

# CSC 170 – Introduction to Computers and Their Applications

## Lecture #1 – Digital Basics

### Data Representation

- ***Data*** refers to the symbols that represent people, events, things, and ideas. Data can be a name, a number, the colors in a photograph, or the notes in a musical composition.
- ***Data Representation*** refers to the form in which data is stored, processed, and transmitted.
- Devices such as smartphones, iPods, and computers store data in digital formats that can be handled by electronic circuitry.

## Data Representation

- **Digitization** is the process of converting information, such as text, numbers, photo, or music, into digital data that can be manipulated by electronic devices.
- The **Digital Revolution** has evolved through four phases, beginning with big, expensive, standalone computers, and progressing to today's digital world in which small, inexpensive digital devices are everywhere.

## Data Representation

- The 0s and 1s used to represent digital data are referred to as binary digits — from this term we get the word *bit* that stands for *binary digit*.
- A bit is a 0 or 1 used in the digital representation of data.
- A digital file, usually referred to simply as a file, is a named collection of data that exists on a storage medium, such as a hard disk, CD, DVD, or flash drive.



## Representing Numbers

DECIMAL (BASE 10)	BINARY (BASE 2)
0	0
1	1
2	10
3	11
4	100
5	101
6	110
7	111
8	1000
9	1001
10	1010
11	1011
1000	1111101000

## Representing Text

- *Character data* is composed of letters, symbols, and numerals that are not used in calculations.
- Examples of character data include your name, address, and hair color.
- Character data is commonly referred to as “text.”

## Representing Text

- Digital devices employ several types of codes to represent character data, including ASCII, Unicode, and their variants.
- ASCII (American Standard Code for Information Interchange, pronounced “ASK ee”) requires seven bits for each character.
- The ASCII code for an uppercase **A** is **1000001**.

## Representing Text

- Extended ASCII is a superset of ASCII that uses eight bits for each character.
- For example, Extended ASCII represents the uppercase letter **A** as **01000001**.
- Using eight bits instead of seven bits allows Extended ASCII to provide codes for **256** characters.

## Representing Text

- **Unicode** (pronounced “YOU ni code”) uses sixteen bits and provides codes for 65,000 characters.
- This is a bonus for representing the alphabets of multiple languages.
- **UTF-8** is a variable-length coding scheme that uses seven bits for common ASCII characters but uses sixteen-bit Unicode as necessary.

## Representing Text

00100000	Space	00110011	3	01000110	F	01011001	Y	01101100	l
00100001	!	00110100	4	01000111	G	01011010	Z	01101101	m
00100010	"	00110101	5	01001000	H	01011011	[	01101110	n
00100011	#	00110110	6	01001001	I	01011100	\	01101111	o
00100100	\$	00110111	7	01001010	J	01011101	]	01110000	p
00100101	%	00111000	8	01001011	K	01011110	^	01110001	q
00100110	&	00111001	9	01001100	L	01011111	_	01110010	r
00100111	'	00111010	:	01001101	M	01100000	`	01110011	s
00101000	(	00111011	;	01001110	N	01100001	a	01110100	t
00101001	)	00111100	<	01001111	O	01100010	b	01110101	u
00101010	*	00111101	=	01010000	P	01100011	c	01110110	v
00101011	+	00111110	>	01010001	Q	01100100	d	01110111	w
00101100	,	00111111	?	01010010	R	01100101	e	01111000	x
00101101	-	01000000	@	01010011	S	01100110	f	01111001	y
00101110	.	01000001	A	01010100	T	01100111	g	01111010	z
00101111	/	01000010	B	01010101	U	01101000	h	01111011	{
00110000	0	01000011	C	01010110	V	01101001	i	01111100	
00110001	1	01000100	D	01010111	W	01101010	j	01111101	}
00110010	2	01000101	E	01011000	X	01101011	k	01111110	~

## Representing Text

- ASCII codes are used for numerals, such as Social Security numbers and phone numbers.
- Plain, unformatted text is sometimes called *ASCII text* and is stored in a so-called text file with a name ending in .txt.
- On Apple devices these files are labeled “*Plain Text*.” In Windows, these files are labeled “*Text Document*”.

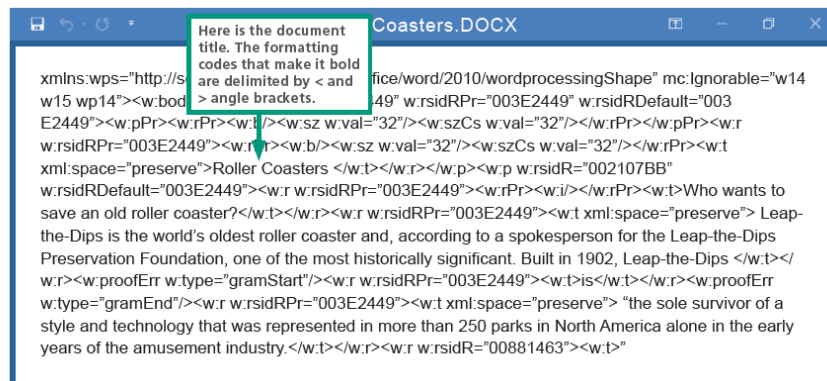
## Representing Text

- ASCII text files contain no formatting.
- To create documents with styles and formats, formatting codes have to be embedded in the text.

## Representing Text

- Microsoft Word produces formatted text and creates documents in **DOCX** format.
- Apple Pages produces documents in **PAGES** format.
- Adobe Acrobat produces documents in **PDF** format.
- HTML markup language used for Web pages produces documents in **HTML** format.

## Representing Text



The screenshot shows a window titled "Coasters.DOCX" with a text area containing XML markup. A text box highlights the sentence: "Here is the document title. The formatting codes that make it bold are delimited by < and > angle brackets." The XML code below shows the corresponding markup for a paragraph about roller coasters, including bold tags and various attributes like w:rsidRPr and w:rsidRDefault.

```
xmlns:wps="http://schemas.microsoft.com/office/word/2010/wordprocessingShape" mc:Ignorable="w14 w15 wp14"><w:body><w:p><w:r><w:t><w:sz w:val="32"/><w:szCs w:val="32"/></w:r><w:pPr><w:rPr><w:rsidRPr="003E2449"><w:rsidRDefault="003E2449"><w:rsidRPr="003E2449"><w:r><w:b/><w:sz w:val="32"/><w:szCs w:val="32"/></w:r><w:t xml:space="preserve">Roller Coasters </w:t></w:p><w:p w:rsidR="002107BB" w:rsidRDefault="003E2449"><w:r w:rsidRPr="003E2449"><w:rPr><w:i/></w:rPr><w:t>Who wants to save an old roller coaster?</w:t></w:r><w:r w:rsidRPr="003E2449"><w:t xml:space="preserve"> Leap-the-Dips is the world's oldest roller coaster and, according to a spokesperson for the Leap-the-Dips Preservation Foundation, one of the most historically significant. Built in 1902, Leap-the-Dips </w:t></w:r><w:proofErr w:type="gramStart"/><w:r w:rsidRPr="003E2449"><w:t>is</w:t></w:r><w:proofErr w:type="gramEnd"/><w:r w:rsidRPr="003E2449"><w:t xml:space="preserve"> "the sole survivor of a style and technology that was represented in more than 250 parks in North America alone in the early years of the amusement industry.</w:t></w:r><w:r w:rsidR="00881463"><w:t>
```



## Bites and Bytes

- All of the data stored and transmitted by digital devices is encoded as *bits*.
- Terminology related to bits and bytes is extensively used to describe storage capacity and network access speed.
- The word **bit**, an abbreviation for **binary digit**, can be further abbreviated as a lowercase *b*.
- A group of eight bits is called a *byte* and is usually abbreviated as an uppercase *B*.

## Bites and Bytes

- When reading about digital devices, you'll frequently encounter references such as 90 kilobits per second, 1.44 megabytes, 2.8 gigahertz, and 2 terabytes.
- *Kilo, mega, giga, tera*, and similar terms are used to quantify digital data.

## Bites and Bytes

Bit	One binary digit
Byte	8 bits
Kilobit	1,024 or $2^{10}$ bits
Kilobyte	1,024 or $2^{10}$ bytes
Megabit	1,048,576 or $2^{20}$ bits
Megabyte	1,048,576 or $2^{20}$ bytes
Gigabit	$2^{30}$ bits
Gigabyte	$2^{30}$ bytes
Terabyte	$2^{40}$ bytes
Petabyte	$2^{50}$ bytes
Exabyte	$2^{60}$ bytes

## Bites and Bytes

- Use bits for data rates, such as Internet connection speeds, and movie download speeds.
- Use bytes for file sizes and storage capacities.
- **104 KB:** Kilobyte (KB or Kbyte) is often used when referring to the size of small computer files.

## Bites and Bytes

- **56 Kbps:** Kilobit (Kb or Kbit) can be used for slow data rates, such as a 56 Kbps (kilobits per second) dial-up connection.
- **50 Mbps:** Megabit (Mb or Mbit) is used for faster data rates, such as a 50 Mbps (megabits per second) Internet connection.

## Bites and Bytes

- **3.2 MB:** Megabyte (MB or MByte) is typically used when referring to the size of files containing photos and videos.
- **100 Gbit:** Gigabit (Gb or Gbit) is used for really fast network speeds.
- **16 GB:** Gigabyte (GB or GByte) is commonly used to refer to storage capacity

## Data Compression

- To reduce file size and transmission times, digital data can be compressed.
- **Data compression** refers to any technique that recodes the data in a file so that it contains fewer bits.
- Compression is commonly referred to as “zipping.”

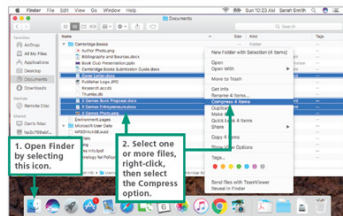
## Data Compression

- Compression techniques divided into two categories: *lossless* and *lossy*
- Lossless compression provides a way to compress data and reconstitute it into its original state; uncompressed data stays exactly the same as the original data
- Lossy compression throws away some of the original data during the compression process; uncompressed data is *not* exactly the same as the original

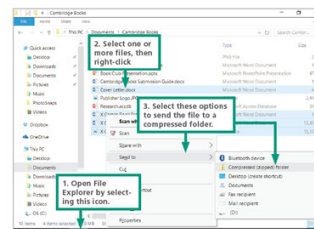
# Data Compression

- Software for compressing data is sometimes referred to as a compression utility or a zip tool.
- On laptops and desktop computers, the compression utility is accessed from the same screen used to manage files.

FIGURE 1-4: COMPRESSING FILES



Compressing files using Finder on a Mac



Compressing files using File Explorer on a PC

# Data Compression

- The process of reconstituting files is called extracting or unzipping.
- Compressed files may end with a .zip, .gz, .pkg, or .tar.gz.

