

Bijlage 9

L^AT_EX & Chinees

Erik Frambach
Rijksuniversiteit Groningen
email: E.H.M.Frambach@eco.rug.nl

abstract

Met behulp van het L^AT_EX-package ‘CJK’ is het mogelijk om tekst te zetten in het Chinees, Japans en Koreaans. In dit artikel laat ik in het kort zien wat ervoor nodig is om Chinese tekst met L^AT_EX — Japans en Koreaans blijven buiten beschouwing. Het installeren van de benodigde software wordt behandeld, evenals het gebruik van dit systeem in de praktijk.

keywords

L^AT_EX, Chinees, CJK, fonts, PostScript, Big 5

Introductie

Chinees is voor computers een lastige taal omdat het een enorm groot karakters omvat, in de orde van grootte van **10.000** tot **50.000**. Een toetsenbord met alle karakters, zoals we met westerse talen gewend zijn, is daarom ondoenlijk.

Maar niet alleen het invoeren van Chinese tekst is daarvoor lastiger, ook voor het opslaan moeten andere afspraken gemaakt worden. Het is duidelijk dat een Chinese tekst niet kan worden gerepresenteerd met 1 byte per teken, zoals bij voorbeeld bij Nederlands. Daarvoor zijn minstens 2 bytes nodig. In de loop van de tijd zijn daarvoor verschillende coderingen bedacht, zoals *BG5*, *Big-5* en *CNF*.

Voor het invoeren bestaan inmiddels redelijk gebruikersvriendelijke programma’s die teksten in verschillende coderingen kunnen im- en exporteren.

CJK

Het L^AT_EX-package ‘CJK’ van Werner Lemberg implementeert een vertaalslag van o.a. BG5- en Big-5-gecodeerde teksten naar T_EX-fonts. Daarnaast biedt CJK een heleboel extra’s die nodig zijn om serieus tekst te kunnen zetten in het Chinees, wat verder gaat dan een enkele kreet. CJK definieert bij voorbeeld teksten als ‘Samenvatting’ en ‘Inhoud’ in het Chinees, maar zorgt er ook voor dat ‘Hoofdstuk 1’ op de Chinese manier wordt geschreven: ‘Eerste hoofdstuk’.

Verder zitten er aardigheden in CJK zoals ‘pinyin’, een transcriptie van Chinees naar Latijns alfabet met accen-

ten, en ‘ruby’s’, kleine hulpjes bij de uitspraak van minder gangbare karakters.

Chinees T_EX-fonts

Een minstens zo belangrijk probleem dat CJK oplost is de vertaling naar T_EX-fonts. T_EX kan in een font maximaal **255** karakters opslaan. Daarom moet voor een compleet Chinees fonts de zaak in stukken worden geknipt. CJK heeft dat zo efficiënt mogelijk gedaan, waardoor maximaal ‘slechts’ **55** (sub)fonts nodig zijn. Om het ook voor T_EX behapbaar te maken worden fonts pas geladen als dat echt nodig is. In de praktijk blijkt dat veel teksten met een stuk of tien (sub)fonts nodig zijn. Intern zijn de fontmetrieken overigens erg simpel. Chinees is niet-proportioneel, dus alle “boxen” zijn even breed.

HBF, TrueType en PostScript

Qua fontmetrieken is het dus eenvoudig, maar qua rendering niet. In principe geeft CJK drie mogelijkheden:

- **hbf2gf**: een zogenaamd HBF-font kan met een eenvoudig programma naar ‘gf’ vertaald worden. ‘gf’ kan op zijn beurt weer naar ‘pk’ vertaald worden, een formaat waar elke div-driver mee overweg kan.
- **ttf2pk**: een TrueType-font kan met een eenvoudig programma direct naar ‘pk’ vertaald worden.
- **ps2pk**: idem maar dan met een PostScript-font.

Chinees HBF-, TrueType- en PostScript-fonts zijn hier en daar op Internet te vinden, maar veel is er niet. Ook zijn er verschillen in de compleetheid van fonts. Sommige fonts kunnen “vrij” gebruikt worden, andere alleen voor niet-commerciële toepassingen.

Verder is er een conversieprogramma (een contributie aan CJK) om volledige TrueType-fonts (dus met *alle* karakters!) in kleinere PostScript subfonts te vertalen. Eén zo’n grote TTF (4 tot 8 megabyte) geeft dan 55 PostScript Type 1 fonts van 100 tot 200 kilobyte per stuk.

Omdat **ps2pk** al volledig geïntegreerd is in 4T_EX heb ik ervoor gekozen via die weg pk-fonts te genereren. Als PostScript als uitvoer wordt gekozen worden direct de PostScript Type 1 fonts gebruikt.

Bij dat laatste moet aangetekend worden dat *partial font downloading* daarvoor van groot belang is. Veel PostScript-printers verslikken zich in de massa fonts die

voor een eenvoudig stuk al gauw nodig zijn. Bovendien wordt uit elk font vaak maar een erg klein deel daadwerkelijk gebruikt, zodat de print file ook onevenredig groot zou worden. Gelukkig ondersteunt DVIPS tegenwoordig partial font downloading niet alleen voor PK-fonts, maar ook voor Type 1 fonts. Een dvi-drivers als em \TeX 's DVIHPLJ stuurt ook alleen de gebruikte karakters naar de printer.

Er is een 8-tal min of meer volledige Chinese TrueType-fonts op Internet te vinden die ik naar PostScript subfonts heb vertaald (kost ca. 10 minuten per stuk op een Pentium 166). Bij die vertaling worden 8×55 Type 1 fonts gegenereerd, naast 8×55 tfm-bestanden. Als al die bestanden op de juiste plek zijn geïnstalleerd hoeft alleen hier en daar nog een \LaTeX -fontdefinitie aangepast te worden in de CJK \TeX input directory.

Van Big-5 naar CJK

Voordat we daadwerkelijk met \LaTeX en CJK aan de slag gaan moeten we eerst de Big-5 tekst bewerken. In de code kunnen namelijk backslashes en tekens zoals '{' en '}' voorkomen die, zoals bekend, in \TeX een bijzondere betekenis hebben.

Een conversieprogramma wordt daarvoor meegeleverd. Dat programma schrijft vooraan het bestand een regeltje zodat CJK weet dat het bestand voorberekt is, en het vangt alle 2-byte karakters door een iets andere representatie die niet met \TeX interfereert. Terugvertaling naar het origineel wordt niet ondersteund, dus maak backups als je de tekst later nog met een Chinese editor verder wilt bewerken.

CJK in de praktijk

Als voorbeeld nemen we een klein stukje Chinese tekst dat met CJK wordt meegeleverd:

```
% CJK example
\documentclass[12pt]{article}
\usepackage{CJK}
\begin{document}
\begin{CJK*}{Bg5}{kai}
\CJKtilde
... onleesbare tekens...\textrm{Big-5}~...
onleesbare tekens ... et cetera...
\end{CJK*}
\end{document}
```

Merk op dat binnen een stuk Chinese tekst gewoon \TeX -commando's uitgevoerd kunnen worden, bv. omschakelen naar romeins of vet.

De tekst in dit bestand is puur Big-5 en moet dus nog voorberekt worden voor CJK. Na voorbereking ziet het er zo uit:

```
\def\CJKpreproc{}% CJK example
\documentclass[12pt]{article}
\usepackage{CJK}
\begin{document}
\begin{CJK*}{Bg5}{ka}
\CJKtilde
...onleesbaar teken 187onleesbaar teken 229...
onleesbaar teken 68... ~\textrm{Big-5} ...
onleesbaar teken 243 ...
\end{CJK*}
\end{document}
```

De precieze betekenis van commando's die CJK allemaal definieert ga ik hier hier uitleggen, daarvoor is een uitgebreide handleiding beschikbaar. De meest basale vermeld ik hieronder.

$\begin{CJK*}{Bg5}{ka}$

Met dit command begin je een \LaTeX -environment waarin tekst staat die Big-5-gecodeerd is. Het font dat gebruikt wordt is 'ka'. Dat betekent dat er een fontdefinitiebestand c00ka.fd moet zijn.

Zo'n fontdefinitiebestand ziet er bij voorbeeld zo uit:

```
\def\fileversion{4.1.3}
\def\filedate{1997/06/20}
\ProvidesFile{c00ka.fd}%
  [\filedate\space\fileversion]
\DeclareFontFamily{C00}{ka}{}
\DeclareFontShape{C00}{ka}{m}{n}%
  {<-> CJK * b5ka}{}
\DeclareFontShape{C00}{ka}{bx}{n}%
  {<-> CJKb * b5ka}{\CJKbold}
```

\LaTeX /CJK verwacht dat de fonts b5ka01 t/m b5ka55 beschikbaar zijn, dat wil zeggen tfm-bestanden. Voor uitvoer zijn overeenkomstige pk- of pfb-bestanden nodig.

Het verschil tussen de CJK en de CJK* omgeving is dat in de laatste spaties en regelovergangen worden genegeerd. Meestal is dat wat je wil. Merk op dat regels niet op willekeurige plekken afgebroken kunnen worden, omdat elk Chinees karakter 2 tekens inneemt. Midden in een karakter breken zal niet goed gaan.

\CJKbold

CJK emuleert desgewenst een vette variant van een willekeurig CJK-font door karakters meerdere keren te printen met heel kleine verschuivingen. Het Chinees kent eigenlijk geen vet of cursief zoals wij dat kennen, maar voor sommige toepassingen is deze methode heel handig. In een stuk Chinese tekst kan nu op de gewone manier (met $\backslash bf$ of $\backslash textbf$) vet geselecteerd worden. Cursief is er niet.

Overigens werken `\small`, `\Large` en dergelijke ook gewoon.

`\CJKcaption`

Met dit commando kan een ‘caption’ bestand geladen worden. Daarmee worden de juiste definities geladen voor zaken als ‘Inhoud’ en ‘Samenvatting’. Voorbeeld: `\CJKcaption{Bg5}`. Eigenlijk is dit nog niet genoeg om Chinese tekst op typografisch verantwoorde wijze te zetten. Om ‘Eerste hoofdstuk’ en dergelijke te krijgen is meer nodig. Met behulp van het \LaTeX -pakket `koma-script` is dit te realiseren. In Werner Lembergs artikel in *Die TeXnische Komödie* wordt in detail uitgelegd hoe dat moet. In ongeveer tien regels aan code is dat te implementeren.

Pinyin

CJK definieert ook de ‘pinyin’ stijl. Dat is een soort transcriptiemethode voor Chinees naar westerse tekens. Het \LaTeX -package heet `pinyin.sty` maar is, in tegenstelling tot de kern van CJK, niet afhankelijk van \LaTeX : in plain \TeX kun je de stijl ook gebruiken door het commando `\input pinyin.sty` te geven.

Een stukje pinyin, met ‘ruby’s’ erboven, ziet er bij voorbeeld zo uit:

我 很 喜 歡 吃 中 國 飯 。
wǒ hěn xǐhuān chī Zhōngguó fàn .

Dit betekent: ik eet graag Chinees. De \TeX -code voor pinyin is voor iedereen die iets van Chinees weet heel voor de hand liggend: `\Wo3 \hen3 \xi3\huan1 \chi1 \Zhong1\guo2 \fan4`

Verticaal schrift

Traditioneel Chinees schrift loopt in kolommen van boven naar onder. De kolommen lopen van rechts naar links. CJK heeft ook hiervoor een oplossing die voor eenvoudige gevallen goed genoeg is: domweg alle karakters een kwart slag draaien! De papieren afdruk draai je vervolgens een kwartslag terug et voilà! Typografen zullen echter snel fouten kunnen aanwijzen, met name bij de interpunctie die in verticale tekst anders kan zijn dan in horizontale.

Voorbeelden

Nu we de belangrijkste kenmerken gezien hebben laat ik wat stukjes zien uit de in het Chinees in \LaTeX geschreven handleiding van CJK.

De titelpagina van de handleiding:

Big-5 碼中文系統下的 CJK $\LaTeX 2_\epsilon$ 排版系統 (For CJK version 4.1.0)

李君宇 (Chun-Yu Lee)

Werner Lemberg

Een stukje uit sectie 5, over het installeren van fonts:

5 字型安裝

本文僅以由 True Type 字型產生的 PK 字型或 PS 字型為說明。字型之取得參考附錄 B 之說明。

PK 字型 (或 PS 字型) 最好安裝在當你 要升級 \TeX/\LaTeX 系統時不會影響的位置, 因為一旦此類字型產生後, 即可長時期的使用; 另外, 如果每次要升級 CJK 系統或 \TeX/\LaTeX 系統時, 也可放心大膽的下手。

Hetzelfde stukje, maar nu niet met het ‘KaiSu’ font maar met ‘ChuYuanTi’:

5 字型安裝

本文僅以由 True Type 字型產生的 PK 字型或 PS 字型為說明。字型之取得參考附錄 B 之說明。

PK 字型 (或 PS 字型) 最好安裝在當你 要升級 \TeX/\LaTeX 系統時不會影響的位置, 因為一旦此類字型產生後, 即可長時期的使用; 另外, 如果每次要升級 CJK 系統或 \TeX/\LaTeX 系統時, 也可放心大膽的下手。

Meer weten

Het CJK-pakket is op CTAN (Comprehensive \TeX Archive Network) te vinden in de directory `language/chinese/-CJK`. Een aantal fonts zijn te vinden in `fonts/CJK`. Op de 4all \TeX cdroms versie 4 is het pakket (althans het Chinese deel) volledig geïnstalleerd.

Als achtergrond is het verstandig te lezen “Das CJK-Paket für $\LaTeX 2_\epsilon$ ” van Werner Lemberg, in *Die TeXnische Komödie* 4/1995. In de nog te verschijnen proceedings van de TUG conferentie 1997 zal een nieuwer en uitgebreider artikel over CJK in het Engels verschijnen.