

Una pagina indirizzata con codici di un byte

Ogni lettera delle tre righe qui sotto costa un byte del flusso di pagina di questo file anziché due. A un font si arriva per indice di glifo salvo diversa indicazione, due byte per glifo, il che mette a portata ogni glifo del carattere. Attraverso una codifica lo si indirizza invece con codici di un byte: al massimo 255, ciascuno per un carattere, e il file indica quale carattere rappresenta ciascuno.

Composto con codici di un byte

HQF Development

Una fattura, una busta paga, un documento di trasporto.

Un byte per lettera, e il file dice quale.

I codici che ha fissato

La codifica è adattata alle parole che disegna questa pagina: lo spazio tiene il codice 32, e ogni altro carattere prende il codice libero più basso dal 33 in su, così niente nel flusso è un byte che il programma di lettura debba proteggere. I primi sono H 33 · Q 34 · F 35 · D 36 · e 37 · v 38 · l 39 · o 40.

Quanto pesa il flusso

Disegnate per indice di glifo, le tre righe qui sopra costano 586 byte di flusso di pagina; con codici di un byte, 362 byte, cioè 224 in meno. Entrambe le cifre sono misurate dal programma che ha disegnato questa pagina, non scritte a mano.

Cosa compra il singolo byte 32

Parole che un programma di lettura può allargare.

Parole che un programma di lettura può allargare.

Le due righe qui sopra hanno gli stessi caratteri. La seconda è disegnata con la spaziatura tra parole impostata, che sposta la penna al singolo byte 32 e a nessun altro codice. Con gli indici di glifo nessun codice è un singolo byte 32, così l'operatore non ha nulla da raggiungere e le parole non si muovono; lo spazio di questa codifica sta al 32, così si muovono.

Una pagina di codici nominata dalla sua norma

Счёт-фактура № 117 — оплата до 30 сентября

La riga qui sopra è composta con la pagina di codici 1251 di Windows, nominata una volta anziché scritta codice per codice: il file indica la codifica Windows e nomina sopra ogni codice che la pagina riempie in modo diverso, così un programma di lettura trova ogni lettera cirillica dove la mette la pagina. I suoi primi byte sono D1 F7 B8 F2 2D F4 E0 EA.

Un byte indirizza 255 caratteri alla volta, così a una pagina di più alfabeti si arriva ancora per indice di glifo. Nemmeno una legatura o un maiuscoletto si disegna attraverso una codifica: sono glifi che nessun carattere rappresenta, e un codice rappresenta un carattere. Un file che dichiara conformità d'archivio indica la codifica Windows come quella su cui si leggono i suoi codici, e questo lo fa.

Trenta codifiche su cui si può leggere un codice

Ciò che il byte 0xC9 rappresenta in ciascuna: il punto di codice del carattere, e il carattere stesso dove questo carattere tipografico sa disegnarlo.

WinAnsi	U+00C9 É	ISO 8859-8	nulla	Windows-1251	U+0419 Ё
Standard	nulla	ISO 8859-9	U+00C9 É	Windows-1253	U+0399 І
PDFDoc	U+00C9 É	ISO 8859-10	U+00C9 É	Windows-1254	U+00C9 É
ISO 8859-1	U+00C9 É	ISO 8859-11	U+0E29	Windows-1255	U+05B9 `
ISO 8859-2	U+00C9 É	ISO 8859-13	U+00C9 É	Windows-1256	U+0629 ̂
ISO 8859-3	U+00C9 É	ISO 8859-14	U+00C9 É	Windows-1257	U+00C9 É
ISO 8859-4	U+00C9 É	ISO 8859-15	U+00C9 É	Windows-1258	U+00C9 É
ISO 8859-5	U+0429 Ш	ISO 8859-16	U+00C9 É	Mac Roman	U+2026 ...
ISO 8859-6	U+0629 ̂	Windows-874	U+0E29	IBM 037	U+0049 І
ISO 8859-7	U+0399 І	Windows-1250	U+00C9 É	IBM 500	U+0049 І

Una codifica fissata a mano

il codice 0x41 rappresenta U+0416, il cui glifo si chiama Zhecyrillic

il codice 0x42 rappresenta U+03A9, il cui glifo si chiama Omegagreek

il codice 0x43 rappresenta U+20AC, il cui glifo si chiama Euro

il codice 0x44 rappresenta U+00C9, il cui glifo si chiama Eacute

ЖΩ€É

La riga qui sopra è disegnata solo con quei quattro codici. La codifica parte da nessuna codifica, e il file nomina 5 codici che ne differiscono.

A un carattere tipografico privo della legatura U+FB01 si dà «fi» al suo posto, e a uno privo del simbolo di ohm U+2126, «Ω».

Ciò che un carattere tipografico disegna per un carattere che non sa disegnare è il suo glifo 0, che ogni sottoinsieme conserva.