Układ QX – kto zacz?

Wiosną tego roku pojawiła się w archiwum GUST-u polonizacja fontu Palladio, pochodzącego z kolekcji fontów URW, dystrybuowanej od jakiegoś czasu z Ghostscriptem. Znaki w spolonizowanym foncie mają kodowanie zgodne z tzw. układem QX. Co to takiego, ten układ QX?

Otóż fonty w postaci elektronicznej przybywają na rynek polski głównie ze Stanów Zjednoczonych i są dostosowane do potrzeb tamtejszego rynku, zarówno jeśli chodzi o zestaw znaków jak i kodowanie znaków. W związku z tym w układzie dystrybuowanych fontów nie znajdują odbicia ani potrzeby Polaków, ani TEX-owców.

Problem ten od paru lat był przedmiotem prac nieformalnej grupy członków GUST-u. W pracach brali udział: Włodek Bzyl, Bogusław Jackowski, Janusz M. Nowacki, Andrzej Odyniec, Piotr Pianowski, Marek Ryćko, Andrzej Tomaszewski, Staszek Wawrykiewicz oraz wielu innych, którzy dobrym lub krytycznym słowem pomagali popychać sprawę do przodu.

Efektem prac jest właśnie układ QX określający dla fontów PostScript-owych zarówno zestaw znaków, jak i tabele kodów do zastosowań windowsowych oraz do zastosowań TEX-owych, gdyż w zamierzeniu układ QX ma zaspokajać potrzeby składu TEXnicznego, z możliwością (ograniczonego) wykorzystywania fontów w środowisku windowsowym.

Założenia. Układ QX nie wziął się „z głowy czyli z niczego” – przeciwnie, wynikł z przyjętych zasad. Zasady mogą się okazać chybione, to pokaże czas, niemniej jednak najpierw sformułowane zostały zasady, a na ich podstawie konstruowany był układ QX.

Oto najważniejsze z tych zasad:

(1) Fonty w miarę możności powinny mieć identyczny układ znaków.
(2) Fonty mają być przeznaczone do używania w TEX-u.
(3) Fonty mają być przeznaczone do składania przede wszystkim polskich tekstów.
(4) Fonty mają się nadawać do składania tekstów technicznych.
(5) Fonty mają się nadawać do składania tekstów zawierających fragmenty w językach europejskich, ze szczególnym uwzględnieniem krajów sąsiadujących z Polską.
(6) Fonty powinny być samowystarczalne (chodzi o to, by znaków dodatkowych, takich jak *copyright*, *trademark*, *bullet*, itp., nie trzeba było czerpać z innych fontów).
(7) Fonty mają być używalne w środowisku Windows, w szczególności w polskiej wersji środowiska.

Ad 1. Oczywiście nie można oczekiwać stuprocentowej zgodności, na przykład niektóre fonty są zaprojektowane bez ligatur związanych z literą „f”. Wydaje się, że brak niektórych znaków jest stosunkowo niegroźny.

Nieprzyjemne konsekwencje może mieć natomiast decyzja zastępowania pewnych znaków innymi. Mimo to taki pomysł przyszedł do głowy nawet samemu Donaldowi E. Knuthowi: w fontach rodziny Computer Modern font cmr5 ma nieco inny zestaw znaków niż pozostałe fonty szeregu CMR – przykra niespodzianka wydaje się wcześniej czy później nieunikniona. Jednak i takie sytuacje mogą być nie do uniknięcia, np. pisanki z reguły mają więcej ligatur niż zwykłe fonty drukarskie. Należy zatem dbać o to, by przynajmniej podstawowe znaki (litery) miały przypisane zawsze te same kody.

Ad 2. TEX operuje jedynie na fontach co najwyżej 256-znakowych – to poważne ograniczenie. Zostało ono oczywiście uwzględnione przy projektowaniu układu QX. Kolejnym ograniczeniem jest limit różnych wysokości i głębokości znaków w foncie (16), w związku z czym znaki w danym foncie nie powinny się zbyt różnić rozmiarami.

Ze względu na powszechność fontów CM zestaw znaków powinien być możliwe zgodny z zestawem znaków fontów tekstowych rodziny CM (fonty matematyczne, jako używalne wyłącznie w kontekście TEX-a, pozostają poza zakresem zainteresowań).

Podobna zasada była ściśle przestrzegana przy projektowaniu pakietu MEX: fonty PL i fonty CM różnią się jedynie znakami o kodach 128 – 255 (polskie znaki diakrytyczne i cudzysłowy). Dzięki temu każde oprogramowanie TEX-owe przeznaczone do pracy z fontami CM działa również z fontami PL, co okazało się w praktyce bardzo korzystne.

Podjęcie takiej decyzji stało niestety w sprzeczności z ustaleniami konferencji EuroTEX’90 w Cork w Irlandii, gdzie przyjęty został jako „nowy standard” tzw. układ EC. Z grubsza rzecz biorąc w układzie EC znaki diakrytyczne umieszczone są w pozycjach o kodach 0 – 32, a narodowe znaki diakrytyczne w pozycjach o kodach 128–256 (w układzie PL pozycje polskich znaków diakrytycznych są takie same jak w układzie EC).

Niezależnie od zalet takiego systemu kodowania użycie fontów EC z formatem plain jest po prostu niemożliwe. Ponadto pula stu dwudziestu ośmiu pozycji na znaki diakrytyczne okazała się zbyt mała. Nie zmieściły się na przykład litewskie litery „į", „Į", „ų" oraz „Ų".

Ad 3. Wolno sądzić, że pakiet MeX zadomowił się w środowisku polskich użytkowników TeX-a. Ze względu na istnienie pakietu MeX jest wskazane, by font był rozszerzeniem układu PL, możliwie mało różniącym się od układu EC. W szczególności oznacza to, że wszystkie polskie znaki diakrytyczne powinny się w foncie znaleźć.

W związku ze składaniem polskich tekstów technicznych przydatne są znaki specyficzne dla polskiej literatury technicznej. Zagadnienie to omówione jest przy okazji analizowania konsekwencji zasady 4.

W fontach anglosaskich powszechnie obecna jest ligatura „fl" (czasem także „ffl"). Nie widać żadnego powodu, dla którego w polskich fontach nie miałoby być ligatury „fk". W polskim języku jest wprawdzie znacznie więcej słów zawierających digram „fl" niż słów zawierających digram „fk", niemniej jednak są takie słowa: *agrafka*, *cofka*, *dryfkotwa*, *fifka*, *Kafka*, *karafka*, *lufka*, *mufka*, *sofka*, *szafka*. Bez ligatury „fk" wyglądają one niezbyt ładnie ze względu na kolizję liter „f" i „k", analogiczną do kolizji liter „f" i „l". Stąd obecność ligatury „fk" w układzie QX.

Ad 4. Zastosowanie TeX-a do składania tekstów sugeruje pewien zakres zastosowań, mianowicie skład tekstów technicznych. Fonty zatem powinny posiadać znaki przydatne do takich zastosowań. Pośród wielu znaków specyficznych dla polskich zwyczajów typograficznych dwa wydają się szczególnie ważne: symbol kąta „∢" (anglearc), oraz symbol średnicy „∅" (diameter). Znaków tych, o ile nam wiadomo, nie posiada żaden font pochodzenia zagranicznego. Również polskie znaki mniejsze-równe i większe-równe, „$\leqslant$" „$\geqslant$", różnią się od symboli stosowanych za granicą („$\leq$" „$\geq$").

W tekstach technicznych często pojawia się promil „‰" (perthousand). Rzecz dziwna, Donald E. Knuth nie przyjął do wiadomości istnienia tego znaku – w fontach CM po prostu go nie ma.

Z punktu widzenia tekstów technicznych pożądana jest obecność niektórych małych greckich liter w foncie. Np. litera „μ" pojawia się w oznaczeniach jednostek (przedrostek mikro). Litera „π" wydaje się również warta uwzględnienia. Najlepiej byłoby móc zawrzeć w foncie cały alfabet grecki. Niestety, szczupłość miejsca i obszerność potrzeb wykluczają taką możliwość.

Jako jednostki używane są czasem symbole „'" (quotesingle, minuta), „"" (quotedbl, sekunda lub cal), „°" (degree, stopień) – ich brak wcześniej czy później zaczyna dokuczać.

Określenie „teksty techniczne" niekoniecznie musi oznaczać teksty o profilu matematycznym lub inżynierskim. Teksty prawne również można zaliczyć do tekstów technicznych. Brak symbolu „§" w standardowych fontach CM (symbol ten występuje jedynie w fontach szeregu CMSY) dawał się dotkliwie odczuć w praktyce.

Ad 5. Większość znaków diakrytycznych daje się uzyskać za pomocą operacji TeX-owych, na ogół za pomocą polecenia \accent. Jeżeli akcent nie styka się z literą (akcenty „acute", „grave", „dieresis" itp.) zabieg taki jest dopuszczalny. Można też skorzystać z bardzo ogólnego mechanizmu konstruowania znaków złożonych, mianowicie z fontów wirtualnych. Dzięki fontom wirtualnym znaki złożone są widziane przez TeX-a jako zwykłe, pojedyncze znaki. Tworzenie fontów wirtualnych jest jednak pracochłonne (brak póki co wygodnych narzędzi) i wymaga sporej wiedzy, tak TeX-owej, jak i typograficznej.

Niestety, niektóre znaki tworzą z elementem diakrytycznym jedną całość. Wymienić tu należy w pierwszym rzędzie polskie i litewskie litery z ogonkiem („ą", „ę", „į", „ų" i odpowiednie majuskuły) oraz litery z cedillą („ç", „ş", „ţ" i odpowiednie majuskuły). Inne znaki diakrytyczne tego rodzaju to islandzkie „þ", „Þ" (thorn) czy „ð" (eth). W dobrze zaprojektowanych fontach takie znaki nie mogą być „składanką" – muszą stanowić spójnie zaprojektowaną całość.

Ad 6. Font powinien zawierać pewne „uświęcone tradycją" znaki dodatkowe. Takie symbole jak „©" (copyright) czy „®" (registered) powinny być zgodne z bieżącym fontem. Teoretycznie wystarczyłoby kółko, w które za pomocą makra TeX-owego wstawiało by się odpowiednie litery. Każdy, kto próbował używać plain-owego makra \copyright, wie, że nie jest to pomysł najszczęśliwszy – pozycjonowanie liter wewnątrz koła musi być przeprowadzone dla każdego fontu (a nawet dla każdej odmiany) osobno, co wyklucza możliwość automatyzacji.

Symbole „†" (dagger) i „‡" (*daggerdbl*) są wprawdzie stosunkowo rzadko używane (ich używanie to w zasadzie zwyczaj anglosaski), jednakże kształt tych symboli jest specyficzny dla danego fontu, tym samym – dla pełności – font powinien je zawierać.

Inne symbole potrzebne sporadycznie to „•" (bullet) i „·" (periodcentered). Na pierwszy rzut oka mogłoby się wydawać, że znaki te nie są w żaden sposób z konkretnym fontem związane. To oczywiście nieprawda, na przykład kształt kropki zależy od fontu. Sięganie do innego fontu po te znaki lub przesuwanie kropki w celu uzyskania drugiego z nich jest zabiegiem niezwykle uciążliwym ze względu na trudności w wyważeniu położenia obu znaków.

Ad 7. Rozpowszechniona obecnie wersja środowiska Windows nakłada dodatkowe ograniczenia na liczbę znaków w foncie: znakom nie można przypisać kodów większych niż 255 i mniejszych niż 32. Redukuje to maksymalną liczbę znaków w foncie dających się równocześnie wykorzystać do 224. W układzie QX przeznaczonym do zastosowań windowsowych kody $\leqslant$32 otrzymały znaki, co do których można ze sporym prawdopodobieństwem założyć, że nie będą potrzebne w tych zastosowaniach (głównie akcenty, symbole i ligatury – p. tabela 2).

Dalsze zmniejszenie puli wolnych miejsc wiąże się ze znakami o niewielkiej użyteczności z punktu widzenia TEX-a, ale niezbędnymi ze względu na pewne tradycje w ich stosowaniu. Takim znakiem jest np. „¶" (*paragraph*), używany jako znak końca akapitu w niektórych edytorach tekstu pracujących w środowisku Windows. Do podobnego celu wykorzystywany bywa znak „¤" (currency).

Pewne aplikacje windowsowe spodziewają się znaleźć pod kodami 32 i 160 odpowiednio odstęp (space) i niełamliwy odstęp (nbspace). Są to znaki absolutnie bezużyteczne z punktu widzenia TEX-a, ale nieusuwalne ze względu na windowsowe tradycje.

Ligatury i kerny. Specyfikacja układu QX w kontekście zastosowań TEX-owych obejmuje także ligatury. Ligatury to ważny choć praktycznie pomijany (z nie do końca jasnych przyczyn) w świecie nieTEX-owym element definicji fontu.

Wersja 0.3 układu QX uwzględnia następujące ligatury:

```
                          f + f → ff
                          f + i → fi
                          f + l → fl
                          f + k → fk
                         ff + i → ffi
                         ff + l → ffl
                hyphen + hyphen → endash
                endash + hyphen → emdash
          quoteleft + quoteleft → quotedblleft
        quoteright + quoteright → quotedblright
                  comma + comma → quotedblbase
              exclam + quoteleft → exclamdown
            question + quoteleft → questiondown
        exclamdown + exclamdown → guillemotleft
    questiondown + questiondown → guillemotright
```

Z dokładnością do ligatury „fk" jest to ten sam zestaw ligatur co w fontach PL, natomiast różni się on od zestawu ligatur zaproponowanego dla układu EC, który obejmował m. in.:

```
              space + l → lslash
              space + L → Lslash
      percent + percent → perthousand
            less + less → guillemotleft
      greater + greater → guillemotright
```

Pierwsze trzy z tych ligatur wydają się nieuzasadnione, dwie pozostałe są nieadekwatne, ze względu na zachowanie w układzie QX konwencji fontów CM umieszczającej hiszpańskie znaki interpunkcyjne w pozycji znaku „<" (znak „¡") oraz „>" (znak „¿").

W definicji układu QX zestaw par kernowych nie jest w żaden sposób specyfikowany, w szczególności dopuszczalne są – nieobecne w układzie EC – kerny między cyframi.

Tabele układu QX. Na dwóch kolejnych stronach zamieszczamy porównanie układu QX (wer. 0.3) z innymi szeroko używanymi układami znaków. Tabela 1 przedstawia porównanie TEX-owych układów – układu PL (przypomnijmy, że znaki o kodach $\leqslant$127 są identyczne w układzie PL i w układzie CM), układu EC oraz układu QX. Z kolei tabela 2 przedstawia porównanie układu QX przeznaczonego do zastosowań windowsowych z układem CP 1250 oraz macintoshowym PLEURO. Jeżeli chodzi o układ QX to – rzecz jasna – w obu tabelach występuje ten sam zestaw znaków, różnice dotyczą jedynie kodów przypisanych poszczególnym znakom.

Redakcja

Tabela 1

Układ QX – zastosowania TEX-owe – porównanie z układem PL i EC

kod	PL (CM)	EC	QX
0	Gamma	grave	α alpha
1	Delta	acute	Δ Delta
2	Theta	circumflex	β beta
3	Lambda	tilde	δ delta
4	Xi	dieresis	π pi
5	Pi	hungarumlaut	Π Pi
6	Sigma	ring	Σ Sigma
7	Upsilon	caron	µ mu
8	Phi	breve	… ellipsis
9	Psi	macron	fk fk
10	Omega	dotaccent	Ω Omega
11	ff	cedilla	ff ff
12	fi	ogonek	fi fi
13	fl	quotesinglbase	fl fl
14	ffi	guilsinglleft	ffi ffi
15	ffl	guilsinglright	ffl ffl
16	dotlessi	quotedblleft	ı dotlessi
17	dotlessj	quotedblright	ȷ dotlessj
18	grave	quotedblbase	` grave
19	acute	guillemotleft	´ acute
20	caron	guillemotright	ˇ caron
21	breve	endash	˘ breve
22	macron	emdash	¯ macron
23	ring	compoundwordmark	˚ ring
24	cedilla	perthousand	¸ cedilla
25	germandbls	dotlessi	ß germandbls
26	ae	dotlessj	æ ae
27	oe	ff	œ oe
28	oslash	fi	ø oslash
29	AE	fl	Æ AE
30	OE	ffi	Œ OE
31	Oslash	ffl	Ø Oslash
32	suppress	visualspace	space
33	exclam	exclam	! exclam
34	quotedblright	quotedbl	” quotedblright
35	numbersign	numbersign	# numbersign
36	dollar	dollar	$ dollar
37	percent	percent	% percent
38	ampersand	ampersand	& ampersand
39	quoteright	quoteright	’ quoteright
40	parenleft	parenleft	(parenleft
41	parenright	parenright	) parenright
42	asterisk	asterisk	* asterisk
43	plus	plus	+ plus
44	comma	comma	, comma
45	hyphen	hyphen	- hyphen
46	period	period	. period
47	slash	slash	/ slash
48	zero	zero	0 zero
49	one	one	1 one
50	two	two	2 two
51	three	three	3 three
52	four	four	4 four
53	five	five	5 five
54	six	six	6 six
55	seven	seven	7 seven
56	eight	eight	8 eight
57	nine	nine	9 nine
58	colon	colon	: colon
59	semicolon	semicolon	; semicolon
60	exclamdown	less	¡ exclamdown
61	equal	equal	= equal
62	questiondown	greater	¿ questiondown
63	question	question	? question
64	at	at	@ at
65	A	A	A A
66	B	B	B B
67	C	C	C C
68	D	D	D D
69	E	E	E E
70	F	F	F F
71	G	G	G G
72	H	H	H H
73	I	I	I I
74	J	J	J J
75	K	K	K K
76	L	L	L L
77	M	M	M M
78	N	N	N N
79	O	O	O O
80	P	P	P P
81	Q	Q	Q Q
82	R	R	R R
83	S	S	S S
84	T	T	T T
85	U	U	U U
86	V	V	V V
87	W	W	W W
88	X	X	X X
89	Y	Y	Y Y
90	Z	Z	Z Z
91	bracketleft	bracketleft	[bracketleft
92	quotedblleft	backslash	“ quotedblleft
93	bracketright	bracketright	] bracketright
94	circumflex	asciicircum	ˆ circumflex
95	dotaccent	underscore	˙ dotaccent
96	quoteleft	quoteleft	‘ quoteleft
97	a	a	a a
98	b	b	b b
99	c	c	c c
100	d	d	d d
101	e	e	e e
102	f	f	f f
103	g	g	g g
104	h	h	h h
105	i	i	i i
106	j	j	j j
107	k	k	k k
108	l	l	l l
109	m	m	m m
110	n	n	n n
111	o	o	o o
112	p	p	p p
113	q	q	q q
114	r	r	r r
115	s	s	s s
116	t	t	t t
117	u	u	u u
118	v	v	v v
119	w	w	w w
120	x	x	x x
121	y	y	y y
122	z	z	z z
123	endash	braceleft	– endash
124	emdash	bar	— emdash
125	hungarumlaut	braceright	˝ hungarumlaut
126	tilde	asciitilde	˜ tilde
127	dieresis	hyphen	¨ dieresis
128		Abreve	
129	Aogonek	Aogonek	Ą Aogonek
130	Cacute	Cacute	Ć Cacute
131		Ccaron	> greater
132		Dcaron	≥ greaterequal
133		Ecaron	≈ approxequal
134	Eogonek	Eogonek	Ę Eogonek
135		Gbreve	Į Iogonek
136		Lacute	< less
137		Lquoteright	≤ lessequal
138	Lslash	Lslash	Ł Lslash
139	Nacute	Nacute	Ń Nacute
140		Ncaron	~ asciitilde
141		Eng	^ asciicircum
142		Ohungarumlaut	ℓ ell
143		Racute	† dagger
144		Rcaron	‡ daggerdbl
145	Sacute	Sacute	Ś Sacute
146		Scaron	Š Scaron
147		Scedilla	Ş Scedilla
148		Tcaron	° degree
149		Tcedilla	Ţ Tcedilla
150		Uhungarumlaut	˛ ogonek
151		Uring	Ų Uogonek
152		Ydieresis	Ÿ Ydieresis
153	Zacute	Zacute	Ź Zacute
154		Zcaron	Ž Zcaron
155	Zdotaccent	Zdotaccent	Ż Zdotaccent
156		IJ	
157		Idotaccent	{ braceleft
158		dbar	} braceright
159		section	§ section
160		abreve	
161	aogonek	aogonek	ą aogonek
162	cacute	cacute	ć cacute
163		ccaron	® registered
164		dquoteright	© copyright
165		ecaron	÷ divide
166	eogonek	eogonek	ę eogonek
167		gbreve	į iogonek
168		lacute	− minus
169		lquoteright	× multiply
170	lslash	lslash	ł lslash
171	nacute	nacute	ń nacute
172		ncaron	± plusminus
173		eng	∞ infinity
174	guillemotleft	ohungarumlaut	« guillemotleft
175	guillemotright	racute	» guillemotright
176		rcaron	¶ paragraph
177	sacute	sacute	ś sacute
178		scaron	š scaron
179		scedilla	ş scedilla
180		tquoteright	• bullet
181		tcedilla	ţ tcedilla
182		uhungarumlaut	
183		uring	ų uogonek
184		ydieresis	ÿ ydieresis
185	zacute	zacute	ź zacute
186		zcaron	ž zcaron
187	zdotaccent	zdotaccent	ż zdotaccent
188		ij	
189		exclamdown	· periodcentered
190		questiondown	" quotedbl
191		sterling	' quotesingle
192		Agrave	À Agrave
193		Aacute	Á Aacute
194		Acircumflex	Â Acircumflex
195		Atilde	Ã Atilde
196		Adieresis	Ä Adieresis
197		Aring	Å Aring
198		AE	\ backslash
199		Ccedilla	Ç Ccedilla
200		Egrave	È Egrave
201		Eacute	É Eacute
202		Ecircumflex	Ê Ecircumflex
203		Edieresis	Ë Edieresis
204		Igrave	Ì Igrave
205		Iacute	Í Iacute
206		Icircumflex	Î Icircumflex
207		Idieresis	Ï Idieresis
208		Eth	Ð Eth
209		Ntilde	Ñ Ntilde
210		Ograve	Ò Ograve
211	Oacute	Oacute	Ó Oacute
212		Ocircumflex	Ô Ocircumflex
213		Otilde	Õ Otilde
214		Odieresis	Ö Odieresis
215		OE	¤ currency
216		Oslash	‰ perthousand
217		Ugrave	Ù Ugrave
218		Uacute	Ú Uacute
219		Ucircumflex	Û Ucircumflex
220		Udieresis	Ü Udieresis
221		Yacute	Ý Yacute
222		Thorn	Þ Thorn
223		Germandbls	| bar
224		agrave	à agrave
225		aacute	á aacute
226		acircumflex	â acircumflex
227		atilde	ã atilde
228		adieresis	ä adieresis
229		aring	å aring
230		ae	_ underscore
231		ccedilla	ç ccedilla
232		egrave	è egrave
233		eacute	é eacute
234		ecircumflex	ê ecircumflex
235		edieresis	ë edieresis
236		igrave	ì igrave
237		iacute	í iacute
238		icircumflex	î icircumflex
239		idieresis	ï idieresis
240		eth	ð eth
241		ntilde	ñ ntilde
242		ograve	ò ograve
243	oacute	oacute	ó oacute
244		ocircumflex	ô ocircumflex
245		otilde	õ otilde
246		odieresis	ö odieresis
247		oe	∢ anglearc
248		oslash	⌀ diameter
249		ugrave	ù ugrave
250		uacute	ú uacute
251		ucircumflex	û ucircumflex
252		udieresis	ü udieresis
253		yacute	ý yacute
254		thorn	þ thorn
255	quotedblbase	germandbls	„ quotedblbase

Tabela 2

Układ QX – zastosowania windowsowe – porównanie z układem CP 1250 i PLEURO

kod	new ANSI (0–127) + Code Page 1250	Macintosh PLEURO	QX
0			
1			´ acute
2			ˆ circumflex
3			¸ cedilla
4			¨ dieresis
5			˘ breve
6			˙ dotaccent
7			¯ macron
8			˚ ring
9			
10			˝ hungarumlaut
11			˛ ogonek
12			ˇ caron
13			
14			ı dotlessi
15			ȷ dotlessj
16			˜ tilde
17			ff ff
18			ffi ffi
19			ffl ffl
20			fk fk
21			≈ approxequal
22			ℓ ell
23			∞ infinity
24			α alpha
25			β beta
26			δ delta
27			π pi
28			Δ Delta
29			Π Pi
30			Σ Sigma
31			Ω Omega
32	space	space	space
33	exclam	exclam	! exclam
34	quotedbl	quotedbl	" quotedbl
35	numbersign	numbersign	# numbersign
36	dollar	dollar	$ dollar
37	percent	percent	% percent
38	ampersand	ampersand	& ampersand
39	quotesingle	quotesingle	' quotesingle
40	parenleft	parenleft	(parenleft
41	parenright	parenright	) parenright
42	asterisk	asterisk	* asterisk
43	plus	plus	+ plus
44	comma	comma	, comma
45	hyphen	hyphen	- hyphen
46	period	period	. period
47	slash	slash	/ slash
48	zero	zero	0 zero
49	one	one	1 one
50	two	two	2 two
51	three	three	3 three
52	four	four	4 four
53	five	five	5 five
54	six	six	6 six
55	seven	seven	7 seven
56	eight	eight	8 eight
57	nine	nine	9 nine
58	colon	colon	: colon
59	semicolon	semicolon	; semicolon
60	less	less	< less
61	equal	equal	= equal
62	greater	greater	> greater
63	question	question	? question
64	at	at	@ at
65	A	A	A A
66	B	B	B B
67	C	C	C C
68	D	D	D D
69	E	E	E E
70	F	F	F F
71	G	G	G G
72	H	H	H H
73	I	I	I I
74	J	J	J J
75	K	K	K K
76	L	L	L L
77	M	M	M M
78	N	N	N N
79	O	O	O O
80	P	P	P P
81	Q	Q	Q Q
82	R	R	R R
83	S	S	S S
84	T	T	T T
85	U	U	U U
86	V	V	V V
87	W	W	W W
88	X	X	X X
89	Y	Y	Y Y
90	Z	Z	Z Z
91	bracketleft	bracketleft	[bracketleft
92	backslash	backslash	\ backslash
93	bracketright	bracketright	] bracketright
94	asciicircum	asciicircum	^ asciicircum
95	underscore	underscore	_ underscore
96	grave	grave	` grave
97	a	a	a a
98	b	b	b b
99	c	c	c c
100	d	d	d d
101	e	e	e e
102	f	f	f f
103	g	g	g g
104	h	h	h h
105	i	i	i i
106	j	j	j j
107	k	k	k k
108	l	l	l l
109	m	m	m m
110	n	n	n n
111	o	o	o o
112	p	p	p p
113	q	q	q q
114	r	r	r r
115	s	s	s s
116	t	t	t t
117	u	u	u u
118	v	v	v v
119	w	w	w w
120	x	x	x x
121	y	y	y y
122	z	z	z z
123	braceleft	braceleft	{ braceleft
124	bar	bar	| bar
125	braceright	braceright	} braceright
126	asciitilde	asciitilde	~ asciitilde
127			
128			Į Iogonek
129			fi fi
130	quotesinglbase	quotesinglbase	fl fl
131		Ydieresis	Ÿ Ydieresis
132	quotedblbase	quotedblbase	„ quotedblbase
133	ellipsis	ellipsis	… ellipsis
134	dagger	dagger	† dagger
135	daggerdbl	daggerdbl	‡ daggerdbl
136		Ecircumflex	Ê Ecircumflex
137	perthousand	perthousand	‰ perthousand
138	Scaron	Scaron	Š Scaron
139	guilsinglleft	guilsinglleft	Ş Scedilla
140	Sacute	Sacute	Ś Sacute
141	Tcaron	OE	Œ OE
142	Zcaron	Zcaron	Ž Zcaron
143	Zacute	Zacute	Ź Zacute
144			į iogonek
145	quoteleft	quoteleft	‘ quoteleft
146	quoteright	quoteright	’ quoteright
147	quotedblleft	quotedblleft	“ quotedblleft
148	quotedblright	quotedblright	” quotedblright
149	bullet	bullet	• bullet
150	endash	endash	– endash
151	emdash	emdash	— emdash
152		ecircumflex	ê ecircumflex
153	trademark	trademark	∢ anglearc
154	scaron	scaron	š scaron
155	guilsinglright	guilsinglright	ş scedilla
156	sacute	sacute	ś sacute
157	tquoteright	oe	œ oe
158	zcaron	zcaron	ž zcaron
159	zacute	zacute	ź zacute
160		nbspace	nbspace
161	caron	exclamdown	¡ exclamdown
162	breve	questiondown	¿ questiondown
163	Lslash	Lslash	Ł Lslash
164	currency	currency	¤ currency
165	Aogonek	Aogonek	Ą Aogonek
166	brokenbar	brokenbar	⌀ diameter
167	section	section	§ section
168	dieresis	dieresis	Ų Uogonek
169	copyright	copyright	© copyright
170	Scedilla	AE	Æ AE
171	guillemotleft	guillemotleft	« guillemotleft
172	logicalnot	Ntilde	Ñ Ntilde
173	hyphen	minus	− minus
174	registered	registered	® registered
175	Zdotaccent	Zdotaccent	Ż Zdotaccent
176	degree	degree	° degree
177	plusminus	plusminus	± plusminus
178	ogonek	twosuperior	⩽ lessequal
179	lslash	lslash	ł lslash
180	acute	acute	⩾ greaterequal
181	mu	mu	μ mu
182	paragraph	paragraph	¶ paragraph
183	periodcentered	periodcentered	· periodcentered
184	cedilla		ų uogonek
185	aogonek	aogonek	ą aogonek
186	scedilla	ae	æ ae
187	guillemotright	guillemotright	» guillemotright
188	Lquoteright	ntilde	ñ ntilde
189	hungarumlaut	onehalf	Ţ Tcedilla
190	lquoteright	threequarters	ţ tcedilla
191	zdotaccent	zdotaccent	ż zdotaccent
192	Racute	Agrave	À Agrave
193	Aacute	Aacute	Á Aacute
194	Acircumflex	Acircumflex	Â Acircumflex
195	Abreve	Atilde	Ã Atilde
196	Adieresis	Adieresis	Ä Adieresis
197	Lacute	Aring	Å Aring
198	Cacute	Cacute	Ć Cacute
199	Ccedilla	Ccedilla	Ç Ccedilla
200	Ccaron	Egrave	È Egrave
201	Eacute	Eacute	É Eacute
202	Eogonek	Eogonek	Ę Eogonek
203	Edieresis	Edieresis	Ë Edieresis
204	Ecaron	Igrave	Ì Igrave
205	Iacute	Iacute	Í Iacute
206	Icircumflex	Icircumflex	Î Icircumflex
207	Dcaron	Idieresis	Ï Idieresis
208	Eth	Eth	Đ Eth
209	Nacute	Nacute	Ń Nacute
210	Ncaron	Ograve	Ò Ograve
211	Oacute	Oacute	Ó Oacute
212	Ocircumflex	Ocircumflex	Ô Ocircumflex
213	Ohungarumlaut	Otilde	Õ Otilde
214	Odieresis	Odieresis	Ö Odieresis
215	multiply	multiply	× multiply
216	Rcaron	Oslash	Ø Oslash
217	Uring	Ugrave	Ù Ugrave
218	Uacute	Uacute	Ú Uacute
219	Uhungarumlaut	Ucircumflex	Û Ucircumflex
220	Udieresis	Udieresis	Ü Udieresis
221	Yacute	Yacute	Ý Yacute
222	Tcedilla	Thorn	Þ Thorn
223	germandbls	germandbls	ß germandbls
224	racute	agrave	à agrave
225	aacute	aacute	á aacute
226	acircumflex	acircumflex	â acircumflex
227	abreve	atilde	ã atilde
228	adieresis	adieresis	ä adieresis
229	lacute	aring	å aring
230	cacute	cacute	ć cacute
231	ccedilla	ccedilla	ç ccedilla
232	ccaron	egrave	è egrave
233	eacute	eacute	é eacute
234	eogonek	eogonek	ę eogonek
235	edieresis	edieresis	ë edieresis
236	ecaron	igrave	ì igrave
237	iacute	iacute	í iacute
238	icircumflex	icircumflex	î icircumflex
239	dquoteright	idieresis	ï idieresis
240	dbar	eth	ð eth
241	nacute	nacute	ń nacute
242	ncaron	ograve	ò ograve
243	oacute	oacute	ó oacute
244	ocircumflex	ocircumflex	ô ocircumflex
245	ohungarumlaut	otilde	õ otilde
246	odieresis	odieresis	ö odieresis
247	divide	divide	÷ divide
248	rcaron	oslash	ø oslash
249	uring	ugrave	ù ugrave
250	uacute	uacute	ú uacute
251	uhungarumlaut	ucircumflex	û ucircumflex
252	udieresis	udieresis	ü udieresis
253	yacute	yacute	ý yacute
254	tcedilla	thorn	þ thorn
255	dotaccent	ydieresis	ÿ ydieresis