

Compilando um arquivo PDF

O conceito de compilação, tomado em sentido largo significa a tradução de uma linguagem para outra. Nesta aula, este processo vai ser estudado como sendo a tradução de uma imagem de página impressa em um arquivo magnético que a descreve. É óbvio que fazer esta tradução tem inúmeras vantagens, das quais algumas estão abaixo escritas.

Mas a principal parece ser que o nosso pequeno planeta certamente não vai aguentar o ritmo de consumo de papel se este continuar o crescimento experimentado ao longo do século XX. (um trabalhador médio de escritório nos EUA costuma imprimir 12.000 páginas/ano. Fonte: VEJA, ed. 1628).

Se você imaginar que a documentação de um Boeing 747 em papel não cabe dentro dele, parece evidente que uma alternativa tem que ser encontrada.

Anatomia de um arquivo PDF

O arquivo PDF é uma coleção de objetos, que descrevem características e conteúdos – tanto gráficos quanto textuais – do material a ser mostrado. Tenha em mente o objetivo do arquivo PDF: ser um padrão de apresentação de conteúdos digitais válido para todas as plataformas possíveis (impressora, gráfica, monitor, celular, tablet, ...). Ele não é um formato de edição de texto (tipo WORD) nem de imagem (tipo COREL), nem é um formato hiper-comprimido de armazenamento de BLOBS (binary large objects). A melhor definição de um arquivo PDF é a do papel virtual.

Embora a grande maioria dos arquivos PDF seja comprimida, valem aqui 2 observações:

- * O arquivo continua perfeitamente legível (via PDF readers) se ele for completamente descomprimido.
- * O que é comprimido é uma parte de alguns objetos do PDF. Como é a maior parte em termos de tamanho, o resultado final é bom. Mas, não impede que o arquivo comprimido seja examinado por um editor plano tipo notepad.

Uma página PDF é formada por um ou mais fluxos de conteúdo definidos pela entrada `/Contents` no objeto da página, além de um conjunto compartilhado de recursos definidos pela entrada `/Resources`. Dentro de `/Contents` há um dicionário de fluxo e depois os dados do fluxo. Dentro de `stream` vem-se inúmeros comandos de produção gráfica.

Gráficos Definido o espaço de trabalho no `/MediaBox` 0 0 595 842, na unidade de pontos (1 ponto = 0.341mm) para uma página A4, com um sistema de eixos cartesianos com o ponto (0,0) localizado no canto inferior esquerdo, com eixo *x* na horizontal saindo para a direita e *y* na vertical em direção ao alto.

m	x y m	movimenta a “pena” até a posição x,y
l	x y l	prepara uma linha desde onde a pena estava até o ponto x,y
S	S	traça a linha
w	t w	altera a largura da linha para o valor <i>t</i>
J	t J	altera o arremate da linha: 0=quadrado, 1=redondo
g	cor g	altera a cor

Veja-se um exemplo:

```
100 100 m 300 200 l 700 100 l
//vai para 100,00 linha em 300,200...
S // traça a linha
8 w // altera a largura da linha de 1 para 8
1 J //altera o arremate da linha para 1
100 200 m 300 300 l 700 200 l // novo path
S // traça
[20] 0 d // altera para tracejados de 20 pontos
```

Fontes Um dos recursos de uma página PDF é o recurso fonte (que vem a ser a coleção de tipos de um alfabeto com o mesmo jeitão visual). Veja-se um exemplo de um recurso fonte:

```
1 0 obj
<< /Type /Font /Subtype /Type1 /BaseFont
/Times-Roman /Firstchar 0 /Lastchar 255
/Widths [255 255 ...]
/FontDescriptor 2 0 R
/Encoding /WinAnsiEncoding
>>
```

`/Widths` fornece em milésimos de pontos a largura de cada caractere.

Ao criar um arquivo PDF as fontes devem ser incorporadas de modo que as descrições e codificações dos glifos estejam disponíveis na apresentação do documento. Ao fazer isto o PDF remove do arquivo da fonte tudo o que for agregado ao PDF e também remove do arquivo as informações dos caracteres que não forem usados. Tudo isto para diminuir o tamanho.

As fontes podem ser diversos tipos:

- * Tipo 1 (`/Subtype /Type1`): Fontes da Adobe originalmente destinadas ao formato postscript. No PDF há 14: Times Roman, Times Bold, Times Italic, Times Bold Italic, Helvetica, Helvetica Bold, Helvetica Oblique, Helvetica Bold Oblique, Courier, Courier Bold, Courier Oblique, Courier Bold Oblique, Symbol e Zapf Dingbats.

- * TrueType (`TrueType`): Fontes da Apple utilizados também no Windows
- * Tipo 3: compostas por fluxos de operadores gráficos PDF. Inclui cor e sombras, mas são inadequados para tamanhos pequenos.
- * Há outros tipos...

Texto

Textos e Fontes

comando cm: posiciona o local base da escrita
translação em dx,dy: 1,0,0,1,dx,dy
redimensionamento (sx,xy): sx,0,0,sy,0,0
rotação anti-horário
em x radianos com centro
em 0,0: cosx,senx,-senx,cosx,0,0.

No caso mais simples: 1. 0. 0. 1. 50. 700. cm

```
BT
/F0 36. Tf
(Ola Mundo Cruel!) Tj
...
ET
```

Alguns comandos de texto

Tj - apresenta o string
T* - avança à próxima linha
1 0 0 0 1 x y Tm - define o início do texto
Tl - espaço entre linhas
Tc - espaço entre caracteres
Tw - espaço entre palavras
Ts - sobe e desce na linha base
TJ - igual a Tj mas aceita ajustes de kerning por exemplo (P)150(J)

coração -> [(cora\030)445(c~)500(ao)] TJ

O que deve ter no programa

```
BT
/F0 36. Tf
50 TL
(123456789012345678901234) Tj T*
(2 ) Tj T*
(3 ) Tj T*
(4 ) Tj T*
(5 ) Tj T*
(6 ) Tj T*
(7 ) Tj T*
(8 ) Tj T*
(9 ) Tj T*
(10 ) Tj T*
(11 ) Tj T*
(12 ) Tj T*
(13 ) Tj T*
(14 ) Tj T*
ET
```

☞ Para você fazer

Você vai receber um arquivo TXT plano, sem nenhuma acentuação em português, contendo registros (linhas) de 24 caracteres. Haverá no máximo 140 linhas nesse arquivo.

Deve escrever um programa denominado `PPDF0001` que construa um arquivo PDF a partir do nada, incluindo páginas obtidas a partir da entrada recém lida. Deve usar a fonte `Courier` com tamanho de letra de 36 pontos. A página A4 deve ser formatada no seu tamanho padrão e o início da escrita na página deve ser no ponto 50 700. A separação entre linhas deve ser de 50 pontos. Com essas características devem caber numa página A4 14 linhas por 24 caracteres. Esse é o volume de dados que caberá em cada página. A primeira coisa que seu programa deve decidir é *quantas páginas de saída serão necessárias*. Note que o número máximo de páginas é 10.

Gerado o primeiro PDF, você poderá usar o PDFTK para acertar os tamanhos e endereços. Feito isso, seu arquivo deverá poder ser aberto e examinado através de um PDF reader convencional. É assim que o seu trabalho vai ser avaliado.

O nome do arquivo que você deve ler é `U10A0001.ENT` e para sua ciência, estas são as duas primeiras linhas do conteúdo

biblioteca cair ou esta
apenas onde na cheese

