COMPUTER SYSTEMS FUNDAMENTALS
(4COSC004W)

Lecture: Week 4. Part 2 of 3

Contact details

- Module Leader:
  - *Noam Weingarten*
  - *See BlackBoard site for further contact details*

In this video we will cover:

- Character codes
  - *ASCII*
  - *UNICODE*

CHARACTER CODES

ASCII & UNICODE

By the end of this unit, you will:

- Understand standards of textual representation for computer systems
  - *ASCII*
  - *UNICODE*

ASCII
American Standard Code Information Interchange

- 7-Bit code:
  - *defining 128 characters*
- Includes:
  - *Roman alphabet (upper and lower case)*
  - *Control characters (CR, LF, ESC)*
  - *Digits*
  - *Punctuation*
  - *Non-printable characters*

ASCII table (abridged)

Den	Char
48	0
49	1
50	2
51	3
52	4
53	5
54	6
55	7
56	8
57	9
58	:

Den	Char
65	A
66	B
67	C
68	D
69	E
70	F
71	G
72	H
73	I
74	J
75	K
76	L
77	M

Den	Char
78	N
79	O
80	P
81	Q
82	R
83	S
84	T
85	U
86	V
87	W
88	X
89	Y
90	Z

Den	Char
97	a
98	b
99	c
100	d
101	e
102	f
103	g
104	h
105	i
106	j
107	k
108	l
109	m

Den	Char
110	n
111	o
112	p
113	q
114	r
115	s
116	t
117	u
118	v
119	w
120	X
121	Y
122	z

ASCII Table with Binary and Hexadecimal

FIGURE 3.25
The American Standard Code for Information Interchange (ASCII).

Char	Bin	Hex	Char	Bin	Hex	Char	Bin	Hex	Char	Bin	Hex
NUL	000 0000	00	SP	010 0000	20	@	100 0000	40	`	110 0000	60
SOH	000 0001	01	!	010 0001	21	A	100 0001	41	a	110 0001	61
STX	000 0010	02	"	010 0010	22	B	100 0010	42	b	110 0010	62
ETX	000 0011	03	#	010 0011	23	C	100 0011	43	c	110 0011	63
EOT	000 0100	04	$	010 0100	24	D	100 0100	44	d	110 0100	64
ENQ	000 0101	05	%	010 0101	25	E	100 0101	45	e	110 0101	65
ACK	000 0110	06	&	010 0110	26	F	100 0110	46	f	110 0110	66
BEL	000 0111	07	'	010 0111	27	G	100 0111	47	g	110 0111	67
BS	000 1000	08	(	010 1000	28	H	100 1000	48	h	110 1000	68
HT	000 1001	09	)	010 1001	29	I	100 1001	49	i	110 1001	69
LF	000 1010	0A	*	010 1010	2A	J	100 1010	4A	j	110 1010	6A
VT	000 1011	0B	+	010 1011	2B	K	100 1011	4B	k	110 1011	6B
FF	000 1100	0C	,	010 1100	2C	L	100 1100	4C	l	110 1100	6C
CR	000 1101	0D	-	010 1101	2D	M	100 1101	4D	m	110 1101	6D
SO	000 1110	0E	.	010 1110	2E	N	100 1110	4E	n	110 1110	6E
SI	000 1111	0F	/	010 1111	2F	O	100 1111	4F	o	110 1111	6F
DLE	001 0000	10	0	011 0000	30	P	101 0000	50	p	111 0000	70
DC1	001 0001	11	1	011 0001	31	Q	101 0001	51	q	111 0001	71
DC2	001 0010	12	2	011 0010	32	R	101 0010	52	r	111 0010	72
DC3	001 0011	13	3	011 0011	33	S	101 0011	53	s	111 0011	73
DC4	001 0100	14	4	011 0100	34	T	101 0100	54	t	111 0100	74
NAK	001 0101	15	5	011 0101	35	U	101 0101	55	u	111 0101	75
SYN	001 0110	16	6	011 0110	36	V	101 0110	56	v	111 0110	76
ETB	001 0111	17	7	011 0111	37	W	101 0111	57	w	111 0111	77
CAN	001 1000	18	8	011 1000	38	X	101 1000	58	x	111 1000	78
EM	001 1001	19	9	011 1001	39	Y	101 1001	59	y	111 1001	79
SUB	001 1010	1A	:	011 1010	3A	Z	101 1010	5A	z	111 1010	7A
ESC	001 1011	1B	;	011 1011	3B	[	101 1011	5B	{	111 1011	7B
FS	001 1100	1C	<	011 1100	3C	\	101 1100	5C	|	111 1100	7C
GS	001 1101	1D	=	011 1101	3D	]	101 1101	5D	}	111 1101	7D
RS	001 1110	1E	>	011 1110	3E	^	101 1110	5E	~	111 1110	7E
US	001 1111	1F	?	011 1111	3F	_	101 1111	5F	DEL	111 1111	7F

48 65 6C 6C 6F

H e l l o

Computer Systems (3.4)

ASCII extension ISO-8859 series

- 8 bit ASCII
  - *Uses 8th Bit for error checking (parity)*
- American standard (no £ sign)
- ISO-8859 series
  - *Uses all 8 Bits for character code*
  - *Extension of ASCII*
  - *Includes other characters; £, €, ….*

UNICODE (UTF-16)

- 16-Bit international standard code
- UNICODE consortium
  - *Adobe, Apple, Facebook, Google, IBM, MS, Oracle, Symantec, ….*
  - *128 ASCII characters*
  - *Latin extended*
  - *….*
- Used by MS Windows, Java, .NET

UNICODE some examples:

Character codes	Decimal range	Hexadecimal
C0 Controls & Basic Latin	0-127	0000-007F
C1 Controls and Latin-1 supplement	128-255	0080-00FF
Greek & Coptic	880-1023	0370-03FF
Hebrew	1424-1535	0590-05FF
Arabic	1536-1791	0600-06FF
….	…	…

ASCII vs. UNICODE

- ASCII
  - *Very limited*
  - *Only Roman alphabet support*
- UNICODE
  - *International 16 bit code*
  - *Support for other alphabets:*
    - Chinese, Thai, Greek, Hebrew, Arabic, ….
  - *Supported as standard by many OS and Development Environments*
  - *Up to 65,536 different characters (2^{16})*
  - *UTF-8 (32-bit)*
    - Preferred encoding for e-mails & web pages
    - Up to 4,294,967,296 different characters (2^{32})

In this video we covered:

- Text storage
  - *ASCII*
  - *UNICODE*