

Michael van Hartskamp

Venuslaan 405, 5632 HM Eindhoven

michael@vanhartskamp.com

T_EX-rubriek

PDFTEX

Het drukprogramma T_EX en de daarvan afgeleide programma's heeft het publiek in staat gesteld om wiskundige teksten op een professionele wijze te zetten. Vaak wordt hierbij gebruik gemaakt van lokale bibliotheken van fonts en macro's die afhankelijk zijn van het computersysteem waar je op werkt. Hoe kan het resultaat op een ander systeem worden bekeken? Michael van Hartskamp, gepromoveerd topoloog, bespreekt deze keer het 'portable document format', dat dit mogelijk moet maken.

De komst van de computer maakte het voor auteurs mogelijk om zelf elektronisch documenten te maken of te bewerken. De grote verschillen in computersystemen enerzijds en professionele afdrukapparaten en thuisprinters anderzijds introduceerde echter nieuwe problemen die een soepele uitwisseling tussen de verschillende apparaten in de weg stonden. Door de opkomst van Internet werd nog een ander facet aan dit probleem toegevoegd: het elektronisch publiceren van documenten voor gebruik door anderen.

Deze keer bespreken we PDF, het 'portable document format' van Adobe. Wat biedt het? Waar komt het vandaan en wat kunnen we ermee vanuit T_EX?

Bij de ontwikkeling van T_EX voorzag Knuth al problemen met de grote verschillen in computers en apparaten die kunnen drukken of printen. De verschillen tussen computers omzeilde hij door het gebruik van een beperkte vorm van de programmeertaal Pascal en door weinig eisen te stellen aan de machine. Zo worden alle T_EX-interne berekeningen uitgevoerd op gehele getallen. Verder introduceerde Knuth het DVI formaat, een "device independent" beschrijving van een document: DVI is de uitvoer van T_EX en de invoer voor de uiteindelijke afdrukapparatuur. DVI is in wezen heel simpel. Het vertelt precies welke teken uit welk lettertype van welke grootte waar op de pagina moet worden geplaatst. DVI heeft echter een groot nadeel. Rasterafbeeldingen (foto's bijvoorbeeld) of zelfs lijnen zijn niet voorzien. Hiervoor zijn uitbreidingen nodig: de zogenaamde `\special` commando's en daarmee begint de ellende

want de specials zijn apparaatafhankelijk en dus is een DVI-bestand voor een iets ingewikkelder document ook apparaatafhankelijk.

Knuth was niet de enige met zijn DVI-oplossing. Het Amerikaanse bedrijf Adobe zag de oplossing in Postscript, haar eigen paginabeschrijvingstaal ook voorzien van programmeermogelijkheden. Postscript is een taal die nog wat verder van de gebruiker afstaat dan T_EX. Postscript werd in 1984 gelanceerd en sprak dankzij de combinatie met de grafische 'Desktop Publishing software' op de nieuwe Apple Macintosh computers vooral in de drukkerswereld erg aan. Al snel werd Postscript een defacto standaard.

Voor T_EX-gebruikers kwam er 'DVIPS' het programma dat DVI omzet in Postscript. Verder was en is er het (gratis) programma Ghostscript. Dit maakt het mogelijk om postscriptbestanden te printen op 'gewone' printers of te bekijken op het scherm. Deze combinatie maakte het mogelijk om postscriptbestanden, bij wetenschappers meestal van artikelen, te verspreiden zo dat ieder er mee overweg kan. Een nadeel was dat de postscriptbestanden vaak erg groot waren. Verder waren Windows gebruikers bijna nooit voorzien van Ghostscript.

Ook John Warnock, mede-oprichter van Adobe en medebedenker van Postscript dacht na over de uitwisseling van documenten. In 1991 leidde dit tot het zogenaamde 'Interchange Postscript': een enerzijds beperkte variant van Postscript die ook weer extra's bood voor interactief gebruik. Uit het hobbyproject van Warnock werd zo Acrobat / PDF geboren: een nieuw documentformaat (PDF) bedoeld voor de uitwisseling van documenten tussen gebruikers. Acrobat is de naam die Adobe gaf aan de lezer (Acrobat Reader) en schrijver (Acrobat Distiller).

Vanaf het moment dat Adobe de Acrobat Reader gratis ging weggeven nam de populariteit toe: het formaat is goed geschikt om (bestaande) documenten met behoud van vormgeving op het web te zetten en het veranderen van een PDF-document is misschien niet onmogelijk, maar zeker niet gemakkelijk. Tegenwoordig kan in een PDF-document een verbod voor de gebruiker zijn opgenomen om het document af te drukken of af te drukken uitsluitend met een slechts geringe kwaliteit toe

	voordeel	nadeel
T _E X-code	te veranderen	lokale macro's geven problemen
DVI	simpel	niet geschikt voor afbeeldingen
Postscript	zeer geschikt voor de drukker	vaak grote bestanden
PDF	zeer geschikt voor lezen van het scherm	minder gemakkelijk om professioneel te printen/drukken

Tabel 1 Voor- en nadelen van verschillende documentformaten

te staan. Dit zijn belangrijke eisen voor sommige gebruikers. Andere functies, die de meeste gebruikers meer zullen aanspreken, zijn de mogelijkheden rond het gebruik van hyperlinks en zelfs het uitvoeren van complete programma's in de taal JavaScript.

Ook PDF heeft zijn nadelen. Het is vaak moeilijker om na te gaan of een document alle noodzakelijke onderdelen bevat voor een goede afdruk: zijn alle lettertypen inbegrepen? Zijn afbeeldingen van de juiste scherpte? Zetten we alles op een rijtje dan krijgen we tabel 1.

Voor wie overtuigd is van de mogelijkheden van PDF geven we nu aan hoe het te gebruiken. Om PDF te generen zijn er drie mogelijkheden. De eerste is de door Adobe verkochte Acrobat Distiller. Deze 'vermomt' zich als een printer. Zodra een printopdracht wordt gegeven, genereert de Distiller een PDF bestand. De Acrobat Distiller is overigens niet gratis.

Wel gratis is het programma PS2PDF dat postscriptbestanden omzet naar PDF. Het spreekt voor zich dat er zo geen winst valt te behalen met de nieuwe mogelijkheden van PDF.

De derde, ook gratis, mogelijkheid is PDF_TE_X. PDF_TE_X is een volledig met T_EX-compatibel nieuw stuk software dat als uitvoer PDF levert in plaats van het klassieke DVI. Daarnaast biedt PDF_TE_X toegang tot de nieuwe interactieve mogelijkheden die het PDF formaat biedt.

De rest van deze aflevering zullen we daarom aan PDF_TE_X besteden. Allereerst de installatie en configuratie van PDF_TE_X. Wie in het bezit is van een recente T_EX-distributie kan PDF_TE_X meestal kant en klaar aantreffen. De eerder besproken Mik_TE_X-distributie (aflevering van juni 2002) bevat PDF_TE_X. Voor wie nog geen PDF_TE_X heeft en het toch wil gebruiken loont het zich wellicht om het gehele T_EX-systeem nog eens na te lopen. PDF_TE_X installeren is namelijk eerder vergelijkbaar met het installeren van een hele TeX distributie dan met het installeren van 'klassiek' T_EX. Dit komt omdat naast het T_EX-deel ook het DVI-deel van de installatie moet worden uitgevoerd waarbij de lettertypes, virtuele lettertypes, en dergelijke moeten worden ingesteld. Voor wie dit pad kiest: Let er op om de zogenaamde outline fonts te gebruiken. Deze geven een veel beter resultaat bij schermweergave dan de standaard T_EX-pk-fonts die voor een bepaalde vaste resolutie worden gegenereerd.

We hebben hier consequent over PDF_TE_X gesproken, er is echter ook PDF_LA_TE_X. Met PDF_LA_TE_X is het mogelijk om PDF-documenten te maken vanuit L_AT_EX. Ook de meeste L_AT_EX-macropakketten werken probleemloos samen met PDF_LA_TE_X. De pakketten die echter gebruik maken van niet-standaard T_EX-trucs in de DVI-driver zoals bijvoorbeeld het veel gebruikte PSTricks dat zijn niet geschikt voor PDF_TE_X. De meest gangbare aanvullende macropakketten voor L_AT_EX werken (inmiddels) wel goed samen met PDF_LA_TE_X.

Hoe genereren we nu PDF? In vroegere versies van PDF_TE_X was het meestal noodzakelijk om voor het eind van de eerste pagina de PDF_TE_X-opdracht `\pdfoutput=1` te geven. Alleen zo werd PDF in plaats van het

standaard DVI gebruikt. Tegenwoordig is PDF standaard voor PDF_TE_X en is voor het maken van DVI de opdracht `\pdfoutput=0` nodig.

We zullen aan de hand van drie kleine voorbeeldtoepassingen het gebruik van PDF_TE_X nader illustreren.

Allereerst het gebruik van afbeeldingen. De normale manier om afbeeldingen in L_AT_EX te gebruiken is middels het pakket `graphics` of `graphicx` en vervolgens op de gewenste plaats de opdracht `\includegraphics{naam.eps}` te plaatsen. Echter niet alle mogelijkheden van Postscript zijn opgenomen in PDF en daarom is het voor PDF_TE_X niet mogelijk om willekeurige EPS-plaatjes te gebruiken. De oplossing bestaat uit het programma `epstopdf` dat een EPS-plaatje in een PDF-bestand omzet. Vervolgens kan via `\includegraphics` het plaatje, nu als PDF-bestand, in het document worden gebruikt. In principe kunnen alle PDF-bestanden worden opgenomen in een ander bestand, maar alleen de eerste pagina wordt gebruikt.

Een erg handige aanvulling bij het werken met PDF_TE_X is het L_AT_EX-macropakket `hyperref`. Dit pakket is bij de meeste T_EX-distributies meegeleverd. Voeg voor gebruik vlak vóór de regel met

```
\begin{document}
```

de volgende regel toe:

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

Zonder verdere werkzaamheden van de auteur worden nu al enkele referenties automatisch aanklikbaar. De lezer kan zo makkelijker door het document bladeren en verwijzingen volgen zoals ook gebruikelijk is op het web.

Als extraatje is de inhoudsopgave zichtbaar in een apart venster. Alle hoofdstukken en secties kunnen worden aangeklikt en de Acrobat Reader zal onmiddellijk de juiste pagina weergeven.

Tot slot de document eigenschappen. Wie in het 'File Menu' van de Acrobat Reader op Document Properties / Summary klikt ziet dat het PDF-bestand gegevens bevat over de titel en auteur van het document, maar ook over onderwerp en keywords. Met het `hyperref` pakket is het eenvoudig mogelijk om deze eigenschappen zelf in te stellen. We gebruiken de opdracht `\hypersetup` om enkele van de vele opties van `hyperref` in te stellen. Een voorbeeld:

```
\hypersetup{
  pdftitle = {Over PDF en PDFTEX},
  pdfauthor = {Michael},
  pdfsubject = {Het gebruik van PDF en PDFTEX},
  pdfkeywords = {PDFTEX, pdf, Acrobat, hyperref},
  pdfcreator = {standaard staat hier TeX},
  pdfproducer = {standaard staat hier pdfTeX}
}
```

Samenvattend, PDF_TE_X is een onmisbare aanvulling op (of eigenlijk een vervanging van) T_EX. Het kan namelijk naast DVI ook PDF genereren. PDF is een geschikt bestandsformaat voor het elektronisch uitwisselen van documenten.

Dankzij het pakket `Hyperref` is het eenvoudig mogelijk om enkele van de nieuwe mogelijkheden van PDF_TE_X te benutten. Over een andere belangrijke toepassing van PDF_TE_X in combinatie met de Acrobat Reader, het maken van presentaties, zullen we het een volgende keer hebben.

