

Zpravodaj Československého sdružení uživatelů TeXu

Ondřej Jakubčík

TrueType fonty, TeX a čeština

Zpravodaj Československého sdružení uživatelů TeXu, Vol. 16 (2006), No. 1, 33–40

Persistent URL: <http://dml.cz/dmlcz/150003>

Terms of use:

© Československé sdružení uživatelů TeXu, 2006

Institute of Mathematics of the Czech Academy of Sciences provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.



This document has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* <http://dml.cz>

Uživatelé T_EXu používají v drtivé většině fonty ve formátu Adobe Type 1, případně fonty v METAFONTu. Bohužel, ne vždy je ten správný font v jednom z těchto formátů k dispozici. TrueType fonty nabízejí poměrně zajímavou alternativu k těmto zavedeným formátům.

Úvod

Přestože T_EX není zdaleka jen záležitostí Unixových systémů, používá i ve Windows hlavně Postscriptové fonty. Jejich výhodou je, že každá laserová tiskárna dokáže Type 1 fonty vykreslit, a tím je výsledný Postscriptový soubor velmi snadno přenositelný. Nevýhodou je, že firma Adobe nebyla ve svém Adobe Standard Encoding k češtině právě štědrá, a pokud některé fonty obsahují i akcentovaná písmena, pak jsou to většinou kompozity, někdy i generované automaticky (typickým příkladem jsou potom písmena vypadající asi takto: *ť, ě, ě*).

Program **a2ac** do značné míry kompenzuje tuto nevýhodu, ovšem uživatelé používající stále rozšířenější formát PDF se musí smířit s tím, že z jejich dokumentů nepůjde kopírovat text. To v případě dokumentů určených jen pro tisk nemusí být na škodu, ale v případě, že je s nimi třeba dále pracovat, jsou akcenty separované od písmen samotných poměrně nepraktické. I přes značnou konfigurovatelnost programu **a2ac** přeci jen není nad to mít odpovídající akcentovaná písmena spíše než kompozity. Toto je obzvláště důležité pro specifická česká (a slovenská) písmena *ť, ě* a *ě*, neboť u těchto písmen se prostě opravdová klíčka dá nahradit apostrofem jen velmi obtížně.

S rozmachem TrueType a OpenType fontů se postupně od Type 1 fontů upouští, a některé fonty v tomto formátu již nejsou ani v katalozích. Pro T_EXovou komunitu by bylo nebezpečné, kdyby zaspala a nevyužila možnosti, které TrueType a OpenType fonty nabízejí.

Abych nemohl být nařčen z toho, že níže popsané metody jsou již známé (viz např. [5], [6], [7]), uvedu, že jsou specifické pro kódování Cork, které trpí podobnými neduhy jako Adobe Standard Encoding. V Čechách se mi kódování XL2, popřípadě XT2 jeví mnohem vhodnější.

Do nedávné doby byl pojem TrueType spojen s Windows. Fonty ve Windows měly jisté neduhy, co se týče ligatur, kerningu a akcentovaných písmen obzvláště. Pokud vím, není mnoho programů pro Windows, které by z písem ligatury a kerning využívaly (snad až na sázecí programy jako je PageMaker, QuarkXPress a InDesign), proto některé fonty vůbec tyto informace neobsahují.

Ovšem pod vlivem $\text{T}_{\text{E}}\text{X}$ u a fontu Computer Modern jsem došel k názoru, že teprve správný kerning dělá font fontem, a že sebekrásnější písmo bez odpovídající metriky je v podstatě pro vážnější práci nepoužitelné.

Dostupná písma

Uživatel Windows 2000/XP má v systému nainstalována písma od firem Monotype a Linotype, která může použít. Protože systém Windows nepoužívám, jal jsem se hledat jinde. Na webových stránkách firmy Artifex [1] jsou volně ke stažení fonty od společnosti URW++. Tyto fonty obsahují všechna akcentovaná písmena (alespoň pro češtinu), ligatury a dostatek kerningových párů. Nemohu se nezmínit také o písmu Lido ze Střešovické písmolijny [2], které je nyní dostupné ve formátu jak TrueType, tak OpenType.

Unicode

TrueType fonty se vyznačují mimo jiné tím, že každý znak má svoje označení v Unicode. To je výhodné, neboť fonty nemusí nést tabulku Postscriptových jmen (a také ji někdy nenesou).

Předpoklady

Pro úspěšné provedení níže popsané procedury je základním předpokladem aktuální distribuce $\text{T}_{\text{E}}\text{X}$ u a nějaký Unixový systém. Já jsem použil Linux a $\text{t}_{\text{E}}\text{X}$ 3.0. Popsané postupy každopádně vyžadují $\text{pdf}_{\text{T}_{\text{E}}\text{X}}$ 1.21a. Příkazy jsou psané v `bash`i, což je, jak doufám, standardní Unix-kompatibilní shell.

Postup asi ve Windows použitelný nebude, možná bude fungovat, pokud máte nainstalované prostředí Cygwin.

Příprava

Nejprve budeme potřebovat kódovací soubory a skript ze stránky [3]. Také budeme potřebovat nějaké TrueTypové fonty, jako příklad použijeme právě fonty od URW. Kódovací soubory uložíme do adresáře, který $\text{T}_{\text{E}}\text{X}$ (resp. `dvips` a `pdftex`) prohledává, a aktualizujeme databázi příkazem `mktextlsr`. Skript `uni-names.sed` uložíme např. do `/usr/local/bin`.

Příprava adresářů

V dočasném adresáři vytvoříme adresáře `afm`, `tfm`, `truetype`, a `vf`. V každém z nich vytvoříme složky `<písmolijna>/<rodina>`. V našem případě použijeme písmo URW Garamond No. 8, takže v Linuxu napíšeme například:

```
cd /tmp
mkdir -p {afm,tfm,truetype,vf}/urw/garamond
Do adresáře /tmp rozbalíme soubor urwfonts-1.41.tar.bz2 a zkopírujeme soubory do odpovídajícího adresáře a pojmenujeme je trochu výhodněji:
tar xvfj /<cesta>/urwfonts-1.41.tar.bz2
cp urwfonts-1.41/GaramondNo8-Reg.ttf \
    truetype/urw/garamond/ugar.ttf
cp urwfonts-1.41/GaramondNo8-Ita.ttf \
    truetype/urw/garamond/ugari.ttf
cp urwfonts-1.41/GaramondNo8-Med.ttf \
    truetype/urw/garamond/ugam.ttf
cp urwfonts-1.41/GaramondNo8-MedIta.ttf \
    truetype/urw/garamond/ugami.ttf
```

Příprava metrik AFM

Nyní je na čase vygenerovat metriky fontu ve formátu AFM. V $\text{T}_{\text{E}}\text{X}$ u je na to utilita `ttf2afm`, kterou použijeme:

```
for i in uga{r,m}-{,i}
do
    ttf2afm -o afm/urw/garamond/${i}.afm truetype/urw/garamond/${i}.ttf
done
Pokud se neobjevilo žádné varování, pak je vše v pořádku a můžeme postoupit k dalšímu kroku. Pokud si však program ttf2afm postěžuje
Warning: ttf2afm (file font.ttf): no names available in the
‘post’ table, print glyph names as indices
pak je nutné příkaz upravit na:
for i in uga{r,m}-{,i}
do
    ttf2afm -u truetype/urw/garamond/${i}.ttf | \
    uni-names.sed > afm/urw/garamond/${i}.afm
done
```

Příprava metrik TFM

Aby byl \TeX schopen s fonty pracovat, potřebuje k nim mít metriku TFM. Pro TrueType fonty je dále nezbytné mít i virtuální font, neboť ten obsahuje ligační tabulky.

Pro převod metrik AFM na metriky TFM slouží program `afm2tfm`. Pomocí následujícího příkazu převedeme metriky na soubory TFM a VF:

```
for i in uga{r,m}{,i}
do
  afm2tfm afm/urw/garamond/${i}.afm -T ntt-xl2.enc -v ${i}8z.vpl \
  l2${i}.tfm >> garamond.map
  vptovf ${i}8z.vpl
  mv ${i}8z.vf vf/urw/garamond
  mv ${i}8z.tfm tfm/urw/garamond
  mv l2${i}.tfm tfm/urw/garamond
  rm ${i}8z.vpl
done
```

Nyní, když máme virtuální fonty a metriky normálních fontů, může se stát, že požadujeme ještě například malé kapitálky. Ty vygenerujeme takto:

```
afm2tfm afm/urw/garamond/ugar.afm -T ntt-xl2.enc -c 0.67 \
  -V ugar8z.vpl l2ugar.tfm
afm2tfm afm/urw/garamond/ugam.afm -T ntt-xl2.enc -c 0.73 \
  -V ugam8z.vpl l2ugam.tfm
vptovf ugar8z.vpl
vptovf ugam8z.vpl
mv *.vf vf/urw/garamond
mv uga*.tfm tfm/urw/garamond
rm l2ugam.tfm l2ugar.tfm
rm ugar8z.vpl ugam8z.vpl
```

Konstanty jsou zjištěny spočítáním poměru nejvyššího velkého písmena a nejvyššího malého písmena ve skutečném kapitálkovém fontu. Obrázek, ze kterého je možno toto spočítat, je dostupný např. na www.myfonts.com, kde vybereme skutečný kapitálkový font, zadáme zkušební text Ix (velké I a malé x) a v libovolném bitmapovém editoru změříme výšky těchto písmen a z nich poměr vy počítáme.

Úprava map

Než budeme moci písma použít, musíme upravit soubor `garamond.map` a přidat na každý řádek odpovídající jméno TrueTypového souboru, tj. změnit řádek `l2ugar GaramondNo8-Reg " XL2encoding ReEncodeFont " \`

```

<ntt-xl2.enc
na
l2ugar GaramondNo8-Reg " XL2encoding ReEncodeFont " \
<ntt-xl2.enc <ugar.ttf
případně, pokud jsme použili skript uni-names.sed,
l2ugar GaramondNo8-Reg " XL2encoding ReEncodeFont " \
<tt-xl2.enc <ugar.ttf

```

Řádky jsou zde rozdělené z důvodu formátování, ale správně mají být vcelku.

Finální úpravy

Nyní přesuneme adresáře `afm`, `tfm`, `truetype`, `vf` a soubor `garamond.map` do adresářové struktury \TeX u (obvykle `<něco>/texmf/fonts`) a zajistíme, aby `pdftex` načetl soubor `garamond.map`, takže např. v $\text{te}\text{\TeX}$ u přidáme řádek

```
Map garamond.map
```

na konec souboru `updmap.cfg` a aktualizujeme jmennou databázi příkazem `mktextlsr` a aktualizujeme mapy příkazem `updmap`. Tím máme písma nainstalovaná.

Formáty písem

Pro srovnání bych rád uvedl několik faktů o čtyřech formátech fontů, které připadají v úvahu pro použití v \TeX u. Existují ještě jiné formáty, které se ale v Čechách nepoužívají, ať již proto, že jsou pro češtinu zbytečné (jako třeba CID fonty s čínštinou), nebo se jedná o interní technologie firem (např. IntelliFont).

METAFONTové fonty jsou známy každému, kdo používá \TeX . U rodiny Computer Modern byla poprvé prezentována myšlenka generování široké škály fontů, lišících se od sebou šířkou, sklonem, výškou atp. Tato myšlenka byla v jistém slova smyslu znovuobjevena v Multiple Master fontech od firmy Adobe [12], které ovšem již také nejsou oficiálně podporovány. Výhodou je, že libovolný matematický výraz může být použit jako vstup pro vykreslení. Ve zdrojových souborech jsou rovnou uvedeny informace o kerningu, ligaturách, akcentech a kódování.

Adobe Type 1 [10] je formát vyvinutý pro Postscriptové tiskárny. Je podporován všemi tiskárnami, které splňují standard PostScript Level 1. Jsou popsány Bézierovými křivkami, částmi kruhových oblouků a úsečkami. Podporují hinting, který umožňuje vykreslit fonty při nižším rozlišení tak, aby vypadaly přijatelně. Vyznačují se tím, že metriky jsou samostatně v souborech AFM nebo PFM.

TrueType [15] fonty byly vyvinuty firmou Apple ve spolupráci s firmou Microsoft. Pro popis znaků používá Bézierovy křivky a úsečky. Používá pro hinting specifické instrukce, které jsou patentovány firmou Apple Computer, Inc. Metriky jsou součástí souboru s fontem.

Zvláštní odnoží písem TrueType je písmo *Type 42* [18], které obsahuje celý soubor s fontem TrueType ve speciální Postscriptové obálce, která umožňuje načíst tento font do Postscriptové tiskárny. Bohužel podpora fontů Type 42 je součástí teprve poměrně nových tiskáren, takže u dokumentů není zaručena stoprocentní přenositelnost.

Několikrát jsem zmínil formát *OpenType* [16]. Je to výsledek společné práce firmy Adobe a Microsoft, a jako takový spojuje výhody TrueType fontů a Type 1 fontů. Může obsahovat buď TrueTypový font nebo font Type 2 či CFF. *CFF* [14] je způsob, který umožňuje uložit více fontů Type 1 do jednoho souboru efektivnějším způsobem, *Type 2* [13] zase umožňuje kompaktnější popis fontu. Výsledkem je formát, který umožňuje snadnou konverzi z předešlých dvou formátů do jednoho univerzálního. Metriky jsou opět součástí souboru s fontem.

Podpora ze strany \TeX u

\TeX podporuje všechna písma, ke kterým je schopen načíst metriky ve formátu TFM. Když získáme metriky z písem (u formátu OpenType je toto zatím složitější, podpora je připravována pro projekt Ω [17], ale neměl jsem tu možnost jej důkladně vyzkoušet), dalším problémem je přidání těchto fontů do dokumentu.

Program *dvips* podporuje přidání (i částečné) fontů Type 1 a Type 42. Pro TrueType a OpenType zatím podpora není.

O něco lépe je na tom *pdf \TeX* , který umožňuje přidat do dokumentu jak Type 1 a Type 42, tak přímo TrueTypový soubor. Tyto fonty umí přidat i částečně. Pokud chceme použít font OpenType, pak musíme fontový soubor přidat celý, což neúměrně zvětšuje soubor PDF. Úplné vložení fontu do dokumentu může způsobit i právní problém, neboť některé písmolijny nepovolují úplné vložení (tvrdí, že soubor s fontem je možné z PDF souboru extrahovat). Pro zájemce o použití fontů OpenType jsou potřebné nástroje k dispozici na CTANu [19].

Pár slov závěrem

V ukázkách byl použit font URW Garamond No. 8, nicméně můžeme použít jakýkoli font. Uživatelům Linuxu bych doporučil zaměřit se například na fonty Luxi, které jsou dodávány společně s X.Org [4]. Soubory *ntt-xl2.enc*, *ntt-xt2.enc*, *tt-xl2.enc* a *tt-xt2.enc* jsou upravené soubory *xl2.enc* a *xt2.enc* pana Zdeňka Wagnera. Od originálu se liší absencí znaku j (dotlessj), který, bohužel, Unicode neobsahuje, a taktéž znak Togonek byl nahrazen znakem Tcedilla. Soubory *ntt** obsahují normální Postscriptová jména, soubory *tt** obsahují znaky ve formátu uni0000, které je možno použít ve stavu nouze, nebo pokud nechceme generovat malé kapitálky a upravovat metriky skriptem. Nevylučuji použití ani

pod Windows, kde se za normálních okolností program `sed` nevyskytuje, a tudíž by skript `uni-names.sed` nefungoval.

Právní doslov

V textu byly použity pojmy, které mohou být registrovanými, obchodními nebo jinými značkami či známkami.

Reference

- [1] <http://www.artifex.com/downloads/>
- [2] <http://www.pismolijna.cz>
- [3] zatím nic
- [4] <http://www.x.org>
- [5] <http://www.radamir.com/tex/ttf-tex.htm>
- [6] <http://www.csit.fsu.edu/~mimi/tex/doc/guides/truetype/>
- [7] <http://www.tug.org/tex-archive/info/TrueType/>
- [8] <http://ipe.compgeom.org/pdftex.html>
- [9] Adobe Systems, Inc.: *PostScript language reference manual*, 3rd edition, Addison-Wesley Publishing, 1999
- [10] Adobe Systems, Inc.: *Adobe Type 1 Font Format*, 3rd printing, Addison-Wesley Publishing, 1993
- [11] Adobe Systems, Inc.: *Adobe Font Metrics File Format Specification*, version 4.1, <http://partners.adobe.com/>
- [12] Adobe Systems, Inc.: *Designing Multiple Master Typefaces*, <http://partners.adobe.com/>, 1995–1997
- [13] Adobe Systems, Inc.: *The Type 2 Charstring Format*, Technical Note #5177, <http://partners.adobe.com/>, 2000
- [14] Adobe Systems, Inc.: *The Compact Font Format Specification*, Technical Note #5176, <http://partners.adobe.com/>, 2003
- [15] Microsoft Corp.: *TrueType 1.0 Font Files Technical Specification*, revision 1.66, <http://www.microsoft.com/typography>, 1995
- [16] Microsoft Corp.: *OpenType specification*, version 1.4, <http://www.microsoft.com/typography>, 2002
- [17] Mehta, A., Bella, G., Haralambous, Y.: *Adapting Ω to OpenType Fonts*, TUGBoat, Volume 24, 2003
- [18] Adobe Systems, Inc.: *The Type 42 Font Format Specification*, Technical Note #5012, <http://partners.adobe.com/>, 1998
- [19] CTAN: [/fonts/utilities/fonttools](#)

Summary: TrueType fonts, T_EX and Czech language

TrueType and OpenType fonts have several advantages compared to the well-known Type 1 fonts, which have been used for years by T_EX users. In this article a way how to use TrueType fonts with T_EX for typesetting in Czech and Slovak language is presented. While instructions are specific for encoding XL2 and XT2 (which are compatible with ISO 8859-2), the encoding vectors can be easily modified to support other encodings, like Cork.

Ondřej Jakubčík

e-mail: ojakubcik@seznam.cz

Zpracování pomocných T_EXových souborů pomocí XSLT 2.0

ZDENĚK WAGNER

Článek nastiňuje možnosti zpracování pomocných T_EXových souborů procesorem XSLT 2.0. Koncept je demonstrován reimplementací programu MakeIndex. Autor se též zamýšlí nad možností reimplementace BibT_EXu pouze pomocí XSLT.

1. Úvod

V moderních operačních systémech se začíná prosazovat jako standard kódování Unicode, resp. UTF-8. V programech se tak projevuje obdoba problému Y2K. Zatímco dosavadní kódování vystačila pro téměř každý jazyk s osmibitovými znaky, nyní potřebujeme šestnáctibitové. Úprava programů však není zcela triviální. Kódování UTF-8 umožňuje úsporu místa na disku tím, že znaky US abecedy zabírají pouze jeden bajt, zatímco akcentované znaky a znaky východoasijských jazyků jsou vícebajtové. Navíc může být program provázán s dalšími pomocnými soubory, např. s fonty. Programátor tak musí řešit komplexní problém.

Stejná potíž postihla též T_EX a jeho podpůrné programy. Zpracování vstupu v UTF-8 lze v T_EXu řešit různými metodami, z nichž nejlepší je encT_EX [1], neboť z hlediska programátora maker se znak stále chová jako znak. Makro nikdy nevidí polovinu znaku. Některé úlohy se však řeší pomocí externích programů a některé z nich ještě na zpracování UTF-8 upraveny nejsou.