

Uma página endereçada por códigos de um byte

Cada letra das três linhas abaixo custa um byte do fluxo de página deste ficheiro em vez de dois. Chega-se a um tipo de letra por índice de glifo salvo indicação em contrário, dois bytes por glifo, o que põe ao alcance cada glifo do tipo. Através de uma codificação é endereçado antes por códigos de um byte: no máximo 255, cada um por um carácter, e o ficheiro indica que carácter cada um representa.

Composto por códigos de um byte

HQF Development

Uma fatura, um recibo de vencimento, uma guia de remessa.

Um byte por letra, e o ficheiro diz qual.

Os códigos que fixou

A codificação ajusta-se às palavras que esta página desenha: o espaço mantém o código 32, e cada um dos outros caracteres toma o código livre mais baixo a partir de 33, por isso nada no fluxo é um byte que o programa de leitura tenha de escapar. Os primeiros são H 33 · Q 34 · F 35 · D 36 · e 37 · v 38 · l 39 · o 40.

O que pesa o fluxo

Desenhadas por índice de glifo, as três linhas acima custam 590 bytes de fluxo de página; por códigos de um byte, 364 bytes, ou seja 226 a menos. Os dois números são medidos pelo programa que criou esta página, não escritos à mão.

O que o byte 32 sozinho permite

Palavras que um programa de leitura pode afastar.

Palavras que um programa de leitura pode afastar.

As duas linhas acima têm os mesmos caracteres. A segunda é desenhada com espaçamento entre palavras, que move a caneta no byte 32 sozinho e em nenhum outro código. Com índices de glifo nenhum código é um byte 32 sozinho, por isso o operador não tem nada a alcançar e as palavras nunca se mexem; o espaço desta codificação está no 32, por isso mexem-se.

Uma página de código nomeada pela sua norma

Счёт-фактура № 117 — оплата до 30 сентября

A linha acima é composta através da página de código 1251 do Windows, nomeada uma vez em vez de detalhada código a código: o ficheiro indica a codificação Windows e nomeia por cima cada código que a página preenche de outra forma, por isso um leitor encontra cada letra ciríllica onde a página a põe. Os seus primeiros bytes são D1 F7 B8 F2 2D F4 E0 EA.

Um byte endereça 255 caracteres de cada vez, por isso a uma página de vários alfabetos chega-se ainda por índice de glifo. Também nenhuma ligadura nem versalete é desenhada através de uma codificação: são glifos que nenhum carácter representa, e um código representa um carácter. Um ficheiro que declara conformidade de arquivo indica a codificação Windows como aquela contra a qual os seus códigos são lidos, e este fá-lo.

Trinta codificações contra as quais um código pode ser lido

O que o byte 0xC9 representa em cada uma: o ponto de código do carácter, e o próprio carácter onde este tipo o sabe desenhar.

WinAnsi	U+00C9 É	ISO 8859-8	nada	Windows-1251	U+0419 Ў
Standard	nada	ISO 8859-9	U+00C9 É	Windows-1253	U+0399 I
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 6
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 6	Windows-874	U+0E29	IBM 037	U+0049 I
ISO 8859-7	U+0399 I	Windows-1250	U+00C9 É	IBM 500	U+0049 I

Uma codificação fixada à mão

o código 0x41 representa U+0416, cujo glifo se chama Zhecyrillic

o código 0x42 representa U+03A9, cujo glifo se chama Omegagreek

o código 0x43 representa U+20AC, cujo glifo se chama Euro

o código 0x44 representa U+00C9, cujo glifo se chama Eacute

ЖΩ€É

A linha acima é desenhada só com esses quatro códigos. A codificação parte de nenhuma codificação, e o ficheiro nomeia 5 códigos que diferem dela.

A um tipo sem a ligadura U+FB01 dá-se «fi» no seu lugar, e a um sem o símbolo de ohm U+2126, «Ω».

O que um tipo desenha para um carácter que não sabe desenhar é o seu glifo 0, que todo o subconjunto conserva.