1 Úvod

Na začátek je třeba říci něco o kódování češtiny. Exituje několik kódování, které lze pro češtinu a slovenštinu použít. Jsou to kódování Kamenických (cp895), cp852 (Latin 2), Windows-1250 (cp1250) a ISO-8859-2 (ISO Latin 2). Ve světě UNIXů a internetu se používá kódování ISO-8859-2. Na převody mezi kódováním existuje několik utilit, vyzdvihnu ale dvě: **GNU recode** a **cstocs**. V rámci sjednocení nástrojů, doporučuji (pokud je to možné) používat **GNU recode**. Například převod z Windows-1250 do ISO Latin 2:

```
recode windows-1250..l2 <soubor.txt >soubor-l2.txt
```

2 Konzole

Aby bylo jasné, co a jak nastavit pro to, aby fungovalo psaní českých písmenek na konzoli, je třeba udělat malý úvod jak to funguje.

Ovladač konzole se skládá ze dvou částí, ovladače klávesnice a ovladače obrazovky. Je třeba vhodně nastavit oboje, abychom byli spokojeni.

2.1 Klávesnice

Ovladač klávesnice umožňuje pracovat (mimo jiné) ve dvou režimech, a to osmibitový výstup a unicodový (UTF-8). Režim přepínáme pomocí příkazu `kbd_mode`. Ovladač je implicitně nastaven v osmibitovém režimu, který budeme používat i my. Ovladač klávesnice může být pro každou konzoli **nastaven zvlášť**.

Klávesové mapy zavádíme pomocí příkazu `loadkeys`. Pro český psací stroj s anglickými „z" a „y" to bude třeba bude:

```
loadkeys cz-lat2.kmap.gz
```

Klávesové mapy lze obvykle nalézt v adresáři */usr/share/keymaps/i386/qwerty* nebo */usr/lib/kbd/keymaps/i386/qwerty*.

K dispozici jsou ještě klávesové mapy *qwerty/cz-lat2-prog* a *qwertz/cz-us-qwertz*. Klávesová mapa je **společná pro všechny konzole**. Mezi českou a anglickou klávesnicí se přepíná pomocí klávesy *Pause*.

Odkazy: **kbd_mode(1)**, **loadkeys(1)**, **keymaps(5)**

2.2 Obrazovka

Ovladač obrazovky, lze nastavit do dvou módů UTF-8 (Unicode) a do osmibitového. Zatím se obvykle používá osmibitový režim.

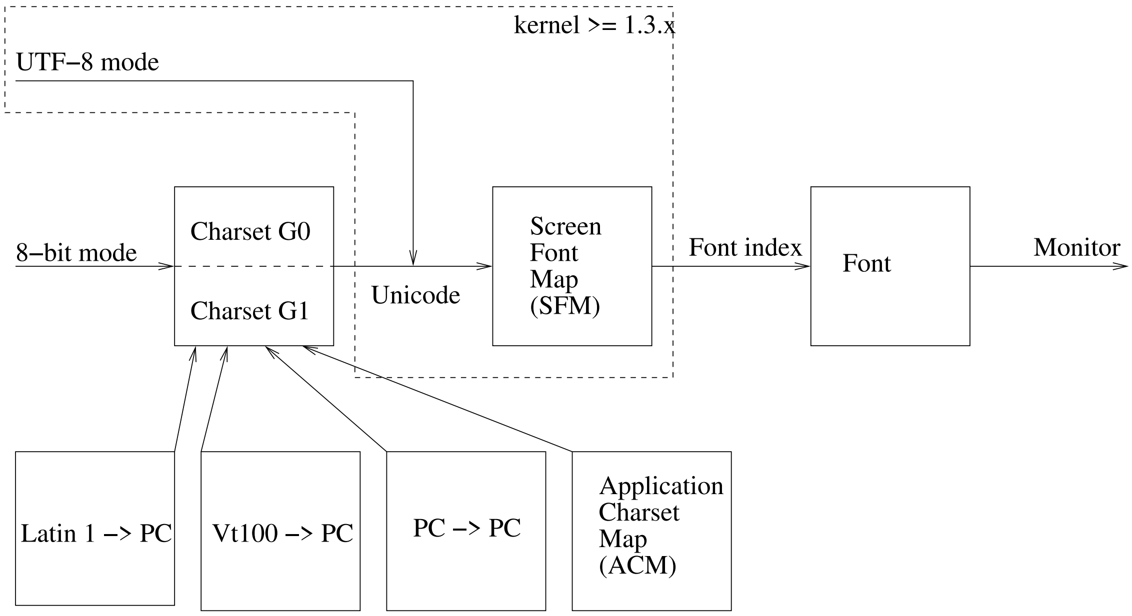
Vnitřně pracuje ovladač v unicode a což nám teoreticky umožňuje zobrazit znaky všech abeced (samozřejmě definovaných v unicode). Bohužel hardware nás omezuje, v případě VGA je to na 512 současně zobrazitelných různých znaků. Proto byl vymyšlen následující mechanismus.

Každý font může mít maximálně 512 znaků. S každým fontem se do jádra nahrává také tabulka, která říká, jakým znakem z fontu se mají zobrazovat znaky z unicodu. Tato tabulka je v anglické literatuře označována jako Screen Font Map (SFM).

Pokud je konzole v UTF-8 režimu, převede se znak z UTF-8 na šestnáctibitový unicodový (UCS2) a pomocí SFM je stanoven znak z fontu, který má být zobrazen (viz obr. 1).

V případě osmibitového režimu, je to trochu složitější. Osmibitový znak je podle aktivního charsetu (G0 nebo G1) překódován do unicodu. Charset může být nastaven do čtyř režimů Latin1 do PC, vt100 grafika do PC, PC do PC a uživatelsky definovaná tabulka ACM.

Tabulky SFM a ACM jsou **globální** tj. společné pro všechny konzole, nastavení režimu (8bit, UTF-8), charsetu (která tabulka) a který charset (G0, G1), je pro **každou konzoli zvlášť**.



Obrázek 1: Schema konverzí znaků, při cestě z aplikace na obrazovku

Pro nastavení klávesnice, fontu a tabulek (Maps) se používal balík **kbd**, který nastavoval klávesnici, font a pouze ACM, jelikož starší jádra měla pouze ACM tabulku. Později byl balík **kbd** nahrazen balíkem **console-tools**, který již umí zavést unicodové fonty a nastavit SFM.

Nastavení fontu:

```
consolechars --acm=iso02 --font=lat2u-16
```

případně `setfont`, používáte-li balík **kbd**. SFM tabulka je někdy (většinou) přibalena ve fontu, a proto ji není třeba zavádět zvlášť.

Pro každou konzoli zvlášť je třeba (například do */etc/profile*):

```
echo -ne '\033%@'  #8bitový režim
echo -ne '\017'    #Použití charsetu G0
echo -ne '\033(K'  #V G0 použít ACM
```

Odkazy: **consolechars(8)**, **console_codes(4)**, **unicode(7)**, **utf-8()**, **setfont(8)**, **mapscrn(8)**

3 Locales

Tato část systému se stará, o definici chování prostředí v různých jazycích a to definice znaků abecedy (co jsou malá písmena, co velká, převody z malých na velké a opačně), pravidla při abecedním řazení, definice zápisu numerických a peněžních údajů, umožnění komunikace programu ve více jazycích.

Locales jsou rozdělena do několika kategorií:

- LC_CTYPE – kategorie znaků, ovlivňuje rozpoznávání jednotlivých znaků programem (co je malé písmeno, co velké, jaké je malé písmeno k velkému a pod.)
- LC_COLLATE – použita při porovnávání znaků, tj. abecedním řazení
- LC_TIME – použita při vstupu a výstupu časových údajů, definuje jejich formát.

- LC_NUMERIC – použita při vstupu a výstupu numerických údajů (čísel). Například pro zapnuté českého prostředí by se číselné údaje měli vypisovat následovně: 1 000 125,00

- LC_MONETARY - definuje výstup peněžních údajů, formát peněžních čísel, označení měny, kde má být umístěna označení umístěno, zda má být odděleno mezerou.

- LC_MESSAGES - v jakém jazyce bude program komunikovat (vypisovat zprávy).

Každou kategorii můžeme nastavit zvlášť a díky tomu můžeme program nastavit třeba tak, aby komunikoval v anglickém jazyce, ale texty, které budeme psát a upravovat považoval za české a třeba i řadil slova podle českých pravidel.

Jednotlivé kategorie nastavujeme proměnými prostředí, jejichž jména jsou shodná s názvy kategorií. Např:

```
export LC_CTYPE=cs_CZ.ISO-8859-2
```

Existují také proměnné, které ovlivňují globální nastavení a to dvě:

- LC_ALL

- LANG

LC_ALL je proměnná s nejvyšší prioritou, pokud ji nastavíme, přebije všechna ostatní nastavení. Proměnná LANG má zase nejmenší prioritu. Pokud není nastavena proměnná LC_ALL nebo proměnná příslušné kategorie, použije se pro tuto kategorii hodnoty z LANG.

Definiční soubory pro locales se obvykle nachází v */usr/share/locale*, kde je pro každý jazyk vytvořen adresář, ve kterém pak jsou definiční soubory. Tyto soubory jsou v binárním tvaru. Zdrojové soubory locales pak jsou v adresáři */usr/share/i18n*.

3.1 Jak proměnné nastavit

Hodnoty proměnných LC_* a LANG by se měly řídit podle pravidla:

zkratkajazyka[_ZKRATKAZEMĚ[.kódování]]

přičemž pro český jazyk je zkratka „cs“ a zkratka České Republiky je „CZ“. Pro slovenský jazyk je zkratka „sk“ a pro Slovenskou Republiku „SK“. Proměnné tedy budeme nastavovat na „cs_CZ“ („sk_SK“) případně „cs_CZ.ISO-8859-2“ („sk_SK.ISO-8859-2“). Zkratky jazyka definuje norma ISO 639 a zkratky zemí definuje norma ISO 3166.

V souboru */usr/share/locale/locale.alias* najdeme aliasy, kterých lze také pro nastavení proměnných použít. Pro češtinu tam najdeme „czech“, pro slovenštinu „slovak“.

Tyto proměnné nastavujeme obvykle buď pro všechny uživatele v */etc/profile* řádkem:

```
export LC_ALL=cs_CZ.ISO-8859-2
```

nebo pro jediného uživatele, stejným řádkem v *.profile* nebo *.bash_profile* v domácím adresáři uživatele.

Nastavení jednotlivých kategorií lze jednoduše zjistit pomocí příkazu `locale`.

Odkazy: **locale(7)**, **locale(1)**, **locale(5)**, *http://www.linux.cz/seminare/jevicko1998/Lokalizace/lokalizace.ps*

4 X Window System

České prostředí v X Window Systemu je trochu více problematická záležitost.

Aby aplikace mohla přijímat české znaky, je třeba, aby knihovna jazyka C, uměla tyto znaky rozpoznávat. K tomu musí být korektně nastavena kategorie LC_CTYPE na „cs_CZ" případně „cs_CZ.ISO-8859-2".

V případě, že spouštíme X server pomocí `startx` může být nastavení provedeno v */etc/profile* nebo */etc/.[bash_]profile*. V případě, že X server je spouštěn z xdm, je třeba nastavení provést buď v *.xsession* (Debian) nebo *.Xclients* (RedHat) před spuštěním správce oken (pro jednoho uživatele), nebo na začátku */etc/X11/Xsession* (Debian) nebo */etc/X11/xinit/Xclients* (RedHat) pro všechny uživatele.

Dále program musí podporovat přijímání českých znaků a to tzv. Input Metody a ještě musí korektně na začátku programu inicializovat internacionalizaci jak Xlib, tak libc, tak přípravků, které používá. Mnoho programů tak nečiní a tak je těžké je přinutit k českému (sloveskému) psaní.

Klávesnici můžeme nastavit dvěma způsoby: přes *Xmodmap* nebo *xkb*. V distribucích se obvykle používá xkb řešení, které je závislé na X serveru. Xmodmap řešení není na X serveru závislé. Nastavení mapy xkb klávesnice se provádí v */etc/X11/XF86Config* v sekci Keyboard, klíčová slova Xkb*. To jaká jména klávesnic pro XkbKeymap můžete použít zjistíte v */usr/lib/X11/xkb/keymap.dir*.

Xmodmap řešení najdete na *http://www.fi.muni.cz/~kabi/linux/linux.html*, kde také najdete knihovnu **libi18n**, kterou lze použít pro vnucení českých znaků do aplikací, které se nechovají internacionálně.

Mezi českou a anglickou klávesnicí se přepíná pomocí klávesy *Scroll Lock*.

5 Tisk

Pokud píšete nový dokument, je nejlépe psát ho v (La)TeXu, jelikož zde je podpora českého tisku zatím vyřešena nejlépe. Jako textové WYSIWYG editory pro X Window System lze také použít Applixware (*http://www.nime.cz/suse_applix_cz.html*) nebo Star Office (*http://www.intersoft.cz/linux/soffice*). Pro LaTeX existuje WYSIWYG nadstavba LyX (*http://www.lyx.org*).

Pro převod českých ASCII textů do PostScriptu pro tisk, lze použít balík **a2ps**. Tisk z X Window Systemu zatím není uspokojivě řešen. Většinou je vygenerovaný PostScript kódován v ISO-8859-2, ale PostScript používá ISO-8859-1 fonty. Někdy je na PostScript možno použít balík **ogonkify** (*http://www.agh.edu.pl/ogonki*), který může problém vyřešit, například tisk z Netscape.

Pro tisk dokumentů z PostScriptu na tiskárnu se používá balík **GhostScript** (*http://www.adobe.com*). Tento balík ve verzi 6.0 už konečně obsahuje fonty, které mají ISO-88959-2 znaky, což by mohlo znamenat konečně blýskání na lepší časy.

6 Nastavení v distribucích

6.1 RedHat

Pokud ve verzi 6.2cz (*ftp://ftp.linux.cz/pub/linux/redhat-cz/6.2/i386*) zatrhnete balík **czech-adaptation**, jsou všechna potřebná nastavení udělána za Vás. Následuje popis, kde je co nastaveno:

V */etc/sysconfig/i18n* je definován font a nastavení proměnných LC_* a LANG. V */etc/inputrc* je nastaveno povolení 8bit znaků na terminálu pro shell **bash**. Implicitní klávesová mapa je zadána v souboru */etc/sysconfig/keyboard*. Konzolové fonty, klávesové mapy a pod nadjete v adresáři */usr/lib/kbd*.

V */etc/X11/xinit/Xkbmap* nebo */etc/X11/XF86Config* definujeme klávesové mapy pro v X Window System. Cestu k fontům definujeme buď v */etc/X11/fs/config* v případě X Font Serveru (implicitně) nebo v */etc/X11/XF86Config*. Fonty pro X server najdeme v adresáři */usr/lib/X11/fonts*.

6.2 Debian

Pro Debian 2.1 nainstalujte balik skcz-slink (*ftp://ftp.debian.cz/debian-czsk/dists/slink/binary-i386*). Docílíte tím instalace klávesnicových map pro X Window System.

Nejprve je třeba vyměnit balík **kbd** za **console-tools**. V */etc/console-tools/config* nastavit SCREEN_FONT na lat2u-16 a APP_CHARSET_MAP na iso02. Pomocí `kbdconfig` nastavit klávesnici. V */etc/inputrc* odkomentovat „convert-meta“. Do */etc/profile* doplnit:

```
case `tty` in /dev/tty[0-9]|/dev/tty[0-9][0-9]
echo -ne '\033%@\017\033(K'
esac
export LANG=czech
export LC\_ALL=czech
```

Po spuštění `/etc/rc.boot/console-tools` (Debian 2.1) nebo `/etc/init.d/console-screen.sh` (Debian 2.2) a `/etc/init.d/keymaps-lct.sh start` a novém přihlášení by vše mělo fungovat. Klávesové mapy pro konzoli najdeme v adresáři */usr/share/keymaps* a konzolové fonty a pod. v adresářích */usr/share/console**.

Klávesové mapy pro X Window System nastavujeme v */etc/X11/XF86Config* v sekci Keyboard, např. na `XkbKeymap "xfree86(us_cz_qwerty)"`. Hodnoty můžete zjistit v souboru */usr/lib/X11/xkb/keymap.dir*. ISO Latin 2 fonty najtete v balících **xfonts-biznet-iso-8859-2-***, **xfonts-intl-european**. Pro Debian 2.1 balík **xfntil2**. Cesty k fontům se nastavují v */etc/X11/XF86Config* sekci Fonts. Pro Debian 2.2 není třeba cesty nastavovat, pro Debian 2.1 je třeba v */etc/X11/XF86 Config* v sekci Files doplnit na začátek cesty k il2 fontům:

```
FontPath    "/usr/X11R6/lib/X11/fonts/misc-il2/"
FontPath    "/usr/X11R6/lib/X11/fonts/75dpi-il2/"
FontPath    "/usr/X11R6/lib/X11/fonts/100dpi-il2/"
```

Je-li X server startován přes xdm (naběhne přihlašovací okno v grafickém režimu na nastartování počítače), musíme LC_CTYPE nastavit v */etc/environment* nebo */etc/X11/Xsession*.

7 Aplikace

Aby fungovali některé aplikace, je třeba jim trošku pomoci. U konzolových je většinou potřeba nastavit akceptování 8bit znaků, u některých X aplikací jim znaky vnutit pomocí nečistých metod (libi18n). Informace o „počeštění“ aplikací najdete v Czech-HOWTO (*http://www.penguin.cz/czech-howto*).

7.1 less

Aby `less` zobrazoval české znaky, je třeba nastavit proměnnou prostředí LESSCHARSET na hodnotu latin1. Například do */etc/profile*:

```
export LESSCHARSET=latin1
```

7.2 ls

Pro koretní zobrazování názvů souborů s českými znaky, je třeba spouštět příkaz ls s přepínačem -N. Tedy do */etc/profile* umístíme:

```
alias ls='/bin/ls -N'
```

7.3 nvi

Editor nvi zobrazuje místo znaků české abecedy jejich hodnotu v hexadecimálním tvaru. Tomuto zamezíme přidáním následujících řádků do */etc/vi.exrc*:

```
set print="áÁčČďĎéÉěĚíÍľĽňŇóÓôÔřŘšŠťŤúÚůŮýÝžŽ"
set print=""
```

Poznámky