

Čeština v Linuxu - UTF-8 kódování

Jana

Osnova:

- 1.** ISO vs. UTF-8 kódování
- 2.** nastavení UTF-8 v debianu
- 3.** Čeština v cslatexu
- 4.** Čtení emailů

Něco málo o kódování

- **Ascii**
 - 7bitové kódování
 - obsahuje 127 znaků, neobsahuje diakritiku
- **ISO-8859-2** (Linux)
 - 8bitů na jeden znak
 - Zahrnuje češtinu, slovenštinu, maďarštinu, ...
- **CP-1250** (MS Windows)
 - 8bitů na znak
 - Pro češtinu se liší se od iso-normy v umístění několika znaků (Ž, Š, Ť, ž, š, ť)
- **UTF-8**
 - Patří mezi unicode – jednotné kódování pro všechny abecedy
 - Nemá pevnou délku znaku
 - Čeština: většina znaků zabírá 1 byte, znaků s diakritikou (2 byty) není mnoho
 - **V současné době preferované kódování**

Nastavení UTF-8 kódování na Debianu

- **V domovském adresáři uživatele soubor *.xsession*** (moje nastavení)

```
export LANG=en_US.UTF-8
export LC_CTYPE=cs_CZ.UTF-8
export LC_MESSAGES=en_US.UTF-8
export LC_TIME=$LC_MESSAGES
export LC_NUMERIC=$LC_MESSAGES
export LANGUAGES="en_US.UTF-8:cs_CZ.UTF-8"
```

- Význam jednotlivých proměnných: `man locale`

- **Nastavení vhodného písma v Xterm**

- Písmo `iso10646-1`

- Univerzální standard pro multi-jazykové texty – čitelné v iso-kódování i v unicode

- Nastavení písma v *.Xresources*

```
XTerm*font: -misc-fixed-medium-r-normal-*-20-200-75-75-c-100-iso10646-1
```

- Defaultní nastavení Xtermu v souboru `/etc/X11/app-defaults/XTerm`

Nastavení editoru Emacs

- Emacs podporuje UTF-8
 - Pokud používám přepínání klávesnice definované v Xterm (nastaveno v souboru `.xsession`), můžeme psát v emacsu bez problémů česky
- Pokud chceme používat jinou klávesnici než v Xterm (vhodné např. pro psaní v latexu), přidáme do souboru `.emacs`

```
(setq locale-coding-system 'utf-8)
```

```
(set-terminal-coding-system 'utf-8)
```

```
(set-keyboard-coding-system 'utf-8)
```

```
(set-selection-coding-system 'utf-8)
```

```
(prefer-coding-system 'utf-8)
```

← Zavedení utf-8

Definice typu klávesnice

```
(setq default-input-method "czech-prog-2")
```

Přepínání pomocí klávesy „Pause“

```
(global-set-key [pause] 'toggle-input-method)
```

```
(define-key isearch-mode-map [pause] 'isearch-toggle-input-method)
```

- Pokud preferujete Vim, tak ten také podporuje UTF-8

Čeština a UTF-8 v (cs)latexu

- **Proč používat cslatex na české texty**
 - Cslatex umí správně dělit česká slova, latex ne!
 - Latex nezná české uvozovky
 - Musely by se nadefinovat
 - V cslatexu příkaz `\uv{Text v uvozovkách}`
- Čeština a UTF-8 kódování
 - **CSLATEX** (balík texlive-lang-czechslovak)
`\usepackage{czech}`
`\usepackage[utf8]{inputenc}`
 - **LATEX**
`\usepackage[czech]{babel}`
`\usepackage[utf8]{inputenc}`

Čtení emailů - alpine

- **Pine**
 - Nepodporuje UTF-8 kódování
 - Přestal se vyvíjet, místo něho nový mail-client **Alpine**
- **Alpine** (balíčky alpine, alpine-pico)
 - Podporuje UTF-8 kódování!
 - Stejný vzhled i ovládání jako pine (nastavení v souboru *.pinerc*)
 - Defaultně nastaveno UTF-8 pro odchozí maily
 - Příchozí maily v UTF-8 kódování se zobrazují bez chyb

```
ALPINE 1.10(962) MESSAGE INDEX Folder: INBOX Message 12 of 12 +
1 10/16/2007 novaj2bm@ipnp.troja.mff.cuni.c (2K) test
2 11/14/2007 dolezal@ipnp.troja.mff.cuni.cz (1K) Bonn
3 01/03/2008 Jana Novakova (3K) Ahoj
4 Yesterday Jana Novakova (2K) test
5 Yesterday Jana Novakova (2K) test
6 Yesterday Tomas Davidek (2K) Linuxovy seminar
7 Yesterday Jana Novakova (2K) test3
8 Yesterday Jana Novakova (2K) test4
9 14:58 Jana Novakova (2K) test - utf8
10 14:59 Jana Novakova (2K) test-alpine
11 15:32 dolezal@ipnp.troja.mff.cuni.cz (2K) Exkurze do Reze
12 20:13 jana novakova (2K) testíček

? Help          FldrList      P PrevMsg      - PrevPage     D Delete       R Reply
O OTHER CMDS   > [ViewMsg]   N NextMsg      Spc NextPage     U Undelete     F Forward
```

Další užitečné věci

- Alternativy k xtermu
 - Uxterm - podporuje UTF-8 kódování
 - Rxvt-inicode (balíček **rxvt-unicode**)
- Převádění textů mezi různými kódováními
 - Příkaz **cstocs** (balíček cstocs)
 - Převody mezi iso (latin2), UTF-8, ascii,
 - `cstocs il2 utf8 jmeno_souboru`
- Kontrola nastavení různých proměnných
 - Příkaz **export**

```
novaj2bm@ipnp-n93:~$ export
declare -x DBUS_SESSION_BUS_ADDRESS="unix:abstract=/tmp/dbus-9WTzJHGFDn,guid=63efdc7b
4f64c7ff4760503e49664d77"
declare -x DISPLAY=":0.0"
declare -x HISTCONTROL="ignoredups"
declare -x HOME="/home/novaj2bm"
declare -x LANG="en_US.UTF-8"
declare -x LANGUAGES="en_US.UTF-8:cs_CZ.UTF-8"
declare -x LC_CTYPE="cs_CZ.UTF-8"
declare -x LC_MESSAGES="en_US.UTF-8"
declare -x LC_NUMERIC="en_US.UTF-8"
declare -x LC_TIME="en_US.UTF-8"
```