

Introduction to Informatics

Lecture 10: Encoding Numbers (Part II)



"Would everyone please phrase their questions in ones and zeros please."

Readings until now



- Lecture notes

- Posted online @ <http://informatics.indiana.edu/rocha/i101>

- *The Nature of Information*
- *Technology*
- *Modeling the World*

- @ *infoport*

- From course package

- Von Baeyer, H.C. [2004]. *Information: The New Language of Science*. Harvard University Press.
 - Chapters 1, 4 (pages 1-12)
- From Andy Clark's book "*Natural-Born Cyborgs*"
 - Chapters 2 and 6 (pages 19 - 67)
- From Irv Englander's book "*The Architecture of Computer Hardware and Systems Software*"
 - Chapter 3: Data Formats (pp. 70-86)

Assignment Situation

- Labs
 - Past
 - Lab 1: Blogs
 - Closed (Friday, January 19): Grades Posted
 - Lab 2: Basic HTML
 - Closed (Wednesday, January 31): Grades Posted
 - Lab 3: Advanced HTML: Cascading Style Sheets
 - Closed (Friday, February 2): Grades Posted
 - Lab 4: More HTML and CSS
 - Closed (Friday, February 9): being graded
 - Lab 5: Introduction to Operating Systems: Unix
 - Due Friday, February 16
 - Next: Lab 6
 - More Unix and FTP
 - Due Friday, February 23
- Assignments
 - Individual
 - First installment
 - Closed: February 9: Being Graded
 - Group Project
 - First installment
 - Presented: February 20, Due: March 9th
- Midterm Exam
 - March 1st (Thursday)



Individual assignment

- Individual Project

- 1st installment

- Presented: February 1st
- Due: February 9th

- 2nd Installment

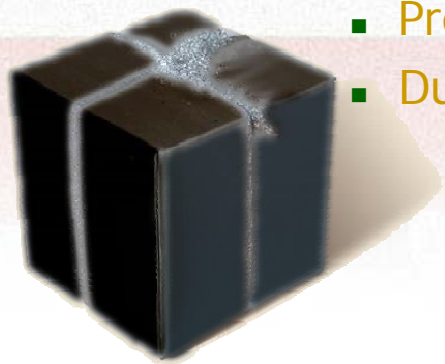
- Presented: February 15th
- Due: March: 2nd

- 3rd Installment

- Presented: March 8th
- Due: March 30th

- 4th Installment

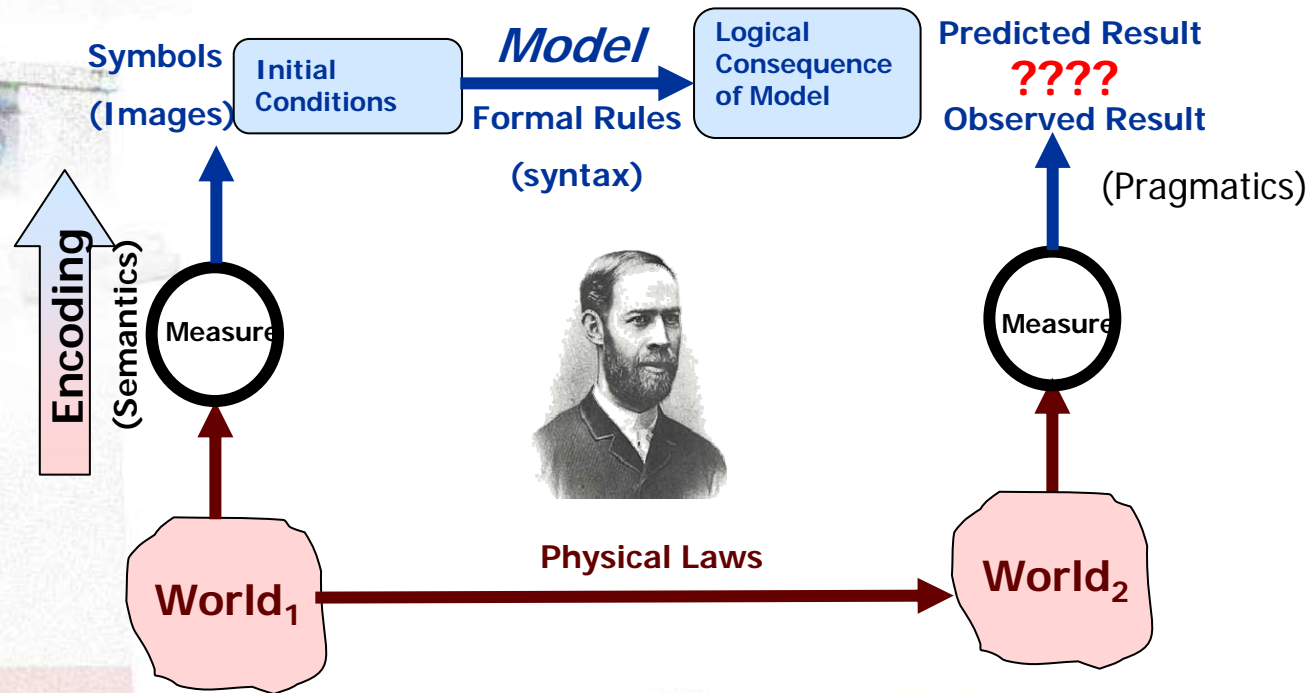
- Presented: April 5th
- Due: April 20th



The Black Box

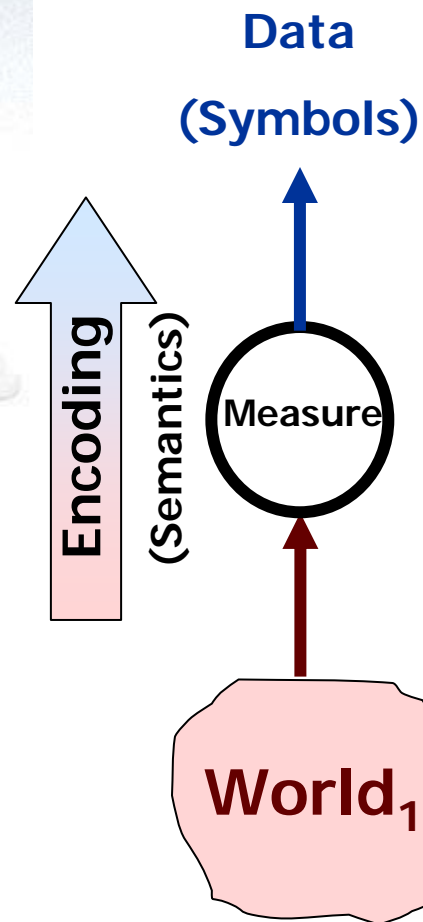


The Modeling Relation



Hertz' Modeling Paradigm

Encoding in the Modeling Relation



- How to encode data?
 - What is data?
 - Information without context and knowledge
 - Part of Syntax
 - Keeping Numbers
 - The most fundamental need for modeling and information

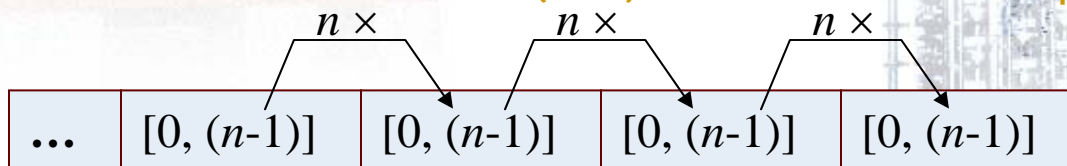
Counting with the Binary System

- Positional number system

- the value of each digit is determined by its position
 - 101 is different from 110
 - The lowest place value is the rightmost position, and each successive position to the left has a higher place value

- Base 2

- The value of each position corresponds to powers of 2
 - $\dots d_4 d_3 d_2 d_1 d_0 = \dots + d_4 \times 2^4 + d_3 \times 2^3 + d_2 \times 2^2 + d_1 \times 2^1 + d_0 \times 2^0$
 - Each digit to the left is 2 times the previous digit.
 - $111100011 (483) = 1 \times 2^8 + 1 \times 2^7 + 1 \times 2^6 + 1 \times 2^5 + 0 \times 2^4 + 0 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 1 \times 2^0$
- To multiply a number by 2 you can simply shift it to the left by one digit, and fill in the rightmost digit with a 0
 - $101 \times 2 = 1010 (5 \times 2 = 10)$
- To divide a number by 2, simply shift the number to the right by one digit (moving the decimal place one to the left).
 - $101 \div 2 = 10.1 (5 \div 2 = 2.5)$
- With n digits, 2^n unique numbers can be represented
 - If $n=8$, 256 ($=2^8$) numbers can be represented 0-11111111.



Comparing Binary with Decimal

- Binary:Decimal
 - 0000:00
 - 0001:01
 - 0010:02
 - 0011:03
 - 0100:04
 - 0101:05
 - 0110:06
 - 0111:07
- Binary:Decimal
 - 1000:08
 - 1001:09
 - 1010:10
 - 1011:11
 - 1100:12
 - 1101:13
 - 1110:14
 - 1111:15



Binary Code

- Language with an alphabet of two symbols
 - "0" and "1", "FALSE" or "TRUE", etc.
- Most economical way of encoding information
 - Example from von Baeyer's [2004] book *"Information: The new language of Science"*
 - Consider a sailor who wants to signal a number between 0 and 127 by means of flags.

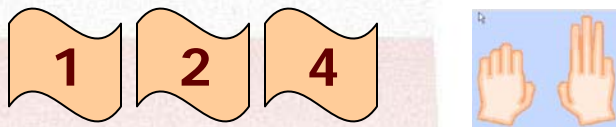
0: 0	4: 100
1: 1	5: 101
2: 10	6: 110
3: 11	7: 111

Message from Ship

System 1: one flag per number



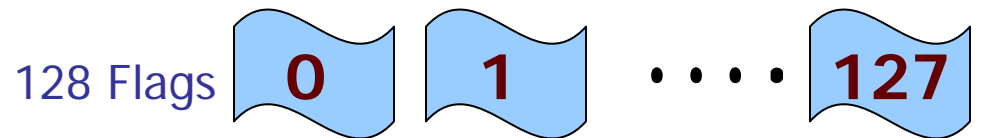
System 2: decimal system



System 2: binary system

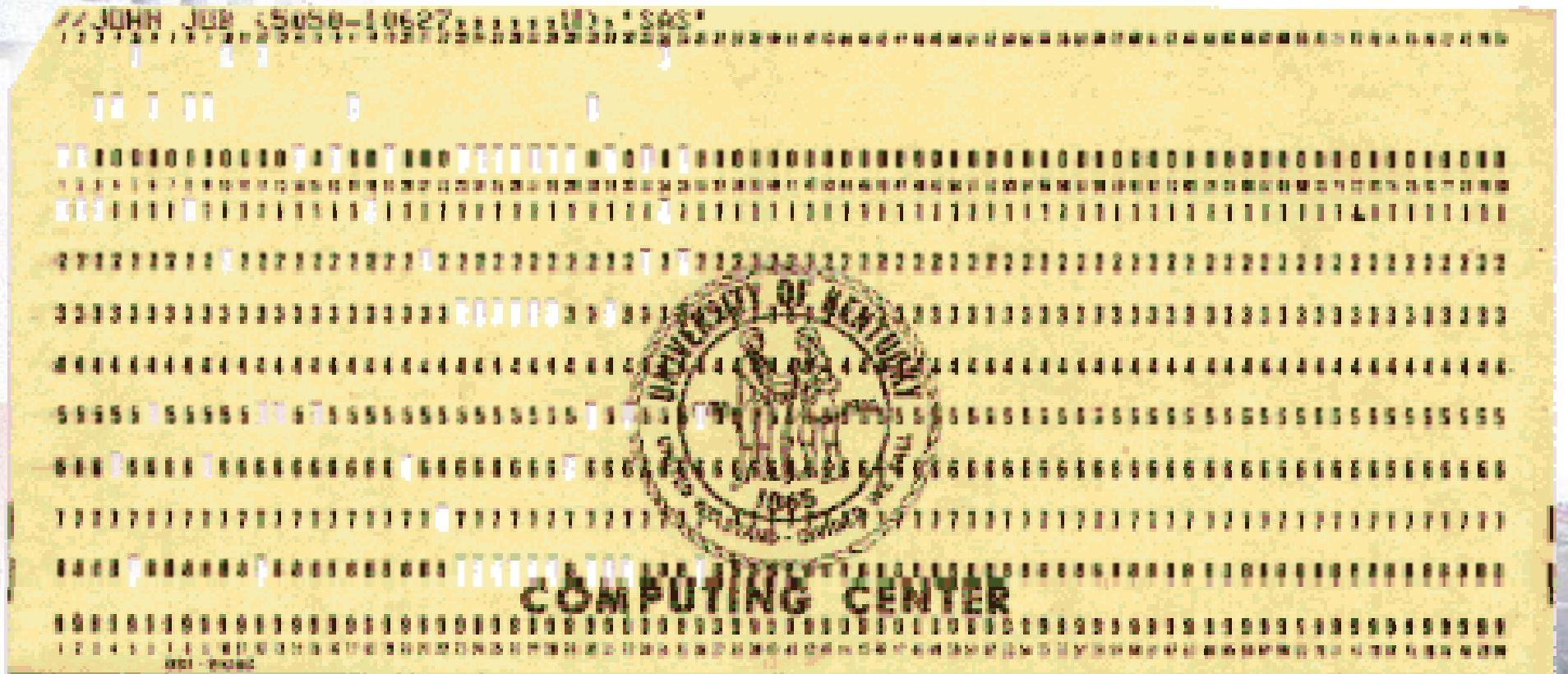


Required Flags



Memory: Punch Card

- Binary Representation
 - Holes denote 1's
 - With 8 holes permissible $2^8 = 256$ numbers possible per column



Converting Binary to Decimal

- $2^8 = 256$

- $2^7 = 128$

- $2^6 = 64$

- $2^5 = 32$

- $2^4 = 16$

- $2^3 = 8$

- $2^2 = 4$

- $2^1 = 2$

- $2^0 = 1$

2^8	2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0
-------	-------	-------	-------	-------	-------	-------	-------	-------

0	1	1	0	0	1	0	0	1
---	---	---	---	---	---	---	---	---

128	+	64
-----	---	----

+	8
---	---

+	1
---	---

201

$$\dots d_4 d_3 d_2 d_1 d_0 =$$

$$\dots + d_4 \times 2^4 + d_3 \times 2^3 + d_2 \times 2^2 + d_1 \times 2^1 + d_0 \times 2^0$$

Base Conversion

■ Decimal to Binary

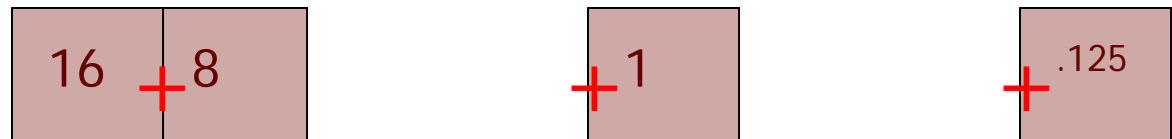
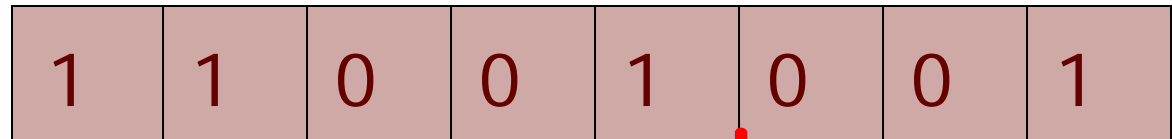
■ Repeated Division by 2

- Divide the decimal number by 2
- If the remainder is 0, on the side write down a 0
- If the remainder is 1, write down a 1
- Continue until the quotient is 0
- Remainders are written beginning at the least significant digit (right) and each new digit is written to left (the most significant digit) of the previous digit.

decimal	quotient	Remain.	binary
58	29	0	0
29	14	1	10
14	7	0	010
7	3	1	1010
3	1	1	11010
1	0	1	111010

Dealing with rational numbers

- $2^4 = 16$
- $2^3 = 8$
- $2^2 = 4$
- $2^1 = 2$
- $2^0 = 1$
- $2^{-1} = 0.5$
- $2^{-2} = 0.25$
- $2^{-3} = 0.125$



25.125

$$\dots d_2 d_1 d_0 . d_{-1} d_{-2} \dots =$$

$$\dots + d_2 \times 2^2 + d_1 \times 2^1 + d_0 \times 2^0 + d_{-1} \times 2^{-1} + d_{-2} \times 2^{-2} + \dots$$

Binary groupings

- Bit
 - Size: 1
 - 0-1 (2 Values)
- Nibble
 - Size: 4
 - 0-15 (16 Values)
 - 1100
- Byte
 - Size: 8
 - 0-255 (256 Values)
 - 10110101
- Word
 - Size: 16
 - 0-65535 (65536)
 - 1100000010100101

Binary Arithmetic

■ Addition Rules

- $0+0 = 0$, with no carry,
- $1+0 = 1$, with no carry,
- $0+1 = 1$, with no carry,
- $1+1 = 0$, and you carry a 1

$$\begin{array}{r} 111 \\ 1010 \\ + 1110 \\ \hline 11000 \end{array}$$

(10+14=24)

$$\begin{array}{r} 1 \\ 1010 \\ + 1100 \\ \hline 10110 \end{array}$$

(10+12=22)

Binary Multiplication

```
1010
1110
-----
111_
0000
1010_
1010_
1010_
-----
10001100
```

(10+14=140)

```
1010
1100
-----
0000
0000_
1010_
1010_
-----
1111000
```

(10×12=120)

Signed Integers

b_7	b_6	b_5	b_4	b_3	b_2	b_1	b_0
1	1	0	1	0	1	0	0
2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0

$$\begin{aligned}x &= (+/-) (b_6 2^6 + b_5 2^5 + \dots + b_0 2^0) \\ &= (-1)(1 \cdot 2^6 + 0 \cdot 2^5 + \dots + 0 \cdot 2^0) \\ &= (-1)(64 + 16 + 4) \\ &= -84\end{aligned}$$

From Cathy Wyss (I308)

Fixed Point Reals

b_9	b_8	b_7	b_6	b_5	b_4	b_3	b_2	b_1	b_0
1	1	0	0	1	0	0	1	0	1
2^5	2^4	2^3	2^2	2^1	2^0	2^{-1}	2^{-2}	2^{-3}	2^{-4}

$$\begin{aligned}x &= b_9 2^5 + b_8 2^4 + \dots + b_0 2^{-4} \\ &= 1 \cdot 2^5 + 1 \cdot 2^4 + \dots + 1 \cdot 2^{-4} \\ &= 32 + 16 + 2 + 0.25 + 0.0625 \\ &= 50.3125\end{aligned}$$

From Cathy Wyss (I308)

Floating Point

$$x = (+/-)(1+F) \cdot 2^{E-B}$$

sign	exponent					fraction			
b ₉	b ₈	b ₇	b ₆	b ₅	b ₄	b ₃	b ₂	b ₁	b ₀
1	1	0	0	1	0	0	1	0	1
	2 ⁵	2 ⁴	2 ³	2 ²	2 ¹	2 ⁰	2 ⁻¹	2 ⁻²	2 ⁻³
(+/-)	<i>E</i>					<i>F</i>			

- Sign (+/-)
 - 0 denotes a positive number
 - 1 denotes a negative number
- Exponent (*E-B*)
 - The exponent base (2) is implicit and need not be stored.
 - A *bias* (B) is added to represent both positive and negative exponents.
 - IEEE single-precision floats B=127.
 - If $E = 127$, exponent is zero
 - If $E = 200$, exponent is $(200-127) 73$.
 - IEEE double precision, exponent field is 11 bits, and bias is 1023.
- Mantissa ($1 + F$)
 - Fraction (*F*) plus an implicit leading digit.

Floating Point Reals

sign	exponent					fraction			
b ₉	b ₈	b ₇	b ₆	b ₅	b ₄	b ₃	b ₂	b ₁	b ₀
1	1	0	0	1	0	0	1	0	1
2 ⁵	2 ⁴	2 ³	2 ²	2 ¹	2 ⁰	2 ⁻¹	2 ⁻²	2 ⁻³	2 ⁻⁴

$$\begin{aligned} X &= (+/-)(1 + F) \cdot 2^{E-B} \\ &= (-)(1 + (0.125 + 0.0625)) \cdot 2^{(2+16)-15} \\ &= (-)(1.1875) \cdot 2^3 \\ &= -8.1875 \end{aligned}$$

- Given:
 - length of exponent
 - bias (here: 15)

From Cathy Wyss (I308)

Luis M. Rocha and Santiago Schnell

Hexadecimal

- Base 16

- 16 symbols: 0,1,2,3,4,5,6,7,8,9, A, B, C, D, E, F
- Easy to convert to and from Binary
 - 16 is a power of 2: $16 = 2^4$
 - It takes 4 binary digits for every hexadecimal one
 - Good to represent binary in compressed form!

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

Encoding Text

■ ASCII

- American Standard Code for Information Interchange
 - between binary numbers and computer and roman symbols
 - Standard to allow computers to communicate textual data
- Uses 7 bits to encode 128 symbols or characters
 - $2^7 = 128$. It fills a byte, but the 8th bit is used to encode additional symbols for other languages and graphics
 - Usually described in hexadecimal
- 4 groups of 32 characters
 - 00 to 1F: **control characters**
 - Mostly printer/display operations: *carriage return* (0Dh), *line feed* (0Ah), *back space* (08h), etc.
 - 20 to 3F: punctuation, numeric, and special characters
 - Space (20h), digits 0-9 (30h-39h)
 - Arranged so that by subtracting 30h from the ASCII code for any digit, we obtain the numeric equivalent of the digit
 - 40 to 5F: uppercase letters, plus some special characters
 - 60 to 7F: lowercase letters, plus some special characters and a control character (DEL)

ASCII Table

Dec	Hex	Oct	Chr	Dec	Hex	Oct	Hex	Chr	Dec	Hex	Oct	Hex	Chr	Dec	Hex	Oct	Hex	Chr
0	0	000	NULL (null)	32	20	040	 	Space	64	40	100	@	@	96	60	140	`	`
1	1	001	SOH (start of heading)	33	21	041	!	!	65	41	101	A	A	97	61	141	a	a
2	2	002	STX (start of text)	34	22	042	"	"	66	42	102	B	B	98	62	142	b	b
3	3	003	ETX (end of text)	35	23	043	#	#	67	43	103	C	C	99	63	143	c	c
4	4	004	EOF (end of transmission)	36	24	044	$	\$	68	44	104	D	D	100	64	144	d	d
5	5	005	ENQ (enquiry)	37	25	045	%	%	69	45	105	E	E	101	65	145	e	e
6	6	006	ACK (acknowledge)	38	26	046	&	&	70	46	106	F	F	102	66	146	f	f
7	7	007	BEL (bell)	39	27	047	'	'	71	47	107	G	G	103	67	147	g	g
8	8	010	BS (backspace)	40	28	050	((72	48	110	H	H	104	68	150	h	h
9	9	011	TAB (horizontal tab)	41	29	051))	73	49	111	I	I	105	69	151	i	i
10	A	012	LF (NL line feed, new line)	42	2A	052	*	*	74	4A	112	J	J	106	6A	152	j	j
11	B	013	VT (vertical tab)	43	2B	053	+	+	75	4B	113	K	K	107	6B	153	k	k
12	C	014	FF (NP form feed, new page)	44	2C	054	,	,	76	4C	114	L	L	108	6C	154	l	l
13	D	015	CR (carriage return)	45	2D	055	-	-	77	4D	115	M	M	109	6D	155	m	m
14	E	016	SO (shift out)	46	2E	056	.	.	78	4E	116	N	N	110	6E	156	n	n
15	F	017	SI (shift in)	47	2F	057	/	/	79	4F	117	O	O	111	6F	157	o	o
16	10	020	DL (data link escape)	48	30	060	0	0	80	50	120	P	P	112	70	160	p	p
17	11	021	DC1 (device control 1)	49	31	061	1	1	81	51	121	Q	Q	113	71	161	q	q
18	12	022	DC2 (device control 2)	50	32	062	2	2	82	52	122	R	R	114	72	162	r	r
19	13	023	DC3 (device control 3)	51	33	063	3	3	83	53	123	S	S	115	73	163	s	s
20	14	024	DC4 (device control 4)	52	34	064	4	4	84	54	124	T	T	116	74	164	t	t
21	15	025	NAK (negative acknowledge)	53	35	065	5	5	85	55	125	U	U	117	75	165	u	u
22	16	026	SYN (synchronous idle)	54	36	066	6	6	86	56	126	V	V	118	76	166	v	v
23	17	027	ETE (end of trans. block)	55	37	067	7	7	87	57	127	W	W	119	77	167	w	w
24	18	030	CAN (cancel)	56	38	070	8	8	88	58	130	X	X	120	78	170	x	x
25	19	031	EM (end of medium)	57	39	071	9	9	89	59	131	Y	Y	121	79	171	y	y
26	1A	032	SOE (substitute)	58	3A	072	:	:	90	5A	132	Z	Z	122	7A	172	z	z
27	1B	033	ESC (escape)	59	3B	073	;	;	91	5B	133	[[123	7B	173	{	{
28	1C	034	FS (file separator)	60	3C	074	<	<	92	5C	134	\	\	124	7C	174	|	
29	1D	035	GS (group separator)	61	3D	075	=	=	93	5D	135]]	125	7D	175	}	}
30	1E	036	RS (record separator)	62	3E	076	>	>	94	5E	136	^	^	126	7E	176	~	~
31	1F	037	US (unit separator)	63	3F	077	?	?	95	5F	137	_	_	127	7F	177		DEL

Source: www.asciitable.com

Luis M. Rocha and Santiago Schnell

Extended ASCII Table

128	Ç	144	É	160	£	176	☒	193	⊥	209	⸮	225	ß	241	±
129	à	145	Ê	161	¸	177	☒	194	⌊	210	⌊	226	Γ	242	≥
130	á	146	Ë	162	ó	178	■	195	⌋	211	⌋	227	π	243	≤
131	â	147	Ì	163	ô	179		196	—	212	⌌	228	Σ	244	∫
132	ã	148	Ó	164	õ	180	†	197	‡	213	⌍	229	σ	245	∫
133	ä	149	Ô	165	¸	181	‡	198	‡	214	⌎	230	μ	246	+
134	å	150	Œ	166	•	182	‡	199	‡	215	‡	231	τ	247	∞
135	ç	151	Ù	167	◊	183	‡	200	⌎	216	‡	232	Φ	248	◊
136	è	152	—	168	¸	184	‡	201	⌎	217	‡	233	⊖	249	.
137	é	153	Ö	169	—	185	‡	202	⌎	218	‡	234	⊖	250	.
138	ê	154	Û	170	¬	186	‡	203	⸮	219	■	235	ø	251	√
139	ë	156	£	171	½	187	‡	204	‡	220	■	236	ω	252	—
140	ì	157	¥	172	¼	188	‡	205	—	221	‡	237	φ	253	±
141	í	158	—	173		189	‡	206	‡	222	‡	238	ε	254	■
142	Ā	159	/	174	α	190	‡	207	⌊	223	■	239	∩	255	
143	Ă	192	L	175	∞	191	‡	208	⌌	224	α	240	=		

Source: www.pubblinet.com

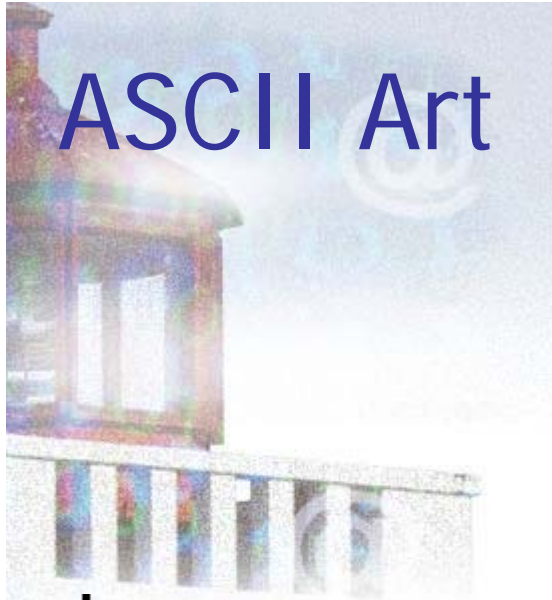
Luis M.Rocha and Santiago Schnell

Alternative Extended ASCII

ASCIIコード表

16進	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	
0	NUL		SP	D	E	F	.	@	—	⊥		—	⌘	⌘	≡	×	0000
1		!	!	!	A	O	a	q	—	T	.	ア	チ	ム	上	円	0080
2		002	!	2	B	R	b	r	■	⊥		イ	ツ	メ	±	年	0040
3			!	3	O	S	o	s	■	⊥		ウ	チ	セ	⊥	月	0010
4		004	!	4	D	T	d	t	■	—	,	エ	ト	セ	▲	日	0400
5			!	5	E	U	e	u	■		.	オ	ナ	ユ	▲	時	0304
6			!	6	F	V	f	v	■		?	カ	ニ	フ	▲	分	0130
7	DEL		!	7	G	W	g	w	■	┘	?	キ	ス	ラ	▲	秒	0111
8	BS	CAN	(8	H	X	h	x		┘	!	ク	ホ	リ	♣	〒	1000
9	HT	EM)	9	I	Y	i	y		┘	?	ケ	ノ	ル	♥	市	1200
A	LF		*	:	J	Z	j	z		┘	!	コ	ハ	レ	♦	区	1310
B	VT	ESC	+	:	K	[k	[■	┘	!	ク	セ	ロ	♣	町	0531
C	FF	FS	-	^	L	W	l	w	■	┘	!	ク	ア	ワ	●	村	1300
D	CR		_	^	M	W	m	w	■	┘	!	ク	ス	ヘ	○	人	1304
E	SO		~	^	N	W	n	w	■	┘	!	ク	セ	ホ	○	人	1210
F	SI		/	^	O	W	o	w	■	┘	!	ク	ソ	マ	○	人	1411
	0000	0004	0008	0011	0400	0453	0470	0470	0000	0001	1000	0073	1000	1101	1110	1411	2000

ASCII Art



I
hi, my honey.
i'm really happy that
my mailbox is full of those
pretty hearts every day. so,
i just thought i would return
the favor, just in case you'd
not yet realized just how i
love you. you are just
so very, very, very
extraordinarily
special and
i adore
you
!



Unicode

- Extends ASCII
 - Much greater support for international characters, glyphs, math symbols, etc.
- Universal Code for Text
 - Each character has a single and unique code in every computer everywhere
 - code point
 - Initially using 16-bits
 - 65536 possible code points
 - Sufficient space to include all the characters for every language on the planet
 - Characters organized into different ranges
 - Greek stored between 880 and 1023 (0x370 and 0x3FF)
 - Accepted by the International Standards Organization (ISO)
 - Version 3.1 in 2001 was expanded to 21-bits
 - over 1 million different code points
 - Logical "planes" contain broad classes of characters

Unicode

<http://www.unicode.org/charts/>

	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F
00	NUL 0000	STX 0001	SOT 0002	ETX 0003	EOT 0004	ENQ 0005	ACK 0006	BEL 0007	BS 0008	HT 0009	LF 000A	VT 000B	FF 000C	CR 000D	SO 000E	SI 000F
10	DLE 0010	DC1 0011	DC2 0012	DC3 0013	DC4 0014	NAK 0015	SYN 0016	ETB 0017	CAN 0018	EM 0019	SUB 001A	ESC 001B	FS 001C	GS 001D	RS 001E	US 001F
20	SP 0020	!	"	#	\$	%	&	'	()	*	+	,	-	.	/
30	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
40	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
50	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
60	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
70	p	q	r	s	t	u	v	w	x	y	z	{		}	~	DEL 007F
80	€ 20AC		/		//	...	†	‡		‰	Š	<	Ś	Ť	Ž	Ž
90		\	/	\"	//	▪	-	-		™	š	>	ś	ť	ž	ž
A0	NBSP 00A0	ˆ	ˇ	Ł	ł	À	Á	Â	Ã	Ä	Å	«	»	-	®	Ž
B0	°	±	²	³	´	µ	¶	·	¸	¸	¸	»	Ł	˜	ı	ž
C0	Ř	Á	Â	Ă	Ä	Í	Ć	Ç	Č	É	Ě	Ë	Ě	Í	Î	Ď
D0	Đ	Ń	Ň	Ó	Ô	Õ	Ö	×	Ř	Ů	Ú	Û	Ü	Ý	Ť	ß
E0	í	á	â	ă	ä	í	ć	ç	č	é	ě	ë	ë	í	î	ď
F0	đ	ń	ň	ó	ô	õ	ö	÷	ř	ů	ú	û	ü	ý	ț	·

Markup Languages

- In addition to the symbols, specify formatting, hyperlinks, images, media, etc.
 - represents text as well as details about the structure and appearance of the text
 - SGML: Standard Generalized Markup Language
 - Specifies a syntax for including the markup in documents, as well as a description of what the markup meant
 - HTML: HyperText Markup Language
 - Does not require a definition of what the markup means
 - XML: Extensible Markup Language
 - Allows the creation of special-purpose markup languages
 - Simplified subset of SGML, also requiring a definition of what the Markup means
 - Can describing many different kinds of data.
 - LaTeX:
 - best way to typeset complex mathematical formulas



More about data representation

- **I308: Information Representation**

- **C.M. Wyss**

- The basic structure of information representation in social and scientific applications.
- Information access and representation on the World Wide Web; object-oriented design and relational databases; AI knowledge representation and discovery.

Next Class!

- Topics
 - Encoding Multimedia
- Readings for Next week
 - Lecture notes Posted online @ <http://informatics.indiana.edu/rocha/i101>
 - *Modeling the World*
 - @ *infoport*
 - Read Binary encoding resources at Infoport!!
 - From course package
 - From Irv Englander's book "*The Architecture of Computer Hardware and Systems Software*"
 - Chapter 3: Data Formats (pp. 70-86)
- Lab 6
 - More Unix and FTP