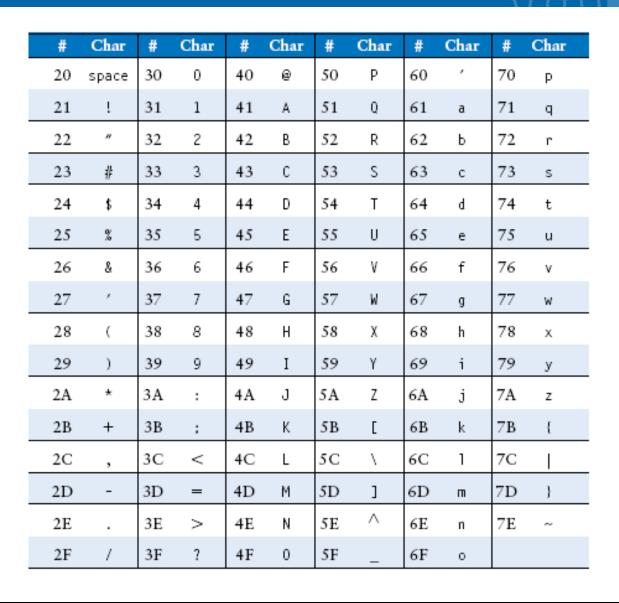- Human readable information must also be represented
  - Characters and text
- ASCII and Unicode code are most common
- Some earlier systems used EBCDIC
- ASCII and EBCDIC are 8-bit codes
  - A single character fits into one byte
  - This allows a maximum of 256 characters
- Unicode uses 2 bytes per symbol
  - Allows many more characters
  - Including user defined symbols

JOHNS HOPKINS UNIVERSITY | **Engineering for Professionals**

#	Char	#	Char	#	Char	#	Char	#	Char	#	Char
20	space	30	0	40	@	50	P	60	'	70	p
21	!	31	1	41	A	51	Q	61	a	71	q
22	"	32	2	42	B	52	R	62	b	72	r
23	#	33	3	43	C	53	S	63	c	73	s
24	$	34	4	44	D	54	T	64	d	74	t
25	%	35	5	45	E	55	U	65	e	75	u
26	&	36	6	46	F	56	V	66	f	76	v
27	'	37	7	47	G	57	W	67	g	77	w
28	(	38	8	48	H	58	X	68	h	78	x
29	)	39	9	49	I	59	Y	69	i	79	y
2A	*	3A	:	4A	J	5A	Z	6A	j	7A	z
2B	+	3B	;	4B	K	5B	[	6B	k	7B	{
2C	,	3C	<	4C	L	5C	\	6C	l	7C		
2D	-	3D	=	4D	M	5D	]	6D	m	7D	}
2E	.	3E	>	4E	N	5E	^	6E	n	7E	~
2F	/	3F	?	4F	O	5F	_	6F	o		

- Strings are arrays of characters
  - the location of the leading byte is the string address

- MIPS uses lbu (load byte) and sb (store byte)

- MIPS has no instruction to manipulate entire strings
  - Loops containing lbu or sb are used
  - Unlike some CISC machines

JOHNS HOPKINS UNIVERSITY

Engineering for Professionals

- The Unicode code space allocation is shown at the right.

- The lowest-numbered Unicode characters comprise the ASCII code.

- The highest provide for user-defined codes.

Character Types	Language	Number of Characters	Hexadecimal Values
Alphabets	Latin, Greek, Cyrillic, etc.	8192	0000 to 1FFF
Symbols	Dingbats, Mathematical, etc.	4096	2000 to 2FFF
CJK	Chinese, Japanese, and Korean phonetic symbols and punctuation.	4096	3000 to 3FFF
Han	Unified Chinese, Japanese, and Korean	40,960	4000 to DFFF
	Han Expansion	4096	E000 to EFFF
User Defined		4095	F000 to FFFE

- Bytes are the smallest addressable items on MIPS
  - It is a byte-addressable machine

- Access to bits require:
  - Using the byte or word containing the desired bits
  - Using masking operations or shifting to access bits

- Contents of a register or memory word could be interpreted as a series of bits (bit string)