Fonts & Encoding

A Never Ending Problem?

Volker RW Schaa

GSI Helmholtzzentrum für Schwerionenforschung GmbH Darmstadt, Germany

JACoW Team Meeting



1

History of fonts and encodings

- In pre-computer time
- What changed in the computer era?
- New standards
- Problem of interpretation

Ponts in print (encoding and subsetting)

- Encoding and naming
- Font character maps
- Font problems part 1 Font not embedded
- Font problems part 2 Font embedded but characters missing

3 Conclusion

- What can we do?
- Thanks

In pre-computer time What changed in the computer era? New standards Problem of interpretation

In pre-computer times fonts never were a problem

Characters were local and known

Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

In pre-computer times fonts never were a problem



In pre-computer time What changed in the computer era? New standards Problem of interpretation

In pre-computer times fonts never were a problem

Characters were local and known

You needed foreign characters — you took them

Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

In pre-computer times fonts never were a problem

Characters were local and known

You needed foreign characters — you took them



Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

In pre-computer times fonts never were a problem

Characters were local and known

You needed foreign characters — you took them





In pre-computer time What changed in the computer era? New standards Problem of interpretation

In pre-computer times fonts never were a problem



Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

In pre-computer times fonts never were a problem



Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

In pre-computer times fonts never were a problem

Characters were placed in the typesetter's tray by frequency of occurrence

German, Swiss and other West-European languages

Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

In pre-computer times fonts never were a problem

Characters were placed in the typesetter's tray by frequency of occurrence

German, Swiss and other West-European languages

Г	A	8	3	С		D			Ξ	F	F		3	ŀ	ł		I	К	
Γ	L	1	4	N		0		Р		Q		R		S			Г	U	
1	2	3	4	5	6	7	8	9	0	-	J	1	/	۷	V	х	Υ	Z	&
à	â	á	Ä		ß			á	ä	č	ö		ü		*	"	"	+	§
é	ê	è	ë										у	z	j	()	[]	1	?
í	î	ì	ï		5	ť		u				`	/	v	v		•	:	;
ó	ô	ò	Ö	ł	١				i						q		•		
ú	û	ù	Ü	Т		n	n		_	ľ	1	0		р		,		1	
Ææ	ÊÉ		k	_{ck} C							,		4		fi	fl	ft		
Œ	ç	С	h				1			е		u		f	ff	g			

used the following distribution

Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Computers introduced problems: Early days

Introduction of ASCII in 1963 (7 Bit, 94 printable characters, 33 control codes, invisible space)

Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Computers introduced problems: Early days

Introduction of ASCII in 1963 (7 Bit, 94 printable characters, 33 control codes, invisible space)

USASCII code chart

B7 D6 D	5 -				· ·	° ° ₀	°°,	° ' o	° , ,	' ° ₀	'°,	' ' ₀	' i ₁
	b 4	b 3 1	b 2 1	ь' 1	Row	0	1	2	3	4	5	6	7
	0	0	0	0	0	NUL .	DLE	SP	0	0	Р	`	P
	00011					SOH	DC1	!	1	Α.	Q	a	q
	0	0	1	0	2	STX	DC2		2	B	R	b	r
	0	0	1	T	3	ETX	DC 3	#	3	C	S	c	5
	0	1	0	0	4	EOT	DC4	1	4	D	т	d	1
	0	0 1 0 1 5				ENQ	NAK	%	5	E	υ	e	U
	0					ACK	SYN	8	6	F	v	f	v
	0	T	1	1	7	BEL	ETB	•	7	G	w	9	w
	1	0	0	0	8	BS	CAN	(8	н	×	h	x
	T	0	0	T.	9	нT	EM)	9	1	Y	i	У
	T I	0	1	0	10	LF	SUB	*	:	J	Z	j	z
	1	0	1	1	11	VT	ESC	+	:	ĸ	C	k	(
	1 1 0 0 12				12	FF	FS		<	L	N	1	1
	1 1 0 1 13				13	CR	GS	-	×	м	C	m	}
	1 1 1 0 14				14	SO	RS		>	N	^	n	\sim
	I	1	L	1	15	S 1	US	1	?	0		0	DEL

Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

- Introduction of ASCII in 1963 (7 Bit, 94 printable characters, 33 control codes, invisible space)
- Introduction of EBCDIC by IBM in 1963 (8 Bit, 94 printable characters, 65 control codes, 3 invisible spaces, softhyphen)

					H in pri	istory nt (er	of fo	nts a ng ar	nd er	ncodii bsetti	ngs ng)		In pre-computer time What changed in the computer era? New standards											
									00	UTF-EB	CDIC		Drahl	000 0	finto	rorote	otion							
			-0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-A	-в	-c	-D	-2	-F						
Com	puters int	0-	NUL 0000 0	SOH 0001 1	STX 0002 2	ETX 0003 3	ST 0090 4	HT 0009 5	SSA 0086 6	DEL 007r 7	EPA 0097 8	RI 0080 9	SS2 008m 10	VT 000m 11	FF 0000 12	CR 000D 13	SO 00011 14	SI 000r 15						
	Intro	1-	DLE 0010 16	DC1 0011 17	DC2 0012 18	DC3 0013 19	0090 20	LF 000A 21	BS 0008 22	ESA 0087 23	CAN 0018 24	EM 0019 25	PU2 0092 26	SS3 008F 27	FS 0010 28	GS 001D 29	RS 001E 30	US 001F 31	characters 33					
15'\1 C	cont	2-	PAD 0080 32	HOP 0081 33	BPH 0082 34	NBH 0083 35	IND 0084 36	NEL 0085 37	ETB 0017 38	ESC 0018 39	HTS 0088 40	HTJ 0089 41	VTS 008A 42	PLD 0088 43	PLU 0080 44	ENQ 0005 45	ACK 0006 46	BEL 0007 47	characters, oo					
J. Ch	 Intro 	3-	DCS 0090 48	PU1 0091 49	SYN 0016 50	STS 0093 51	CCH 0094 52	MW 0095 53	SPA 0096 54	EOT 0004 55	SOS 0098 56	SGCI 0099 57	SCI 009A 58	CSI 0098 59	DC4 0014 60	NAK 0015 61	PM 0095 62	SUB 001A 63	4 printable					
icon .	char	{ ≁-	SP 0020 64											002m 75	< 00.3c 76	(0028 77	+ 0028 78	007c 79	softhyphen)					
F===>.	Let 1	5-	& 0026 80										! 0021 90	\$ 0024 91	* 002A 92) 0029 93	003m 94	005m 95	247. 500					
	×~~~	6-	0020 96	0025 97										002c 107	% 0025 108	005r 109	> 003E 110	? 003r 111	in the second					
120-1	8.01	7-										0060 121	: 003A 122	# 0023 123	@ 0040 124	0027 125	= 003D 126	0022 127	The start					
1 CM		8-		a 0061 129	b 0062 130	C 00.63 131	d 0064 132	e 0065 133	f 0066 134	g 0067 135	h 0068 136	i 0069 137							Let & Cli					
Green		9-		ј 006а 1 45	k 0068 146	1 0060 147	m 006D 148	n 006m 149	0 006r 150	p 0070 151	q 0071 152	r 0072 153							AL Some					
	6	٨-		~ 007m 161	S 0073 162	t 0074 163	U 0075 164	V 0076 165	W 0077 166	X 0078 167	У 0079 168	Z 007A 169				[005m 173			SUL ES					
(r) 🛋	No.	в-] 0050 189								
1. 16	80	c-	{ 007m 192	A 0041 193	B 0042 194	C 0043 195	D 0044 196	E 0045 197	F 0046 198	G 0047 199	H 0048 200	I 0049 201							- 23					
F CIN	2.11	D-) 0070 208	J 004A 209	K 004n 210	L 004c 211	M 004 D 212	N 004m 213	0 004r 214	P 0050 215	Q 0051 216	R 0052 217							AN FOR					
1		E-	005c 224		S 0053 226	T 0054 227	U 0055 228	V 0056 229	W 0057 230	X 0058 231	Y 0059 232	Z 005a 233							90 × 2					
re-cor	TP	F-	0 0030 240	1 0031 241	2 0032 242	3 0033 243	4 0034 244	5 0035 245	б 0036 246	7 0037 247	8 0038 248	9 0039 249						APC 0097 255	SUL E					
	144		-0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-A	-в	-c	-D	-2	-9	r.					
Con Cont		1		CH AS	5	6						S. Land			Cast		12.300		Time Time					

Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

- Introduction of ASCII in 1963 (7 Bit, 94 printable characters, 33 control codes, invisible space)
- Introduction of EBCDIC by IBM in 1963 (8 Bit, 94 printable characters, 65 control codes, 3 invisible spaces, softhyphen)
 From the Jargon book:

In pre-computer time What changed in the computer era? New standards Problem of interpretation

- Introduction of ASCII in 1963 (7 Bit, 94 printable characters, 33 control codes, invisible space)
- Introduction of EBCDIC by IBM in 1963 (8 Bit, 94 printable characters, 65 control codes, 3 invisible spaces, softhyphen)
- From the Jargon book: Today, IBM claims to be an open-systems company, but IBM's own description of the EBCDIC variants and how to convert between them is still internally classified top-secret, burn-before-reading.

In pre-computer time What changed in the computer era? New standards Problem of interpretation

- Introduction of ASCII in 1963 (7 Bit, 94 printable characters, 33 control codes, invisible space)
- Introduction of EBCDIC by IBM in 1963 (8 Bit, 94 printable characters, 65 control codes, 3 invisible spaces, softhyphen)
- A popular joke:

In pre-computer time What changed in the computer era? New standards Problem of interpretation

- Introduction of ASCII in 1963 (7 Bit, 94 printable characters, 33 control codes, invisible space)
- Introduction of EBCDIC by IBM in 1963 (8 Bit, 94 printable characters, 65 control codes, 3 invisible spaces, softhyphen)
- A popular joke: Professor: "So the American government went to IBM to come up with an encryption standard, and they came up with—"

In pre-computer time What changed in the computer era? New standards Problem of interpretation

- Introduction of ASCII in 1963 (7 Bit, 94 printable characters, 33 control codes, invisible space)
- Introduction of EBCDIC by IBM in 1963 (8 Bit, 94 printable characters, 65 control codes, 3 invisible spaces, softhyphen)
- A popular joke: Professor: "So the American government went to IBM to come up with an encryption standard, and they came up with—" Student: "EBCDIC!"

In pre-computer time What changed in the computer era? New standards Problem of interpretation

- Introduction of ASCII in 1963 (7 Bit, 94 printable characters, 33 control codes, invisible space)
- control codes, invisible space)
 Introduction of EBCDIC by IBM in 1963 (8 Bit, 94 printable characters, 65 control codes, 3 invisible spaces, softhyphen)
- Both widely spread 'standards' were implemented for US-English only

In pre-computer time What changed in the computer era? New standards Problem of interpretation

- Introduction of ASCII in 1963 (7 Bit, 94 printable characters, 33 control codes, invisible space)
- Introduction of EBCDIC by IBM in 1963 (8 Bit, 94 printable characters, 65 control codes, 3 invisible spaces, softhyphen)
- Both widely spread 'standards' were implemented for US-English only
- Even ASCII wasn't usable for British-English because the '£' was missing, Canada used an own version which supported French characters

In pre-computer time What changed in the computer era? New standards Problem of interpretation

- Introduction of ASCII in 1963 (7 Bit, 94 printable characters, 33 control codes, invisible space)
- Introduction of EBCDIC by IBM in 1963 (8 Bit, 94 printable characters, 65 control codes, 3 invisible spaces, softhyphen)
- Both widely spread 'standards' were implemented for US-English only
- Even ASCII wasn't usable for British-English because the '£' was missing, Canada used an own version which supported French characters
- The later standard ISO/IEC 646, like ASCII, was a 7-bit character set. It did not make any additional codes available, so the same code points encoded different characters in different countries.

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Computers introduced problems: Early days

- Introduction of ASCII in 1963 (7 Bit, 94 printable characters, 33 control codes, invisible space)
- Introduction of EBCDIC by IBM in 1963 (8 Bit, 94 printable characters, 65 control codes, 3 invisible spaces, softhyphen)
- Both widely spread 'standards' were implemented for US-English only
- Even ASCII wasn't usable for British-English because the '£' was missing, Canada used an own version which supported French characters
- The later standard ISO/IEC 646, like ASCII, was a 7-bit character set. It did not make any additional codes available, so the same code points encoded different characters in different countries.
- a German, French, or Swedish, etc., programmer had to get used to reading and writing

a aÄiÜ='Ön'; ü
instead of
{ a[i]='\n'; }

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Computers introduced problems: Later times

 In April 1984 an eight-bit standard was adopted as ECMA-94 (later as ISO/IEC 8859). This derived from the character sets DEC-MCS and Mac OS Roman, developed as true extensions of ASCII. It leaves the original character-mapping intact, and adds additional character definitions after the first 128 (i.e., 7 bit) characters.

History of fonts and encodings

Fonts in print (encoding and subsetting) Conclusion In pre-computer time What changed in the computer era? New standards Problem of interpretation

Com	puters int					b	0	0	0	0	0	0 1	0	0	1	1	1	1	1	1	1	1		
						b	Ō	0	1	1	0	0	1	1	Ŭ 0	0	1	1	0	0	1	1		
	 In Apple 		l.	6.		D	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	CMA-94	
G	(late	0	0,	0,	0	00			SP	0	ລ	Ρ	•	р			NBSP	0	À	Ð	à	ð	ter sets	
613	DEC	0	0	0	1	01			!	1	A	Q	а	q			i	±	Á	Ñ	á	ñ	tensions of	
in in	ASC	0	0	1	0	02			"	2	в	R	b	r			¢	2	Â	ò	â	ò	and adds	
	addi	0	0	1	1	03			#	3	С	S	С	s			£	3	Ã	Ó	ã	ó	7 bit)	
	char	0	1	0	0	04			\$	4	D	Т	d	t			¤	'	Ä	Ô	ä	ô) (F 🍙 🖌	
	20	0	1	0	1	05			%	5	Е	U	е	u			¥	μ	Å	õ	a	õ	1 10 10	
	4 1 1	0	1	1	0	06			&	6	F	۷	f	v			1	٩	Æ	ö	æ	ö	Co I	
621		0	1	1	1	07			'	7	G	W	g	W			§	•	Ç	×	ç	÷		
con		1	0	0	0	08			(8	Н	Х	h	х					È	Ø	è	ø	Tren -	
		1	0	0	1(09)	9	Ι	Y	i	у			©	1	É	Ù	é	ù	to the south	
		1	0	1	0	10			*	:	J	Z	j	z			ā	<u>0</u>	Ê	Ú	ê	ú		
	66	1	0	1	1	11			+	;	К	E	k	{			*	≫	E	Û	ë	û		
<u>9</u> 1 - 1		1	1	0	0	12			'	<	L	\	ι					1/4	Ì	Ú	i	ů		
G/A		4	1	0	1	13			-	=	Μ.	1	m	}			SHY	1/2	I	Y	1	y L	1 01 1	
		Ĥ	1	1	0	14			•	>	N		n				(19)	·/4	I	P 0	î	P Ü	E E E	
		1	1	1	1	15			/	?	0	-	0					6	I	В	1	У		
					2					1	10.			F					13	101.			A CONTRACT	

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Computers introduced problems: Later times

- In April 1984 an eight-bit standard was adopted as ECMA-94 (later as ISO/IEC 8859). This derived from the character sets DEC-MCS and Mac OS Roman, developed as true extensions of ASCII. It leaves the original character-mapping intact, and adds additional character definitions after the first 128 (i.e., 7 bit) characters.
- With ECMA-94 (ISO/IEC 8859-1/-2/-3/-4) we are now able to write and encode the following European languages: Albanian, Catalan, Czech, Danish, Dutch, English, Estonian, Faeroese, Finnish, French, Galician, German, Greenlandic, Hungarian, Icelandic, Irish, Italian, Lappish, Latvian, Lithuanian, Maltese, Norwegian, Polish, Portuguese, Rumanian, Serbo-croatian, Slovak, Slovene, Spanish, Swedish, and Turkish (plus Afrikaans and Esperanto).

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Computers introduced problems: Later times

- In April 1984 an eight-bit standard was adopted as ECMA-94 (later as ISO/IEC 8859). This derived from the character sets DEC-MCS and Mac OS Roman, developed as true extensions of ASCII. It leaves the original character-mapping intact, and adds additional character definitions after the first 128 (i.e., 7 bit) characters.
- With ECMA-94 (ISO/IEC 8859-1/-2/-3/-4) we are now able to write and encode the following European languages: Albanian, Catalan, Czech, Danish, Dutch, English, Estonian, Faeroese, Finnish, French, Galician, German, Greenlandic, Hungarian, Icelandic, Irish, Italian, Lappish, Latvian, Lithuanian, Maltese, Norwegian, Polish, Portuguese, Rumanian, Serbo-croatian, Slovak, Slovene, Spanish, Swedish, and Turkish (plus Afrikaans and Esperanto).
- So now we have ways of defining different characters in character sets, but...

In pre-computer time What changed in the computer era? New standards Problem of interpretation

New standards: Problems solved?

 Now imagine, somebody types (or even better cuts/pastes) an 'Abstract' text into the submission form of SPMS for the conference where you are the editor.

Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

- Now imagine, somebody types (or even better cuts/pastes) an 'Abstract' text into the submission form of SPMS for the conference where you are the editor.
- Now we looking at the input text as the computer does. What do we (or the computer) see?

In pre-computer time What changed in the computer era? New standards Problem of interpretation

New standards: Problems solved?

	6.6	fe	ff	20	20	20	04	1f	04	40	04	3e	04	34	04	30	04	35	04	42	04		
1	Now	41	04	4f	20	004	37	04	35	04	3c	04	3b	04	4£	20	04	41	04	35	04	an	
G	7	3b	04	4c	04	45	04	3e	04	37	20	004	3d	04	30	04	37	04	3d	04	30		
	'Abs	04	47	04	35	04	3d	04	38	04	4f	20	20	20	20	04	3e	04	42	20	04		
	conf	41	04	3e	04	31	04	41	04	42	04	32	04	35	04	3d	04	3d	04	38	04		
cri l	COIII	3a	04	30	20	04	21	04	20	04	1e	04	27	04	1d	04	1e	0a	0c	0a	0c	200	177
	 Now 	04	23	04	47	04	30	04	41	04	42	04	3a	04	38	3a	32	34	04	33	04	at do	1.15
		30	02	62	20	37	33	32	30	04	33	04	30	2b	31	32	39	30	04	33	04		
	we (30	2b	31	33	36	31	04	33	04	30	04	26	04	35	04	3d	04	30	3a	38		
		30	30	30	04	40	04	43	04	31	04	3b	04	35	04	39	20	04	37	04	30		E. See
8 J. 1	S. (1)	20	04	33	04	30	20	32	38	38	30	20	04	40	04	43	04	31	04	Зb	04	A 1991	5.
	i I	35	04	39	2f	04	41	04	3e	04	42	04	3a	04	30	29	04	22	04	1e	04		
Chi		20	04	13	0a	0c	0a	0c	04	23	04	47	04	30	04	41	04	42	04	3a	04	CI	
		38	20	04	32	04	4b	04	34	04	35	04	3b	04	35	04	3d	04	4b	2b	20		Congret
		04	41	04	32	04	38	04	34	04	35	04	42	04	35	04	3b	04	4c	04	41		
	12	04	42	04	32	04	30	20	04	23	04	24	04	20	04	21	20	04	Зf	04	3e		17
		04	3b	04	43	04	47	04	35	04	3d	04	4b	1b	20	04	20	04	30	04	41		
6		04	3f	04	3e	04	Зb	04	3e	04	36	04	35	04	3d	04	38	04	35	3a	20	C	
	Second Second	04	23	04	3b	04	4c	04	4f	04	3d	04	3e	04	32	04	41	04	3a	04	30		1 Part
A 188	0 0	04	4f	20	04	3e	04	31	04	3b	04	30	04	41	04	42	04	4c	2b	04	18	A TELE	
- 1		04	3d	04	37	04	35	04	3d	04	41	04	3a	04	38	04	39	20	04	40	04	1	
CI.	1.	30	04	39	04	3e	04	3d	1b	0a	0c	04	14	04	3e	04	3a	04	43	04	3c	Gr.	1.
12		04	35	04	3d	04	42	04	4b	20	04	Зf	04	3e	20	04	37	04	30	04	3f	12	
		04	40	04	3e	04	41	04	43	1e	20	0a	0c	04	1a	04	3e	04	3d	04	42		
11	1000	04	30	04	3a	04	42	04	4b	3a	20	20	2b	20	37	20	32	38	20	39	49	242	E D
		36	32	39	20	39	39	04	27	2d	04	17	20	36	0a	0c	2d	36	31	2b	20		
12.01	그런지	38	30	35	32	35	31	31	32	38	30	61	32	39	20	67	6d	61	69	6b	2e		
	12 come	63	6f	6d	0a	0c	04	1c	04	30	04	3a	04	41	04	38	04	3c	20	0a	0c		- Ale

Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

- Now imagine, somebody types (or even better cuts/pastes) an 'Abstract' text into the submission form of SPMS for the conference where you are the editor.
- Now we looking at the input text as the computer does. What do we (or the computer) see?
- Perfect! But what does it mean?

In pre-computer time What changed in the computer era? New standards Problem of interpretation

- Now imagine, somebody types (or even better cuts/pastes) an 'Abstract' text into the submission form of SPMS for the conference where you are the editor.
- Now we looking at the input text as the computer does. What do we (or the computer) see?
- Perfect! But what does it mean?
- We have (or the computer has) no idea where this might come from. We only know that we accept 8 Bit values. No idea whether it's a Chinese scientist at CERN using his laptop or a French who is sitting at KEK, or ...

In pre-computer time What changed in the computer era? New standards Problem of interpretation

- Now imagine, somebody types (or even better cuts/pastes) an 'Abstract' text into the submission form of SPMS for the conference where you are the editor.
- Now we looking at the input text as the computer does. What do we (or the computer) see?
- Perfect! But what does it mean?
- We have (or the computer has) no idea where this might come from. We only know that we accept 8 Bit values. No idea whether it's a Chinese scientist at CERN using his laptop or a French who is sitting at KEK, or ...
- No knowledge about used character set or character encoding.

In pre-computer time What changed in the computer era? New standards Problem of interpretation

- Now imagine, somebody types (or even better cuts/pastes) an 'Abstract' text into the submission form of SPMS for the conference where you are the editor.
- Now we looking at the input text as the computer does. What do we (or the computer) see?
- Perfect! But what does it mean?
- We have (or the computer has) no idea where this might come from. We only know that we accept 8 Bit values. No idea whether it's a Chinese scientist at CERN using his laptop or a French who is sitting at KEK, or ...
- No knowledge about used character set or character encoding.
- If the computer switches to a different interpretation of the input, it might look like this.

In pre-computer time What changed in the computer era? New standards Problem of interpretation

New standards: Problems solved?

•

0

0

\ufeff \u041f\u0440\u043e\u0434\u0430\u0435\u0442\u0441\u044f \u0437\u0435 \u043c\u043b\u044f \u0441\u0435\u043b\u044c\u0445\u043e\u0437 \u043d\u0430\u0437 \u043d\u0430\u0447\u0435\u043d\u0438\u044f

\u043e\u0442 \u0441\u043e\u0431\u0442\u0432\u0435\u043d\u043d \u0438\u043a\u0430 \u0421\u0420\u041e\u0427\u041d\u041e

\u0423\u0447\u0430\u0441\u0442\u043a\u0438: 24 \u0433\u0430, 732 \u0433\u0430,

1290 \u0433\u0430, 1361 \u0433\u0430 \u0426\u0435\u043d\u0430: 8000 \u0440\u0443\u0431\u043b\u0435\u0439 \u0437\u0430 \u0433\u0430 (80 \u0440\u0443\u0431\u043b\u0435\u0439/\u0441\u043e\u0442\u043a \u0430) \u0422\u041e\u0420\u0413

\u0423\u0447\u0430\u0441\u0442\u043a\u0438 \u0432\u044b\u0434\u0435\u043b\u0435 \u043d\u044b, \u0441\u0432\u0438\u0434\u0435\u0442\u0435\u043b\u044c\u0441\u0442 \u0432\u0430\u0423\u0423\u0424\u0420\u0421 \u043f\u043e\u043b\u0443\u0447\u0435\u043b \u044b.

\u0420\u0430\u0431\u043f\u043e\u043b\u043e\u0436\u0436\u0435\u043d\u0438\u0435: \u0423 \u043b\u044c\u044f\u043d\u043e\u0432\u0441\u043a\u0430\u044f \u043e\u0431\u043b \u0430\u0441\u042\u044c, \u0418\u043d\u0437\u0435\u043d\u0441\u043a\u0438\u0438 u0430\u0441\u0430\u0439\u043e\u043d.

\u0414\u043e\u043a\u0443\u043c\u0435\u043d\u0442\u044b \u043f\u043e \u0437\u0430 \u043f\u0440\u043e\u0441\u0443.

\u041a\u043e\u043d\u0442\u0430\u043a\u0442\u044b: +7 (916) 99\u0427-\u04176-61, 8052511(a)gmail.com

\u041c\u0430\u043a\u0441\u0438\u043c
In pre-computer time What changed in the computer era? New standards Problem of interpretation

New standards: Problems solved?

- Now imagine, somebody types (or even better cuts/pastes) an 'Abstract' text into the submission form of SPMS for the conference where you are the editor.
- Now we looking at the input text as the computer does. What do we (or the computer) see?
- Perfect! But what does it mean?
- We have (or the computer has) no idea where this might come from. We only know that we accept 8 Bit values. No idea whether it's a Chinese scientist at CERN using his laptop or a French who is sitting at KEK, or ...
- No knowledge about used character set or character encoding.
- If the computer switches to a different interpretation of the input, it might look like this.
- It helps a bit, but we still have no idea about the characters not being ASCII (or in the range of 32...127 or x'20'...'7e')...

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

To get an idea about what it could mean we switch to
 West-European languages and select 1S08859-1:

Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

ODIAÁAÔÓŃ ÚAÍÌŃ ÓAÌØÈÏÚ ÍAÚÍAĮAÎÉŃ IÔ ÓIÂÓÔ×AÎÎÉËÁ óòiţîï

õĮÁÓÔËÉ: 24 ĢÁ, 732 ĢÁ, 1290 ĢÁ, 1361 ĢÁ ãÅÎÁ: 8000 ÒÕÂÌÅĘ ÚÁ ĢÁ (80 ÒÕÂÌÅĘ/ÓÏÔËÁ) ôïòç

õĮÁÓÔËÉ ×ÙÄÅÌÅÎÙ, Ó×ÉÄÅÔÅÌØÓÔ×Á õcòó ĐÏÌÕĮÅÎÙ. òÁÓĐÏÌÏÖÅÎÉÅ: õÌØŃÎÏ×ÓËÁŃ ÏÂÌÁÓÔØ, éĨÚÅÎÓËÉĘ ÒÁĘÏÎ.

äÏËÕÍÅÎÔÙ ÐÏ ÚÁÐÒÏÓÕ. ëÏÎÔÁËÔÙ: +7 (9I6) 99į-ú6-61, 8052511(a)gmail.com íÁËÓÉÍ

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

To get an idea about what it could mean we switch to
 West-European languages and select IS08859-1:
 Not a West-European language!

Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

To get an idea about what it could mean we switch to
the Greek language and select IS08859-7:

Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

Π ΦΟΔΑΕΤΣΡ ΪΕΝΜΡ ΣΕΜΨΘΟΪ ΞΑΪΞΑήΕΞΙΡ ΟΤ ΣΟΒΣΤΧΕΞΕΙΛΑ σςοώξο

υήΑΣΤΛΙ: 24 ΗΑ, 732 ΗΑ, 1290 ΗΑ, 1361 ΗΑ γΕΞΑ: 8000 �YBMEK ΪΑ ΗΑ (80 �YBMEK/ΣΟΤΛΑ) τοςη

υήΑΣΤΛΙ ΧΩΔΕΜΕΞΩ, ΣΧΙΔΕΤΕΜΨΣΤΧΑ υζςσ ΠΟΜΥήΕΞΩ. ςΑΣΠΟΜΟΦΕΞΙΕ: υΜΨΡΞΟΧΣΛΑΡ ΟΒΜΑΣΤΨ, ιΞΪΕΞΣΛΙΚ �ΑΚΟΞ.

δΟΛΥΝΕΞΤΩ ΠΟ ΪΑΠ�ΟΣΥ. λΟΞΤΑΛΤΩ: +7 (9Ι6) 99ώ-ï6-61, 8052511(a)gmail.com νΑΛΣΙΝ

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

To get an idea about what it could mean we switch to
the Greek language and select IS08859-7:
Doesn't look Greek to me!

Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

To get an idea about what it could mean we switch to
 the Baltic languages and select IS08859-13:

Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

šŅĻÄĮÅŌÓŃ ŚÅĶĢŃ ÓÅĢŲČĻŚ ĪĮŚĪĮŽÅĪÉŃ ĻŌ ÓĻĀÓŌ×ÅĪĪÉĖĮ óņļžīļ

õŽįÓŌĖĖ: 24 ĒĮ, 732 ĒĮ, 1290 ĒĮ, 1361 ĒĮ ćÅĪĮ: 8000 ŅÕĀĢÅŹ ŚĮ ĒĮ (80 ŅÕĀĢÅŹ/ÓĻŌĖĮ) ōļņē

öŽĮÓŌĖĖ ×ŁÄÅĢÅĪŁ, Ó×ĖÄÅŌÅĢŲÓŌ×Į õęņó ŠĻĢÕŽÅĪŁ. nĮÓŠĻĢĻÖÅĪĖÅ: õĢŲŃĪĻ×ÓĖĮŃ ĻĀĢĮÓŌŲ, ĖĪŚÅĪÓĖĖŹ ŅĮŹĻĪ.

äĻĖÕĶÅĪŌŁ ŠĻ ŚĮŠŅĻÓÕ. ėĻĪŌĮĖŌŁ: +7 (9I6) 99ž-ś6-61, 8052511(a)gmail.com ķĮĖÓĖĶ

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

To get an idea about what it could mean we switch to
the Baltic languages and select IS08859-13:
Not a Baltic language!

Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

To get an idea about what it could mean, let's try
 the Thai language and select IS08859-11:

Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

๐าฯถุมลีาั ฺลอฟั ําลฟุัศฯฺฮมฺฮม�ลฮษั ฯ ําฯยำลฮฮษหม ๓๒๐�๎๐

๕�มำิหษ: 24 วม, 732 วม, 1290 วม, 1361 วม ใลฮม: 8000 าียพัลสุม วม (80 าียพัลส/าฯิหม) ๔๏๒์

๙♥มาิหษ ูีถุลพัลฮู, "าษถุลิลฟุ้าม ๕ๆ๒๓ ะฯฟั♥ลฮู. ๒มาะฯฟฯีลฮษล: ๕ฟุ้ฮฯาหมั ฯยฟมๅิ, ้ฮุลฮำหษส ามสฯฮ.

ไฯหือลฮู๊ ะฯ ุมะาฯาี. ๋ฯฮิมหู๊: +7 (9I6) 99�-๚6-61, 8052511(a)gmail.com ํมหำษอ

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

To get an idea about what it could mean, let's try
 the Thai language and select IS08859-11:
 Looks funny but perhaps not Thai! (See question marks in black diamonds?)

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

• To get an idea about what it could mean we switch to the Armenian language and select IS08859-16:

Volker RW Schaa

Fonts & Encoding

ՊկխղծժԿՂմ Հկ ՅըՀՁկձղ. սկԿՂըխՂմ: +7 (9I6) 99՝-06-61, 8052511(a)gmail.com վըխձլծ

ւՇըձՂխլ ճմԺժԾժԿմ, ձճլԺժՂժԾՄձՂճը ւՋՑց ՀկԾղՇժԿմ. ՑըձՀկԾկՃժԿլժ: ւԾՄհԿկճձխըհ կԹԾըձՂՄ, ռԿՅժԿձխլԽ ՁըԽկԿ.

ւՇըձՂխլ: 24 իը, 732 իը, 1290 իը, 1361 իը չժԿը: 8000 ՁղԹԾժԽ Յը իը (80 ՁղԹԾժԽ/ձկՂխը) ՒտՑջ

ՐՁկժըժՂձհ ՅժծԾհ ձժԾՄԼկՅ ԿըՅԿըՇժԿլհ կՂ ձկԹձՂճժԿԿլխը ցՑտ՝Տտ

Let's try to uncover the secret

History of fonts and encodings Fonts in print (encoding and subsetting) Conclusion In pre-computer time What changed in the computer era? New standards Problem of interpretation

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

To get an idea about what it could mean we switch to
the Armenian language and select IS08859-16:
Looks good, but is it Armenian?

Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

To get an idea about what it could mean, let's stay for a moment
 in Europe and try Romanian and select MAC-Romanian:

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

lŢ¡"'À...: 24 «¡, 732 «¡, 1290 «¡, 1361 «¡ "≈Œ¡: 8000 ``'ÑÃ≈ /¡ «¡ (80 ``'ÑÃ≈ /"œ'À¡) ÙÔÚÁ

lŢ¡″'À… ◊Ÿf≈Ã≈ŒŸ, ″◊…f≈'≈Ãÿ″'◊¡ lĘÚŨ −œÃ'Ţ≈ŒŸ. Ú¡″−œÃœ÷≈Œ…≈: lÃÿ─œœ◊″À¡─ œÑá″'ÿ, ÈŒ∕≈Œ″À… ``; œŒ.

‰œÀ'Ő≈Œ'Ÿ -œ ∕¡-"œ″'. ΜŒ'¡À'Ÿ: +7 (9I6) 99[°],-.6-61, 8052511(a)gmail.com ̦À″…Ő

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

To get an idea about what it could mean, let's stay for a moment
 in Europe and try Romanian and select MAC-Romanian:
 Surely not Romanian (even on a MAC)

Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

To get an idea about what it could mean, let's try
 Arabic and switch to IS08859-6:

Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret



Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret



Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

- To get an idea about what it could mean, let us assume it's our Chinese colleague at CERN,
- so switch to Chinese-Simplified (ISO-2022-CN):

Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

鹨夏憭杂� 谂吞� 优特认� 瘟谖赁盼裳 显 酉掠宰盼紊肆 篁稔铒

蹀劣运�: 24 橇, 732 橇, 1290 橇, 1361 橇 闩瘟: 8000 艺绿攀 诹 橇(80 艺绿攀/酉运�)麸蜱

蹀劣运� 踪呐膛钨,幼赡旁盘赜宰� 蹑蝮 邢陶夼钨. 蛄有咸现盼膳: 跲匮蜗子肆� 下塘釉�,槲谂斡松� 伊氏�.

Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

 To get an idea about what it could mean, let us assume it's our Chinese colleague at CERN,

so switch to Chinese-Simplified (ISO-2022-CN):

Still many black diamonds! Not simple Chinese!

Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

• To get an idea about what it could mean, what about traditional Chinese, let's try Chinese-Traditional:

Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

導花◆唐◆ 情性◆ 苑坷缶◆ 氛悸◆欠呐 近 衍◆哥欠沽芒 檔點複

輾�剛�: 24 共, 732 共, 1290 共, 1361 共 嗇氛: 8000 皆�日 惋 共 (80 皆�日/衍剛�) 縹錄

輾�剛� 耿千咀治,負呆氏月迸哥� 邂雕 垠咎喜治. 謂衫芷邸欠呃: 謠追泓耄芒� 芟呵訂�,漣情況豕� 狡抑�.

積迄怔沮� 垠 惋城返�. 錂沮�唆: +7 (9I6) 99�-�6-61, 8052511(a)gmail.com 樅迤吻

Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

To get an idea about what it could mean, what about
 traditional Chinese, let's try Chinese-Traditional:
 Different but still wrong!

Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

• To get an idea about what it could mean, what about Farsi? We use Mac-Farsi:

Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret



Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret



In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

 To get an idea about what it could mean, what about the fifth most spoken language

Hindi? We switch to Mac-Hindi:

Volker RW Schaa

Fonts & Encoding

ोरभशयधय़व ऱर ाणऱळरऴश. �रय़वणभव: +7 (9I6) 99�-९6-61, 8052511(a)gmail.com �णभऴफय

४ूणऴवभफ सदधमधय, ऴसफदधवधमहऴवसण ४ौ१२ ऱरमशूधय. १णऴऱरमरषधयफध: ४महलयरसऴभणल रतमणऴवह, याधयऴभफब ळणबरय.

४ूणऴवभफ: 24 ऩण, 732 ऩण, 1290 ऩण, 1361 ऩण धऱाण: 8000 ळशतमधब ाण ऩण (80 ळशतमधब/ऴरवभण) ३�१ॉ

छिठरदणधवळल ाधयमल ऴधमहपरा यणायण्धयफल रव ळरतळवसधययफभण २१

Let's try to uncover the secret

History of fonts and encodings Fonts in print (encoding and subsetting) Conclusion In pre-computer time What changed in the computer era? New standards Problem of interpretation

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

• To get an idea about what it could mean, what about the fifth most spoken language

Hindi? We switch to Mac-Hindi:

Still black diamonds!

Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

 To get an idea about what it could mean, what about our Japanese colleagues?

Let's try Japanese in Shift_JIS encoding:

Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

?マトチナヤモム レナヘフム モナフリネマレ ホチレホチ*ナホノム マヤ モマツモヤラナホホノヒチ ?・ i

?チモヤヒノ: 24 ヌチ, 732 ヌチ, 1290 ヌチ, 1361 ヌチ 翡ホチ: 8000 メュッフナハ レチ ヌチ (80 メュッフナハ/モマヤヒチ) ??

?チモヤヒノ ラルトナフナホル, モラノトナヤナフリモヤラチ ?? ミマフェ[×]ナホル. ?モミマフマヨナホノナ: ?リムホマラモヒチム マツフチモヤリ, 鯰シナホモヒノハ メチハマホ.

湊とユヘナホヤル ミマ レチミメマモユ.

・ホヤチヒヤル: +7 (916) 99?-・-61, 8052511(a)gmail.com 嘻ヒモノヘ
In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

• To get an idea about what it could mean, what about our Japanese colleagues?

Let's try Japanese in Shift_JIS encoding:

Anybody can read this?

Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

• To get an idea about what it could mean, we now try Unicode Let's switch to UTF-16:

Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

To get an idea about what it could mean, we now try Unicode
 Let's switch to UTF-16:

+++?쿄쇅퓓턠◆췌턠폅쳘죏�컁�쇞엎집႞ठ+++++쿔ェ쿂폔?컎짋선?0?@ण`?�폔 쭙?約?섬™淼?섬‱仳@쟁?ᠠ??선_일ण?컁??~|헂쳅쨠�?선?@틕싌엊?쿔쯁 ???@ण`?�폔쭙→�엌엎�ェ?쓅퓅쳘폔????탏쳕�컙??쇓탏쳏훅컉씺?쳘퇎쿗폋쇑?싌 쇓퓘??�컓쭙쨠틁쫏츮_일ण`?쿋헍엎퓙←켠�탒쿓픮_일ण?컔쇋퓙????????广+???乂愩 杭慩氮捯洠_일ण?쫗짍_{일ण일}ण

Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret



Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

OK, we left out a language which is the eighth most spoken one:
 Russian! Let's switch to Cyrillic in IS08859-5 encoding:

Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

Продается земля сельхоз назначения от собственника СРОЧНО

Участки: 24 га, 732 га, 1290 га, 1361 га Цена: 8000 рублей за га (80 рублей/сотка) ТОРГ

Участки выделены, свидетельства УФРС получены. Расположение: Ульяновская область, Инзенский район.

Документы по запросу. Контакты: +7 (9I6) 994-36-61, 8052511(a)gmail.com Максим

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret



Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

So in the end it's not an 'Abstract' but some SPAM message.
And this is the English translation (OK, mostly English):

Volker RW Schaa

In pre-computer time What changed in the computer era? New standards Problem of interpretation

Let's try to uncover the secret

The ground сельхоз purposes is on sale From the proprietor URGENTLY

Sites: 24 hectares, 732 hectares, 1290 hectares, 1361 hectares The price: 8000 roubles for hectares (80 roubles/cotka) the TENDER

Sites are allocated, certificates YOPC are received. Arrangement: the Ulyanovsk area, area Inzensky.

Documents by inquiry. Contacts: 7 (916) 994-36-61, 8052511 (a) gmail.com Maxim

Volker RW Schaa Fonts & Encoding

Encoding and naming Font character maps Font problems part 1 – Font not embedded Font problems part 2 – Font embedded but characters missing

Encoding

 We now have understood that Encoding is important as it points to the right character: x'a4' + IS08859-15 ⇒ €



Volker RW Schaa

Encoding and naming Font character maps Font problems part 1 - Font not embedded Font problems part 2 - Font embedded but characters missing

- We now have understood that Encoding is important as it points to the right character: x'a4' + IS08859-15 ⇒ €
- If we would have used: $x'a4' + IS08859-14 \Rightarrow \dot{C}$



- We now have understood that Encoding is important as it points to the right character: x'a4' + IS08859-15 ⇒ €
- If we would have used: $x'a4' + IS08859-14 \Rightarrow \dot{C}$
- But how are Symbol fonts encoded? Is there a unique way to define an encoding for them?



- We now have understood that Encoding is important as it points to the right character: x'a4' + IS08859-15 ⇒ €
- If we would have used: x'a4' + IS08859-14 $\Rightarrow \dot{C}$
- But how are Symbol fonts encoded? Is there a unique way to define an encoding for them?
- Short answer: before Unicode some had names (PostScript Level 3 named ≈180 including Greek letters but nothing specific for Math Symbols)



- We now have understood that Encoding is important as it points to the right character: x'a4' + IS08859-15 ⇒ €
- If we would have used: x'a4' + IS08859-14 $\Rightarrow \dot{C}$
- But how are Symbol fonts encoded? Is there a unique way to define an encoding for them?
- Short answer: before Unicode some had names (PostScript Level 3 named ≈180 including Greek letters but nothing specific for Math Symbols)
- In Unicode it look like this:



pt becific

Encoding

R C

R C

• We now have understood that Encoding is important as it points
to the right character: x ' a4 ' + IS08859-15 $\Rightarrow \in$
• If we would have used: x'a4' + IS08859-14 $\Rightarrow \dot{C}$
• Multiplication and division sign operators

2A2F ×		VECTOR OR CROSS PRODUCT
		\rightarrow 0.0D7 \times multiplication sign

2A32
$$\times$$
 SEMIDIRECT PRODUCT WITH BOTTOM CLOSED

Encoding and naming Font character maps Font problems part 1 – Font not embedded Font problems part 2 – Font embedded but characters missing

Mapping

• A typical font (Type 1, TrueType and OpenType) contains a map with character names and its location in the font.

Volker RW Schaa

- A typical font (Type 1, TrueType and OpenType) contains a map with character names and its location in the font.
- Nowadays fonts have far more than the possible 192 characters (standard is 250, but there are fonts with more 2 000 characters).

- A typical font (Type 1, TrueType and OpenType) contains a map with character names and its location in the font.
- Nowadays fonts have far more than the possible 192 characters (standard is 250, but there are fonts with more 2 000 characters).
- The PostScript (printer) driver looks up the requested font and
 - accesses the character map table of this fonts,
 - it scans the text to typeset and finds out about the requested Encoding and
 - all different characters that appear in the text.

- A typical font (Type 1, TrueType and OpenType) contains a map with character names and its location in the font.
- Nowadays fonts have far more than the possible 192 characters (standard is 250, but there are fonts with more 2 000 characters).
- The PostScript (printer) driver looks up the requested font and
 - accesses the character map table of this fonts,
 - it scans the text to typeset and finds out about the requested Encoding and
 - all different characters that appear in the text.
- Now it copies the needed Encoding vector and, depending on the parameters, copies either the whole set of characters, just the needed ones or nothing (Embedded, Embedded Subset, nada).

- A typical font (Type 1, TrueType and OpenType) contains a map with character names and its location in the font.
- Nowadays fonts have far more than the possible 192 characters (standard is 250, but there are fonts with more 2 000 characters).
- The PostScript (printer) driver looks up the requested font and
 - accesses the character map table of this fonts,
 - it scans the text to typeset and finds out about the requested Encoding and
 - all different characters that appear in the text.
- Now it copies the needed Encoding vector and, depending on the parameters, copies either the whole set of characters, just the needed ones or nothing (Embedded, Embedded Subset, nada).
- If it finds characters which do not fit into one Encoding vector, the driver may copy a second map and the corresponding characters.

Encoding and naming Font character maps Font problems part 1 – Font not embedded Font problems part 2 – Font embedded but characters missing

Mapping and character copying

 How does a Encoding map looks like? An example of a Custom map which encodes more than the standard ISOLatin1Encoding by using some of the code point between x '00' and x '1f'.

Encoding and naming Font character maps Font problems part 1 – Font not embedded Font problems part 2 – Font embedded but characters missing



Encoding and naming Font character maps Font problems part 1 – Font not embedded Font problems part 2 – Font embedded but characters missing

Mapping and character copying

- How does a Encoding map looks like? An example of a Custom map which encodes more than the standard ISOLatin1Encoding by using some of the code point between x '00' and x '1f'.
- The code which makes up a font Subset may look like this:

Encoding and naming Font character maps Font problems part 1 – Font not embedded Font problems part 2 – Font embedded but characters missing



more than the of the code point

look like this:

Encoding and naming Font character maps Font problems part 1 – Font not embedded Font problems part 2 – Font embedded but characters missing

Mapping and character copying

- How does a Encoding map looks like? An example of a Custom map which encodes more than the standard ISOLatin1Encoding by using some of the code point between x '00' and x '1f'.
- The code which makes up a font Subset may look like this:
- So this all looks fine, but why do we face problems so often?

Encoding and naming Font character maps Font problems part 1 – Font not embedded Font problems part 2 – Font embedded but characters missing

Mapping and character copying

- How does a Encoding map looks like? An example of a Custom map which encodes more than the standard ISOLatin1Encoding by using some of the code point between x '00' and x '1f'.
- The code which makes up a font Subset may look like this:
- So this all looks fine, but why do we face problems so often?
- Let us look into the Document Properties of a typical PDF file:

	Fonts in print (encoding and subsetting) Font character maps		
	Document Properties	ers missing	
	Fonts in print (encoding and subsetting) Font character maps Document Properties Description Security Fonts Initial View Custom Advanced Fonts Used in this Document Initial View Custom Advanced Image: Comparison of the second state of the second st	rs missing ne oint 1? file:	
e Ad	Type: TrueType Encoding: Roman		
6	Help OK Cancel		22.3

Volker RW Schaa

Encoding and naming Font character maps Font problems part 1 – Font not embedded Font problems part 2 – Font embedded but characters missing

Mapping and character copying

- How does a Encoding map looks like? An example of a Custom map which encodes more than the standard ISOLatin1Encoding by using some of the code point between x '00' and x '1f'.
- The code which makes up a font Subset may look like this:
- So this all looks fine, but why do we face problems so often?
- Let us look into the Document Properties of a typical PDF file:
- We find
 - 3 different Encodings (Ansi, Identity-H, and Roman) for one font,
 - the 2 pairs of character sub-maps (Identity-H, Ansi) are surely not the same,
 - and there are 3 different font format embedded: Type 1, TrueType, and TrueType (CID)

Encoding and naming Font character maps Font problems part 1 – Font not embedded Font problems part 2 – Font embedded but characters missing

Mapping and character copying

- How does a Encoding map looks like? An example of a Custom map which encodes more than the standard ISOLatin1Encoding by using some of the code point between x '00' and x '1f'.
- The code which makes up a font Subset may look like this:
- So this all looks fine, but why do we face problems so often?
- Let us look into the Document Properties of a typical PDF file:
- We find
 - 3 different Encodings (Ansi, Identity-H, and Roman) for one font,
 - the 2 pairs of character sub-maps (Identity-H, Ansi) are surely not the same,
 - and there are 3 different font format embedded: Type 1, TrueType, and TrueType (CID)

 Positive: all fonts are embedded as subsets!
 But: we surely will have problems to change or edit such a file with Pitstop or Acrobat!

Encoding and naming Font character maps Font problems part 1 – Font not embedded Font problems part 2 – Font embedded but characters missing

Why is a font not embedded? Just for the statistics: in \approx 35% of the PDF files uploaded for a JACoW conference one or more fonts are missing.

• A font was requested which is one of the 15 base fonts and the switch always embed fonts had not been set in the PostScript produced by the author.

Encoding and naming Font character maps Font problems part 1 – Font not embedded Font problems part 2 – Font embedded but characters missing

Why is a font not embedded? Just for the statistics: in \approx 35% of the PDF files uploaded for a JACoW conference one or more fonts are missing.

- A font was requested which is one of the 15 base fonts and the switch always embed fonts had not been set in the PostScript produced by the author.
- In the PostScript driver setup the option for TrueType fonts was set as "Substitute with Device Font".

History of fonts and encodings

Encoding and naming



Why is a font not embedded? Just for the statistics: in \approx 35% of the PDF files uploaded for a JACoW conference one or more fonts are missing.

- A font was requested which is one of the 15 base fonts and the switch always embed fonts had not been set in the PostScript produced by the author.
- In the PostScript driver setup the option for TrueType fonts was set as "Substitute with Device Font".
- A font was requested in the PostScript file which is not installed on the editor's computer, therefore a substitute font is used for display and PDF generation, but the font will not be included in the final PDF.

Why is a font not embedded? Just for the statistics: in \approx 35% of the PDF files uploaded for a JACoW conference one or more fonts are missing.

- A font was requested which is one of the 15 base fonts and the switch always embed fonts had not been set in the PostScript produced by the author.
- In the PostScript driver setup the option for TrueType fonts was set as "Substitute with Device Font".
- A font was requested in the PostScript file which is not installed on the editor's computer, therefore a substitute font is used for display and PDF generation, but the font will not be included in the final PDF.
- A Truetype, CID or OpenType font was used that contained the flag /FSType <x> def with an x-value ≠ 0 (i.e. 2 [restricted license embedding], 4 [embedding allowed, but not editable], 256 [no subset embedding], 512 [Bitmap embedding only])

Encoding and naming Font character maps Font problems part 1 – Font not embedded Font problems part 2 – Font embedded but characters missing

Why are characters missing (or shown as □)?

 A font was requested with a high version (newer release date containing more characters) than the font installed on the editor's computer (therefore missing in display and/or print) or on the printer (missing in print).

Why are characters missing (or shown as \Box)?

- A font was requested with a high version (newer release date containing more characters) than the font installed on the editor's computer (therefore missing in display and/or print) or on the printer (missing in print).
- Missing Characters in PDF in http://americanprinter.com/mag/printing_best_pdfs/ On the Mac, as any user knows, fonts can be found just about anywhere and it's very easy to have duplicate fonts. To make matters worse, OS X includes Apple's new dfonts, which are simply the Mac version of the most commonly used fonts, like Helvetica and Times. When it comes to PDF creation, this variability can result in the wrong version of a font being embedded in a PDF file with the possible outcome of different spacing, kerning or even inclusion of a glyph that is entirely different from the one originally set in the page layout.
Why are characters missing (or shown as \Box) cont.?

 In the PostScript driver setup the option for TrueType fonts was set as "Substitute with Device Font" but the printer device has a font with different layout or less characters.



 Make sure when installing Word on PC or Mac that the newest fonts are installed and that there is only one version on the computer.



Volker RW Schaa

Fonts & Encoding

- Make sure when installing Word on PC or Mac that the newest fonts are installed and that there is only one version on the computer.
- Make sure that the newest firmware is loaded on printers used in the editorial office.

- Make sure when installing Word on PC or Mac that the newest fonts are installed and that there is only one version on the computer.
- Make sure that the newest firmware is loaded on printers used in the editorial office.
- The editors should check twice that fonts are embedded.



- Make sure when installing Word on PC or Mac that the newest fonts are installed and that there is only one version on the computer.
- Make sure that the newest firmware is loaded on printers used in the editorial office.
- The editors should check twice that fonts are embedded.
- We should provide a set of fonts (on JACoW) which are not distributed with the OS any longer (i.e. original Times family) for installation on computers in the editorial office.



- Make sure when installing Word on PC or Mac that the newest fonts are installed and that there is only one version on the computer.
- Make sure that the newest firmware is loaded on printers used in the editorial office.
- The editors should check twice that fonts are embedded.
- We should provide a set of fonts (on JACoW) which are not distributed with the OS any longer (i.e. original Times family) for installation on computers in the editorial office.
- Ease the configuration of Tools & Utilities



- Make sure when installing Word on PC or Mac that the newest fonts are installed and that there is only one version on the computer.
- Make sure that the newest firmware is loaded on printers used in the editorial office.
- The editors should check twice that fonts are embedded.
- We should provide a set of fonts (on JACoW) which are not distributed with the OS any longer (i.e. original Times family) for installation on computers in the editorial office.
- Ease the configuration of Tools & Utilities
- Listen to Raphael's talk about a new step to ease configuration woes with a "Generic PostScript Printer for JACoW"!

History of fonts and encodings Fonts in print (encoding and subsetting) Conclusion

What can we do? Thanks

cannot be found

torem toxum dalar

dolare POSsiver citta kand out sanctus amer. conscience ing. Lorem manum labore of Sec diam Annaly Pero Oay Seve Cictural Oha. 94317 minim tation utioncoper

Thanks

for your

attention!

Volker RW Schaa Fonts & Encoding

400