Las fuentes en el formato PDF

Gustavo Sánchez Muñoz

(Septiembre de 2022)

Parte de la complejidad de los textos y fuentes en el formato PDF tiene una razón cronológica. La tipografía digital y el formato PDF han tenido una larga evolución y algunas soluciones técnicas han sido temporales para caer abandonadas en favor de otras más flexibles.

El formato PDF nació para poder transferir los contenidos impresos sin alteraciones en su forma. Desde su primera versión, donde lo que se buscaba era la integridad de la presentación, cada revisión del formato ha concedido creciente importancia a mantener también la integridad del contenido y, a ser posible su estructura. Eso afecta particularmente a los textos.

Carácter (concepto)	Glifos (formas gráficas)				
Letra a	a	a	a	a	a
Letra m	m	m	m	m	m
Número 4	4	4	4	4	4
Letra z	z	z	z	z	z

Los textos están compuestos por caracteres (cada uno de los signos básicos que componen una escritura). Los caracteres de una escritura son conceptos que se reflejan gráficamente de forma concreta en glifos *(glyphs).*

Estos glifos se organizan por grupos en programas que no pueden funcionar de forma autónoma y que sirven para informar a otros programas como representar los caracteres de los textos. Estos programas o agrupaciones de glifos son las fuentes tipográficas (la muestra de todos los caracteres que componen una fuente es una póliza).

En el formato PDF una fuente es un recurso que va dentro de un diccionario especial, donde se identifica y se proporcionan los datos necesarios para su uso en el documento.

En PDF, los glifos se representan en cada página invocando un diccionario de fuentes. Cada cadena de caracteres, se interpreta como glifos concretos de esa fuente gracias a unos códigos que establecen la relación entre caracteres y glifos en esa fuente concreta.

La incrustación de las fuentes en un PDF

	,	-	.	?	B	C	E
U+0020 CID 32 space	U+002C CID 44 comma	U+002D CID 45 hyphen	U+002E CID 46 period	U+003F CID 63 question	U+0042 CID 66 B	U+0043 CID 67 C	U+0045 CID 69 E
H	L	M	T	U	a	b	c
U+0048 CID 72 H	U+004C CID 76 L	U+004D CID 77 M	U+0054 CID 84 T	U+0055 CID 85 U	U+0061 CID 97 a	U+0062 CID 98 b	U+0063 CID 99 c
d	e	f	g	h	i	l	m
U+0064 CID 100 d	U+0065 CID 101 e	U+0066 CID 102 f	U+0067 CID 103 g	U+0068 CID 104 h	U+0069 CID 105 i	U+006C CID 108 l	U+006D CID 109 m
n	o	p	q	r	s	t	u
U+006E CID 110 n	U+006F CID 111 o	U+0070 CID 112 p	U+0071 CID 113 q	U+0072 CID 114 r	U+0073 CID 115 s	U+0074 CID 116 t	U+0075 CID 117 u
v	x	“	”				
U+0076 CID 118 v	U+0078 CID 120 x	U+201C CID 147 quotedblleft	U+201D CID 148 quotedblright				

Glifos de una fuente usados en un PDF (subjuego)

Cuando en el diccionario se incluye toda la poliza de una fuente, se dice que la fuente se ha incrustado completa, pero cuando sólo van los caracteres que se han usado y ninguno más, se dice que la fuente se ha incrustado como subjuego. Cuando no se incrusta ningún dato sino que se hace referencia a un archivo externo (que debe estar en el sistema), se dice que la fuente está sin incrustar.

Dependiendo de cómo se haya creado el PDF, una misma fuente puede estar incrustada varias veces como subjuegos distintos (que pueden tener la misma codificación o varias distintas).

Si una fuente no está incrustada y no se encuentra en el sistema, en el formato PDF se han previsto mecanismos de seguridad para intentar representar los textos de forma razonable.

Las 14 fuentes básicas o estándar

En todas las versiones del formato PDF se define la existencia de 14 variantes de fuentes, correspondientes a 5 tipografías, que se llaman las 14 fuentes básicas o estándar *(Base 14 fonts* o *Standard 14 fonts).* Las fuentes básicas son una monoespaciada (Courier), una con serifa (Times), una para expresiones matemáticas (Symbol) y una con pequeñas florituras (Zapf Dingbats), en detalle:

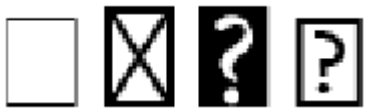
- Courier (Courier, Courier Bold, Courier Oblique, Courier BoldOblique).
- Helvetica (Helvetica, Helvetica Bold, Helvetica Oblique, Helvetica BoldOblique).
- Times Roman (Times Roman, Times Bold, Times Italic, Times BoldItalic).
- Symbol.
- Zapf Dingbats.

Cuando Adobe lanzó el formato PDF (1993), se pensó que definir un conjunto básico de fuentes e incluirlas en todos los programas lectores (que por entonces eran sólo suyos) evitaría sobrecarga de tamaño en los documentos y en su transmisión. Las 14 fuentes eran ese conjunto básico, “que se garantiza que estarán presentes al instalar Acrobat Exchange y Acrobat Reader” (según la misma Adobe en su día). Por eso antiguamente no se incluían en los documentos PDF.

La herencia de esta práctica es que existen muchos PDF en los que estas fuentes no están incrustadas sino meramente referenciadas por el nombre. Como en la actualidad, la presencia de estas fuentes no es siempre segura, es muy conveniente incrustarlas total o parcialmente.

Los glifos faltantes o no definidos (.notdef)

En el formato PDF (como herencia del PostScript), cuando una codificación de texto apunta a un glifo que no está definido en una fuente, el sitio no debe quedar vacío, sino que lo debe ocupar un glifo especial llamado .notdef (de “[glifo] no definido”) para indicar esa ausencia.



Todas las fuentes deben contener un glifo 0, que es ese glifo `.notdef`. Su forma no es obligatoria pero la convención es usar las que se ven arriba o similares.

Principales tipos de fuentes

A día de hoy, el formato PDF admite sólo los siguientes formatos de fuentes: TrueType, Tipo 1, Tipo 3, Fuentes compuestas (de Tipo 0, que son CIDFontType0 (Tipo 1) y CIDFontType2 (TrueType), Multiple Master y OpenType (a partir del nivel 1.6 del formato PDF). La práctica usual con otro tipo de fuentes al crear un PDF es transformarlas en uno de esos tipos admitidos; por ejemplo, Tipo 3.

De hecho, las fuentes que están en un documento no siempre pasan directamente, tal cual, al PDF, sino que pasan por algún tipo de proceso y conversión. Muchos programas no pasan la fuente con su nombre original, sino que les asignan un nombre nuevo con algún identificador único o, como en el caso de Acrobat Distiller, usando su nombre PostScript *(PostScript name)* original, si está disponible).

Estos son los principales tipos de fuentes que podemos encontrar, aunque, como ya hemos indicado, varios de ellos no se pasan *tal cual* a un PDF:

Fuentes de mapas de bits

En este formato, cada carácter es un rectángulo de píxeles donde los valores de cada píxel dibujan la forma del glifo, que al estar formado de esa manera, tiene una resolución óptima de reproducción. Son fuentes bastante antiguas y de uso muy limitado. Por eso su codificación también puede plantear problemas.

Los programas usualmente las incorporan a un PDF convirtiéndolas fuentes de Tipo 3 (aunque los programas de menos calidad las rasterizaron directamente, con lo cual caen fuera de esta discusión porque ya no son textos).

Tipo 1

Éste es el primer formato de fuentes vectoriales que acompañó a la revolución de la autoedición en 1984 y están presentes en el formato PDF desde su inicio. Pese a su nombre, las fuentes de Tipo 1 sólo pueden usar algunas características del lenguaje PostScript. Por razones comerciales, su uso inicial estaba reservado a firmas comerciales, por lo que solían tener bastante calidad y eran sinónimo de "imprimir sin problemas".

Cada variante de una familia está compuesta de dos archivos y no son multiplataforma (es decir, no se puede usar una fuente de Macintosh en Windows o viceversa). Se pueden usar directamente en un documento, (incrustadas como subjuego, completas o sin incrustar).

Al igual que con las fuentes TrueType, las fuentes Tipo 1 van desapareciendo en favor de OpenType. Adobe ha anunciado que sus programas dejarán completamente de operar con este formato a inicios de 2023 (los PDF que tengan incrustadas fuentes de este Tipo seguirán mostrándose e imprimiéndose sin problemas).

Tipo 3

Hasta hace poco, para encontrar una fuente de Tipo 3 en un PDF, debíamos examinar documentos muy antiguos y era algo inusual. En origen, las fuentes PostScript de Tipo 3 se usaba especialmente para fuentes decorativas, ya que admitía más características del lenguaje PostScript que el formato Tipo 1, además de admitir la incorporación de mapas de bits. Por razones histórico comerciales también era posible encontrar algunas fuentes *normales* que se habían creado como Tipo 3.

Sin embargo, con la aparición relativamente reciente de fuentes como las OpenType SVG, encontrar documentos con fuentes de Tipo 3 ya no es tan raro. El motivo es que actualmente el formato PDF no admite la incrustación directa de fuentes SVG, por lo que lo hace considerándolas y tratándolas como fuentes Tipo 3.

(Cuando los programas creadores encuentran fuentes de formatos no admitidos directamente que incorporan órdenes, las consideran y tratan como Tipo 3, lo que las convierte en una especie de cajón de sastre).

Una fuente de este tipo puede incluir mapas de bits (objetos de píxeles), máscaras, degradados, patrones, órdenes de color y similares. En general, Por sus características especiales, su rareza y sus características pueden dar problemas técnicos o de calidad en su procesado para impresión.

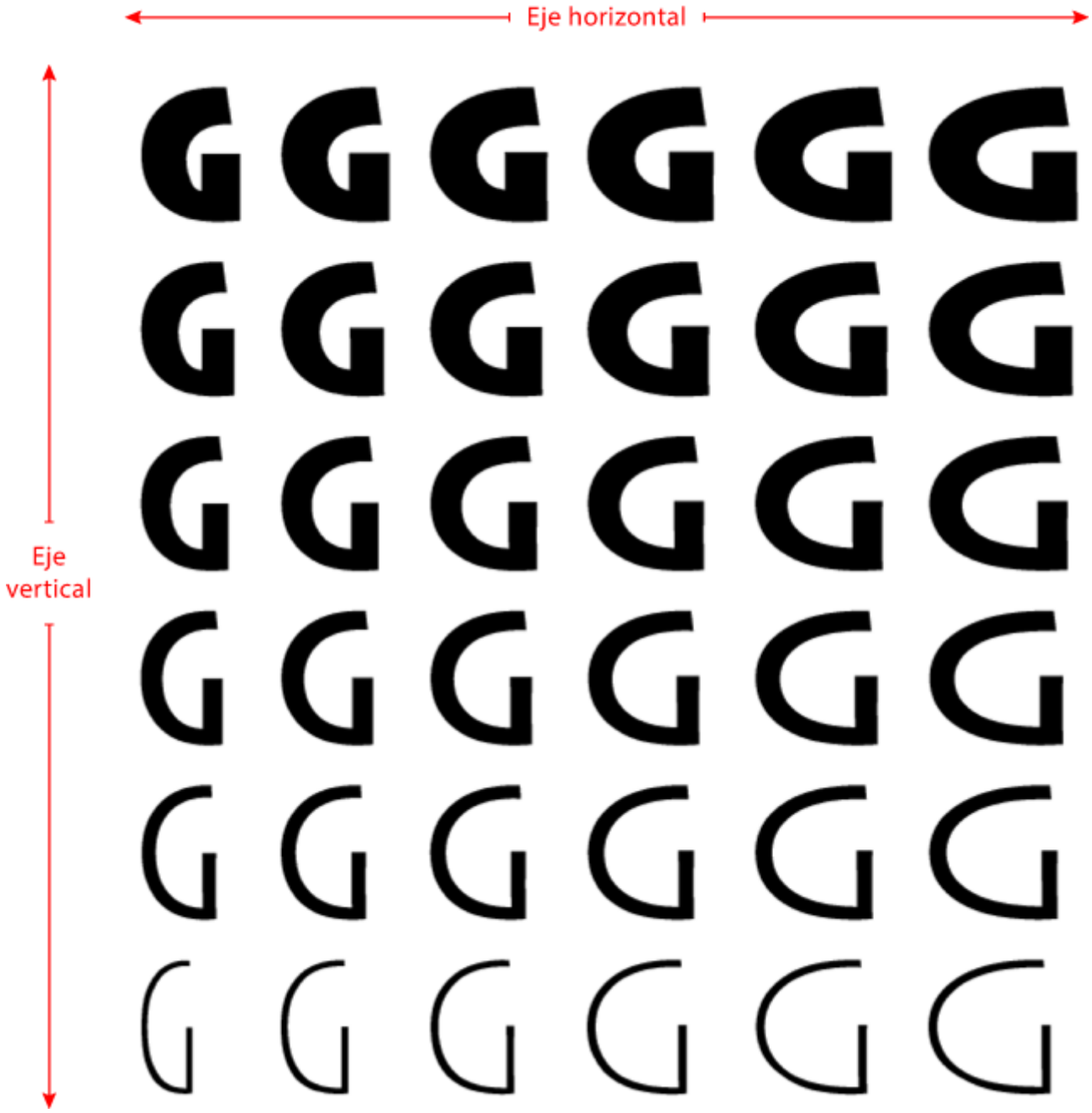
TrueType

Es un formato de fuentes vectoriales lanzado en 1991 por Apple y apoyado por Microsoft para competir con las PostScript Tipo 1. Cada variante de una fuente está formada por un único archivo multiplataforma (se puede usar indistintamente en Windows o Macintosh).

Por razones que no vienen al caso TrueType ha sido el formato más usual en los ordenadores personales durante muchos años y la mayoría de las tipografías gratuitas creadas por aficionados o directamente versiones piratas de baja calidad recurren a este formato. Eso (y no una supuesta baja calidad técnica del formato en si, que no es cierta) hace que muchos de los problemas de fuentes de baja calidad se den con este formato, lo que ha hecho que numerosos profesionales lo miren con recelo a priori.

Al igual que con las fuentes PostScript Tipo 1, TrueType va desapareciendo en favor de OpenType (las fuentes OpenType de base TrueType conservan la extensión de archivo "`.ttf`").

Multiple Master



Es un tipo de fuentes vectoriales creadas por Adobe en 1992 como extensión de las PostScript Tipo 1. Su principal característica era que, interpolando valores a partir de algunos ejes de unas *fuentes maestras,* permitía a los usuarios la creación y uso inmediato de tipografías de peso y anchura variables. Su desarrollo se abandonó con la aparición de las fuentes OpenType. Sus recientes sucesoras son las fuentes OpenType variables.

Compuestas (Tipo 0)

Este tipo de fuentes *(composite fonts)* fueron una forma de solucionar las limitaciones que tenían en origen las fuentes PostScript, que sólo admitían 256 caracteres o glifos.

Para poder atender las necesidades de escrituras e idiomas como el chino, japonés o coreano, que tienen una cantidad de caracteres muy superior, Adobe desarrolló la idea de una fuente que agrupa y contiene varias fuentes (de ahí que sean fuentes “compuestas”).

Las fuentes compuestas aparecieron en el nivel 2 del lenguaje PostScript pero su comportamiento en el interior del formato PDF tiene algunas leves diferencias. En el formato PDF una fuente compuesta va en un diccionario en cuyo interior se organizan jerárquicamente, como si fuera un árbol, las fuentes que van incluidas. La fuente *madre* en la jerarquía se identifica como Tipo 0. Las fuentes no compuestas se consideran “fuentes simples”.

Las fuentes *hijas* dentro de una fuente compuesta pueden ser simples u otras fuentes compuestas. A las primeras fuentes que siguen este concepto se las denomina “fuentes compuestas originales” *(Original Composite Fonts: OCF).*

Fuentes CID

Con el nivel 3 de PostScript apareció un tipo nuevo de fuentes compuestas llamadas “fuentes CID” o “fuentes con clave CID” *(CID-keyed fonts).* Este tipo de fuentes se pensó sobre todo para escrituras asiáticas con un gran número de caracteres, como la china, por ejemplo.

Esas fuentes se basan en una colección de caracteres que se relacionan mediante números identificadores de caracteres *(Character Identifier: CID)* con un conjunto de glifos. En este sistema, los glifos no tienen nombre sino ese identificador numérico.

Lo principal a efectos de este resumen es que lo que en las acciones de PitStop se denomina “`fuentes compuestas T1`” son las de tipo CidFontType0, que aplican el concepto de fuentes compuestas CID a las fuentes PostScript de Tipo 1, mientras que las llamadas “`fuentes compuestas TT`” son las de tipo CidFontType2, que hacen lo mismo con fuentes TrueType.

Nada impide que las fuentes OpenType usen este método (y entonces se consideran fuentes de Tipo 2, porque recurren a un formato de descripción de glifos llamado Formato Compacto de Fuente (CFF), propio de ese tipo).

Por decisiones de programación diversas, algunas versiones de algunos programas pasan al PDF fuentes originalmente no compuestas como fuentes compuestas.

OpenType

Este formato de fuentes vectoriales es una evolución a partir de los formatos TrueType y Tipo 1, obra conjunta de Microsoft y Adobe, publicada en 1996 para superar las limitaciones de ambos formatos.

Como formato ha ido transformándose y, de momento, es el tipo de fuentes digitales más recomendable , que pretende recoger todas las mejoras de la tipografía digital.

Todas las fuentes OpenType se atienen el estándar Unicode, ideado para una amplísima capacidad multidioma y multiescritura estandarizada. Su capacidad de incluir muchos glifos hace que los archivos puedan ser comparativamente muy grandes y complejos. La aplicación concreta de Unicode en cada fuente depende del fabricante y del nivel de ambos estándares (OpenType y Unicode) en el momento de su creación.

Sin entrar en detalles técnicos, una fuente OpenType puede tener una estructura interna similar a las fuentes TrueType o a las fuentes de Tipo 1 (como formato CFF (Compact Font Format), no son fuentes Tipo 1). Los archivos de fuentes del primer tipo (TrueType) mantienen la extensión "`.ttf`", mientras que los de formato CFF *herederos* del Tipo 1 tienen la extensión "`.otf`"

La incrustación nativa de fuentes OpenType, como tales sólo es posible a partir del nivel 1.6 del formato PDF y a día de hoy los principales programas no lo hacen. La alternativa es pasar la fuente a los diccionarios del PDF como formato TrueType o Tipo 1, dependiendo de su origen (así lo hacen Illustrator o InDesign, por ejemplo). Por eso, al examinar un PDF en el que se han usado fuentes OpenType es muy inusual encontrarlas como tales.

Lo que suelen hacer los programas al crear un PDF, es tomar los datos del los glifos que están en formato CFF dentro de una fuente OpenType e incluirlos en el archivo sin el resto de datos propios de la fuente, que son innecesarios. Esa reducción y que los datos sean CFF es lo que hace que los programas de lectura de PDF identifiquen el recurso de fuente resultante como Tipo 1 o Tipo 0

(dependiendo se su estructura original).

OpenType-SVG

Este formato es una creación de Adobe y Mozilla que se estabilizó hacia 2016. De momento hay varios formatos, todos ellos compuestos de un único archivo por cada variante tipográfica. Como base para dibujar los glifos usa archivos SVG, que son vectoriales pero admiten la incrustación de imágenes de píxeles, el uso de degradados, patrones y colores (de ahí que también se las llame *Color Fonts).* Se atienen a las especificaciones OpenType y siguen la codificación Unicode en su variante UTF-8.

Aunque está en estudio la posibilidad, el formato PDF no las incorpora directamente. En el momento de la creación de un documento, los programas las convierten a fuentes de Tipo 3.

Advertencia: Este formato de fuentes pueden producir problemas en el procesado para impresión por las características complejas que algunas de ellas tienen. Y no siempre se puede recurrir a la solución de convertirlas a trazados o rasterizarlas automáticamente.

Fuentes variables OpenType

Este formato es un desarrollo de las fuentes OpenType que recoge con mejoras la posibilidad de las fuentes Multiple Master de modificar el peso, inclinación, anchura y otros parámetros de una fuente para crear variantes intermedias.

En principio pasan al PDF como fuentes de Tipo 1 o TrueType, dependiendo de su origen.