Lokalizace z pohledu správce a programátora

Na úvod

Několik používaných zkratek:

- i18n – internationalization – i+18 písmen+n, to jak psát lokalizované programy - API
- l10n – localization – l+10 písmen+n, lokalizační data

Počeštění konzole

Nastavení 8 bit vstupu a výstupu, můžeme provést v souborech /etc/inputrc nebo $HOME/.inputrc.

Do souboru /etc/profile je také nutné přidat, pro přepnutí na správný font:

```
case `tty` in /dev/tty[0-9]|/dev/tty[0-9][0-9])
  printf "\033(K" ;;
esac
```

Odkazy

- **ucw** – http://atrey.karlin.mff.cuni.cz/sinmj/linuxcs.html
- **cstools** – ftp://ftp.fi.muni.cz/pub/localization/*/obsolete
- **Czech-HOWTO** – http://www.penguin.cz/czech-howto

Lokalizace glibc

Binární lokalizační soubory jsou umístěny v adresáři /usr/shale/locale. Definice lokalizací a znakové sady jsou umístěny v adresářích /usr/share/i18n/locales a /usr/share/i18n/charmaps. To zda tomu tak je, lze zjistit příkazem `localedef --help`.

Zdrojový tvar /usr/share/i18n/locales/cs_CZ je nutno převést do několika binárních souborů pomocí příkazu:

localedef -i cs_CZ -f "ISO-8859-2" cs_CZ

Pokud tak neuděláte nebo někdo za Vás, systém nebude tuto lokalizaci znát.

Kategorie lokalizace a jejich nastavení

Kategorie jsou následující:

LC_COLLATE – Abecední řazení

LC_CTYPE – Kategorie znaků, např. mezery, malá písmena, velká písmena, ...

LC_MESSAGES – Výpisy zpráv

LC_MONETARY – Peněžní údaje

LC_NUMERIC – Formátování čísel

LC_TIME – Formáty datumů, názvy dní v týdnu, měsíců

Kategorie nastavujeme těmito proměnnými prostředí:

LC_COLLATE

LC_CTYPE

LC_MESSAGES

LC_MONETARY

LC_NUMERIC

LC_TIME

LC_ALL – všechny kategorie najednou, přebije ostatní nastavení

LANG – implicitní hodnota pro kategorie

Lokalizační proměnné prostředí mají tuto prioritu: LC_ALL, LC_*, LANG.

Tyto proměnné prostředí se obvykle na UNIXových systémech nastavují na

jazyk_ZEMĚ.znaková_sada

kde označení jazyka odpovídá ISO 639 a označení země odpovídá ISO 3166. Pro náš případ tedy platí Čeština – cs, Česká Republika – CZ, Znaková sada – ISO-8859-2 nebo iso88592. Plné označení tedy zní cs_CZ.ISO-8859-2.

V souboru /usr/share/locale/locale.alias, lze nadefinovat přezdívky pro jednotlivé lokalizace. Implicitně je definována přezdívka „czech“, takže proměnné prostředí lze nastavovat i na „czech“.

Průběh nastavování kategorie

Glibc vykonává při nastavování lokalizace některé kategorie tyto kroky:

1. vyzvednutí označení z proměnných prostředí
2. hledání, zda to není přezdívka, je-li nahrazení originálem
3. hledání potřebného souboru na disku

Hledání lokalizačních souborů

Glibc hledá lokalizační soubory v /usr/share/locale v těchto adresářích a v tomto pořadí:

1. jazyk_ZEMĚ.CHARSET – původní tvar z prom. prostř., bez přezdívek (cs_CZ.ISO-8859-2)
2. jazyk_ZEMĚ.charset – označení znakové sady je unifikováno, pouze písmena a čísla (cs_CZ.iso88592)
3. jazyk_ZEMĚ (cs_CZ)
4. jazyk.CHARSET – označení znakové sady původní (cs.ISO-8859-2)
5. jazyk.charset – unifikované označení znakové sady (cs.iso88592)
6. jazyk

Z toho plyne jak by měli být nazvány adresáře. Adresář pro české lokalizační soubory tedy pojmenujeme cs_CZ nebo cs_CZ.iso88592, adresář pro katalogy zpráv cs nebo cs.iso88592.

Programy typu man

Man bere název adresáře pro přeložené manuálové stránky primárně z proměnné prostředí LANG. Podobně se někdy chová i X Window System při hledání app-defaults.

LANG je tedy třeba nastavit na název adresáře, ve kterém jsou data, například české manuálové stránky.

Například:

```
export LANG=cs -> /usr/man/cs
export LANG=cs.iso88592 -> /usr/man/cs.iso88592
export LANG=czech -> /usr/man/czech
```

Internacionalizovaný program

- musí zpracovávat 8-bitové znaky
- volá setlocale(LC_ALL, "") nebo pro potřebnou kategorii
- používá lokalizačně závislé funkce pro
  - výpisy čísel
  - peněžních údajů
  - abecední řazení
  - pro identifikaci třídy znaku
  - pro výpis zpráv

Bližší informace:

- **Linuxové noviny 6/98** – Jak psát lokalizované programy
- **Czech HOWTO** – http://www.penguin.cz/czech-howto
- **ISO-Programing** – ftp://ftp.vse.cz/pub/.ccd3c/linux/Czech/ISO-programming

Funkce setlocale

Syntaxe:

char * setlocale(int kategorie, char *hodnota)

hodnota:

- "" – převzata z proměnných prostředí
- "xxxx"
- NULL – funkce vrátí aktuální nastavení. Vrací název adresáře, ve kterém byly nalezeny lokalizační soubory.

Pro kategorii LC_ALL může být hodnota "kategorie1=hodnota1;kategorie2=hodnota2; …". Volání setlocale s hodnotou NULL, může dopadnout:

```
setlocale(LC_ALL, NULL) -> "hodnota"
setlocale(LC_ALL, NULL) -> "kategorie1=hodnota1; ..."
setlocale(LC_ALL, NULL) -> "hodnota1 hodnota2 ..."
```

Druhý případ nastane, nejsou-li všechny kategorie nastaveny stejně. Třetí případ je obvyklý pro některé UNIXy, např: Digital UNIX.

Umístění lokalizačních souborů může být funkci setlocale() oznámeno proměnnou prostředí „LOCPATH“.

Katalogy zpráv

Pro lokalizaci zpráv vypisovaných programy, existují dva standardy:

- X/Open, XPG
- Sun - GNU Gettext

X/Open katalogy

Programy, které používají tento způsob, hledají katalogy v adresáři zadaném v proměnné prostředí NLSPATH. Pokud jsou neuspěšné, pak ještě hledají v těchto adresářích:

- /etc/locale/LC_MESSAGES
- /usr/lib/locale/LC_MESSAGES
- /usr/lib/locale/Locale/LC_MESSAGES

NLSPATH lze nastavit například na:

NLSPATH=/usr/share/locale/%L/LC_MESSAGES/%N.cat:./%N.cat

V NLSPATH se mohou vyskytnout následující sekvence, které jsou nahrazovány:

- **%L** – název lokalizace
- **%N** – jméno katalogu požadovaného funkcí catopen
- **%l** – (el) jazyk
- **%t** – země
- **%c** – znaková sada

Fragment programu používajícího X/Open katalogy:

```
setlocale(LC_MESSAGES, "");
cfd = catopen("test", MCLoadBySet);

printf("%s\n", catgets(cfd, TestlistSet,
  TestlistWelcome, "Welcome in test program"));
printf("%s\n", catgets(cfd, TestlistSet,
  TestlistAuthor, "Author: Vladimir Michl"));
catclose(cfd);
```

Příklad X/Open katalogu: test.m:

```
$set Testlist
$quote "
$ Poznámka
Welcome "Vítejte v testovacím programu"
Author "Autor: Vladimír Michl"
```

Převod katalogu do binární podoby a generaci hlavičkových souborů, můžeme provést příkazem:

gencat –new -H test-cat.h test.cat test.m

S programem gencat dodávaným s libc-5 byla možná ještě jedna forma syntaxe katalogu: test.m:

```
$set 10 #Testlist
$ #Welcome
# Vítejte v~testovacím programu
$ #Author
# Autor: Vladimír Michl
```

Kompilace katalogu a generace hlavičkového souboru byla příkazem:

gencat -new -lang C test.cat test.m -h test-cat.h

Sunovské katalogy - Gettext

Adresář, kde jsou hledány katalogy je obvykle /usr/share/locale/Locale/LC_MESSAGE/. Tento adresář bývá zadán při kompilaci programu. „Locale“ je označení lokalizace. Lokalizaci danou systémem lze přebít proměnnou prostředí „LANGUAGE“. Tato proměnná je GNU rozšířením.

Program který využívá gettext může vypadat následovně: test.c:

```
setlocale(LC_MESSAGES, "");
bindtextdomain (PACKAGE, LOCALEDIR);
textdomain (PACKAGE);

printf("%s\n", gettext("Welcome in test program"));
printf("%s\n", gettext("Author: Vladimir Michl"));
```

Příklad katalogu: test.po:

```
msgid ""
msgstr ""
"Project-Id-Version: test 1.0\n"
"POT-Creation-Date: 1997-05-02 16:28+0000\n"
"PO-Revision-Date: 1997-05-02 18:00+0200\n"
"Last-Translator: Vladimir Michl <Vladimir.Michl@upol.cz>\n"
"Language-Team: Czech <cs@li.org>\n"
"MIME-Version: 1.0\n"
"Content-Type: text/plain; charset=ISO-8859-2\n"
"Content-Transfer-Encoding: 8bit\n"

#: test.c:13
msgid "Welcome in test program"
msgstr "Vítejte v~testovacím programu"

#: test.c:14
msgid "Author: Vladimir Michl"
msgstr "Autor: Vladimír Michl"
```

Překlad do binární podoby provedeme příkazem

msgfmt -v -f -o test.mo test.po

Emacs - po-mód

ENTER – editace překladu (ukončení C-c C-c nebo C-c C-k), # – editace poznámky, s – zobrazení zprávy ve zdrojovém kódu, Shift-v – kontrola správnosti katalogu a aktualizace hlavičky katalogu, Shift-e – editace celého katalogu jako textu (návrat „M-x po-mode“), TAB – odstranění „fuzzy“, DEL – označení „fuzzy“, t, u, f, o – na další zprávu (přeloženou, nepřeloženou, „fuzzy“, zastaralou), h – nápověda.

X Window System

Pro korektní funkci je třeba:

- fungující lokalizace libc
- nainstalované ISO-8859-2 fonty
- případně česká klávesnice

Kde najít ISO-8859-2 fonty?

- Většinou v každé distribuci
- ftp://ftp.vslib.cz/pub/unix/X11/fonts/iso8859-2/
- **IL2 + CP1250** – http://www.ttc.cz/links/ftp/pub/win/csfonts/unix-pcf/
- balík xfntil2
- http://www.cg.tuwien.ac.at/sinprikryl/linux/bdftab.html
- **Czech HOWTO** – http://www.penguin.cz/czech-howto, kap. 4

Aby fonty X Server používal je třeba následujícího:

- mkfontdir – vytvoření seznamu fontů (soubor fonts.dir)
- případné přidání přezdívek (soubor fonts.alias)
- přidání adresářů s fonty do XF86Config

Příklad sekce adresářů souboru XF86Config:

```
FontPath   "/usr/X11R6/lib/X11/fonts/misc-il2:unscaled"
FontPath   "/usr/X11R6/lib/X11/fonts/75dpi-il2:unscaled"
FontPath   "/usr/X11R6/lib/X11/fonts/100dpi-il2:unscaled"
FontPath   "/usr/X11R6/lib/X11/fonts/misc:unscaled"
FontPath   "/usr/X11R6/lib/X11/fonts/75dpi:unscaled"
FontPath   "/usr/X11R6/lib/X11/fonts/Type1"
FontPath   "/usr/X11R6/lib/X11/fonts/Speedo"
FontPath   "/usr/X11R6/lib/X11/fonts/PEX"
```

Aby fungovala česká klávesnice je třeba mít ji nainstalovánu. A to buď XKB nebo Xmodmap klávesnici. Navíc aby byli správně rozpoznávány ISO-8859-2 znaky, musí být nastavena proměnná prostředí „LC_CTYPE“.

Pro korektní funkci lokalizace doporučuji upravit soubor /usr/lib/X11/locale/locale.alias na:

```
cs                       cs_CZ.ISO8859-2
cs_CZ                    cs_CZ.ISO8859-2
cs_CZ.iso88592           cs_CZ.ISO8859-2
cs_CZ.ISO-8859-2         cs_CZ.ISO8859-2
```

X toolkit a lokalizace

Aby vše správně fungovalo, měl by začátek programu vypadat takto:

```
Widget top;
XtSetLanguageProc(NULL, NULL, NULL);
top = XtOpenAplication(...)
```

Prototyp:

XtSetLanguageProc(XtAppContext, XtLanguageProc, XtPointer)

Implicitní jazyková funkce XtLanguageProc vypadá asi takto:

```
String XtLanguageProc(Display *, String, XtPointer) {
  if (!(s = setlocale(LC_ALL, "")))
    XtWarning("I18N: Locale not supported by C library, "
              "locale unchanged\n");
  if (!XSupportsLocale()) {
    char msg[256];
    sprintf(msg, "I18N: Locale '%s' not supported by Xlib, "
                 "locale set to C\n", s);
    XtWarning(msg);
    s~= setlocale(LC_ALL, "C");
  }
  if (!XSetLocaleModifiers(""))
  XtWarning("I18N: X locale modifiers not supported, "
            "using default\n");

  return setlocale(LC_ALL, NULL);
}
```

Pokud je knihovna X11 přeložena s makrem X_LOCALE, tak knihovna předpokládá nefunkčnost funkce setlocale() a implementuje ji sama. Následující opravou se tomu dá zamezit.

```
#ifndef X_LOCALE
#include <X11/Xlocale.h>
#endif

...

#ifdef X_LOCALE
 if (!_Xsetlocale(0 /* LC_ALL */, ""))
 XtWarning("I18N: Locale not supported by Xlib library, "
           "locale unchanged\n");
#endif
```

App-defaults se hledají v tomto pořadí:

- $XUSERFILESEARCHPATH
- /usr/X11R6/lib/X11/%L/app-defaults/%N
- /usr/X11R6/lib/X11/%l/app-defaults/%N
- /usr/X11R6/lib/X11/app-defaults/%N

Pokud není proměnná XUSERFILESEARCHPATH definována, pak v tomto pořadí:

- $HOME/%L/app-defaults/%N
- $HOME/%l/app-defaults/%N
- $HOME/app-defaults/%N
- /usr/X11R6/lib/X11/%L/app-defaults/%N
- /usr/X11R6/lib/X11/%l/app-defaults/%N
- /usr/X11R6/lib/X11/app-defaults/%N

Substituce prováděné v hodnotě proměnné XUSERFILESEARCHPATH:

- **%N** – jméno aplikace
- **%L** – název lokalizace vracený XtLanguageProc() nebo LANG
- **%l** – (el) jazyk
- **%t** – země
- **%c** – znaková sada

Výsledky této substituce se také liší, podle toho, zda aplikace volala funkci XtSetLanguageProc() nebo ne. Jestliže tato funkce volána nebyla, je použita hodnota proměnné LANG.

Podrobnosti najdete v xbooks/hardcopy/Xt/intrinsics.ps, kap. 2.2, 2.3, 2.4, str. 30.

Aplikaci lze také explicitně určit jaký jazyk má používat a to pomocí přepínače xnllanguage:

aplikace -xnllanguage označení_pro_libc

Lokalizace GTK

U GTK stačí na začátek programu umístit volání funkce gtk_set_locale().

Lokalizace a Xlib

Pro vstupy a výstupy přímo přes Xlib, by se mělo využívat vstupních a výstupních metod a kontextů. Bližší informace najdete v xbook/hardcopy/X11/*, funkce XmbLookupString, XOpenIM, XOpenIC a XOpenOC, XOpenOM.

Odkazy

- ftp://ftp.fi.muni.cz/pub/localization
- http://www.penguin.cz/czech-howto
- http://linux-cz.inecnet.cz
- http://www.inf.upol.cz/ michlv/gnu-loc.shtml
- xbook/hardcopy/X11/*
- xbook/hardcopy/Xt/*
- xbook/hardcopy/Xaw/*