CS2211b

Software Tools and Systems Programming

Week 5a

Data Representation & Binary

Data Representation & Binary

Assigned Reading

Binary Tutorial:

- https://ryanstutorials.net/binary-tutorial/

Parts 1, 2 and 4

Not required to know Binary Arithmetic or Octal

What is Data?

- A set of values of qualitative or quantitative variables
- Individual pieces of information
- Measured phenomena
- The pattern of organization of matter and energy
- Portion of the entire information environment available to a sensing organism that is taken in, or processed, by that organism

Where Does Data Come From?

- Measurement
- Transactions
- Computation
- Databases

How is Data Represented?

By Humans

- Writing (Numbers, Letters)
- Drawings (Technical, Art)
- Speech (Talking, Lecturing, etc.)
- Body Language (Counting on fingers, etc.)

By Computers

- ?

How is Data Represented?

By Computers

- **Location:** In memory, in files
- **Representation:** Bits
- **Interpretation**: Binary

How is Data Represented?

By Computers

- **Location:** In memory, in files
- **Representation:** Bits
- **Interpretation**: Binary

How is Data Represented?

Unsigned Integers

- Represented in binary (base 2)
- Only positive
- Example #1:

$$1101_2 = 1 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0$$
$$= 1 \times 8 + 1 \times 4 + 0 \times 2 + 1 \times 1$$
$$= 8 + 4 + 0 + 1$$
$$= 13_{10}$$

How is Data Represented?

Signed Integers

- How to represent negative numbers?
- Use most significant bit as the sign bit
- Example #2:

$$11010_2 = -1 \times (1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 0 \times 2^0)$$

$$= -1 \times (1 \times 8 + 0 \times 4 + 1 \times 2 + 0 \times 1)$$

$$= -1 \times (8 + 0 + 2 + 0)$$

$$= -10_{10}$$

How is Data Represented?

Signed Integers

- How to represent negative numbers?
- Use most significant bit as the sign bit
- Example #3:

$$01010_2 = 1 \times (1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 0 \times 2^0)$$

$$= 1 \times (1 \times 8 + 0 \times 4 + 1 \times 2 + 0 \times 1)$$

$$= 1 \times (8 + 0 + 2 + 0)$$

$$= 10_{10}$$

How is Data Represented?

Unsigned

For 8 bits:

Min: 0

Max: 255

Signed

For 8 bits:

Min: -127

Max: 127

Both 10000000 and 00000000 equal 0!

How is Data Represented?

Can we do better for signed numbers?

- 2's Complement
- Still have sign bit as the most significant binary digit. However, its place/value matters.
- Positive numbers are represented in the same way as unsigned numbers.
- Negative numbers are the result of adding the negative (sign) bit with the other bits.

How is Data Represented?

2's Complement

Example #4:

$1011_2 = -1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 1 \times 2^0$

$= -1 \times 8 + 0 \times 4 + 1 \times 2 + 1 \times 1$

$= -8 + 0 + 2 + 1$

$= -5_{10}$

How is Data Represented?

2's Complement

Example #5:

$0110_2 = -0 \times 2^3 + 1 \times 2^2 + 1 \times 2^1 + 0 \times 2^0$

$= -0 \times 8 + 1 \times 4 + 1 \times 2 + 0 \times 1$

$= -0 + 4 + 2 + 0$

$= 6_{10}$

How is Data Represented?

2's Complement

Unsigned

For 8 bits:

Min: 0

Max: 255

Signed

For 8 bits:

Min: -128

Max: 127

One more value than before.

How is Data Represented?

2's Complement

- Requires correct padding on binary number so sign bit is in correct place.
- This is how Integers are represented in C (and in most computer applications). Will be important later when we get in to C programming.
- Covered in more depth in assigned reading:

https://ryanstutorials.net/binary-tutorial/binary-negative-numbers.php

How is Data Represented?

Convert Back to Decimal (unsigned)

- Repeated division by 2

For 18_{10}

18 / 2 = 9 (no remainder)	0
9 / 2 = 4**.5** (remainder)	1
4 / 2 = 2 (no remainder)	0
2 / 2 = 1 (no remainder)	0
1 / 2 = 0**.5** (remainder)	1

$18_{10} = 10010_2$

Read bottom up!

For 31_{10}

31 / 2 = 15.5	1
15 / 2 = 7.5	1
7 / 2 = 3.75	1
3 / 2 = 1.5	1
1 / 2 = 0.5	1

$18_{10} = 11111_2$

How is Data Represented?

Other Bases

- Works for other bases
- For example, base 16 (Hexadecimal):

$$A7_{16} = A \times 16^1 + 7 \times 16^0$$
$$= 10 \times 16^1 + 7 \times 16^0$$
$$= 10 \times 16 + 7 \times 1$$
$$= 160 + 7$$
$$= 167_{10}$$

A = 10
B = 11
C = 12
D = 13
E = 14
F = 15

How is Data Represented?

Converting Back is Harder

- Easier to convert to binary first
- Example:

$C4F_{16}$

Base 16:	C	4	F
Base 2:	1100	0100	1111

$C4F_{16} = 110001001111_2$

Decimal	Binary	Hexadecimal
0	0000	0
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

How is Data Represented?

Converting Back is Harder

- Now convert from binary to decimal
- Example:

$$C4F_{16} = 110001001111_2$$

$$= 2^{11}+2^{10}+2^6+2^3+2^2+2^1+2^0$$

$$= 3151_{10}$$

How is Data Represented?

What about letters?

- American Standard Code for Information Interchange (ASCII)
- Table for converting numbers to characters

How is Data Represented?

ASCII TABLE

Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char
0	0	[NULL]	32	20	[SPACE]	64	40	@	96	60	`
1	1	[START OF HEADING]	33	21	!	65	41	A	97	61	a
2	2	[START OF TEXT]	34	22	"	66	42	B	98	62	b
3	3	[END OF TEXT]	35	23	#	67	43	C	99	63	c
4	4	[END OF TRANSMISSION]	36	24	$	68	44	D	100	64	d
5	5	[ENQUIRY]	37	25	%	69	45	E	101	65	e
6	6	[ACKNOWLEDGE]	38	26	&	70	46	F	102	66	f
7	7	[BELL]	39	27	'	71	47	G	103	67	g
8	8	[BACKSPACE]	40	28	(	72	48	H	104	68	h
9	9	[HORIZONTAL TAB]	41	29	)	73	49	I	105	69	i
10	A	[LINE FEED]	42	2A	*	74	4A	J	106	6A	j
11	B	[VERTICAL TAB]	43	2B	+	75	4B	K	107	6B	k
12	C	[FORM FEED]	44	2C	,	76	4C	L	108	6C	l
13	D	[CARRIAGE RETURN]	45	2D	-	77	4D	M	109	6D	m
14	E	[SHIFT OUT]	46	2E	.	78	4E	N	110	6E	n
15	F	[SHIFT IN]	47	2F	/	79	4F	O	111	6F	o
16	10	[DATA LINK ESCAPE]	48	30	0	80	50	P	112	70	p
17	11	[DEVICE CONTROL 1]	49	31	1	81	51	Q	113	71	q
18	12	[DEVICE CONTROL 2]	50	32	2	82	52	R	114	72	r
19	13	[DEVICE CONTROL 3]	51	33	3	83	53	S	115	73	s
20	14	[DEVICE CONTROL 4]	52	34	4	84	54	T	116	74	t
21	15	[NEGATIVE ACKNOWLEDGE]	53	35	5	85	55	U	117	75	u
22	16	[SYNCHRONOUS IDLE]	54	36	6	86	56	V	118	76	v
23	17	[ENG OF TRANS. BLOCK]	55	37	7	87	57	W	119	77	w
24	18	[CANCEL]	56	38	8	88	58	X	120	78	x
25	19	[END OF MEDIUM]	57	39	9	89	59	Y	121	79	y
26	1A	[SUBSTITUTE]	58	3A	:	90	5A	Z	122	7A	z
27	1B	[ESCAPE]	59	3B	;	91	5B	[	123	7B	{
28	1C	[FILE SEPARATOR]	60	3C	<	92	5C	\	124	7C	|
29	1D	[GROUP SEPARATOR]	61	3D	=	93	5D	]	125	7D	}
30	1E	[RECORD SEPARATOR]	62	3E	>	94	5E	^	126	7E	~
31	1F	[UNIT SEPARATOR]	63	3F	?	95	5F	_	127	7F	[DEL]

How is Data Represented?

What about letters?

- Example:

01001000_2 01100101_2 01101100_2 01101100_2 01101111_2

= 72_{10} 101_{10} 108_{10} 108_{10} 111_{10}

= H e l l o

How is Data Represented?

Other Text Encodings

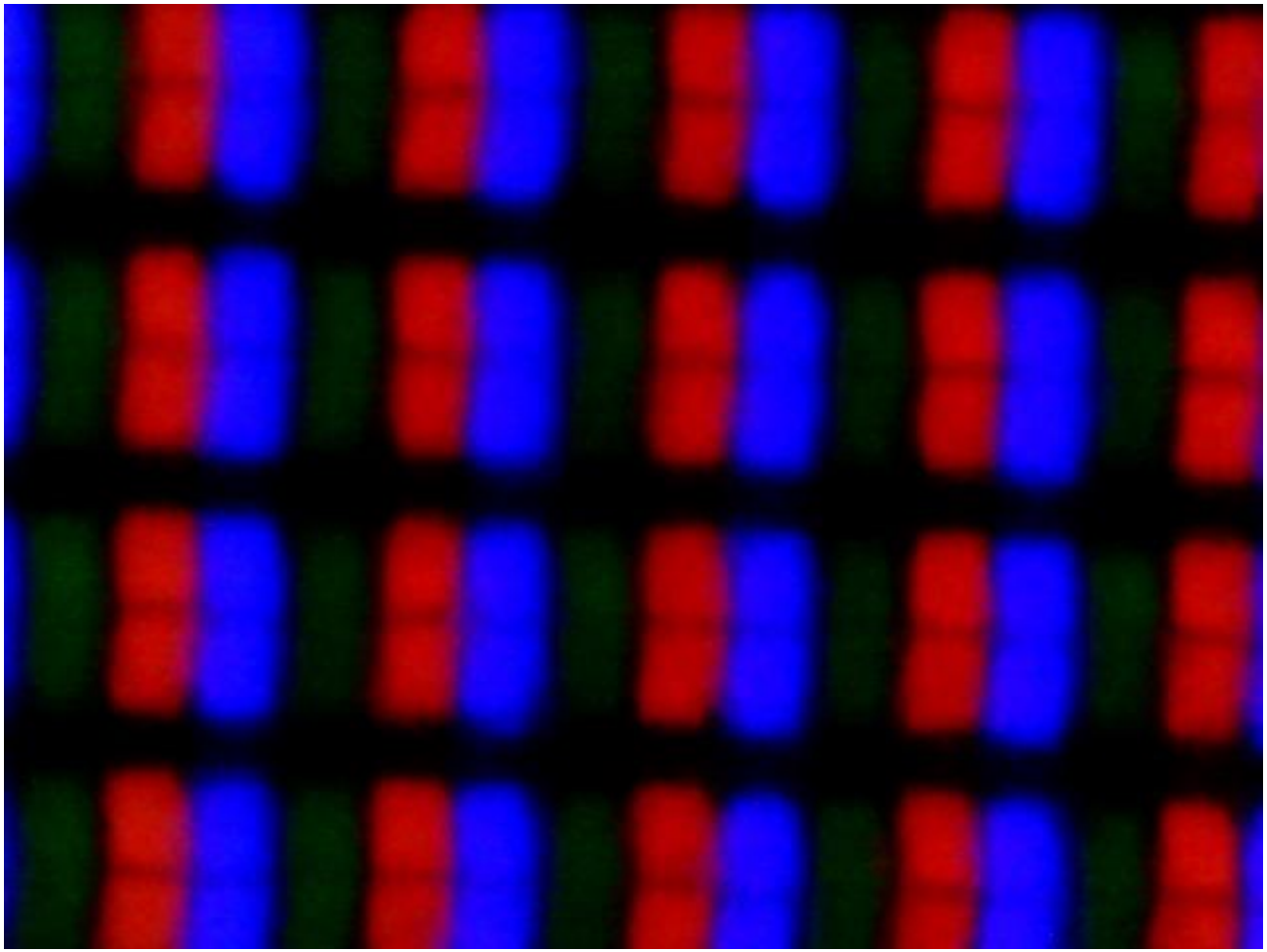
- Other encodings than ASCII exist:
  - UTF
  - Unicode
- Newer and more modern standards but take up more space (bits)
- Offer more characters:
  - Other languages
  - Math symbols
  - Emoji
  - And more!

How is Data Represented?

What about images?

How is Data Represented?

What about images?



How is Data Represented?

What about images?

- Encoded as pixels
- 3 bytes (24 bits)
- Each bit represents a color (red, green and blue)
- Values are turned into light and mixed together to form a single color
- Example: 79_{10}, 38_{10}, 131_{10} =

- Often represented in hex:

#4F2683

How is Data Represented?

In Files?

- How is data represented in files?
- Also in bits.
- Same way as in memory, but storage and interpretation may be different.
- Some file formats:
  - Text
  - Binary
  - XML

How is Data Represented?

Text Files

- Bits encoded a ASCII, UTF or Unicode
- Not always machine readable

ASCII Text File

```
Hello World!

This is a text file.
```

ASCII Binary

```
01001000 01100101 01101100
01101100 01101111 00100000
01010111 01101111 01110010
01101100 01100100 00100001
00100000 00001010 00001010
01010100 01101000 01101001
01110011 00100000 01101001
01110011 00100000 01100001
00100000 01110100 01100101
01111000 01110100 00100000
01100110 01101001 01101100
01100101 00101110 00001010
```

How is Data Represented?

Text Files

- Bits encoded a ASCII, UTF or Unicode
- Not always machine readable

ASCII Text File

```
Hello World!

This is a text file.
```

ASCII Binary

```
01001000 01100101 01101100
01101100 01101111 00100000
01010111 01101111 01110010
01101100 01100100 00100001
00100000 00001010 00001010
01010100 01101000 01101001
01110011 00100000 01101001
01110011 00100000 01100001
00100000 01110100 01100101
01111000 01110100 00100000
01100110 01101001 01101100
01100101 00101110 00001010
```

How is Data Represented?

Text Files

- Have invisible characters to denote things like spaces and line breaks

ASCII Text File (showing invisible chars)

```
Hello␣World!␣\n
\n
This␣is␣a␣text␣file.\n
```

ASCII Binary

```
01001000 01100101 01101100
01101100 01101111 00100000
01010111 01101111 01110010
01101100 01100100 00100001
00100000 00001010 00001010
01010100 01101000 01101001
01110011 00100000 01101001
01110011 00100000 01100001
00100000 01110100 01100101
01111000 01110100 00100000
01100110 01101001 01101100
01100101 00101110 00001010
```

How is Data Represented?

Text Files

- Easy to construct and to read
- Easy for people to check and debug
- Used for both human readable files and files for programs to read, Ex:
  - Comma-Separated Values (CSV)
  - Tab-Separated Values (TSV)
  - Shell Scripts
  - Most log files in /var/log
  - XML, HTML, etc.

How is Data Represented?

Binary Files

- Define their own format/encoding
- Not human readable

PNG Image: 

Viewed as ASCII

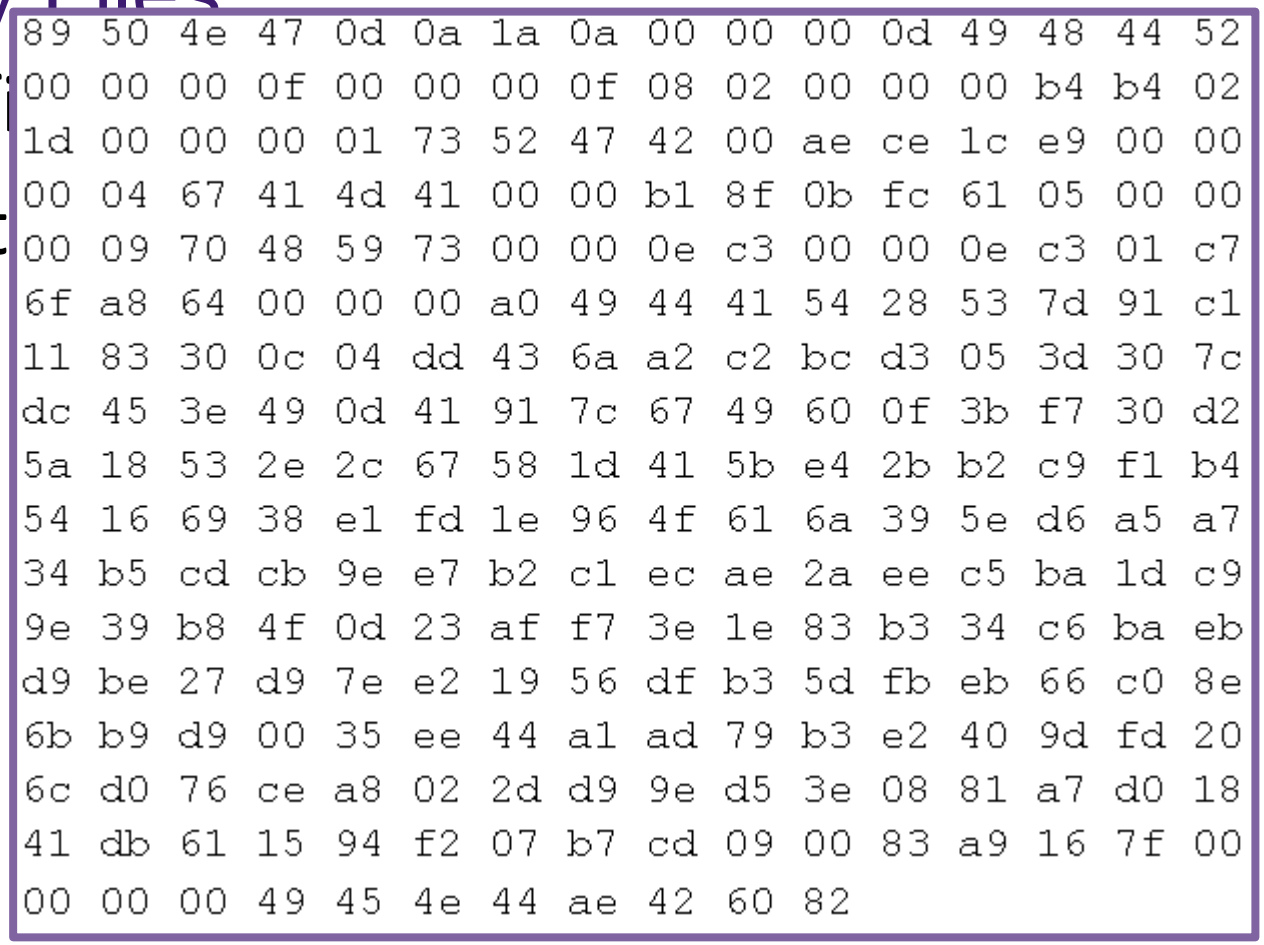
```
‰PNG

IHDR          ´´   sRGB ®Îé   gAMA  ±
üa     pHYs  Ã  ÃÇo¨d    IDAT(S}‘Áƒ0ÝCj¢Â¼Ó=0|ÜE>I
A‘|gI`;÷0ÒZS.,gXA[ä+²Éñ´Ti8áý–Oaj9^Ö¥§4µÍËžç²Áì®*îÅºÉž9¸O
#¯÷>ƒ³4ÆºëÙ¾'Ù~âVß³]ûëfÀŽk¹Ù 5îD¡y³â@ lÐvÎ¨-ÙžÕ>ÐAÛa”ò·Í
ƒ©    IEND®B`,
```

How is Data Represented?

Binary Files

- Defi
- Not



```
89 50 4e 47 0d 0a 1a 0a 00 00 00 0d 49 48 44 52
00 00 00 0f 00 00 00 0f 08 02 00 00 00 b4 b4 02
1d 00 00 00 01 73 52 47 42 00 ae ce 1c e9 00 00
00 04 67 41 4d 41 00 00 b1 8f 0b fc 61 05 00 00
00 09 70 48 59 73 00 00 0e c3 00 00 0e c3 01 c7
6f a8 64 00 00 00 a0 49 44 41 54 28 53 7d 91 c1
11 83 30 0c 04 dd 43 6a a2 c2 bc d3 05 3d 30 7c
dc 45 3e 49 0d 41 91 7c 67 49 60 0f 3b f7 30 d2
5a 18 53 2e 2c 67 58 1d 41 5b e4 2b b2 c9 f1 b4
54 16 69 38 e1 fd 1e 96 4f 61 6a 39 5e d6 a5 a7
34 b5 cd cb 9e e7 b2 c1 ec ae 2a ee c5 ba 1d c9
9e 39 b8 4f 0d 23 af f7 3e 1e 83 b3 34 c6 ba eb
d9 be 27 d9 7e e2 19 56 df b3 5d fb eb 66 c0 8e
6b b9 d9 00 35 ee 44 a1 ad 79 b3 e2 40 9d fd 20
6c d0 76 ce a8 02 2d d9 9e d5 3e 08 81 a7 d0 18
41 db 61 15 94 f2 07 b7 cd 09 00 83 a9 16 7f 00
00 00 00 49 45 4e 44 ae 42 60 82
```

Viewed as hexadecimal

How is Data Represented?

Binary Files

- Pros:
  - Store data more compactly
  - Less space required to store data
  - Faster for computers to read and write
  - Faster to send over network
- Cons:
  - Unforgiving format
  - Harder to program and read (for humans)
  - Usually specific to one program or family of programs

How is Data Represented?

Binary Files

- Examples:
  - Images: PNG, JPG, GIF, BMP
  - Video: MP4, AVI, FLV, WMV
  - Audio: MP3, WAV, AAC, FLAC
  - Executables

Practice Problems

Practice Problems

Binary

Convert these unsigned base 2 numbers to base 10:

1101_2 = _____$_{10}$

0010111_2 = _____$_{10}$

1101010_2 = _____$_{10}$

Convert these unsigned base 2 numbers to base 16:

00001101_2 = _____$_{16}$

10010111_2 = _____$_{16}$

01101010_2 = _____$_{16}$

Practice Problems

Decimal

Convert these unsigned base 10 numbers to base 2:

10_{10} = _____$_2$

162_{10} = _____$_2$

0060_{10} = _____$_2$

Convert these unsigned base 10 numbers to base 16:

285_{10} = _____$_{16}$

141_{10} = _____$_{16}$

096_{10} = _____$_{16}$

Practice Problems

Hexadecimal

Convert these unsigned base 16 numbers to base 2:

DA_{16} = _____$_2$

85_{16} = _____$_2$

$B1F_{16}$ = _____$_2$

Convert these unsigned base 16 numbers to base 10:

FB_{16} = _____$_{10}$

51_{16} = _____$_{10}$

$11A_{16}$ = _____$_{10}$

Appendix

Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char
32	20	[SPACE]	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	k
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	[DEL]

Decimal	Binary	Hexadecimal
0	0000	0
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