

HTML \longrightarrow L^AT_EX \longrightarrow PDF, of de intrede van T_EX in het hypertext tijdperk*

Yannis Haralambous

187, rue Nationale, 59800 Lille, Frankrijk
yannis.haralambous@univ-lille1.fr

Abstract

Wij beschrijven hier het produktieproces van elektronische hyper-documenten met behulp van L^AT_EX en *Adobe Acrobat*. Na een algemene beschrijving van de voor- en nadelen van L^AT_EX terzake geven we een omschrijving van elke stap, alsook van zekere te nemen voorzorgen met als doel efficiënte *Acrobat* documenten te bekomen.

De lezer zal in dit artikel een beschrijving vinden van de software tools DVIPS *reperere* en *recticrt*, en eveneens de basisprincipes van het formaat PDF.

Sleutelwoorden: Acrobat, DVIHPS, PDF, *reperere*, *recticrt*

1 Inleiding

1.1 Hypertext en L^AT_EX wat is het verband?

In tegenstelling met *hypermarkt*, *hypertentie*, *hyperactiviteit*, *hypertrofie* en de *hypersnelheid*, waar het voorvoegsel ‘*hyper*’ de hoge, zelfs overdadige hoeveelheid aanduidt, is *hypertext* geen zeer uitgebreide tekst, maar een tekst die gekenmerkt wordt door een interne structuur die aan bepaalde schermbesturingsprogramma’s toelaat om door de pagina’s van het document te navigeren.

Er is dus een triviaal verband tussen het begrip hypertext en de markeringen in een L^AT_EX document: beide voegen structuur toe aan het document. Bijvoorbeeld, het begrip van indexeren binnen L^AT_EX stemt volledig overeen met het begrip van hypertext-verband.

Het essentieel verschil tussen beide concepten is de desinteresse van T_EX ten opzichte van de schermuitvoer. Inderdaad, T_EX manipuleert hokjes welke bestemd zijn om karakters of tekeningen te bevatten. De taak van het invullen van deze hokjes met karakters/tekeningen valt ten deel aan de verschillende drivers voor scherm, fax of printer. T_EX is in de eerste plaats een werktuig voor typografische opmaak, het scherm dient enkel om te ‘previewen’, en een afbeelding op scherm wordt nooit als einddoel van een T_EX-compilatie beschouwd.

Deze desinteresse van T_EX t.o.v. het scherm is des te belangrijker vermits de PostScript constructies, welke in een DVI file ingevoerd zijn met behulp van het SPECIAL commando, in het algemeen niet zichtbaar zijn op het scherm.¹ De schermafbeelding is in die gevallen des te meer ontgoochelend vergeleken met het afgedrukte resultaat.

Er blijkt dat, voor de eerste maal in zijn bestaan, T_EX uitermate nuttig is voor het aanmaken van documenten die bedoeld zijn om vanaf het scherm te worden gelezen. L^AT_EX is inderdaad volledig geschikt voor de automatische produktie van hypertext verbanden, en de methodes waarover wij zullen spreken binnen dit artikel laten een automatische conversie naar een hypertext document toe van elk reeds bestaand L^AT_EX document! Het is passend te benadrukken dat een dergelijk document zijn volledige typografische L^AT_EX kwaliteit bewaart en afgedrukt kan worden op dezelfde wijze als voordien.

1.2 Het algemeen schema

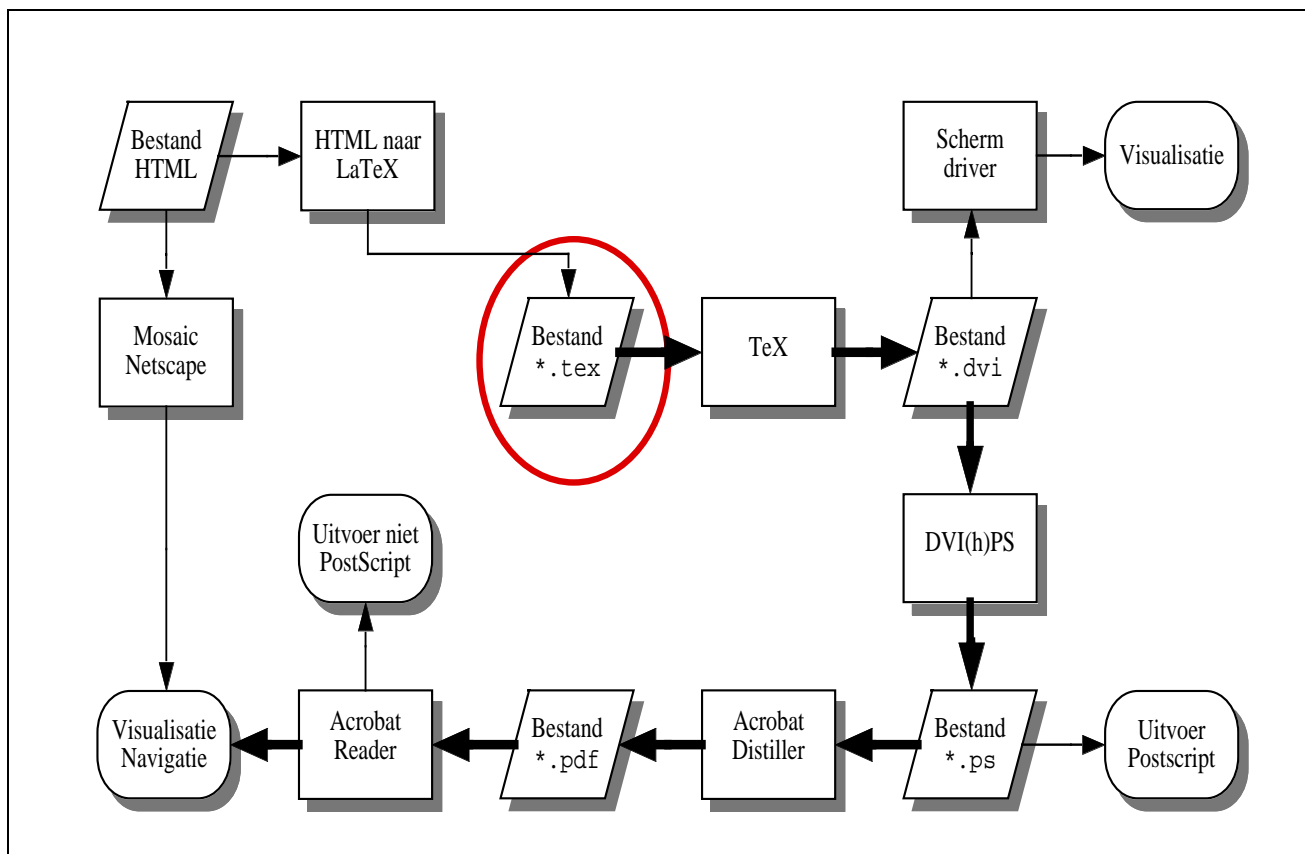
T_EX en L^AT_EX lezen een bestand in dat voorzien is van structurele en visuele markeringen en creëren een tweede bestand dat de te drukken pagina met een grote nauwkeurigheid omschrijft. Dit uitvoerbestand van T_EX wordt DVI genoemd (*DeVice Independent* = platform-onafhankelijk) omdat het enkel abstracte gegevens bevat: de plaatsing van elk karakter op de pagina, de naam van het font waarin de driver de pixels zal vinden van dit karakter, zijn code binnen dit font, enzovoort.

*Dit artikel stemt overeen met de presentatie gegeven te Nanterre (F) op 19 januari 1995 tijdens de dag *Verspreiding van elektronische documenten*.

Eerder gepubliceerd in **Cahiers GUTenberg**, #19, Januari 1995.

Vertaling van het Frans naar het Nederlands: **Philippe Vanoverbeke**, Hameau de Penfrat, F-29160 Crozon, Frankrijk; Langenhoeckstraat 28, B-8210 Veldegem, België. Met veel dank van de MAPS redactie aan Philippe voor het moeilijke vertaalwerk, vermits ‘Ik heb in sommige termen gewoonweg in het Engels vertaald vermits dit zo gebruikelijk is (dacht ik). In Frankrijk heeft men echter wel voor alles (godbeterdhet) een Franse uitdrukking. . . ’, aldus Philippe.

¹Met uitzondering van de gelukkigen onder ons die een operating system gebruiken voorzien van *Display PostScript*.



Figuur 1: *Het algemene schema*

Het visualiseren/afdrukken van een DVI bestand veronderstelt dus de beschikbaarheid over een aantal fonts. Dit is altijd mogelijk in het kader van informaticasystemen verbonden via een netwerk (werkstations, mainframes), doch dit wordt problematisch in het kader van de personal computers. De situatie wordt nog erger wanneer men documenten elektronisch wenst te verspreiden: een document bedoeld om gevisualiseerd en afgedrukt te worden door een groot aantal personen kan moeilijk in DVI formaat worden verspreid (dit zou enkel de \TeX gemeenschap betreffen, en bovendien zou men zich dienen te beperken tot het voorbijgestreefde CM font dat nog altijd zeer verspreid is in de V.S.). Meer bepaald zou het onmogelijk zijn een document op een floppy te plaatsen met voldoende utilities opdat het document onmiddellijk zou kunnen worden gevisualiseerd of afgedrukt worden zonder dat een \TeX systeem is geïnstalleerd.

Tenslotte zijn de hypertext verbanden niet voorzien binnen de syntax van het DVI formaat; elke poging tot ontwikkeling van een hypertext driver die het DVI formaat gebruikt zal er ons toe leiden een nieuw formaat 'Hyper-DVI' te definiëren, met alle problemen van compatibiliteit van dien en de terughoudendheid van de \TeX gemeenschap die fier is op de stabiliteit van zijn werkinstrumenten. Er blijkt dus dat het DVI formaat niet de ideale kandidaat is voor een bestandsformaat dat snel uit te baten en voldoende interactief is om de integratie van hypertext verbanden toe te laten.

Wat te doen dan? De voor de hand liggende keuze van een dergelijke kandidaat — vandaag tenminste — is het PDF formaat (*Portable Document Format*) van Adobe Systems. Er bestaat een uitbreiding van de PostScript taal, zeer nauw aansluitend met de syntax van bestanden aangemaakt met het programma *Illustrator*, met twee grote vernieuwingen t.o.v. de pure PostScript taal: het visualiseren en afdrukken op eender welk platform, onafhankelijk van de gebruikte fonts binnen het document en de integratie van de functionaliteit van hypertext.

We zullen het nog later hebben over het formaat PDF (zie 6.1). Laten we bekijken hoe het gebruik van dit formaat past binnen het proces van de aanmaak van een document (elektronisch of gedrukt). In het schema van figuur 1 zijn de in-uit bestanden aangeduid met schuine symbolen, de programma's met rechthoeken en de bewerkingen (visualisatie, afdrukken enz.) door symbolen met afgeronde hoeken. De vet afgedrukte pijlen geven de te volgen weg aan waaraan dit artikel is gewijd.

Het `*.tex` bestand is ons vertrekpunt (aangeduid met een cirkel). Vertrekkend van dit bestand (alsook van de style bestanden en de fontgegevens) zal \TeX een `*.dvi` bestand produceren. Indien men gebruik heeft gemaakt van het \LaTeX formaat en men de extensie `hyperref` erin opgenomen heeft, zullen de commando's voor cross-references, bibliografische aanduidingen, samenvatting en indexen de hypertext verbanden produceren binnen het `*.dvi` bestand met behulp van `\special` invoegingen.

Een andere te beschouwen mogelijkheid als vertrekpunt is een HTML bestand (bovenaan links op ons schema). Een HTML bestand kan inderdaad zeer gemakkelijk worden geconverteerd in een L^AT_EX bestand en de hypertext verbanden binnen deze eerste kunnen bewaard blijven binnen de tweede.

Laten we teruggaan naar ons *.dvi bestand. Dit kan meteen uitgebaat worden d.m.v. een scherm driver of een niet-PostScript printer. Maar men kan het tevens converteren naar PostScript m.b.v. het programma DVIPS. Een uitgebreide versie van DVIPS, met name DVIHPS ('DVI naar hyperPostScript') bewaart de eventuele hypertext verbanden welke het *.dvi bevat. Eenmaal het PostScript bestand aangemaakt kan men dit afdrukken op een PostScript printer (of op een niet-PostScript printer d.m.v. Ghostscript) of converteren naar PDF formaat m.b.v. het programma Adobe Acrobat Distiller. Deze laatste zal de hypertext verbanden interpreteren en ze opnemen in het PDF document.

Om tenslotte het PDF document te visualiseren, gebruikt men Adobe Acrobat Reader, een programma dat gratis wordt verspreid en geschikt is voor Macintosh, Windows, DOS en Solaris. Dit programma laat ons tevens toe om te navigeren in documenten en deze af te drukken op een niet-PostScript printer.

Men merkt dat als het vertrekpunt een HTML document is, de hypertext functionaliteit bewaard is gebleven, maar dat men er bovenop de typografische presentatie van L^AT_EX er bij heeft gekregen. Een PDF document is een waarheidsgetrouwe kopie van het gedrukte document (het kan op film gezet worden en zo een professionele afdruk leveren met kleurenfoto's, grafieken enz.) die als extra de hypertext navigatiemogelijkheid biedt in het document of over het net (de navigatie over het net is enkel mogelijk met versie 2 van de Adobe Acrobat software).

1.3 Conclusies

De documentstructuur vereist door L^AT_EX wordt binnen steeds meer domeinen toegepast: het meest frappante voorbeeld is ongetwijfeld deze van de vocale synthesiser (zie [5]), voor gebruik door niet-zienden, die in staat is een wiskundige formule uit te spreken zoals een wiskundige en toelaat doorheen het document te navigeren met behulp van speciale toetsen op de recorder.

In dit artikel zullen we een andere applicatie omschrijven: de creatie van elektronische boeken waarvan de presentatie niet hoeft onder te doen voor traditionele boeken (vermits ze kunnen worden afgedrukt zonder verlies van kwaliteit) en die een minimum aan interactiviteit bieden: hypertext navigatie tussen samenvatting, index, bibliografische referentielijst en tekst binnen de tekst zelf met behulp van L^AT_EX cross-references, en tenslotte navigatie tussen documenten aanwezig op eenzelfde machine of op het netwerk.

In de rest van dit artikel volgen alle stappen van het proces, aangeduid door vetgedrukte pijlen op het schema van fig. 1 (en op deze van fig. 2 die er een uitbreiding van vormt).

2 HTML \longrightarrow L^AT_EX

Het HTML mark-up systeem is gedefinieerd volgens de SGML norm (de lezer zal de volledige DTD vinden in [3]). Het betreft een klein aantal markeringen, hoofdzakelijk bedoeld voor de weergave op een scherm. Hierdoor vind men diverse tekst stijlen, zowel logische (emphasized, very emphasized, citation, address enz.) als visuele (schuin, vet, onderlijnd enz.), maar alsook markeringen voor fundamentele structuren als voetnoten, tabellen, enz.

Het spreekt voor zich dat L^AT_EX een oneindig rijker mark-up systeem is dan HTML en bijgevolg is de conversie 'HTML \longrightarrow L^AT_EX' triviaal. Inderdaad, het betreft eenvoudigweg de conversie van:

1. bepaalde markeringen binnen L^AT_EX zoals <CITE> in \begin{quotation} en </CITE> en \end{quotation}
2. andere markeringen in L^AT_EX commando's met een argument: een woord in \emph{een woord}, enz.
3. enkele zeldzame markeringen in L^AT_EX commando's zonder argument zoals <P> die geconverteerd kan worden in een lege regel of in \par
4. geaccentueerde letters in L^AT_EX syntax: ´ zal \'{e} geven, ¸la; geeft \c{C} enzovoort.

Twee particuliere gevallen doen zich voor:

1. De HTML markeringen die geen onmiddellijk L^AT_EX equivalent hebben, bijvoorbeeld die een sterk emphasized omgeving aanduidt. In dit geval zou de definitie van nieuwe L^AT_EX omgevingen de aan te raden oplossing zijn, eventueel met behoud van de dezelfde benaming en de keuze van de visuele presentatie ervan aan de gebruiker over te laten, dit met een default presentatie voorhanden.

Zo zou bijvoorbeeld

```
<STRONG>very emphasized</STRONG>
```

kunnen geconverteerd worden in:

```
\begin{STRONG}sterk benadrukt\end{STRONG}
```

deze omgeving zou kunnen gedefinieerd worden als

```
\newenvironment{STRONG}{%
```

```
\bfseries\itshape}{\}
```

in een bestand html2ltx.sty en die dus 'sterk benadrukt' oplevert in een L^AT_EX document.

2. De markeringen van hypertext functies. In dit geval volstaat het om de syntax, aangewend door de extensie hyperref, die we vervolgens (zie 3) zullen gebruiken om precies deze informatie in het DVI bestand op te nemen. Hier volgen de expliciete commando's:

1. om een *doel* te definiëren (een 'anker' in het HTML jargon), zal men onder HTML de gekozen tekst omgeven door de markeringen (waar sleutelwoord het sleutelwoord is gegeven aan het doel in kwestie), en <\A>. Het overeenkomstig L^AT_EX commando zal \hyperdef{}{sleutelwoord}{}{...} zijn waar ... de gekozen tekst is;

2. om een *verband* met een doel te leggen binnen hetzelfde document zal men in HTML de gekozen tekst omgeven met de markeringen `` (waar `sleutelwoord` het sleutelwoord is gegeven aan het doel in kwestie), en `<\A>`. Het overeenkomstig L^AT_EX commando zal `\hyperref{{sleutelwoord}}{...}` zijn waar ... de gekozen tekst is;
3. om een *verband* te leggen naar een ander document zal men in HTML de gekozen tekst omgeven met de markeringen `` (waar `adres` het attpadres is van het document in kwestie), en `<\A>`. Het overeenkomstig L^AT_EX commando zal `\hyperref{adres}{...}` zijn waar ... de gekozen tekst is.

3 L^AT_EX \longrightarrow DVI

Laten we het meteen vertellen: elk L^AT_EX_{2 ϵ} , zonder uitzondering, kan een elektronisch document produceren door de toevoeging van één regel:

```
\usepackage{hyperref}
```

vooraan in het bestand. Deze uitbreiding werd ontwikkeld door Tanmoy Bhattacharya en gewijzigd door Sebastian Rahtz. De L^AT_EX commando's die zullen dienen om hypertext verbanden te creëren zijn:

- `\label`, `\ref` en `pageref` (cross-references);
- `\chapter`, `\section`, `\subsection`, etc. (indeling van het document);
- `\index` (aanmaak van een index);
- `\cite` (aanmaak van een lijst met bibliografische referenties).

Er dient dus niets gewijzigd te worden in de brontekst van het L^AT_EX document, tenzij men er nieuwe hypertext functionaliteiten wenst aan toe te voegen waarbij in dit geval de interne commando's `hypertarget` en `\hyperlink` kunnen gebruikt worden, of door nieuwe commando's te definiëren gebruik makend van de 'private' commando's `hyper@anchor` en `hyper@link` (zie [4] voor meer details).

3.1 De aanmaak van een PDF bestand met merkteken

Zoals men kan zien op fig. 3 geeft het programma Acrobat Reader ons de mogelijkheid een inhoudstafel te hebben die hiërarchisch is en interactief op het linkergeedeelte van het Acrobat venster. Het programma DVIHPS — tenminste in zijn huidige versie — maakt deze inhoudstafel niet automatisch aan. Om dit tekort te verzachten heeft de auteur een 'utility' ontwikkeld, `reperere` genaamd, samen met een L^AT_EX style (`reperere.sty`). Een 'executable' van `reperere` op het eigen platform te bekomen dient de gebruiker eerst de 'utility' Flex (versie \geq 2.4.6) los te laten op `reperere.lex`, met als optie `-8` in de commandoregel:

```
flex -8 reperere.lex
```

en vervolgens een C compilatie van het resultaatbestand van Flex (`lex.yy.c` onder UNIX of Macintosh, `lexyy.c` onder DOS enzovoort).

In fig. 2 zal de lezer het algemeen schema van fig. 1 vinden waaraan de tussenstap `reperere` is toegevoegd.

Ziehier de exacte procedure om een PDF inhoudstafel aan te maken:

- men voegt het commando `\usepackage{reperere}` toe in het begin (preamble) van het document, *na* het commando `\usepackage{hyperref}`;
- men vervangt de commando's `\part`, `\section`, `\subsection` en `\subsubsection` die men in de inhoudstafel wil opnemen door respectievelijk `\Part`, `\Section`, `\Subsection` en `\Subsubsection`;
- L^AT_EX zal dan een nieuw auxiliair bestand aanmaken met extensie `.rep`;
- Na DVIHPS start men `reperere` op met als invoerbestand vervolgens:
 1. het bestand `foo.rep` aangemaakt door L^AT_EX;
 2. het bestand `foo.ps` aangemaakt door DVIHPS;
 3. en opnieuw het bestand `foo.rep`.

Onder UNIX (en onder elk ander operating system dat over 'pipes' beschikt) kan dit heel eenvoudig gebeuren met behulp van een commandoregel van het type

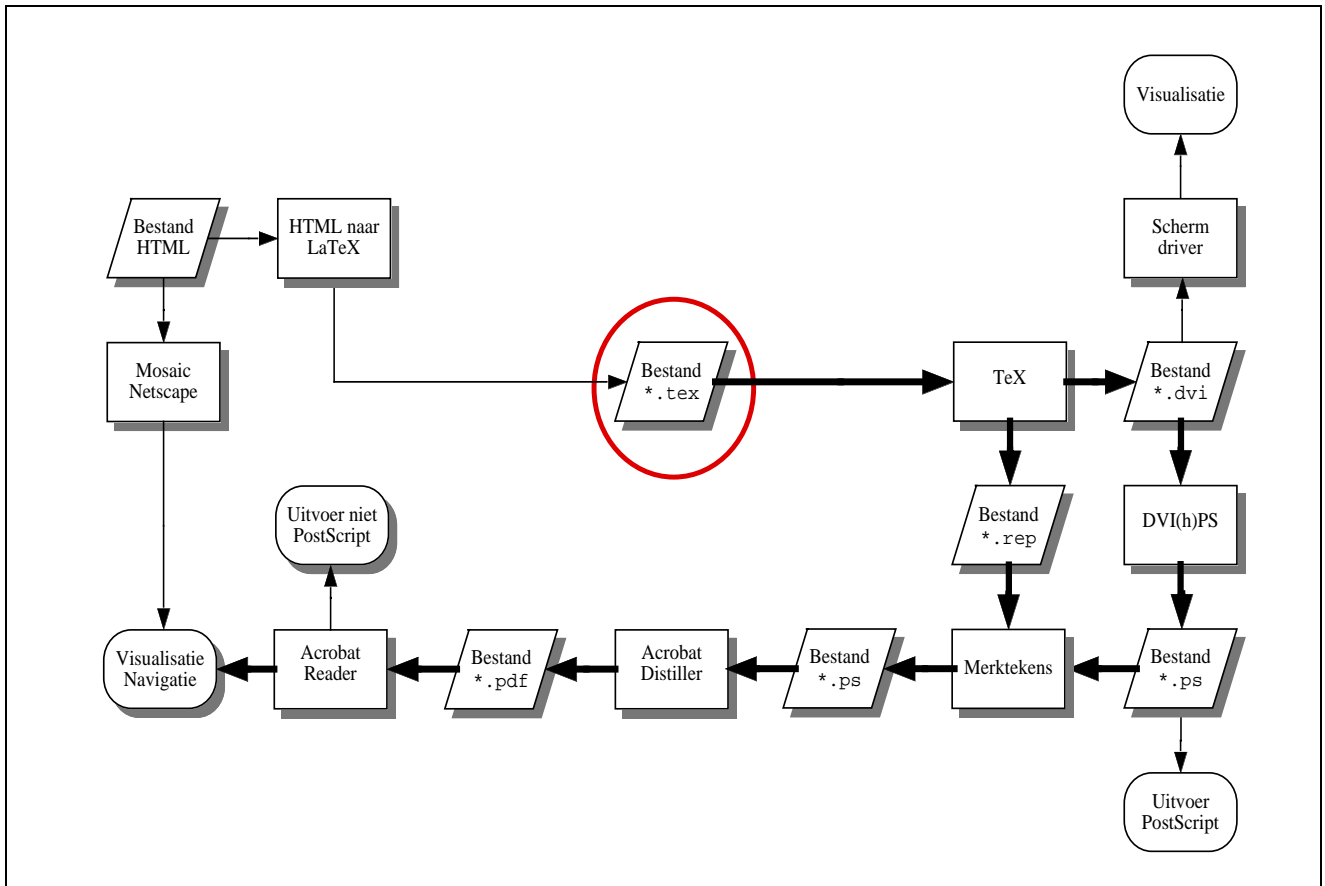
```
cat foo.rep foo.ps foo.rep | reperere >
```

- Het bestand `new-foo.ps` aangemaakt door `reperere` is een exacte kopie van het bestand `foo.ps` met enkele extra regels welke Acrobat Distiller toelaten om een nieuwe PDF inhoudstafel te creëren.

Het meest delicate gedeelte van deze operatie is de conversie van de verschillende tekstonderverdelingen naar de PDF codering. Inderdaad, deze codering is een samenraapsel van Macintosh Standard, Adobe Standard en Windows ANSI coderingen. De moeilijkheid vloeit voort uit het feit dat L^AT_EX een uitbreiding van de commando's maakt wanneer hij reeksen karakters wegschrijft in het bestand `*.rep\verb`. Een geaccentueerd karakter kan dus op verschillende wijzen omschreven worden, afhankelijk van de codering komende van L^AT_EX de accent-commando's enz. De 'utility' `reperere` kan moeilijk elke schrijfwijze herkennen en deze vervangen door het overeenstemmende 8-bits PDF karakter. Het kan dus voorvallen dat de gebruiker enkele correcties dient uit te voeren in de titels van de inhoudstafel van het PDF bestand. Het is dus mogelijk dat de gebruiker enkele correcties dient uit te voeren in de titels van de onderverdelingen van de PDF inhoudstafel. Gelukkig heeft dit geen enkele invloed op de interactiviteit: de hypertext relaties tussen deze inhoudstafel en de tekst van het PDF document zullen bewaard blijven.

3.2 Voorzorgsmaatregelen te nemen op het T_EX niveau

Het feit dat een DVI bestand totnogtoe bijna uitsluitend diende voor het afdrukken, verplicht ons bepaalde voorzorgen te nemen bij de voorbereiding van dit bestand wanneer



Figuur 2: Algemeen schema om een bestand PDF met merktekens te verkrijgen

het onze bedoeling is dit later te converteren naar het PDF formaat.

Deze voorzorgen betreffen hoofdzakelijk de keuze van het in het document gebruikte lettertype. Inderdaad, het grootste probleem voor een programma zoals Acrobat, dat beweert eender welk PostScript bestand te kunnen vertonen en afdrukken, is de beschikbaarheid van de PostScript lettertypes die in het document gebruikt worden. 99% van de bestaande PostScript lettertypes (er zijn er duizenden...) zijn commercieel en het gebruik ervan houdt dus een voorafgaandelijke transactie in tussen de gebruiker en het bedrijf dat de rechten ervan bezit. Wat te doen dan om een dergelijk document te verdelen? Hoe kan men zeker zijn dat elke lezer de beschikking heeft over dezelfde PostScript lettertypes? Wat te doen indien dit niet het geval is?

Adobe heeft dit probleem opgelost door een nieuwe techniek voor lettertypes te ontwikkelen: de *Multiple Master* lettertypes. Het principe is aardig verwant met METAFONT: het betreft meta-lettertypes² en door de karakters enigszins te wijzigen is de scherm- en printerdriver (Super-ATM) in staat niet-beschikbare PostScript lettertypes na te bootsen. Twee Multiple Master lettertypes vergezellen Acrobat Reader: een 'roman' lettertype en een 'sans serif'. Elk onbeschikbaar lettertype zal vervangen worden door een van deze twee 'generieke' lettertypes.

Elk onbeschikbaar lettertype? Ja, elk onbeschikbaar lettertype, inclusief alle wiskundige symbolen of speciale karakters (fonetisch, niet-latijnse, enzovoort). Door gebruik te maken van niet-standaard lettertypes kan men dus snel tot rampzalige resultaten komen.

Om hieraan te verhelpen stelt Acrobat Distiller twee oplossingen voor:

1. de opname van het 'exotisch' PostScript lettertype binnen het PDF bestand. Dit voorkomt de substitutie door Multiple Master maar stelt eventueel problemen op het vlak van het copyright;
2. het gebruik van (niet-PostScript) PK fonts in het geval van \TeX . Er is geen copyright probleem vermits enkel enige bitmaps opgenomen zullen worden in het PDF bestand maar de schermafdruck onder Acrobat Reader lijdt hieronder. Inderdaad, Acrobat Reader is niet geoptimaliseerd om bitmap karakters te vertonen en moet vaak overgaan op een grote vergroting om het ontcijferen toe te laten. Daarentegen gebeurt het afdrukken zonder probleem (vermits uiteindelijk de resolutie van de bitmap karakters opgenomen in het bestand eerder overeenkomen met de resolutie van de printer dan deze van de monitor).

Een derde oplossing bestaat uit het vermijden van het probleem door enkel 'universele' lettertypes te gebruiken, t.t.z.

²In feite zijn hun meta-eigenschappen lachwekkend vergeleken bij deze van de METAFONT lettertypes.

ongewijzigd aanwezig onder elk operating system: zodoende is men niet aangewezen op de opname van lettertypes in het document (en bijgevolg stellen er zich geen copyright problemen) maar men behoudt de PostScript vertoning op het scherm die gevoelig beter is dan de vertoning van bitmap karakters. Deze lettertypes zijn Times, Helvetica, Courier, Symbol, Palatino, New Century Schoolbook en Bookman.

Uiteindelijk een ultieme voorzorg: men dient het gebruik van *virtuele fonts welke gebruik maken van samenvoegingen van tekens* te vermijden. Anders gezegd: er dient gebruik gemaakt worden van her-coderingen van lettertypes ‘à la DVIPS’ eerder dan gebruik te maken van de samenstelling van geaccentueerde karakters door de combinatie van de letter en het accent zoals `fontinst.sty` dit doet. De reden hiervoor is dat Acrobat toelaat om karakter-strings op te zoeken in een document: indien de geaccentueerde karakters in werkelijkheid samenvoegingen zijn van niet-geaccentueerde letters en accenten zal men geen zoekslag kunnen uitvoeren op basis van 8-bits geaccentueerde karakters. Het woord ‘dénégré’ bijvoorbeeld zal gecodeerd worden als

```
de<accent aigu>ge<accent aigu>ne<accent
      aigu>re<accent aigu>
```

in het PDF document en het opzoeken van de karakterstring `d\`eg\`en\`er\`e` (waar ‘\`e’ een 8-bits Macintosh of Windows of ISO Latin-1 karakter is) is tot mislukken gedoemd.

Om virtuele fonts te bekomen *die zich beperken tot her-codering van karakters*, kan men gebruik maken van de utility `afm2tfm` van de DVIPS distributie, met de commandoregeloedopties `-T extex.enc` (Cork encoding) en `-v` gevolgd door de naam van het gewenste VPL bestand. In het configuratiebestand `psfonts.map` van DVIPS moeten bijgevolg de volgende commando’s bijgevoegd worden:

```
" ExtendedTeXEncoding ReEncodeFont " < extex.enc
```

aan de font definities waarvan sprake (veronderstellend dat het bestand `extex.enc` zich in dezelfde directory bevindt als `psfonts.map`, anders dient `extex.enc` vervangen te worden door zijn volledig pad).

Ziehier een voorbeeld: de auteur heeft het reëel font `rtimes.tfm` en het virtueel font `vtimes.vf` (met metriek bestand `vtimes.tfm`) verkregen vanuit het bestand `AFM Times-Roman.AFM`, met behulp van de volgende commando’s:

```
afm2tfm Times-Roman.afm -T extec.enc
      -v vtimes.vpl rtimes.tfm
vftovp vtimes.vpl -o vtimes.vf
```

Om het te gebruiken heeft hij de regel

```
rtimes Times-Roman " ExtendedTeXEncoding
      ReEncodeFont " <extec.enc
```

toegevoegd aan het bestand `psfonts.map`, en heeft hij een bestand `Tlvtime.fd` gecreëerd dat, onder andere, de volgende regels bevat:

```
\DeclareFontFamily{Tl}{vtime}{}
\DeclareFontShape{Tl}{vtime}{m}{n}{<->vtimes}{}
\endinput
```

Om deze nieuwe font familie te gebruiken volstond het de volgende regels op te nemen:

```
\usepackage{t1enc}
\def\rmdefault{vtime}
```

in het begin van het L^AT_EX bestand.

Voor de wiskundige formules kan men de fonts Times en Symbol gebruiken: meerdere makro- en virtuele fontpakketten laten dit toe, onder andere *MathTimes* van Michael Spivak.

4 DVI \longrightarrow (Hyper)PostScript

Zoals T_EX is DVIPS een typisch voorbeeld van een public domain programma van zeer hoge kwaliteit. Unaniem erkend als de beste vertaler van DVI naar PostScript zal DVIPS hoogstwaarschijnlijk de norm worden op dit vlak. Overigens bieden steeds meer T_EX systemen een scherm-driver aan (of een niet-PostScript printer) specifiek voor het onderliggende operating system en een implementatie van DVIPS voor PostScript afdruk.

Mark Doyle heeft het programma DVIPS (van Tom Rokicki) gewijzigd (aangespoord door [*sic*] Paul Ginsparg en Arthur Smuth) opdat het het beheren van hypertext verbanden zou toestaan. Was een dergelijke ‘chirurgische’ ingreep noodzakelijk? Het blijkt van wel en we zullen zien waarom.

Om een hypertext verband te definiëren heeft Acrobat Distiller (op zijn minst) twee gegevens nodig: de actieve zone bij het begin en het scherm bij aankomst van het verband. Deze informatie dient hem medegedeeld te worden onder de vorm van PostScript coördinaten (1/72 inches vanuit de oorsprong van de PostScript coördinaten, onderaan links van de bladzijde).

Om de exacte coördinaten van het doel van een hypertext verband te determineren, welke zich meerdere pagina’s na de actieve beginzone kunnen bevinden *moet* men post-processing toepassen op het DVI bestand: het zou zonet niet onmogelijk dan wel op zijn minst enorm ingewikkeld zijn deze coördinaten op het niveau van L^AT_EX te bekomen. Vermits een DVI post-processor benodigd is kan men evengoed meteen DVIPS gebruiken, deze werkt met PostScript coördinaten en houdt rekening met alle schaalveranderingen en translaties van de oorsprong welke kunnen voorkomen.

Zodoende heeft Mark Doyle een ‘opgeknapte’ versie van DVIPS ontwikkeld welke hij DVIHPS noemt (‘DVI naar HyperPostScript’). De auteur hoopt dat deze wijzigingen binnenkort deel zullen uitmaken van de DVIPS standaard.

Praktisch gezien wordt DVIHPS op dezelfde wijze gebruikt als DVIPS op een uitzondering na: indien de gebruiker een PostScript bestand wenst aan te maken waarin de hypertext informatie ten behoeve van Acrobat Distiller is opgenomen, dan zal hij de optie `z` (kleine `z`) dienen toe te voegen aan de commandoregel.

Er dient opgemerkt dat de header van het aangemaakte PostScript bestand PostScript code zal bevatten die alle

hypertext commando's desactiveert indien het bestand geïnterpreteerd wordt door een andere PostScript interpreter dan Acrobat Distiller. Diezelfde code zal tevens de versie van Acrobat Distiller nagaan (bijvoorbeeld: versie 1 van dit programma staat geen verbanden toe tussen verschillende documenten: enkel deze verbanden worden gedesactiveerd gedurende de interpretatie).

5 PostScript \longrightarrow PDF

Deze etappe, welke zeker de langste is op het vlak van de benodigde tijd voor de uitvoering ervan, wordt volledig afgehandeld door het programma Acrobat Distiller (een eerder duur programma, doch ongelukkiglijk onmisbaar om degelijke PDF documenten te produceren). Acrobat Distiller is tegelijkertijd een zeer goede 'debugger' voor PostScript code en een goede interpreter: zonder een PostScript kleurenprinter tot zijn beschikking te hebben, noch het Display PostScript systeem om PostScript te kunnen previewen op een scherm, gebruikt de auteur Acrobat Distiller om T_EX documenten in kleur te visualiseren.

6 Previewen, navigatie en afdruk van het PDF bestand

Deze operaties gebeuren met behulp van het programma Adobe Acrobat Reader, hetgeen gratis wordt verdeeld.³ Meerdere functies voor zoeken, navigatie, kopiëren van text⁴, enz. worden de gebruiker aangeboden onder de vorm van 'buttons' en menus. Er dient opgemerkt dat dit programma werkt op Macintosh, alsook onder DOS, Windows en UNIX (Solaris en binnenkort andere systemen), met exact dezelfde gebruikers interface.

Om de scherm-afdruk van een T_EX bestand te vergelijken kan de gebruiker in figuur 3 een kopie van een scherm van Acrobat Reader op Macintosh vinden en in figuur 4 de afgedrukte versie van hetzelfde document. De L^AT_EX uitbreiding `hyperref` laat de gebruiker toe de presentatie van de actieve zone van de hypertext link te kiezen (by default in het rood) alsook de doel-zones (by default groen). Het PDF formaat laat onder andere toe de actieve zones te omlijnen.

6.1 Enkele inlichtingen betreffende het PDF formaat

Voor de gewone sterveling is het PDF formaat nog ontoegankelijk en onleesbaarder dan de PostScript taal zelf. Niettemin is het interessant om enige kennis te hebben over de structuur ervan om, indien nodig, enige kleine aanpassingen te kunnen uitvoeren aan een dergelijk bestand (het PDF formaat is zeer recent van oorsprong en er is een zeer groot gebrek aan utilities om PDF documenten te kunnen wijzigen).

Een PDF bestand is een tekstbestand, geheel geschreven in 7-bit. Het bestaat uit vier delen: de *header*, de *body*, de *cross-reference table* en de *trailer*. In de versie 1 be-

vat de header één enkele regel: `%PDF-1.0`. De body is samengesteld uit objecten: elke pagina is een object, de verbanden, de nota's, de markeringen, de fontcodering, de fontomschrijvingen alsook de omschrijvingssystemen t.b.v. kleur zijn objecten. Het voordeel om objecten te gebruiken vloeit voort uit het feit dat men hierdoor pagina's kan bijvoegen, verwijderen, de volgorde ervan kan wijzigen, zonder de bestaande hypertext verbanden te verbreken: de paginavolgorde is bewaard in de cross-references tabel, verwijderde pagina's blijven dus in het document behouden en worden enkel 'virtueel' verwijderd. Elke wijziging brengt met zich de aanmaak van een nieuwe tabel met cross-references mee in dit in de trailer van het document. De applicaties voor de visualisatie van PDF documenten beginnen dus een dergelijk document vanaf het einde in te lezen en halen in de cross-references tabel de pointers op naar de diverse objecten van het document.

De meeste objecten zijn gecomprimeerd en daarna gecoedeerd in 7-bit code: vier comprimeringsmethodes kunnen aangewend worden: Lempel-Ziv, *run length*, CCITT Fax groep 3 of 4, JPEG; voor de codering naar 7-bit code kunnen twee methodes gebruikt worden: de hexadecimale notatie of de 'modulo 85' notatie.

Met Acrobat Distiller kan dit comprimeren gedesactiveerd worden, doch dit biedt weinig voordelen vermits geen enkele utility voorhanden is die het comprimeren *a posteriori* van gewijzigde PDF documenten toelaat.

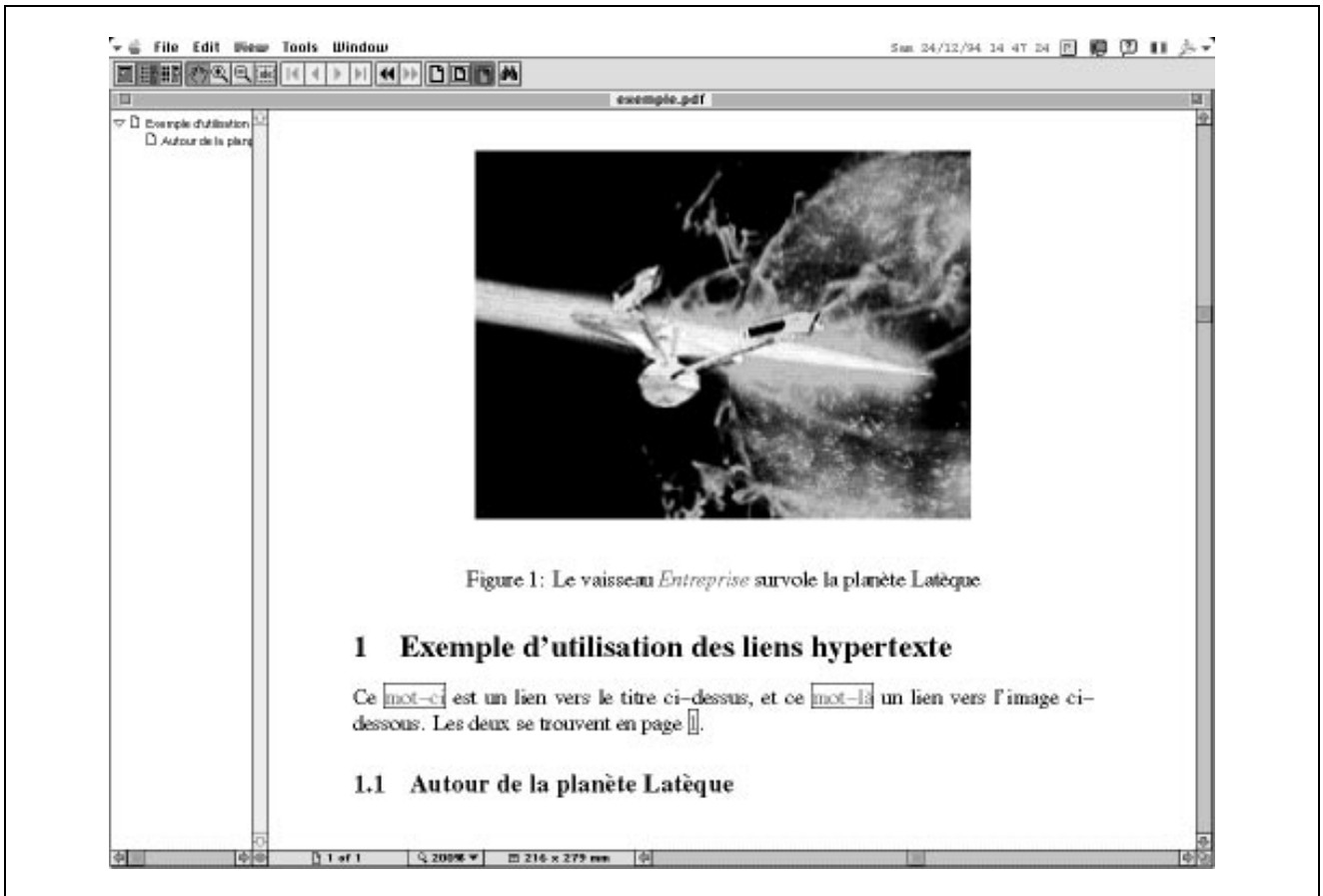
We beschrijven hier verder uitsluitend enkele niet-gecomprimeerde objecten welke bijgevolg door de gebruiker vrijelijk kunnen gewijzigd worden. Niettemin dient te worden opgemerkt dat elke wijziging van het PDF bestand (met één uitzondering, die we verder behandelen) de verplichting met zich meebrengt om de cross-references tabel aan te passen: inderdaad, deze tabel bevat voor elk PDF object zijn verschuiving in positie van *bytes* ten opzichte van het begin van het bestand. Elk object heeft een nummer dat tevens het eerste gegeven is van een object. De objecten zijn in het PDF bestand niet noodzakelijk geordend in stijgende lijn van dit nummer.

De tabel met cross-references bevat een regel voor elk object; deze regel bevat de verschuiving van het object t.o.v. het begin van het bestand (getal met 10 cijfers) gevolgd door een lege spatie en een getal met 5 cijfers dat aanduidt hoe vaak het betreffende object gewijzigd is geworden, opnieuw gevolgd door een lege spatie en dan de letter 'n'. Indien het object gewist word is het objectnummer dus beschikbaar op deze regel doch de syntax wordt gewijzigd: het getal met 10 cijfers duidt het nummer aan van het volgende vrije object in de tabel (=0 indien het hier het laatste vrije object betreft) en de letter op het einde van de regel is 'f'.

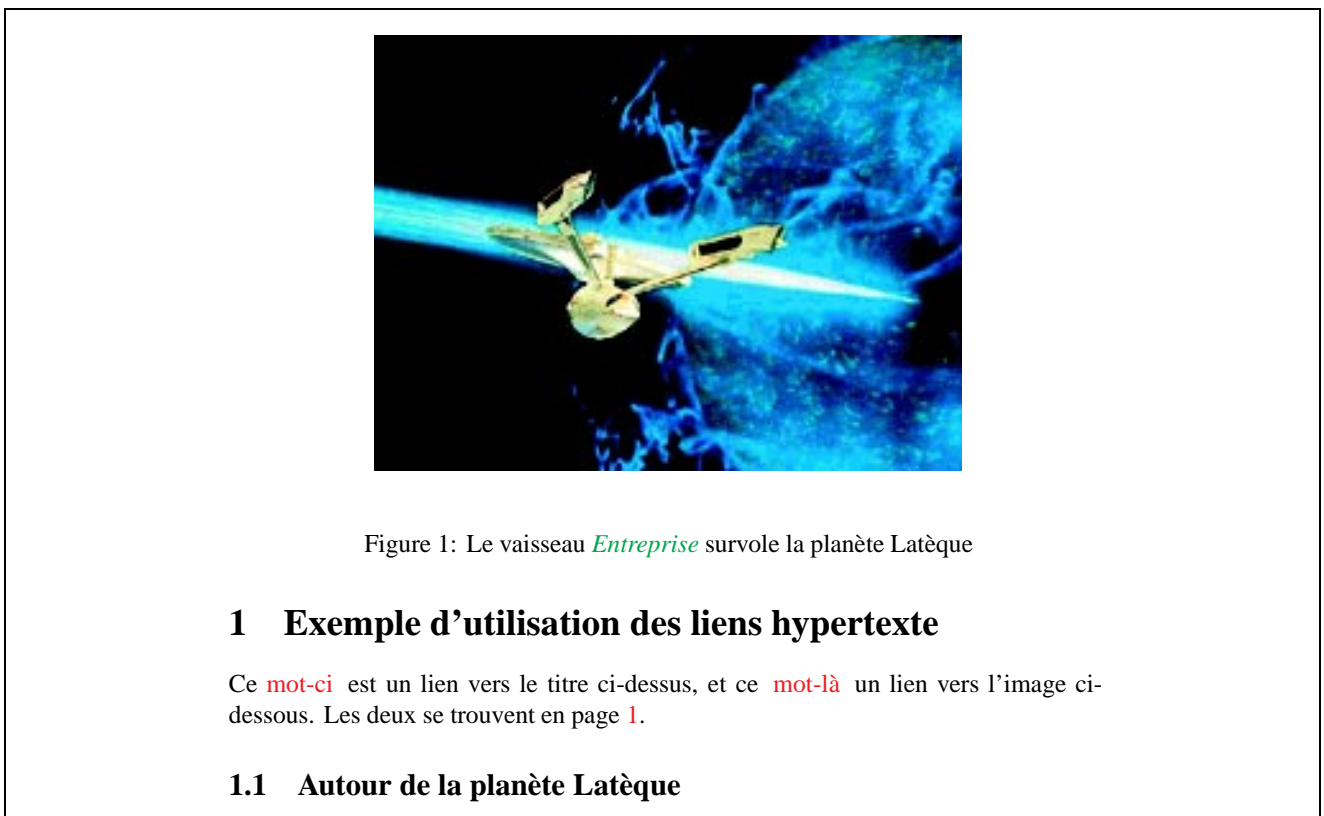
Voor elke wijziging uitgevoerd aan een object dient de resulterende verschuiving doorgegeven te worden aan de opeenvolgende objecten in het bestand. Tevens dient een

³En op de Adobe demo CD-ROM welke bij deze MAPS is bijgesloten.

⁴En het zal de eerste maal zijn dat men kopieën van tekst kan maken vanuit een afgeleide van een DVI bestand.



Figuur 3: Schermkopie Macintosh systeem: een voorbeeld van een PDF bestand aangemaakt onder \LaTeX



Figuur 4: Resultaat van de afbeelding van het PDF bestand

getal op het einde van het bestand gewijzigd te worden, dit getal duidt de verschuiving aan van de cross-references tabel ten opzichte van het begin van het bestand.

Een voorbeeld: het bestand uit fig. 3 bevat 20 objecten; ziehier de twee laatste, de cross-reference table en de trailer:

```

1 0 obj
<<
/CreationDate (Dim 25 D\`ec 1994)
/Producer (Acrobat Distiller 1.0 pour Macintosh)
>>
endobj
3 0 obj
<<
/Pages 5 0 R
/Outlines 16 0 R
/Type /Catalog
>>
endobj
xref
0 21
0000000000 65535 f
0000043515 00000 n
0000042654 00000 n
0000043617 00000 n
0000040662 00000 n
0000042693 00000 n
0000000010 00000 n
0000039522 00000 n
0000040847 00000 n
0000042235 00000 n
0000042360 00000 n
0000042487 00000 n
0000042777 00000 n
0000042897 00000 n
0000043017 00000 n
0000043211 00000 n
0000043137 00000 n
0000043373 00000 n
0000039543 00000 n
0000040641 00000 n
0000042612 00000 n
trailer
<<
/Size 21
/Root 3 0 R
/Info 1 0 R
>>
startxref
43683
%%EOF

```

Het getal 43683 duidt de verschuiving aan van de cross-references tabel ten opzichte van het begin van het bestand.

We zullen het object 1 wijzigen: vooreerst zullen we een fout van Acrobat Distiller voor de (Franse) Macintosh versie verbeteren: de datum bevat 8-bit karakters (de ‘é’ in ‘Déc’) die Distiller verzuimd heeft te wijzigen naar de PDF codering. Het dialoogvenster met als doel bepaalde inlichtingen te vertonen, waaronder de datum van aanmaak van het document, zal een foutief karakter vertonen in plaats van de ‘é’. Dit 8-bit karakter moet veranderd worden door de code voor ‘é’ in PDF code: in PDF octale notatie, \351. We zullen dus een ‘é’ vervangen door viermaal \351: de verschuiving wordt +3.

Het betreft object 1, en het enige object die hierop volgt in het bestand is het object 3. De derde regel in de cross-

reference tabel dient dus gewijzigd te worden (we rekenen vanaf 0: de regel

```
0000000000 65535 f
```

is dus de nul-regel. De eerste regel van de tabel is

```
0000043515 65535 n
```

deze stemt overeen met het object dat de informatie bevat. De derde regel is

```
0000043617 65535 n
```

en deze stemt overeen met het enige object dat zich bevindt na het object dat we zullen wijzigen (het nummer 43617 is inderdaad het hoogste in de tabel, het betreft hier wel degelijk het laatste object in het bestand). We zullen 3 optellen bij het getal 43617, alsook bij het getal 43683 op het einde van het bestand (de cross-reference tabel is ook met 3 verschoven). Het bestand wordt dus:

```

1 0 obj
<<
/CreationDate (Dim 25 D\351c 1994)
...
xref
0 21
0000000000 65535 f
0000043515 00000 n
0000042654 00000 n
0000043620 00000 n
...
startxref
43686
%%EOF

```

Hetzelfde geldt indien we enige extra informatie opnemen in object 1, we berekenen de uiteindelijke verschuiving en corrigeren vervolgens de cross-reference tabel en zijn pointer

```

1 0 obj
<<
/CreationDate (Dim 25 D\351c 1994)
/Producer (Acrobat Distiller 1.0 voor Macintosh)
/Author (Yannis Haralambous)
/Title (Voorbeeld van een PDF bestand)
/Subject (Conversie HyperLaTeX naar PDF)
/Creator (LaTeX, DVIHPS, en anderen...)
>>
endobj
...
xref
0 21
0000000000 65535 f
0000043515 00000 n
0000042654 00000 n
0000043762 00000 n
...
startxref
43828
%%EOF

```

Om de verbetering aan de cross-references tabel automatisch te laten gebeuren heeft de auteur het programma flex ontwikkeld, genaamd `recticrt`. Om een executable te verkrijgen voor de eigen omgeving dient de gebruiker vooreerst Flex (versie \geq 2.4.6) op te starten voor `recticrt`, met als optie `-8` op de commandoregel.

```
flex -8 recticrt.lex
```

en vervolgens het uitvoerbestand van Flex te compileren met zijn C compiler (`lex.yy.c` onder UNIX of Macintosh, `lexyy.c` onder DOS enzovoort). Het gebruik van `rectict` is zeer eenvoudig:

```
rectict < oud bestand > nieuw bestand
```

Indien de cross-reference tabel correct is zal `rectict` deze intact laten.

Een andere courante wijziging bestaat uit het onzichtbaar maken van de omkadering van de actieve zones van de hypertext verbanden. Elk hypertext verband is gedefinieerd door een PDF object van het type

```
12 0 obj
<<
/Type /Annot
/Subtype /Link
/Rect [ 148 399 177 411 ]
/Border [ 1 1 1 ]
/Dest [ 4 0 R /FitH 607 ]
>>
endobj
```

De commando's `/Rect` en `/Dest` bepalen de coördinaten van de actieve begin- en eindzones. De parameters van `/Border` zijn:

1. de horizontale straal
2. de verticale straal
3. de lijndikte

De defaultwaardes zijn 1 1 1. Door de derde parameter te wijzigen door 0 maakt men het kader onzichtbaar. Indien met oplet dat er geen spaties toevoegt of wegneemt wordt er ook geen verschuiving veroorzaakt. Het betreft dus een wijziging die geen aanpassing vereist van de cross-references tabel en zijn pointer.

De lezer kan de volledige beschrijving van het PDF formaat vinden in [1].

7 Beschikbaarheid

Adobe Acrobat Reader is beschikbaar op diverse public servers. Adobe Acrobat Distiller kan bekomen worden via Adobe Systems Inc.

DVIHPS (de gewijzigde versie van DVIPS) is — op het ogenblik dat dit artikel geschreven werd — in het stadium van β -test. De beschikbaarheid ervan (of de integratie van de functionaliteiten ervan in DVIPS, die in tegenstelling met T_EX en METAFONT nog niet bevroren is) zullen aangekondigd worden via de gebruikelijke kanalen.

De utility `repere`, de overeenkomstige L^AT_EX style en de utility `rectict` kunnen worden bekomen via de public server `ftp.ens.fr`, ze bevinden zich in de directory `/pub/tex/yannis/acrobat`. De lezer zal er tevens de voorbeelden beschreven in dit artikel kunnen vinden, alsook het artikel zelf (in Times, om de redenen vermeld in 3.2) onder de vorm `.tex`, `.bbl`, `.rep`, `.dvi`, `.ps` en `.pdf`.

Referenties

- [1] Adobe Systems Inc, *Portable Document Format Reference Manual*. Addison Wesley, 1st edition, 1993.
- [2] M. Goossens, 'L^AT_EX — HTML aller et retour', *Cahiers GUTenberg n° 19*, janvier 1995, 98–120.
- [3] M. Goossens, 'Introduction pratique à SGML', *Cahiers GUTenberg n°19*, janvier 1995, 27–58.
- [4] S. Rahtz, Hypertext marks in L^AT_EX. 1994.
- [5] T.V. Raman, An audio view of T_EX documents. In *13th annual T_EX Users Group Meeting, Portland, 1992*, 1992.