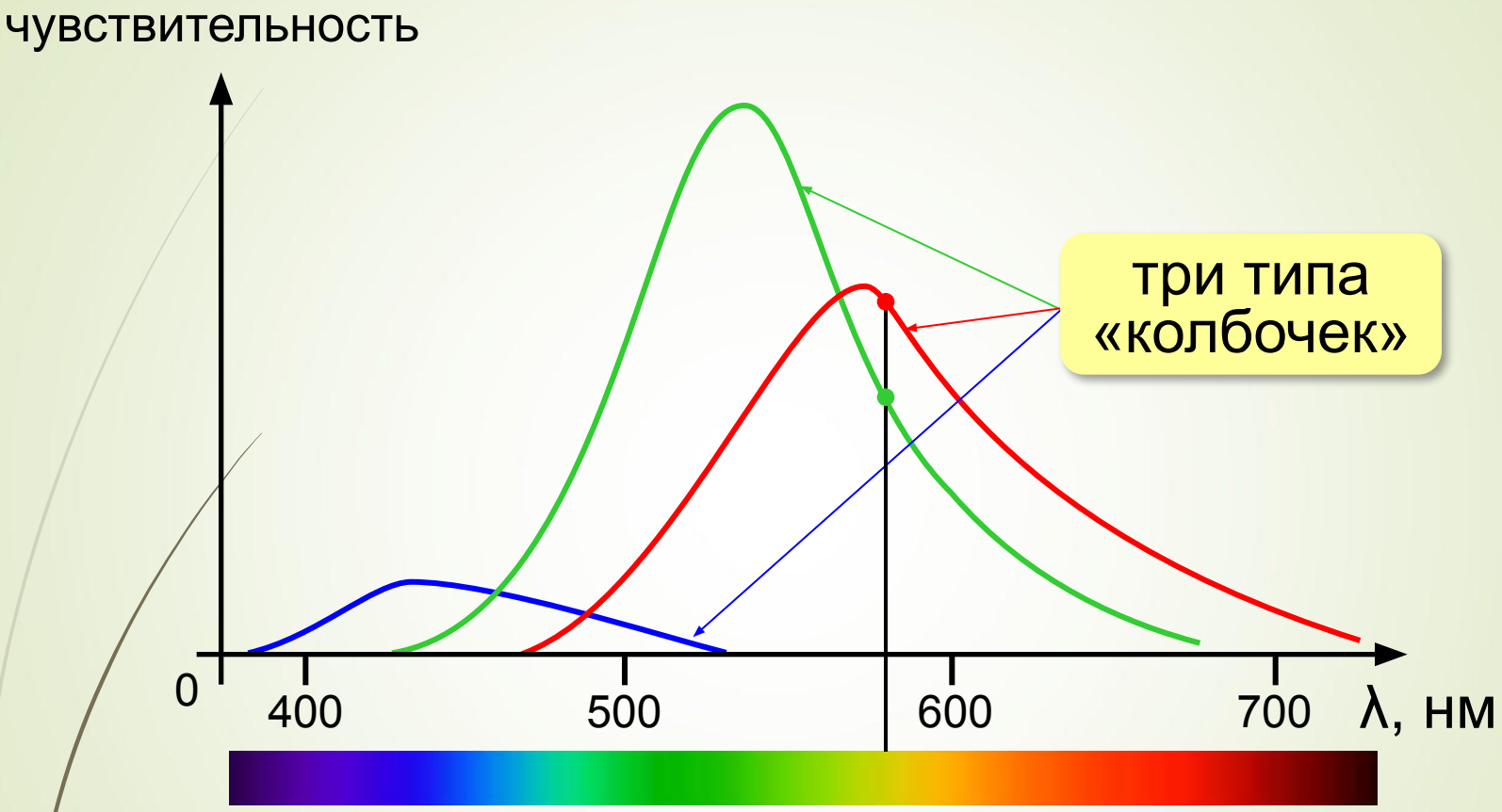
Задача 2. Закодированный рисунок имеет размеры 5760 × 3840 пикселей и разрешение 600 ppi. Какой размер будет у изображения, отпечатанного на принтере?

ширина $\dfrac{5760\text{ пикселей} \times 2{,}54\text{ см}}{600\text{ пикселей}} \approx \mathbf{24{,}4\text{ см}}$

высота $\dfrac{3840\text{ пикселей} \times 2{,}54\text{ см}}{600\text{ пикселей}} \approx \mathbf{16{,}3\text{ см}}$

© К.Ю. Поляков, Е.А. Ерёмин, 2013 http://kpolyakov.spb.ru

Кодирование цвета. Теория цвета Юнга-Гельмгольца



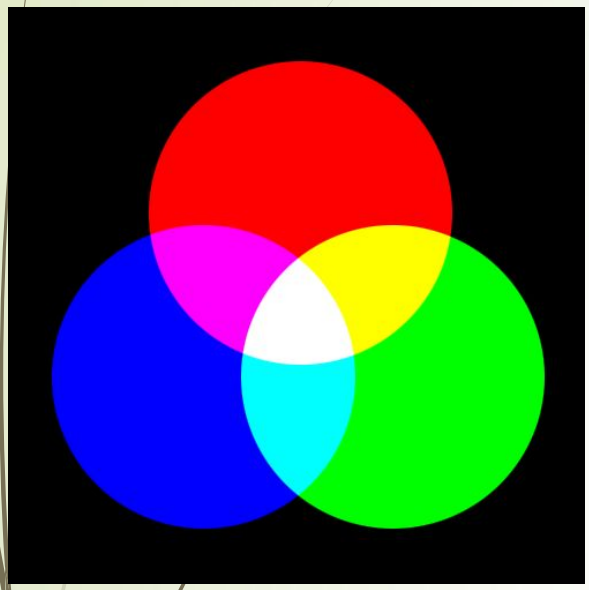
Свет любой длины волны можно заменить на красный, зелёный и синий лучи!

Цветовая модель RGB

Д. Максвелл, 1860

цвет = (R, G, B)

red	***green***	***blue***
красный	зеленый	синий
0..255	0..255	0..255



- (0, 0, 0)
- (255, 255, 255)
- (255, 0, 0)
- (255, 150, 150)
- (0, 255, 0)
- (255, 255, 0)
- (0, 0, 255)
- (100, 0, 0)

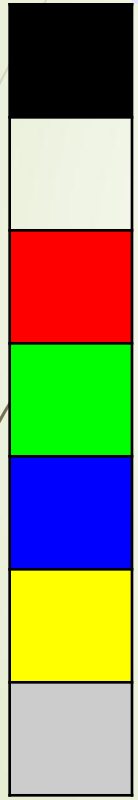
? Сколько разных цветов можно кодировать?

256·256·256 = **16 777 216** (*True Color,* «истинный цвет»)

! RGB – цветовая модель для устройств, излучающих свет (мониторов)!

Цветовая модель RGB

(255, 255, 0) → #FFFF00



RGB	Веб-страница
(0, 0, 0)	#000000
(255,255,255)	#FFFFFF
(255, 0, 0)	#FF0000
(0, 255, 0)	#00FF00
(0, 0, 255)	#0000FF
(255, 255, 0)	#FFFF00
(204,204,204)	#CCCCCC

Глубина цвета

Глубина цвета — это количество битов, используемое для кодирования цвета пикселя.

Сколько памяти нужно для хранения цвета 1 пикселя в режиме *True Color*?

R (0..255) $256 = 2^8$ вариантов 8 битов = 1 байт

R G B: 24 бита = 3 байта

True Color (истинный цвет)

Задача. Определите размер файла, в котором закодирован растровый рисунок размером 20×30 пикселей в режиме истинного цвета (*True Color*)?

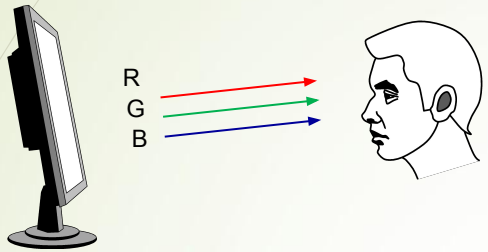
Основная формула I = i* m * n, где I – размер файла в битах, i – битовая глубина цвета, m * n – размер изображения в пикселях

Дано	Решение
m * n =20x30 i= 3 байта	I = i* m * n 20 · 30 · 3 байта = **1800 байт**

Растровые рисунки: форматы файлов

Формат	True Color	Палитра	Прозрачность	Анимация
BMP	✔	✔	⛔	⛔
JPG	✔	⛔	⛔	⛔
GIF	⛔	✔	✔	✔
PNG	✔	✔	✔	⛔

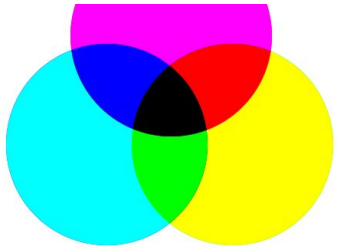
Кодирование цвета при печати (CMYK)



Белый – красный = голубой **C = Cyan**

Белый – зелёный = пурпурный **M = Magenta**

Белый – синий = желтый **Y = Yellow**



Модель CMY

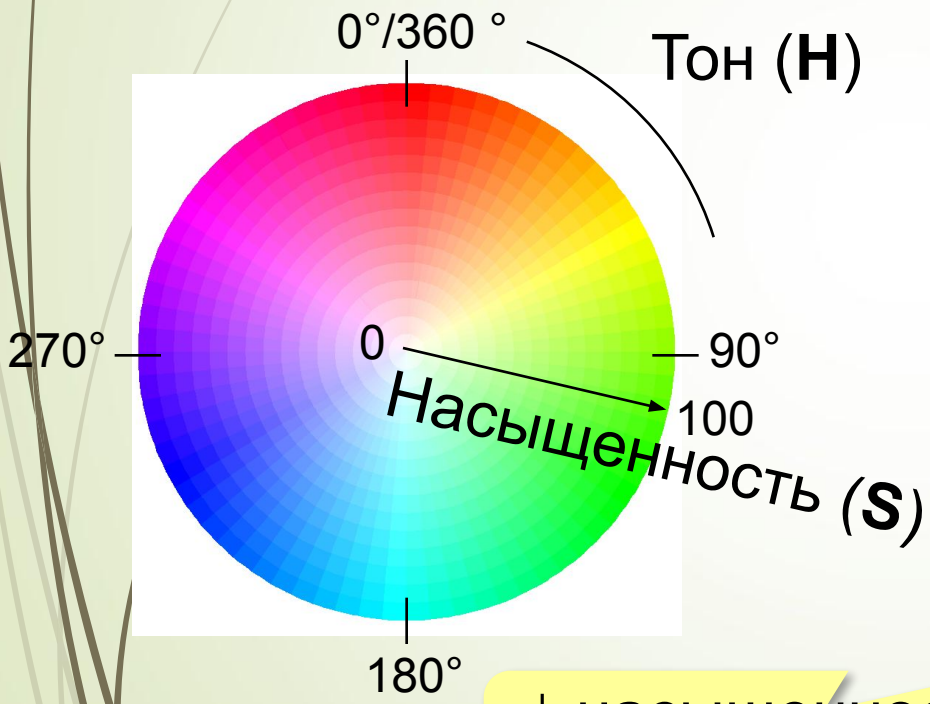
C	M	Y	
0	0	0	
255	255	0	
255	0	255	
0	255	255	
255	255	255	

Модель CMYK: + ***Key color***

- меньший расход краски и лучшее качество для чёрного и серого цветов

Цветовая модель HSB (HSV)

HSB = ***H****ue* (тон, оттенок)
S*aturation* (насыщенность)
B*rightness* (яркость) или ***V****alue* (величина)



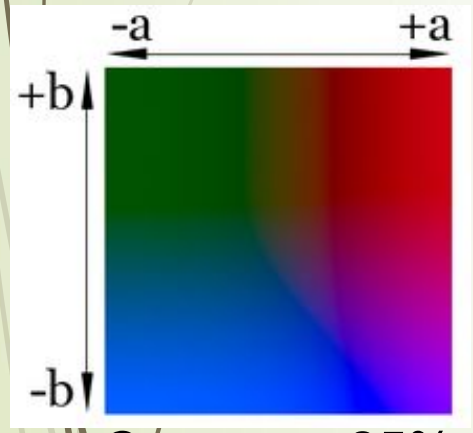
↓ яркость – добавить чёрного

↓ насыщенность – добавить белого

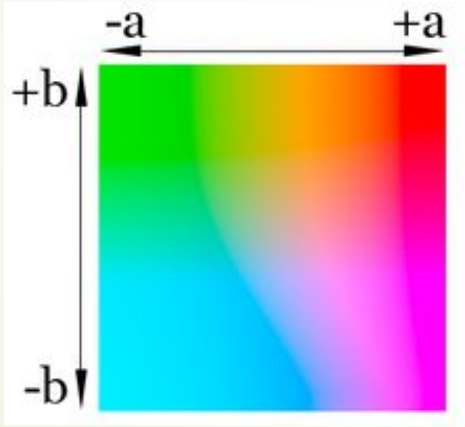
Цветовая модель Lab

Международный стандарт кодирования цвета, независимого от устройства (1976 г.)
Основана на модели восприятия цвета человеком.

Lab = *Lightness* (светлота)
a, b (задают цветовой тон)



Светлота 25%



Светлота 75%

- для перевода между цветовыми моделями: RGB → Lab → CMYK
- для цветокоррекции фотографий

Растровое кодирование: итоги

- универсальный метод (можно закодировать любое изображение)
- единственный метод для кодирования и обработки размытых изображений, не имеющих чётких границ (фотографий)

- есть **потеря информации** (почему?)
- при изменении размеров цвет и форма объектов на рисунке **искажается**
- **размер файла** не зависит от сложности рисунка (а от чего зависит?)

Решение задач

Задача 1. В процессе преобразования растрового графического файла количество цветов уменьшилось с 512 до 8. Во сколько раз уменьшился информационный объем файла?

Дано:

$N_1=512$

$N_2=8$

Найти I_1/I_2

Решение

1) Найдем глубину цвета i_1 для 512 цветов

$$2^{i1} = N_1,\ 2^{i1} = 512,\ i_1 = 9 \text{(бит)}$$

2) Найдем глубину цвета i_2 для 8 цветов

$$2^{i2} = N_2,\ 2^{i2} = 8,\ i_2 = 3 \text{(бит)}$$

3) Изменение информационного объема

$$i_1/i_2 = 9/3 = 3 \text{ (раза)}$$

Ответ: информационный объем уменьшился в 3 раза.

Задача 2. Для хранения растрового изображения размером 128 x 128 пикселей отвели 4 килобайта памяти. Каково максимально возможное число цветов в палитре изображения?

Дано:

$m*n=128*128$

$I=4$ Кбайт

Найти N

Основные формулы

Объем изображения $I=m*n*i$, где i - глубина цвета в бит

Количество цветов в палитре $N=2^i$

Решение

$i=I/(m*n)=4*1024*8/(128*128)=2^{15}/2^{14}=2$ (бита)

$2^2=4$ (цветов)

Ответ: макс возможное кол-во цветов в палитре 4

Векторное кодирование

Рисунки из геометрических фигур:

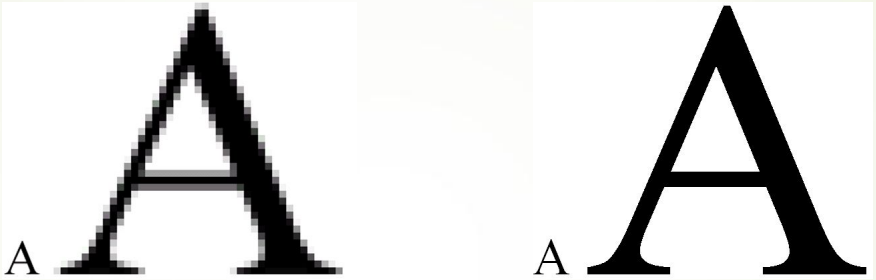
- отрезки, ломаные, прямоугольники
- окружности, эллипсы, дуги
- сглаженные линии (кривые Безье)

Для каждой фигуры в памяти хранятся:

- размеры и координаты на рисунке
- цвет и стиль границы
- цвет и стиль заливки (для замкнутых фигур)

Векторное кодирование (итоги)

- лучший способ для хранения **чертежей, схем, карт**
- при кодировании **нет потери информации**
- при изменении размера **нет искажений**



растровый рисунок | векторный рисунок

- меньше **размер файла**, зависит от сложности рисунка

- неэффективно использовать для **фотографий** и размытых изображений

Векторное кодирование: форматы файлов

- **WMF** (*Windows Metafile*)
- **EMF** (*Windows Metafile*)
- **CDR** (программа *CorelDraw*)
- **AI** (программа *Adobe Illustrator*)
- **SVG** *(Scalable Vector Graphics,* масштабируемые векторные изображения*)*

для веб-страниц

Практическая работа №4 Кодирование текстовой и графической информации

30

А •—	И ••	Р •—•	Ш ————
Б —•••	Й •———	С •••	Щ ——•—
В •——	К —•—	Т —	Ъ •——•—•
Г ——•	Л •—••	У ••—	Ы —•——
Д —••	М ——	Ф ••—•	Ь —••—
Е •	Н —•	Х ••••	Э ••—••
Ж •••—	О ———	Ц —•—•	Ю ••——
З ——••	П •——•	Ч ———•	Я •—•—

Задание 1.

Используя шифр Цезаря с шагом 3,закодировать сообщение **УТРО ВЕЧЕРА МУДРЕНЕЕ**

Шифр Цезаря. Способ кодировки, при котором буква заменяется на другую букву, отстоящую от первой на определенный шаг.

Например, если шаг равен 3, то буква А меняется на Г, буква Б – на Д, а буква Ю – на букву Б.

Оформление задания

Исходная	У	Т	Р	О		В	Е	Ч	Е	Р	А		М	У	Д	Р	Е	Н	Е	Е
Закодированная																				

Задание 2.

Для кодирования сообщения используется таблица

А	Б	В	Г	Д
10	11	001	010	01

Приведите все варианта декодирования сообщения 0101110010110.

Ответ оформить в виде последовательности букв. Например, АГДБВ

Задание 3.

Текст длиной 73728 символов закодирован с помощью алфавита, содержащего 8 символов. Сколько килобайт занимает в памяти этот текст?

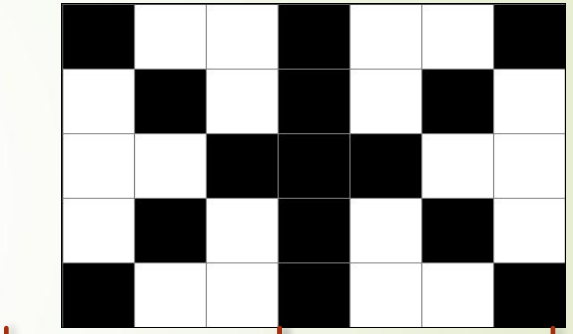
Практическая работа №4 Кодирование текстовой информации

Задание 4.

Текст длиной 73728 символов закодирован с помощью алфавита, содержащего 8 символов. Сколько килобайт занимает в памяти этот текст?

Задание 5.

Закодируйте рисунок с помощью двоичного и шестнадцатеричного кода, используя образец и справочную таблицу (Слайды 13, 14). Для решения задачи дополните рисунок столбцом белых клеток слева



Задание 6.

Разрешение экрана монитора – 1024 x 768 точек, глубина цвета – 16 бит. Каков необходимый объем видеопамяти для данного графического режима?