

Una página direccionada con códigos de un byte

Cada letra de las tres líneas de abajo cuesta un byte del flujo de página de este archivo en vez de dos. A una fuente se llega por índice de glifo salvo que se diga otra cosa, dos bytes por glifo, lo que da acceso a cada glifo de la fuente. A través de una codificación se la direcciona en cambio con códigos de un byte: 255 como mucho, cada uno por un carácter, y el archivo indica qué carácter representa cada uno.

Compuesto con códigos de un byte

HQF Development

Una factura, una nómina, un albarán.

Un byte por letra, y el archivo dice cuál.

Los códigos que fijó

La codificación se ajusta a las palabras que dibuja esta página: el espacio conserva el código 32, y cada uno de los demás caracteres toma el código libre más bajo desde el 33, así que nada en el flujo es un byte que el programa de lectura tenga que escapar. Los primeros son H 33 · Q 34 · F 35 · D 36 · e 37 · v 38 · l 39 · o 40.

Lo que pesa el flujo

Dibujadas por índice de glifo, las tres líneas de arriba cuestan 510 bytes de flujo de página; con códigos de un byte, 324 bytes, que es 186 menos. Las dos cifras las mide el programa que dibujó esta página, no se escriben a mano.

Lo que da el byte 32 solo

Palabras que un programa de lectura puede separar.

Palabras que un programa de lectura puede separar.

Las dos líneas de arriba tienen los mismos caracteres. La segunda se dibuja con espaciado de palabras, que mueve la pluma en el byte 32 solo y en ningún otro código. Con índices de glifo ningún código es un byte 32 solo, así que el operador no tiene nada que alcanzar y las palabras no se mueven; el espacio de esta codificación está en el 32, así que sí se mueven.

Una página de códigos nombrada por su norma

Счёт-фактура № 117 — оплата до 30 сентября

La línea de arriba se compone con la página de códigos 1251 de Windows, nombrada una vez en vez de detallada código a código: el archivo indica la codificación Windows y nombra además cada código que la página de códigos asigna de otra forma, así que un programa de lectura encuentra cada letra cirílica donde la pone la página. Sus primeros bytes son D1 F7 B8 F2 2D F4 E0 EA.

Un byte direcciona 255 caracteres a la vez, así que a una página de varios alfabetos se sigue llegando por índice de glifo. Tampoco se dibuja ninguna ligadura ni versalita a través de una codificación: son glifos que ningún carácter representa, y un código representa un carácter. Un archivo que declara conformidad de archivo indica la codificación Windows como aquella con que se leen sus códigos, y este lo hace.

Treinta codificaciones con que puede leerse un código

Lo que representa el byte 0xC9 en cada una: el punto de código del carácter, y el carácter mismo donde esta fuente puede dibujarlo.

WinAnsi	U+00C9 É	ISO 8859-8	nada	Windows-1251	U+0419 Ё
Standard	nada	ISO 8859-9	U+00C9 É	Windows-1253	U+0399 Ĭ
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 ˆ
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 Щ	ISO 8859-16	U+00C9 É	Mac Roman	U+2026 ...
ISO 8859-6	U+0629 ̐	Windows-874	U+0E29	IBM 037	U+0049 Ĩ
ISO 8859-7	U+0399 Ĭ	Windows-1250	U+00C9 É	IBM 500	U+0049 Ĩ

Una codificación fijada a mano

el código 0x41 representa U+0416, cuyo glifo se llama Zhecyrillic
el código 0x42 representa U+03A9, cuyo glifo se llama Omegagreek
el código 0x43 representa U+20AC, cuyo glifo se llama Euro
el código 0x44 representa U+00C9, cuyo glifo se llama Eacute

ЖΩ€É

La línea de arriba se dibuja solo con esos cuatro códigos. La codificación parte de ninguna codificación, y el archivo nombra 5 códigos que difieren de ella.

A una fuente sin la ligadura U+FB01 se le da «fi» en su lugar, y a una sin el signo ohmio U+2126, «Ω».

Lo que dibuja una fuente para un carácter que no puede dibujar es su glifo 0, que todo subconjunto conserva.