2ο μέρος τμήμα 1ο

Data Representation

Αναπαράσταση Δεδομένων
Αναλογικά και Ψηφιακά δεδομένα

Analog and Digital Information

Information can be represented in one of two ways: **analog** or **digital**

Analog data

A continuous representation, analogous to the actual information it represents

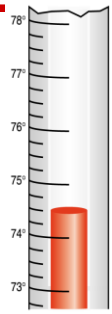
Digital data

A discrete representation, breaking the information up into separate elements

2

Analog and Digital Information

Thermometer is an analog device



A mercury thermometer continually rises in direct proportion to the temperature

3

Analog and Digital Information

Computers cannot work well with analog data, so we digitize the data

Digitize

Breaking data into pieces and representing those pieces separately

Why do we use binary to represent digitized data?

4

Electronic Signals

Important facts about electronic signals

- An analog signal continually fluctuates in voltage up and down
- A digital signal has only a high or low state, corresponding to the two binary digits
- All electronic signals (both analog and digital) degrade as they move down a line
- The voltage of the signal fluctuates due to environmental effects

5

Electronic Signals *(Cont'd)*

Periodically, a digital signal is **reclocked** to regain its original shape

Figure 3.2
An analog and a digital signal

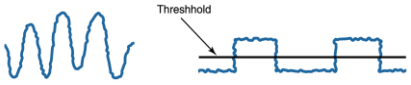


Figure 3.3
Degradation of analog and digital signals

6

Binary Representations

One bit can be either 0 or 1
One bit can represent two things *(Why?)*
Food (sweet or sour – 0 or 1)
Two bits can represent four things *(Why?)*
Gears (neutral, park, drive, reverse – 00,01,10,11)
How many things can three bits represent?
How many things can four bits represent?
How many things can eight bits represent?

7

Binary Representations

1 Bit	2 Bits	3 Bits	4 Bits	5 Bits
0	00	000	0000	00000
1	01	001	0001	00001
	10	010	0010	00010
	11	011	0011	00011
		100	0100	00100
		101	0101	00101
		110	0110	00110
		111	0111	00111
			1000	01000
			1001	01001
			1010	01010
			1011	01011
			1100	01100
			1101	01101
			1110	01110
			1111	01111
				10000
				10001
				10010
				10011
				10100
				10101
				10110
				10111
				11000
				11001
				11010
				11011
				11100
				11101
				11110
				11111

Counting with binary bits

Figure 3.4

Converting Binary to Octal

Binary	Octal	Decimal
0	0	0
1	1	1
10	2	2
11	3	3
100	4	4
101	5	5
110	6	6
111	7	7
1000	10	8
1001	11	9
1010	12	10

- Mark groups of *three* (from right)
- Convert each group

10101011 → 10 101 011
2 5 3

10101011 is 253 in base 8

9

Converting Binary to Hexadecimal

- Mark groups of *four* (from right)
- Convert each group

10101011 → 1010 1011
A B

10101011 is AB in base 16

0	0000
1	0001
2	0010
3	0011
4	0100
5	0101
6	0110
7	0111
8	1000
9	1001
A	1010
B	1011
C	1100
D	1101
E	1110
F	1111

10

Binary and Computers

Byte
8 bits

The number of bits in a word determines the word length of the computer, but it is usually a multiple of 8

- 32-bit machines
- 64-bit machines etc.

11

Representing Text

What must be provided to represent text?

There are finite number of characters to represent, so list them all and assign each a binary string

Character set
A list of characters and the codes used to represent each one

Computer manufacturers agreed to standardize

12

The ASCII Character Set

ASCII stands for American Standard Code for Information Interchange

ASCII originally used seven bits to represent each character, allowing for 128 unique characters

Later extended ASCII evolved so that all eight bits were used

How many characters could be represented?

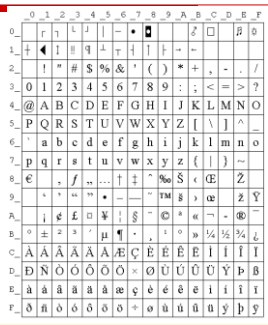
The ASCII Character Set

σε δεκαδικούς κώδικες

Left Digit(s) \ Right Digit	ASCII 0	1	2	3	4	5	6	7	8	9
0	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT
1	LF	VT	FF	CR	SO	SI	DLE	DC1	DC2	DC3
2	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS
3	RS	US	□	!	"	#	$	%	&	'
4	(	)	*	+	,	-	.	/	0	1
5	2	3	4	5	6	7	8	9	:	;
6	<	=	>	?	@	A	B	C	D	E
7	F	G	H	I	J	K	L	M	N	O
8	P	Q	R	S	T	U	V	W	X	Y
9	Z	[	\	]	^	_	`	a	b	c
10	d	e	f	g	h	i	j	k	l	m
11	n	o	p	q	r	s	t	u	v	w
12	x	y	z	{	|	}	~	DEL		

The ASCII Character Set

σε δεκαεξαδικούς κώδικες



The Unicode Character Set

Extended ASCII is not enough for international use

Unicode uses 16 bits per character

How many characters can UNICODE represent?

Unicode is a superset of ASCII

The first 256 characters correspond exactly to the extended ASCII character set

The Unicode Character Set

Code (Hex)	Character	Source
0041	A	English (Latin)
042F	Я	Russian (Cyrillic)
0E09	ฉ	Thai
13EA	Ꮺ	Cherokee
211E	℞	Letterlike Symbols
21CC	⇌	Arrows
282F	⠯	Braille
345F	㑟	Chinese/Japanese/ Korean (Common)

Figure 3.6 A few characters in the Unicode character set

Text Compression

Assigning 16 bits to each character in a document uses too much file space

We need ways to store and transmit text efficiently

Text compression techniques
- keyword encoding
- run-length encoding
- Huffman encoding

Keyword Encoding

Replace frequently used words with a single character

Word	Symbol
as	^
the	~
and	+
that	$
must	&
well	%
these	#

Keyword Encoding

Given the following paragraph,

> We hold these truths to be self-evident, that all men are created equal, that they are endowed by their Creator with certain unalienable Rights, that among these are Life, Liberty and the pursuit of Happiness. ↳ That to secure these rights, Governments are instituted among Men, deriving their just powers from the consent of the governed, ↳ That whenever any Form of Government becomes destructive of these ends, it is the Right of the People to alter or to abolish it, and to institute new Government, laying its foundation on such principles and organizing its powers in such form, as to them shall seem most likely to effect their Safety and Happiness.

Keyword Encoding

The encoded paragraph is

> We hold # truths to be self-evident, $ all men are created equal, $ ~y are endowed by ~ir Creator with certain unalienable Rights, $ among # are Life, Liberty + ~ pursuit of Happiness. — $ to secure # rights, Governments are instituted among Men, deriving ~ir just powers from ~ consent of ~ governed, — $ whenever any Form of Government becomes destructive of # ends, it is ~ Right of ~ People to alter or to abolish it, + to institute new Government, laying its foundation on such principles + organizing its powers in such form, ^ to ~m shall seem most likely to effect ~ir Safety + Happiness.

Keyword Encoding

What did we save?

- Original paragraph
  - 656 characters
- Encoded paragraph
  - 596 characters
- Characters saved
  - 60 characters
- Compression ratio
  - 596/656 = 0.9085

Could we use this substitution chart for all text?

Run-Length Encoding

A single character may be repeated over and over again in a long sequence

Replace a repeated sequence with

- a flag character
- repeated character
- number of repetitions

*n8

- * is the flag character
- n is the repeated character
- 8 is the number of times n is repeated

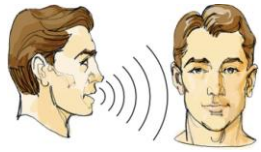
Run-Length Encoding

Original text
bbbbbbbbjjjkllqqqqqq+++++
Encoded text
*b8jjjkll*q6*+5 *(Why isn't l encoded? J?)*
The compression ratio is 15/25 or .6

Encoded text
*x4*p4l*k7
Original text
xxxxpppplkkkkkkk

This type of repetition doesn't occur in English text; can you think of a situation where it might occur?

Representing Audio Information



We perceive sound when a series of air compressions vibrate a membrane in our ear, which sends signals to our brain

Representing Audio Information

A stereo sends an electrical signal to a speaker to produce sound

This signal is an analog representation of the sound wave

The voltage in the signal varies in direct proportion to the sound wave

Representing Audio Information

Digitize the signal by sampling

- periodically measure the voltage
- record the numeric value

How often should we sample?

A sampling rate of about 40,000 times per second is enough to create a reasonable sound reproduction

Representing Audio Information

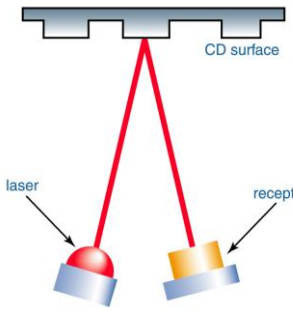
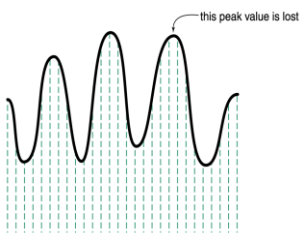


Figure 3.9
A CD player reading binary information

Representing Audio Information



Some data is lost, but a reasonable sound is reproduced

Figure 3.8 Sampling an audio signal

Representing Audio Information

CDs store audio information digitally

On the surface of the CD are microscopic pits that represent binary digits

A low intensity laser is pointed as the disc The laser light reflects strongly if the surface is smooth and poorly if the surface is pitted

Audio Formats

Audio Formats

- WAV, AU, AIFF, VQF, and MP3

MP3 (MPEG-2, audio layer 3 file) is dominant

- analyzes the frequency spread and discards information that can't be heard by humans
- bit stream is compressed using a form of Huffman encoding to achieve additional compression

Representing Images and Graphics

Color

Perception of the frequencies of light that reach the retinas of our eyes

Retinas have three types of color photoreceptor cone cells that correspond to the colors of red, green, and blue

Representing Images and Graphics

Color is expressed as an RGB (red-green-blue) value--three numbers that indicate the relative contribution of each of these three primary colors

An RGB value of (255, 255, 0) maximizes the contribution of red and green, and minimizes the contribution of blue, which results in a bright yellow

Representing Images and Graphics

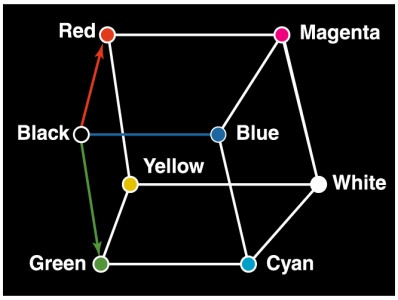


Figure 3.16 Three-dimensional color space

Representing Images and Graphics

color depth

The amount of data that is used to represent a color

HiColor

A 16-bit color depth: five bits used for each number in an RGB value with the extra bit sometimes used to represent transparency

TrueColor

A 24-bit color depth: eight bits used for each number in an RGB value

Representing Images and Graphics

RGB Value			Color
Red	Green	Blue	
0	0	0	black
255	255	255	white
255	255	0	yellow
255	130	255	pink
146	81	0	brown
157	95	82	purple
140	0	0	maroon

A few TrueColor RGB values and the colors they represent

Representing Graphics

- There are two ways of representing graphics
  - Bit Mapped Graphics
  - Vector Graphics
- We will only study bit-mapped graphics but a brief explanation of vector graphics is included

Digitized Images and Graphics

Digitizing a picture

Representing it as a collection of individual dots called pixels

Resolution

The number of pixels used to represent a picture

Raster Graphics

Storage of data on a pixel-by-pixel basis

Bitmap (BMP), GIF, JPEG, and PNG are raster-graphics formats

38

Digitized Images and Graphics

Bitmap format

Contains the pixel color values of the image from left to right and from top to bottom

GIF format (indexed color)

Each image is made up of only 256 colors

JPEG format

Averages color hues over short distances

PNG format

Like GIF but achieves greater compression with wider range of color depths

39

Digitized Images and Graphics

Whole picture

Figure 3.12 A digitized picture composed of many individual pixels

40

Digitized Images and Graphics

Magnified portion of the picture

See the pixels?

Figure 3.12 A digitized picture composed of many individual pixels

41

Vector Graphics

Vector graphics

A format that describes an image in terms of lines and geometric shapes

A vector graphic is a series of commands that describe a line's direction, thickness, and color

The file sizes tend to be smaller because not every pixel is described

42

Vector Graphics

The good side and the bad side…

Vector graphics can be resized mathematically and changes can be calculated dynamically as needed

Vector graphics are *not* good for representing real-world images

Bit Mapped Graphics

- Any graphic is made up from a series of pixels (Picture Elements).
- Each pixel is an individual point on the screen

Vector Vs Bit-Mapped

Advantages of vector graphics (draw packages)

- Images can be enlarged without losing resolution
- Objects can be edited by changing their attributes
- Objects can be layered on top/behind
- Images take up less disc space
- Ideal for drawing plans; use library of objects

Advantages of bit-mapped graphics (paint packages)

- Each pixel can be altered
- More realistic when used for photos/real life

Representing Video

Video codec COmpressor/DECompressor
Methods used to shrink the size of a movie to allow it to be played on a computer or over a network

Almost all video codecs use lossy compressions to minimize the huge amounts of data associated with video

Representing Video

Temporal compression

A technique based on differences between consecutive frames: If most of an image in two frames hasn't changed, why should we waste space to duplicate all of the similar information?

Spatial compression

A technique based on removing redundant information within a frame: This problem is essentially the same as that faced when compressing still images