

# Slovenčina a podpora UTF-8 kódovania v L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub>.

Róbert Vaško

1. február 2005

**Obsah**

1	Úvod	2
2	Inštalácia balíka <code>unicode</code>	2
3	Použitie balíka <code>unicode</code> s balíkom <code>babel</code>	2
4	Použitie balíka <code>unicode</code> s formátom <code>CSL<sup>A</sup>T<sub>E</sub>X</code>	3
5	Úprava balíka <code>DipTeX</code>	3
6	Slovenské písmena	4

## 1 Úvod

Linux od verzie 2.4<sup>1</sup> využíva kódovanie UNICODE presnejšie UTF-8, pričom L<sup>A</sup>T<sub>E</sub>X nemá priamu podporu UTF-8 kódovania. Podpora pre UTF-8 kódovanie je možná tromi spôsobmi, a to:

1. Formátom resp. sadou šablón `PassiveTEX`, ktorý sa používa pri spracovaní XML dokumentov, ktorého výsledkom je XSL transformácia do formátovacích objektov.
2. Balíkom `utf8`, ktorý je doplňujúcim resp. rozširujúcim balíkom pre balík `inputenc`.
3. Balíkom `unicode`, taktiež známy aj ako `ucs` alebo `utf8x`, ktorý je taktiež rozširujúcim pre balíkom pre balík `inputenc`.

Prvá možnosť sa zdá užitočná iba pri písaní XML dokumentov, druhá možnosť sa zdá byť ideálna, ale je iba vo vývojovom štádiu a zatiaľ nepodporuje všetky UTF-8 znaky. Tretia možnosť je najvhodnejšia, pretože podporuje všetky UTF-8 znaky (vrátane tzv. CJK znakov) a dá sa použiť s balíkom `babel`, ako aj C<sup>S</sup>L<sup>A</sup>T<sub>E</sub>X-om.

## 2 Inštalácia balíka unicode

Balík `unicode` nie je štandardným balíkom te<sub>E</sub>X-u, ale dá sa stiahnuť z CTAN archívu alebo priamo z adresy: <http://www.unruh.de/DniQ/latex/unicode/>. Inštalácia je triviálna a pozostáva z kopírovania niekoľkých súborov do priečinka, kde T<sub>E</sub>X hľadá štandardne balíky. Taktiež je potrebné obnoviť `ls-R` databázu.

## 3 Použitie balíka unicode s balíkom babel

Pri použití balíka `babel` nie je implicitne zakomponovaná podpora pre delenie slovenských slov. Najprv je potrebné odkomentovať riadok `skhyp.tex`<sup>2</sup> v súbore `language.dat` a rebuild-núť formáty. Potom sa už balík `unicode` používa veľmi jednoducho. Napríklad:

```
\documentclass{article}
\usepackage[utf8x]{inputenc}
\usepackage[slovak,english]{babel}
\usepackage{palatino,verbatim}
\newcommand\sktxt[1]{\foreignlanguage{slovak}{#1}}
\begin{document}
\sktxt{ Slovenský text s~diakritikou. }
\end{document}
```

Je možné písať dokument aj bez príkazu `\sktxt`, pokiaľ využívate iba jeden jazyk.

<sup>1</sup>Tento text je napísaný programom KWrite, pod OS Linux Fedora Core 2 (2.6.5-1.358 (i686)).

<sup>2</sup>Alebo `skhyp2e.tex` pokiaľ nemáte nainštalovaný C<sup>S</sup>L<sup>A</sup>T<sub>E</sub>X.

## 4 Použitie balíka unicode s formátom $\LaTeX$

Použitie balíka je nasledovné:

```
\documentclass[a4paper,12pt]{article}
\usepackage{ucs,slovak}
\usepackage{mathpple} % Písmo palatino pre matematické
                      % a textové prostredie
\usepackage[latin2,utf8x]{inputenc}
%% PreAmbula ...
\begin{document}
Dokument písaný v \cs\LaTeX{u s~d' t\ / ň l' písmenkami.
```

Poznámka:

Keď v dokumente použijeme písmo `\emph{palatino}`, za písmenom `t` sa odporúča použiť kurzívovú korekciu (medzeru) `\ /`.

```
\end{document}
```

Preklad zdrojového súboru sa nevykoná klasickým príkazom `cslatex` ale príkazom:

```
[work@localhost ltx]# tex -fmt=cslatex diptexhc.tex
```

pretože `cslatex` využíva tabuľku<sup>3</sup> na preklad z ISO Latin 2 kódovania na interné texovské kódovanie<sup>4</sup>.

## 5 Úprava balíka DipTeX

Balík `DipTeX` slúži na písanie diplomových prác pre študentov FEI TU v Košiciach. Úprava tohto balíka pre použitie s balíkom `unicode` je veľmi jednoduchá. Najprv je potrebné konvertovať súbory `feidiphc.tex`, `feidipp.tex` a `feidipsp.tex` do kódovania UTF-8 a potom upraviť tieto súbory tak, aby používali balík `unicode`. Úprava spočíva v použití príkazu:

```
\usepackage[utf8x]{inputenc}
```

Podrobnejšie sme o tom písali v kapitole 3 a 4. Konverzia súborov do formátu UTF-8 sa vykoná napríklad príkazom:

```
[work@localhost]# iconv -f ISO_8859-2 -t UTF-8 feidiphc.tex >
output.tex
```

Prvý parameter predstavuje vstupné kódovanie, druhý predstavuje výstupné kódovanie a tretí je názov súboru. Pokiaľ je vstupné kódovanie `cp1250cs`<sup>5</sup>, použije sa parameter `WINDOWS-1250`.

<sup>3</sup>Presnejšie súbor `il2-cs.tcx`.

<sup>4</sup>Tuším, že je to OT1 kódovanie.

<sup>5</sup>Windows verzia balíka `DipTeX`.

## 6 Slovenské písmena

A nakoniec zoznam písmien s diakritikou. . .

Í š č ť ž ý á í é ó ň ä ú ô Í  
Ľ Š Č Ť Ž Ý Á Í É Ó Ň Ú Ľ.

A to je všetko priatelia. Ďakujem za čítanie.