

---

---

# Programming data for display, the PDF Story

Papers We Love Conf 2017  
Chas Emerick — @cemerick  
PDFDATA.io

---

---

# Prologue

On the day of PWLConf 2017 (September 28th), one of the scheduled speakers fell ill. The PWL organizers (Zeeshan Lakhani & Darren Newton) asked me to step in approximately 1h30m prior to the newly-open timeslot.

I do plan to complete and refine these materials for future presentations...but aside from this prologue, this deck is as it was that afternoon.

Darren, Zeeshan: Thanks ❤️

---

Input

Computation /  
Communication

Output

Input

Computation /  
Communication

Display

---

# Display via...

- Telegraph / teletypes
  - Dynamic displays
    - Oscilloscope/CRT/vector
    - Raster displays (LCDs)
  - Printers
    - Fax machines
    - Laser printers
    - Offset printers
  - CAD/CAM
-

---

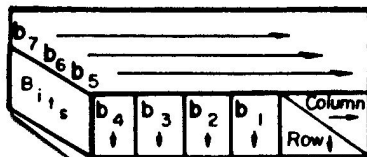
# “Page Description Language”

*Any characterization of the layout and contents of a page (and collectively, a document) that is more efficient and/or expressive than a visually-equivalent bitmap of that page.*

- DVI (via LaTeX)
  - Interpress
  - PostScript
  - PDF
  - ...dozens more...
-

# ASCII, the ~first page description language

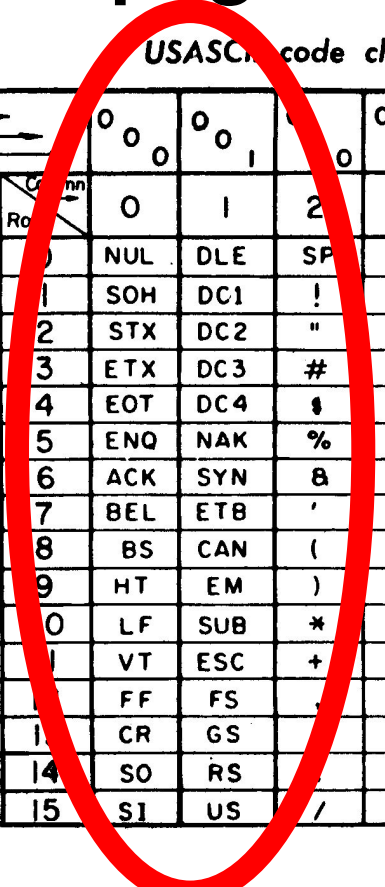
USASCII code chart



Column	Row	b <sub>7</sub>	b <sub>6</sub>	b <sub>5</sub>	b <sub>4</sub>	b <sub>3</sub>	b <sub>2</sub>	b <sub>1</sub>	b <sub>0</sub>	Character
0	0	0	0	0	0	0	0	0	0	NUL
0	1	0	0	0	0	0	0	0	1	DLE
0	2	0	0	0	0	0	0	1	0	SP
0	3	0	0	0	0	0	0	1	1	0
0	4	0	0	0	0	0	1	0	0	@
0	5	0	0	0	0	0	1	0	1	P
0	6	0	0	0	0	0	1	1	0	\
0	7	0	0	0	0	0	1	1	1	p
0	8	0	0	0	0	1	0	0	0	SOH
0	9	0	0	0	0	1	0	0	1	DC1
0	10	0	0	0	0	1	0	1	0	!
0	11	0	0	0	0	1	0	1	1	1
0	12	0	0	0	0	1	1	0	0	A
0	13	0	0	0	0	1	1	0	1	Q
0	14	0	0	0	0	1	1	1	0	a
0	15	0	0	0	0	1	1	1	1	q
0	16	0	0	0	1	0	0	0	0	STX
0	17	0	0	0	1	0	0	0	1	DC2
0	18	0	0	0	1	0	0	1	0	"
0	19	0	0	0	1	0	0	1	1	2
0	20	0	0	0	1	0	1	0	0	B
0	21	0	0	0	1	0	1	0	1	R
0	22	0	0	0	1	0	1	1	0	b
0	23	0	0	0	1	0	1	1	1	r
0	24	0	0	0	1	1	0	0	0	ETX
0	25	0	0	0	1	1	0	0	1	DC3
0	26	0	0	0	1	1	0	0	1	#
0	27	0	0	0	1	1	0	1	0	3
0	28	0	0	0	1	1	0	1	1	C
0	29	0	0	0	1	1	1	0	0	S
0	30	0	0	0	1	1	1	0	1	c
0	31	0	0	0	1	1	1	1	0	s
0	32	0	0	0	1	1	1	1	1	4
0	33	0	0	1	0	0	0	0	0	EOT
0	34	0	0	1	0	0	0	0	1	DC4
0	35	0	0	1	0	0	0	0	1	5
0	36	0	0	1	0	0	0	1	0	D
0	37	0	0	1	0	0	0	1	1	T
0	38	0	0	1	0	0	1	0	0	d
0	39	0	0	1	0	0	1	0	1	t
0	40	0	0	1	0	0	1	1	0	%
0	41	0	0	1	0	0	1	1	1	5
0	42	0	0	1	0	1	0	0	0	E
0	43	0	0	1	0	1	0	0	1	U
0	44	0	0	1	0	1	0	1	0	e
0	45	0	0	1	0	1	0	1	1	u
0	46	0	0	1	0	1	1	0	0	ACK
0	47	0	0	1	0	1	1	0	1	SYN
0	48	0	0	1	0	1	1	1	0	8
0	49	0	0	1	0	1	1	1	1	6
0	50	0	0	1	0	1	1	1	1	F
0	51	0	0	1	1	0	0	0	0	V
0	52	0	0	1	1	0	0	0	1	v
0	53	0	0	1	1	0	0	0	1	7
0	54	0	0	1	1	0	0	1	0	G
0	55	0	0	1	1	0	0	1	1	W
0	56	0	0	1	1	0	1	0	0	g
0	57	0	0	1	1	0	1	0	1	w
0	58	0	0	1	1	0	1	1	0	(
0	59	0	0	1	1	0	1	1	1	8
0	60	0	0	1	1	1	0	0	0	H
0	61	0	0	1	1	1	0	0	1	X
0	62	0	0	1	1	1	0	0	1	h
0	63	0	0	1	1	1	0	1	0	x
0	64	0	0	1	1	1	0	1	1	)
0	65	0	0	1	1	1	1	0	0	9
0	66	0	0	1	1	1	1	0	1	I
0	67	0	0	1	1	1	1	0	1	Y
0	68	0	0	1	1	1	1	1	0	i
0	69	0	0	1	1	1	1	1	1	y
0	70	0	1	0	0	0	0	0	0	LF
0	71	0	1	0	0	0	0	0	1	SUB
0	72	0	1	0	0	0	0	0	1	*
0	73	0	1	0	0	0	0	1	0	:
0	74	0	1	0	0	0	0	1	1	J
0	75	0	1	0	0	0	1	0	0	Z
0	76	0	1	0	0	0	1	0	1	j
0	77	0	1	0	0	0	1	1	0	z
0	78	0	1	0	0	1	0	0	0	VT
0	79	0	1	0	0	1	0	0	1	ESC
0	80	0	1	0	0	1	0	0	1	+
0	81	0	1	0	0	1	0	1	0	;
0	82	0	1	0	0	1	0	1	1	K
0	83	0	1	0	0	1	1	0	0	[
0	84	0	1	0	0	1	1	0	1	k
0	85	0	1	0	0	1	1	1	0	{
0	86	0	1	0	1	0	0	0	0	FF
0	87	0	1	0	1	0	0	0	1	FS
0	88	0	1	0	1	0	0	0	1	,
0	89	0	1	0	1	0	0	1	0	<
0	90	0	1	0	1	0	0	1	1	L
0	91	0	1	0	1	0	1	0	0	\
0	92	0	1	0	1	0	1	0	1	l
0	93	0	1	0	1	0	1	1	0	
0	94	0	1	0	1	1	0	0	0	-
0	95	0	1	0	1	1	0	0	1	=
0	96	0	1	0	1	1	0	0	1	M
0	97	0	1	0	1	1	0	1	0	J
0	98	0	1	0	1	1	0	1	1	m
0	99	0	1	0	1	1	1	0	0	}
0	100	0	1	0	1	1	1	0	1	.
0	101	0	1	0	1	1	1	0	1	>
0	102	0	1	0	1	1	1	1	0	N
0	103	0	1	0	1	1	1	1	1	^
0	104	0	1	1	0	0	0	0	0	SO
0	105	0	1	1	0	0	0	0	1	RS
0	106	0	1	1	0	0	0	0	1	~
0	107	0	1	1	0	0	0	1	0	n
0	108	0	1	1	0	0	0	1	1	~
0	109	0	1	1	0	0	1	0	0	/
0	110	0	1	1	0	0	1	0	1	?
0	111	0	1	1	0	0	1	1	0	O
0	112	0	1	1	0	0	1	1	1	—
0	113	0	1	1	0	1	0	0	0	o
0	114	0	1	1	0	1	0	0	1	DEL

# ASCII, the ~first page description language

USASCII code chart



Bits				Row	Column													
b <sub>7</sub>	b <sub>6</sub>	b <sub>5</sub>	b <sub>4</sub>	b <sub>3</sub>	b <sub>2</sub>	b <sub>1</sub>	b <sub>0</sub>		0	1	2	3	4	5	6	7		
0	0	0	0	0	0	0	0	0	NUL	DLE	SP	0	@	P	\	p		
0	0	0	0	0	0	0	1	1	SOH	DC1	!	1	A	Q	a	q		
0	0	0	0	0	1	0	0	2	STX	DC2	"	2	B	R	b	r		
0	0	0	0	0	1	0	1	3	ETX	DC3	#	3	C	S	c	s		
0	0	0	0	1	0	0	0	4	EOT	DC4	\$	4	D	T	d	t		
0	0	0	0	1	0	0	1	5	ENQ	NAK	%	5	E	U	e	u		
0	0	0	0	1	0	1	0	6	ACK	SYN	&	6	F	V	f	v		
0	0	0	0	1	0	1	1	7	BEL	ETB	'	7	G	W	g	w		
0	0	0	1	0	0	0	0	8	BS	CAN	(	8	H	X	h	x		
0	0	0	1	0	0	0	1	9	HT	EM	)	9	I	Y	i	y		
0	0	0	1	0	0	1	0	10	LF	SUB	*	:	J	Z	j	z		
0	0	0	1	0	0	1	1	11	VT	ESC	+	;	K	[	k	{		
0	0	0	1	0	1	0	0	12	FF	FS	,	<	L	\	l			
0	0	0	1	0	1	0	1	13	CR	GS	-	=	M	]	m	}		
0	0	0	1	0	1	1	0	14	SO	RS	.	>	N	^	n	~		
0	0	0	1	0	1	1	1	15	SI	US	/	?	O	_	o	DEL		



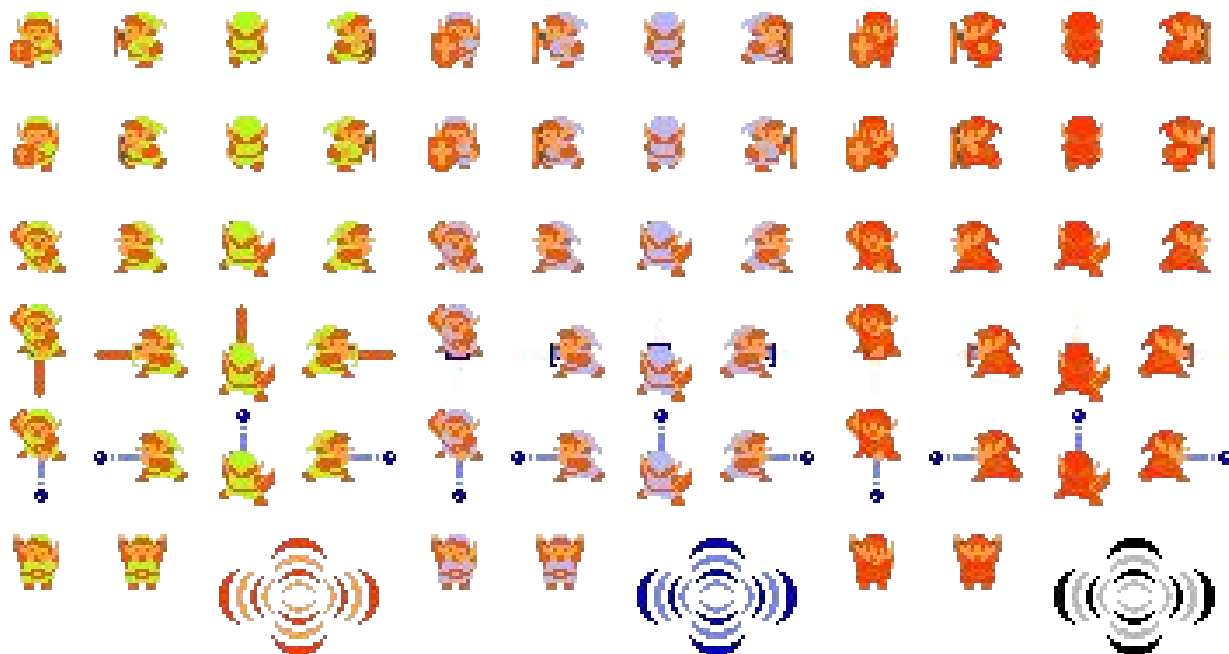
Table 1. Rank order of lessons according to mean probability of a correct response.

Rank	Lesson	Type	Predominant Form	Number of Subjects	Proportion of Successes
1	2	Mixed	$(a \ o \ b) \ o' \ c = d \times \_$	22	.54
2	4	Mixed	$(a \ o \ b) \ o' \ c = \_$	39	.58
{ 3	1	Mixed	$(a \ o \ b) \ o' \ c = \_$	23	.66
{ 3	3	Units of Measure		19	.66
{ 3	15	Units of Measure		39	.66
6	8	Mixed		36	.69
7	13	Mixed	$a \times b = \_$	33	.73
{ 8	11	Word		39	.78
{ 8	14	Mixed		38	.78
10	12	Mixed	$a \times b = \_$	36	.80
11	7	Multn	$a \times b = \_, a = 8, 9, 10$	39	.84
12	5	Multn	$a \times b = \_, a = 8, 9, 10$	40	.85
13	6	Multn	$a \times b = \_, a = 10, 100$	24	.90
14	10	Multn	$a \times b = \_, a = 10, 100$	39	.94
15	9	Multn	$a \times b = \_, a = 10, 100$	39	.98

# Vector Graphics



# Raster Images



# Text & Typography

A B C Č Ď Đ E F G H I J  
 K L M N O P Q R S Š T U  
 V W X Y Z Ž a b c č d e f g  
 h i j k l m n o p q r s š t u v w x y  
 z ž Ā Â Ê Ô Ő Ů ě â ê ô σ υ 1 2  
 3 4 5 6 7 8 9 0 ' ? ' " ! " ( % ) [   
 # ] { @ } / & \ < - + ÷ × = >   
 ® © \$ € £ ¥ ¢ : ; , . \*

---

# How PDF came to be

- Evans & Sutherland
  - Chuck Geschke and John Warnock => Xerox
  - Interpress @ Xerox
  - Chuck Geschke and John Warnock => Adobe\*
  - PostScript @ Adobe
  - PDF @ Adobe
-

---

From: Brian Reid <reid@Glacier>

This essay offers a comparison of two modern schemes for controlling what laser printers print. One scheme, called PostScript, is offered by Adobe Systems, Inc.; the other scheme, called Interpress, is offered by the Xerox Corporation. A discussion of these two schemes has provoked a considerable amount of interest in this forum recently. I have for some time been promising (threatening?) to provide my interpretation of the difference between the two systems. It is long enough and detailed enough that you will certainly never want to read another word on the topic after you read it, but given the nature of computer mail systems you almost certainly will be given the opportunity.

---

To a first order, PostScript and Interpress are indistinguishable. What I mean by that is that by comparison with all other current techniques for page image representation, the two can be considered to be nearly identical. I believe that it is worth looking at how they got to be that way; their similarities and differences can best be understood with a proper historical perspective.

Reid, Brian. *PostScript and Interpress: a comparison*. Newsgroup posting, March 1985. <http://bit.ly/2hzPHoK>

---

---

## The Camelot Project J. Warnock

This document describes the base technology and ideas behind the project named “Camelot.” This project’s goal is to solve a fundamental problem that confronts today’s companies. The problem is concerned with our ability to communicate visual material between different computer applications and systems. The specific problem is that most programs print to a wide range of printers, but there is no universal way to communicate and view this printed information electronically. The popularity of FAX machines has given us a way to send images around to produce remote paper, but the lack of quality, the high communication bandwidth and the device specific nature of FAX has made the solution less than desirable. What industries badly need is a universal way to communicate documents across a wide variety of machine configurations, operating systems and communication networks. These documents should be viewable on any display and should be printable on any modern printers. If this problem can be solved, then the fundamental way people work will change.

---

---

# PostScript

Stack-based interpreted language

Embedded bitmaps, referencing fonts externally

Interpreters pushed to the “edge”, i.e. printers, graphics display drivers on workstations

---



---

---

# PDF

(vamping)

---