IMPLEMENTACJE

Rendez-vous z PDFTEX

Tomasz Przechlewski

Program PDFTEX, zaprezentowany w poprzednim numerze biuletynu [3], umożliwia generowanie pliku PDF (*Portable Document Format*) zamiast pliku DVI. W niniejszym artykule opisano jak zainstalować i jak rozpocząć pracę z PDFTEX-em.

Instalacja i uruchomienie

Dystrybucja PDFTEX-a znajduje się w katalogu `systems/pdftex` każdego archiwum CTAN. Źródła programu wraz z niezbędnymi dodatkowymi plikami zawarte są w pliku `pdftex.tar.gz`; programy wykonywalne umieszczono w podkatalogu `bin/`, w tym m.in.:

- `DJGPP/pdftex-DJGPP.zip` zawiera wersję dla Windows 95/NT lub DOS-a (do uruchomienia programu potrzebny jest emulator pamięci `cwsdpmi.exe`);
- `win32` zawiera tylko plik README, w którym można przeczytać, że reszta plików znajduje się w: `ftp://ftp.ese-metz.fr/pub/TeX/win32/pdftex-win32.exe.gz`. Jest to wersja na Windows 95/NT przygotowana przez Fabrice'a Popineau;
- `linux/pdftex-bin.tar.gz` zawiera wersję dla Linuxa.

Do zainstalowania programu potrzebny jest plik `pdftex.tar.gz` oraz plik zawierający odpowiedni program wykonywalny. W `pdftex.tar.gz` interesuje nas – jeżeli nie chcemy kompilować programu – tylko zawartość katalogu `web2c/libpdf/` etc. W szczególności znajdują się tam następujące pliki (opisane bardziej szczegółowo poniżej): `texmf.cnf`, `pdftex.map`, `pdfcolor.tex` oraz kilkadziesiąt plików `*.enc`. Pozostałe kilka plików z rozszerzeniem `*.tex` z tego katalogu zawiera przykłady wykorzystania PDFTEX-a.

Konfigurację rozpoczynamy od nadania prawidłowych wartości kilku zmiennym środowiskowym. W systemie DOS/Windows zmienne te można ustalić bezpośrednio poleceniem `SET` lub też nadać im odpowiednie wartości w pliku konfiguracyjnym `texmf.cnf`. Jeżeli zmienne podajemy w pliku `texmf.cnf` to wystarczy jedna zmienna środowiskowa wskazująca katalog (katalogi), zawierające ten plik, np: `SET TEXMFCNF=c:\tex\tex2pdf`.

Plik `texmf.cnf` zawiera kilkadziesiąt różnych zmiennych ale do rozpoczęcia pracy wystarczy nadanie prawidłowych wartości następującym:

`t1fonts` – zmienna wskazująca na katalogi zawierające fonty. Zapis `c:\psfonts\\` oznacza, że przeszukiwane są katalogi `\psfonts` i wszystkie jego podkatalogi (odpowiednik konwencji dwóch znaków wykrzyknika `!!` z dystrybucji emTEX-a). PDFTEX może wykorzystać tylko PostScript-owe fonty typu pierwszego, nie obsługuje innych formatów. W praktyce nie jest to wielkie ograniczenie, ponieważ wykorzystanie tradycyjnych, bitmapowych fontów TEX-owych nie daje dobrych rezultatów.

`vffonts` – zmienna wskazująca na katalogi zawierające fonty wirtualne.

`tfmfonts` – zmienna wskazująca na katalogi zawierające fonty `tfm`.

`texinputs` – zmienna określająca katalogi zawierające pliki dołączane poleceniem `\input`.

`texformats` – zmienna wskazująca na katalogi zawierające formaty.

`texpool` – zmienna wskazująca na katalog zawierający plik `pdftex.pool`. Podczas „normalnej" pracy z programem zmienna `texpool` nie jest wykorzystywana, jest ona potrzebna tylko podczas generowania formatu. Uwaga dla użytkowników Windows 95: PDFTEX z dystrybucji `win32` szuka pliku `pdftex.pool` a nie `pdftex.poo`! Jeżeli np. podczas kopiowania obetniemy rozszerzenie nazwy do trzech znaków to nie uda się wygenerować formatu.

`texpsheaders` – zmienna wskazująca na katalogi zawierające `pdftex.map` oraz rysunki w formacie png, który jest jedynym formatem graficznym obsługiwanym przez PDFTEX.

Następnie odpowiednio modyfikujemy zawartość pliku konfiguracyjnego `pdftex.map`. Podobnie jak w wypadku pliku `psfonts.map` dla sterownika `dvips`, informacje zawarte w `pdftex.map` są niezbędne do prawidłowego korzystania z fontów. Plik ten zawiera wiersze postaci:

```
qplr QuasiPalladio-Regular 4
                 <qplr.pfb qpl.enc
plr10 PLR10 4 </gslib/pl/plr10.pfb
```

Pierwsze pole określa nazwę TEX-ową fontu, czyli nazwę pliku `tfm`, drugie to nazwa PostScript-owa, kolejne pole zawiera liczbę określającą typ fontu. Znaczenie tej liczby jest opisane w [1] (punkt 7.9.2 *Font descriptor flags*, s. 164). Regułą kciuka jest wstawienie tu 4 (*script font*) co powinno upewnić przeglądarkę, że font jest nietypowy i ma go po prostu wyświetlić korzystając z informacji dołączonej do pliku PDF, zamiast starać się szukać jeszcze lepszego rozwiązania. Jeżeli używamy innego kodowania fontu niż wbudowane, należy jako ostatnie pole podać ⟨*plik.enc*⟩, który zawiera wektor kodowania (w powyższym przykładzie font `qplr` jest przekodowany, a `plr10` nie). W dystrybucji znajduje się kilkadziesiąt plików `*.enc` dla różnych publicznie dostępnych fontów.

W dystrybucji programu znajduje się także zestaw makr ułatwiających korzystanie z kolorów, `pdfcolor.tex`. Kopiujemy ten plik do odpowiedniego katalogu zawierającego makra.

Instalowanie PDFTEX-a w systemie Linux wygląda prawie identycznie. Zakładając, że mamy zainstalowaną dystrybucję teTEX-a, kopiujemy `pdftex` do katalogu `$TETEXDIR/bin/i586-linux`[1] natomiast `pdftex.pool` do katalogu, w którym znajdują się pliki `pool`, tj. `$TETEXDIR/texmf/web2c`. W miarę potrzeb modyfikujemy pliki `pdftex.map` i `texmf.cnf`, korzystając z pliku `texmf.cnf` z dystrybucji PDFTEX.

Ostatnią czynnością jest wygenerowanie formatu. W wypadku systemu Linux przy generowaniu formatu problemem może być zbyt mała domyślna wartość zmiennej `pool_size` w oryginalnym (teTEX-owym) pliku `texmf.cnf`. Aby uniknąć wystąpienia błędu, edytujemy plik `texmf.cnf` wstawiając `pool_size=150000`. Format generujemy poleceniem, np.:

```
pdftex -i \input mex.tex \dump
```

W ten sposób otrzymujemy plik `mex.fmt`, polski odpowiednik formatu `plain`. Ponieważ powstały plik nazywa się `mex.fmt`, a nie jest „normalnym” formatem, należy zmienić jego nazwę, np. na `pdfmex.fmt`. Tworzenie formatu LATEX jest analogiczne. Formaty kopiujemy do odpowiednich katalogów. W systemie Linux jest to zazwyczaj katalog `$TETEXDIR/texmf/web2c`. Należy pamiętać o utworzeniu odpowiednich dowiązań.

PDFTEX jest w istotny sposób zależny od systemu `web2c`, w związku z czym nie ma możliwości przekodowania znaków na wejściu/wyjściu tak, jak ma to miejsce np. w dystrybucji emTEX-a. Polskie znaki diakrytyczne wprowadzamy za pomocą notacji ciachowej lub korzystamy z rozwiązań typu LATEX-owy `inputenc.sty`[2].

Tworzenie dokumentów hipertekstowych

Program może pracować w dwóch trybach. W trybie pełnej zgodności z TEX-em oraz w trybie rozszerzonym. W pierwszym trybie, w wyniku działania PDFTEX-a, powstaje normalny plik DVI; w drugim – plik PDF. Oczywiście tylko w trybie rozszerzonym możliwe jest wykorzystanie dodatkowych instrukcji wbudowanych (*primitives*), służących do definiowania: odsyłaczy, zakładek, notek i innych obiektów związanych z formatem PDF. Przejście do trybu rozszerzonego jest deklarowane za pomocą:

```
\pdfoutput=1
```

Deklaracja ta powinna być wykonana *przed* wysłaniem przez PDFTEX-a pierwszej strony do pliku wynikowego.

Ponadto w preambule dokumentu należy umieścić deklarację `\pdfcompresslevel` = ⟨n⟩, gdzie ⟨n⟩ $\in \{0,\ldots,9\}$ określa stopień kompresji pliku PDF. Skrajna wartość 0 powoduje, że plik nie jest kompresowany, wartość 9 powoduje maksymalną kompresję (zalecane). Z uwagi na ograniczenia patentowe, związane z algorytmem LZW, nie jest on zaimplementowany. PDFTEX wykorzystuje metodę określaną jako *flate encode*, która zresztą przeciętnie daje dużo lepsze rezultaty[3]. Kompresja *flate encode* jest zdefiniowana w specyfikacji formatu PDF 1.2, ale nie 1.1, co oznacza, że np. Acrobat

1: `$TETEXDIR` jest korzeniem drzewa instalacyjnego teTeX-a. W przypadku dystrybucji Redhat jest to `/usr/lib/texmf`.

2: Piotr Bolek doniósł, że PDFTEX działa także bez problemów z patchem Petra Olšaka (encTEX – `ftp://math.feld.cvut.cz/pub/olsak/enctex`), realizującym przekodowanie literek ze standardu zewnętrznego na wewnętrzny.Wymaga to oczywiście skompilowania PDFTEX-a, ale jest opłacalne, bo znacznie wygodniej się go potem używa.

3: Jest to ten sam algorytm kompresji, który jest wykorzystywany w programie gzip itp.

Reader w wersji 2.1 i wcześniejszej nie będzie umiał wyświetlić pliku.

Jeżeli używamy formatu LaTeX, to wykonanie dokumentu w formacie PDF wymaga dodania do preambuły dokumentu deklaracji:

```
\usepackage[pdftex]{hyperref}
```

Użytkownicy MeX-a muszą radzić sobie sami. Plik `example.tex` zawiera przykłady wykorzystania wszystkich instrukcji wbudowanych, związanych z PDF-em. Można także spróbować wykorzystać pakiet `tp-pdf`, opisany w poprzednim numerze biuletynu, który także potrafi od pewnego czasu współpracować z programem PDFTeX.

Grafika

Najpoważniejszym ograniczeniem PDFTeX-a jest możliwość dołączania grafiki tylko i wyłącznie w formatach PNG, PDF oraz plików PostScript-owych wygenerowanych METAPOST-em. Z uwagi na to, że standardem w świecie TeX-owym jest raczej format EPS, konieczna jest odpowiednia konwersja, przy czym jakość tej konwersji z reguły nie jest najlepsza. Do konwersji plików EPS do formatu PNG można zastosować programy `Ghostscript` lub `netpmb`.

Pliki PNG są jednynym „naturalnym" formatem, który rozumie PDFTeX, natomiast rysunki zapisane w formacie PDF i wynikowe METAPOST-a mogą być wstawiane przy pomocy pakietu makr TeX-owych(!) autorstwa Hansa Hagena. Jako przykład przedstawię, jak można wykonać PDF-ową wersję deklaracji członkowskiej GUST-u. Rozpoczynamy od konwersji (za pomocą `Ghostscript-a`) pliku `gustlogo.eps` do formatu PDF:

```
ps2pdf gustlogo.eps  gustlogo.pdf
```

Konwersja taka daje najlepszy wynik, jeżeli plik PostScript-owy zawiera tylko obwiednie, a nie tekst, czy mapy bitowe. Następnie w preambule dokumentu umieszczamy

```
\input supp-pdf
\input supp-mis
```

Przygotowane uprzednio logo wstawiamy w odpowiednie miejsce instrukcją:

```
\convertPDFtoPDF{gustlogo.pdf}
 {.195}{.195}
 {0pt}{0pt}{112.79pt}{71.09pt}
```

Argumenty drugi i trzeci określają odpowiednio przeskalowanie w poziomie i pionie, czwarty i piąty to przesunięcie poziome i pionowe i wreszcie szósty i siódmy to szerokość i wysokość rysunku. Do konwersji wynikowego pliku PostScript-owego METAPOST-a służy instrukcja:

\convertMPtoPDF{⟨*plik*⟩}
  {⟨*przeskalowanie-x*⟩}{⟨*przeskalowanie-y*⟩}

Makra `supp-pdf` są już w tej chwili częścią LaTeX-owego pakietu `graphics`. W preambule dokumentu należy umieścić deklaracje:

```
\usepackage[pdftex]{graphicx}
\DeclareGraphicsExtensions{.pdf,.png,.mps}
```

Rysunki są dołączane za pomocą standardowej komendy pakietu:

\includegraphics{⟨*plik*⟩}

Porównanie z innymi programami

Stopień kompresji otrzymanego pliku wynikowego jest jednym z podstawowych kryteriów oceny programów generujących pliki w formacie PDF. Z uwagi na potencjalne zastosowanie sieciowe, im mniejszy plik, tym lepiej.

Deklaracja GUST-u wygenerowana za pomocą PDFTeX-a z wykorzystaniem pakietu makr `supp-pdf` miała 217 kB, podczas gdy odpowiedni plik wygenerowany za pomocą Distillera 126 kB.

Wyniki dla przykładu, którym był artykuł Davida Carlisle z tego numeru biuletynu zamieszczono w tabeli:

Program	Wielkość pliku (kB)
Acrobat 3.0	110
Acrobat 2.1	169
PDFTeX	228
Ghostscript 5.03	1668
Ghostscript 5.10	338
PostScript	596

Trzecim dokumentem testowym był „Wstęp do METAPOST-a" w tłumaczeniu J. Marszałkowskiej, dostępny w `www.gust.org.pl`. Dokument ten zawiera 9 rysunków dołączonych, w przypadku generowania pliku programem PDFTeX, za pomocą makr z pakietu `supp-pdf`. Wielkość pliku PDF wygenerowanego za pomocą różnych programów przedstawiono w tabeli na następnej stronie.

Widać wyraźnie, że najlepsze wyniki osiągnięto w przypadku użycia Acrobat Distillera – plik wykonany za pomocą PDFTeX-a jest prawie dwa razy

Program	Wielkość pliku (kB)
Acrobat 3.0	–*
Acrobat 2.1	302
PDFTEX	640
Ghostscript 5.03	2743
Ghostscript 5.10	750
PostScript	1200

* Pliku nie udało się wygenerować. Distiller zgłasza błąd w pliku PostScript-owym.

większy. Ghostscript w wersji 5.03 tworzy plik PDF w bardzo prosty sposób oraz nie potrafi prawidłowo dołączać fontów, zamieniając je po prostu na mapy bitowe (za wyjątkiem czternastu podstawowych, tj. Times, Helvetica i Courier), stąd ogromna wielkość pliku. Nowa wersja tego programu (5.10), aczkolwiek dalej nie potrafi prawidłowo dołączyć fontów, tworzy znacznie mniejsze pliki. Autorzy Ghostscript-a zapowiadają radykalną poprawę już w wersji 5.60, spodziewanej w pierwszej połowie 1998 r.

Podsumowanie

PDFTEX to bardzo interesujący program, który jednak w obecnej wersji ma kilka poważnych braków. O dwóch już wspomniałem: niemożność korzystania z grafiki PostScript-owej oraz generowanie zauważalnie większych plików niż np. Acrobat Distiller. Brakiem jest także uproszczona w stosunku do oryginalnej specyfikacji formatu [1] implementacja odsyłacza. Pominięto możliwość zdefiniowania odesłania do dowolnego miejsca na stronie (czemu odpowiada nadanie obiektowi `Dest` wartości `XYZ` [1, s. 95]), co akurat jest najbardziej precyzyjnym sposobem zdefiniowania, w jaki ma być wyświetlona strona, do której nastąpi odesłanie.

Poważny problem zauważył Thierry Bouche, który testował PDFTEX-a na komputerach Macintosh i stwierdził, że „sposób w jaki fonty są dołączane do pliku PDF powoduje, że kombinacja programów ATM / Acrobat Reader w ich obecnych wersjach (na Maca) nie umie ich prawidłowo wyświetlić”. W rezultacie, każdy znak w foncie tekstowym nie jest wyświetlany, albo zamiast niego wyświetlany jest znak `bullet` z fontu Helvetica. Wyświetlane są tylko symbole matematyczne i ligatury.

Twórcy PDFTEX-a nie bardzo wiedzą jak to poprawić, zresztą wcale nie musi to być związane z błędem PDFTEX-a ale raczej z działaniem programów ATM / Acrobat, które – wszystko na to wskazuje – mogą poprawnie działać tylko z fontami typu pierwszego uwzględniającymi pewne dodatkowe założenia, oprócz tych wynikających ze specyfikacji formatu [2].

Drobniejsze błędy / niezgodności z działaniem TEX-a są nieuniknionym rezultatem modyfikowania jego „niemodyfikowalnego” źródła. Przykładowo w trybie rozszerzonym operator `\mag` jest ignorowany.

Wszystkim zainteresowanym wykorzystaniem PDFTEX-a polecam listę dyskusyjną `pdftex@mail.tug.org`, która jest dobrym miejscem do zadawania pytań i wymiany doświadczeń.

Bibliografia

1. Adobe Systems Incorporated, *Portable Document Format Reference Manual version 1.2*, November 27, 1996.
2. Adobe Systems Incorporated, *Adobe Type 1 Font Format*, Addison-Wesley 1990. Dostępny także w `www.adobe.com`.
3. Sojka P., H. The Thanh, J. Zlatuška, *The Joy of TEX2PDF*, Biuletyn GUST nr 8 (1997) (przedruk).

◇ Tomasz Przechlewski
`ekotp@univ.gda.pl`