The Linux Cyrillic HOWTO

Alexander L. Belikoff, (belikoff@netvision.net.il) v2.8, 27 March 1997

This document describes how to set up your Linux box to typeset, view and print the documents in the Russian language.

Contents

1 General notes **2**

1.1 Introduction . . . 2

1.2 Availability and feedback . . . 3

1.3 Acknowledgments and copyrights . . . 3

2 Characters and codesets **4**

3 Text mode setup **5**

3.1 Linux Console . . . 5

4 The X Window System **6**

4.1 The X fonts. . . . 6

4.2 The input translation . . . 8

4.2.1 The table of characters . . . 8

4.2.2 The mode switching rules . . . 10

5 Cyrillic support in TeX and LaTeX **13**

5.1 Using the Washington Cyrillic . . . 13

5.2 Using the cmcyralt package for LaTeX . . . 14

5.3 Using the CyrTUG package . . . 15

6 Miscellaneous utilities setup **15**

6.1 bash . . . 15

6.2 csh/tcsh . . . 15

6.3 emacs . . . 16

6.4 ispell . . . 17

6.5 ksh . . . 17

6.6 less . . . 17
6.7 mc (The Midnight Commander) . . . 17
6.8 netscape . . . 17
6.9 rlogin . . . 18
6.10 sendmail (aka "The Doom of a Sysadmin") . . . 18
6.11 zsh . . . 18

7 Printing **18**

8 Useful Tools **18**

8.1 Conversion Utilities . . . 18
8.2 Programmer's tools . . . 18

9 Summary of the various useful resources **19**

1 General notes

1.1 Introduction

This document covers the things you need to successfully typeset, view, and print documents in Russian under Linux. Although this document assumes your using Linux as an operating system, most of information presented is eqally applicable to many other Unix flavours. I shall try to keep the distinction as visible as possible.

There are a number of popular Linux distributions. As an example system I describe the RedHat 3.0.3 Linux (Picasso) - the one I am personally using. Nevertheless, I shall try to highlight the differences, if they exist, in the Slackware Linux setup.

Since such setup directly modifies and extends the Operating System, you should understand, what you are doing. Even though I tried to keep things as easy as possible, having some experience with a given piece of software is an advantage. I am not going to describe what the X Window System is or how to typeset the documents with TeX and LaTeX, or how to install printer in Linux. Those issues are covered in other documents.

For the same reason, in most cases I describe a system-wide setup, by default requiring *root* privileges. Still, if there is a possibility for user-level setup, I'll try to mention it.

NOTE: The X Window System, TeX and other Linux components are complex systems with a sofisticated configuration. If you do something wrong, you can not only fail with Russian setup, but to break the component as well, if not the entire system. This is not to scare you off, but merely to make you understand the seriousness of the process and be careful. Preliminary backup of the config files is **highly** recommended. Having a guru around is also advantageous.

1.2 Availability and feedback

This document is available at sunsite.unc.edu or tsx-11.mit.edu as a part of the *Linux Document Project*. Also, it may be available at various FTP sites containing Linux. Moreover, it may be included as a part of Linux distribution.

You may also get it directly from the author at ftp.netvision.net.il.

If you have any suggestions or corrections regarding this document, please, don't hesitate to contact me via e-mail. Any new and useful information about cyrillic support in various Unices is *highly appreciated*. Remember, it will help the others.

1.3 Acknowledgments and copyrights

Many people helped me (and not only me) with valuable information and suggestions. Even more people contributed software to the public community. I am sorry if I forgot to mention somebody.

So, here they go: Bas V. de Bakker, David Daves, Serge Vakulenko, Sergei O. Naoumov, Winfried Truemper.

This document is Copyright (C) 1995,1997 by Alexander L. Belikoff. It may be used and distributed under the usual Linux HOWTO terms described below.

The following is a Linux HOWTO copyright notice:

> *Unless otherwise stated, Linux HOWTO documents are copyrighted by their respective authors. Linux HOWTO documents may be reproduced and distributed in whole or in part, in any medium physical or electronic, as long as this copyright notice is retained on all copies. Commercial redistribution is allowed and encouraged; however, the author would like to be notified of any such distributions.*
>
> *All translations, derivative works, or aggregate works incorporating any Linux HOWTO documents must be covered under this copyright notice. That is, you may not produce a derivative work from a HOWTO and impose additional restrictions on its distribution. Exceptions to these rules may be granted under certain conditions; please contact the Linux HOWTO coordinator at the address given below.*
>
> *In short, we wish to promote dissemination of this information through as many channels as possible. However, we do wish to retain copyright on the HOWTO documents, and would like to be notified of any plans to redistribute the HOWTOs.*

If you have questions, please contact Greg Hankins, the Linux HOWTO coordinator, at `gregh@sunsite.unc.edu`. You may finger this address for phone number and additional contact information.

Unix is a technology trademark of the X/Open Ltd.; MS-DOS, Windows, Windows 95, and Windows NT are trademarks of the Microsoft Corp.; The X Window System is a trademark of The X Consortium Inc. Other trademarks belong to the appropriate holders.

2 Characters and codesets

In order to understand and print characters of various languages, the system and software should be able to distinguish them from other characters. That is, each unique character must have a unique representation inside the operating system, or the particular software package. Such collection of all unique characters, that the system is able to represent at once, is called a *codeset*.

At the time of the most operating system's creation, nobody cared about software being multilingual. Therefore, the most popular codeset was (and actually is) an **ASCII** (American Standard Code for Information Interchange).

The *standard ASCII* (aka 7-bit ASCII) comprises 128 unique codes. Some of them ASCII defines as real printable characters, and some are so-called *control characters*, which had special meanings in the old communication protocols. Each element of the set is identified by an integer *character code* (0-127). The subset of printable characters represents those found on the typewriter's keyboard with some minor additions. Each character occupies 7 least significant bits of a byte, whereas the most significant one was used for control purposes (say, transmission control in old communication packages).

The 7-bit ASCII concept was extended by 8-bit ASCII (aka *extended ASCII*). In this codeset, the characters' codes' range is 0-255. The lower half (0-127) is pure ASCII, whereas the upper one contains 127 more characters. Since this codeset is backward compatible with the ASCII (character still occupies 8 bit, the codes correspond the old ASCII), this codeset gained wide popularity.

Although the extended ASCII doesn't define the contents of the upper half of the codeset, the most popular and widespread implementation of it is the **Latin 1** codeset. In Latin 1, the upper half of the table defines various characters which are not part of the English alphabet, but are present in various european languages (german *umlauts*, french *accentes* etc). Another popular extended ASCII implementation is **IBM** (named after some computer company, that developed this codeset for it's infamous personal computers). This one contains pseudo-graphic characters in the upper half.

Software, that doesn't make any assumptions about the 8-th bit of the ASCII data is called *8-bit clean*. Some older programs, designed with 7-bit ASCII in mind are not 8-bit clean and may work incorrectly with your extended ASCII data. Most of packages, however, are able to deal with the extended ASCII by default, or require some very basic setup. **NOTE:** before posting the question *"I did all setup right, but I cannot enter/view cyrillic characters!"*, please consult the section 6 for the notes on the program, you are using.

For information about making your software 8-bit clean, see section 8.2.

Since on most systems character occupies 8 bits, there is no way to extend ASCII more and more. The way to implement new symbols in ASCII-based codesets is creation of other extended ASCII implementations. This is the way, the Cyrillic ASCII set is implemented.

Although, there were many of them, nowadays there are three. The most popular and widespread are only two. One is the *Alt* codeset (so-called *"alternative codeset"*); the other one is *KOI-8*. This one is specified in *RFC 1489* `<file://ds.internic.net/rfc/rfc1489.txt>` (*"Registration of a Cyrillic Character Set"*).

These two standards differ only in positions of the cyrillic characters in the table (that is in cyrillic character codes).

The principal difference is that the Alt codeset is used by MS-DOS users only, whereas KOI-8 is used in Unix, as well as in MS-DOS (though in the latter KOI-8 is much less popular). Since we are doing the right thing (namely working in the Unix operating system), we shall focuse mostly on KOI-8

There are other standards, which are different from ASCII and much more flexible. **Unicode** is most known. However, they are not implemented as good as the basic ones in Unix in general and Linux in particular. Therefore, I am not describing them here.

3 Text mode setup

Generally, the text mode setup is the easiest way to show and input cyrillic characters. There is one significant complication, however: the text mode fonts and keyboard layout manipulations depend on terminal driver implementation. Therefore, there is no portable way to achieve the goal accross different systems.

Right now, I describe the way to deal with the Linux console driver. Thus, if you have another system, dont expect it to work for you. Instead, consult your terminal driver manual. Nevertheless, send me any information you find, so I'll be able to include it in futher versions of this document.

3.1 Linux Console

The Linux console driver is quite a flexible piece of software. It is capable of changing fonts as well as keyboard layouts. To achieve it, you'll need the kbd package. Both RedHat and Slackware install kbd as part of a system.

The kbd package contains keyboard control utilities as well as a big collection of fonts and keyboard layouts.

Cyrillic setup with **kbd** usually involves two things:

1. Screen font setup. This is performed by the `setfont` program. The fonts files are located in `/usr/lib/kbd/consolefonts`.

   NOTE: Never run the `setfont` program under X because it will hang your system. This is because it works with low-level video card calls which X doesn't like.

2. Load the appropriate keyboard layout with the `loadkeys` program.

NOTE: In RedHat 3.0.3, `/usr/bin/loadkeys` has too restrictive access permissions, namely 700 (`rwx------`). There are no reasons for that, since everyone may compile his own copy and execute it (the appropriate system calls are not root-only). Thus, just ask your sysadmin to set more reasonable permissions for it (for example, 755).

The following is an excerpt from my `cyrload` script, which sets up the Cyrillic mode for Linux console:

```
if [ notset.$DISPLAY != notset. ]; then
    echo "`basename $0`:  cannot run under X"
```

```
    exit
fi

loadkeys /usr/lib/kbd/keytables/ru.map
setfont /usr/lib/kbd/consolefonts/koi-8x16
echo "Use the right Ctrl key to switch the mode..."
```

Now you probably want to test it. Do the appropriate bash or tcsh setup, rerun it, then press the right `Control` key and make sure you are getting the cyrillic characters right. The '`q`' key must produce russian "`short i`" character, '`w`' generates "`ts`", etc.

If you've screwed something up, the very best thing to do is to reset to the original (that is, US) settings. Execute the following commands:

```
loadkeys /usr/lib/kbd/keytables/defkeymap.map
setfont /usr/lib/kbd/consolefonts/default8x16
```

NOTE: unfortunately enough, the console driver is not able to preserve it's state (at least easily enough), while running the X Window System. Therefore, after you leave the X (or switch from it to a console), you have to reload the console russian font.

4 The X Window System

Like the console mode, the X environment also requires some setup. This involves setting up the input mode and the X fonts. Both are being discussed below.

4.1 The X fonts.

First of all, you have to obtain the fonts having the Cyrillic glyphs at the appropriate positions.

If you are using the most recent X (or XFree86) distribution, chances are, that you already have such fonts. In the late 1995, the X Window System incorporated a set of cyrillic fonts, created by Cronyx. Ask your system administrator, or, if *you* are the one, check your system, namely:

1. run '`xlsfonts | grep koi8`'. If there are fonts listed, your X server is already aware about the fonts.
2. otherwise, run

   ```
   find -name crox\*.pcf\*
   ```

   to find the location of the cyrillic fonts in the system. You'll have to `enable` those fonts to the X server, as I explain below.

If you haven't found such fonts installed, you'll have to do it yourself.

1. Download the fonts from any FTP site, containing the X distribution. For example, from the XFree86 FTP site, or directly from the creators of the fonts - Cronyx.

   It is important to note, that it is not the only collection. Of the older ones, there is a *vakufonts* package, created by Serge Vakulenko**, which was the base for the one in the X distribution. There are also a number of others. The important point is that the fonts' names in the old collection were not strictly conforming to the standard. The latter is fine in general, but sometimes it may cause various weird errors. Fow example, I had a bad experience with Maple V for Linux, which crashed misteriously with the **vakufonts** package, but ran smoothly with the "standard" ones.

2. Now when you have the fonts, you create some directory for them. It is generally a bad idea to put new fonts to the already existing font directory. So, place them, to, say, `/usr/lib/X11/fonts/cyrillic` for a system-wide setup, or just create a private directory for personal use.

3. If the new fonts are in BDF format (`*.bdf` files), you have to compile them. For each font do:

   ```
   bdftopcf -o <font>.pcf <font>.bdf
   ```

   If your server supports compressed fonts, do it, using the *compress* program:

   ```
   compress *.pcf
   ```

   Also, if you do want to put the new fonts to an already existing font directory. you have to concatenate the old and the new files named `fonts.alias` in the case both of them exist.

4. Each font directory in the X must contain a list of fonts in it. This list is stored in the file `fonts.dir`. You don't have to create this list manually. Instead, do:

   ```
   cd <new font directory>
   mkfontdir .
   ```

5. Now you have to make this font directory known to the X server. Here, you have a number of options:

   - System-wide setup for XFree86. If you are running this version of X, then append the new directory to the list of directories in the file `XF86Config`. To find the location of this file, see output of *startx*. Also, see **XF86Config(4/5)** for details.
   - System-wide setup through `xinit`. Add the new directory to the *xinit* startup file. See **xinit(1x)** and the next option for details.
   - Personal setup. You have a special start-up file for the X - `~/.xinitrc` (or `~/.Xclients`, or `~/.xsession` for the RedHat users). Add the following commands to it:

   ```
   xset +fp <new font directory>
   xset fp rehash
   ```

6. Now restart your X. If you have done everything right, the tests in the beginning of the section will be successful. Also, play with `xfontsel` to make sure you are able to select the cyrillic fonts.

4.2 The input translation

The switching between the different input translations is set up by the `xmodmap` program. This program allows customization of codes emitted by various characters and their combinations. It sets the things up based on the file containing the translation table.

If you don't want to deal with all these tricks and you prefer having a solution right away, either download an appropriate xmodmap table, available at many sites dealing with the Cyrillic, for example, kiae or uunet. Also, I made the table, described below, available at my FTP site.

Another alternative is to install the *xrus* package, which allows you to configure most of the input translation parameters without having to know about `xmodmap`.

The following is a simplified description of input customization. If you want to do more sophisticated tricks, refer to the **xmodmap(1)** or, even better, wait for the next major X release, which will hopefully address the current input problems.

In our case, the translation table should define two things:

- the character codes emitted by the alphanumeric characters, and
- the mode switching rules

4.2.1 The table of characters

This is basically a sequence of directives which assign the certain keysyms to a specified keycodes. The general syntax is the following:

`keycode` *code* = *sym1 sym2 sym3 sym4*

where *code* is the numerical code of the given key on the keyboard (refer to the standard table for your system. In my case it is stored in the file `/usr/lib/X11/etc/xmodmap.std`). The *sym*s define the keysyms emitted by that key in different conditions. *Sym1* is the keysym emitted by the key in a regular state, *sym2* corresponds the key in *shifted* state (usually when `Shift` is held down). *Sym3* and *sym4* define the keysyms emitted when the `Mode_switch` is active for the normal and shifted states respectively (*group 2*, according the X Protocol Specification). In our case, the active `Mode_switch` corresponds to the Cyrillic input mode.

These should be either hexadecimal codes or the symbolic constants from `/usr/include/X11/keysymdef.h` (without leading "XK_").

Thus, if we wanted the key corresponding to the Latin 'a' generate the Russian 'a' in the alternative mode, we would write the following:

```
keycode 38  =  a  A  0xC1 0xE1
```

The reader might be curious why I don't use the `Cyrillic_a` and `Cyrillic_A` constants respectively. The answer is that it didn't work for me. I am not very familiar with the guts of the X Window System specification, but I've got the following explanation. The symbolic constants above have the values `0x6C1`

and `0x6E1` respectively. This means that in really multi-lingual environment they could be successfully used without overlapping with any other character set. However the KOI-8 standard is not well suited for such environment. Thus, since we want to retain compatible with the past, we will violate the rules of multi-lingual support in the X Window System.

The following is a table for the most popular russian JCUKEN keyboard layout (these tables are derived from the ones in the `VakuFonts` package):

```
keysym  4             = 4             dollar          4       quotedbl
keysym  5             = 5             percent         5       colon
keysym  6             = 6             asciicircum     6       comma
keysym  7             = 7             ampersand       7       period
keysym  q             = q             Q               0xCA    0xEA
keysym  w             = w             W               0xC3    0xE3
keysym  e             = e             E               0xD5    0xF5
keysym  r             = r             R               0xCB    0xEB
keysym  t             = t             T               0xC5    0xE5
keysym  y             = y             Y               0xCE    0xEE
keysym  u             = u             U               0xC7    0xE7
keysym  i             = i             I               0xDB    0xFB
keysym  o             = o             O               0xDD    0xFD
keysym  p             = p             P               0xDA    0xFA
keysym  bracketleft   = bracketleft   braceleft       0xC8    0xE8
keysym  bracketright  = bracketright  braceright      0xDF    0xFF
keysym  a             = a             A               0xC6    0xE6
keysym  s             = s             S               0xD9    0xF9
keysym  d             = d             D               0xD7    0xF7
keysym  f             = f             F               0xC1    0xE1
keysym  g             = g             G               0xD0    0xF0
keysym  h             = h             H               0xD2    0xF2
keysym  j             = j             J               0xCF    0xEF
keysym  k             = k             K               0xCC    0xEC
keysym  l             = l             L               0xC4    0xE4
keysym  semicolon     = semicolon     colon           0xD6    0xF6
keysym  apostrophe    = apostrophe    quotedbl        0xDC    0xFC
keysym  grave         = grave         asciitilde      0xA3    0xB3
keysym  z             = z             Z               0xD1    0xF1
keysym  x             = x             X               0xDE    0xFE
keysym  c             = c             C               0xD3    0xF3
keysym  v             = v             V               0xCD    0xED
keysym  b             = b             B               0xC9    0xE9
keysym  n             = n             N               0xD4    0xF4
keysym  m             = m             M               0xD8    0xF8
keysym  comma         = comma         less            0xC2    0xE2
keysym  period        = period        greater         0xC0    0xE0
```

Also, for those using the russian YAWERTY layout, I include the following table:

```
keysym  q             = q             Q             0xD1    0xF1
keysym  w             = w             W             0xD7    0xF7
keysym  e             = e             E             0xC5    0xE5
keysym  r             = r             R             0xD2    0xF2
keysym  t             = t             T             0xD4    0xF4
keysym  y             = y             Y             0xD9    0xF9
keysym  u             = u             U             0xD5    0xF5
keysym  i             = i             I             0xC9    0xE9
keysym  o             = o             O             0xCF    0xEF
keysym  p             = p             P             0xD0    0xF0
keysym  bracketleft   = bracketleft   braceleft     0xDB    0xFB
keysym  bracketright  = bracketright  braceright    0xDD    0xFD
keysym  a             = a             A             0xC1    0xE1
keysym  s             = s             S             0xD3    0xF3
keysym  d             = d             D             0xC4    0xE4
keysym  f             = f             F             0xC6    0xE6
keysym  g             = g             G             0xC7    0xE7
keysym  h             = h             H             0xC8    0xE8
keysym  j             = j             J             0xCA    0xEA
keysym  k             = k             K             0xCB    0xEB
keysym  l             = l             L             0xCC    0xEC
keysym  z             = z             Z             0xDA    0xFA
keysym  x             = x             X             0xD8    0xF8
keysym  c             = c             C             0xC3    0xE3
keysym  v             = v             V             0xD6    0xF6
keysym  b             = b             B             0xC2    0xE2
keysym  n             = n             N             0xCE    0xEE
keysym  m             = m             M             0xCD    0xED
keysym  backslash     = backslash     bar           0xDC    0xFC
keysym  grave         = grave         asciitilde    0xC0    0xE0
keysym  equal         = equal         plus          0xDE    0xFE
keysym  3             = 3             numbersign    3       0xDF
keysym  4             = 4             dollar        4       0xFF
```

4.2.2 The mode switching rules

This is the trickiest part of the X Cyrillic setup. You should define the conditions in which the current mode is switched between the regular and the Cyrillic one.

There are two ways to achieve that in Linux. One is XFree86-specific, while the other is more general (well, not too much, as I'll show below).

The XFree86-specific way is the following. There are two virtual actions which can be assigned to the keys

in the `XF86Config` file: `ModeShift` which changes to the mode alternative to the regular one without locking, and `ModeLock` which does the same but with locking. In the first case the keys will emit the alternative keysyms only when the key generating the `ModeShift` is held down, whereas in the latter case the user needs to press the key generating the `ModeLock` keysym only once and the keyboard will be generating the alternative keysyms until that key is pressed for a second time. You should assign the `ModeShift` and `ModeLock` keysyms to the keys you want to work the mode switches.

Thus, if one wants to assign the `ModeShift` action to the right Alt key, she should place the following directive in her XF86Config:

```
RightAlt        ModeShift
```

Similarly, if the action required was ModeLock, the directive would be:

```
RightAlt        ModeLock
```

See the **XF86Config(4/5)** for more details.

The other way is, again, to use the `xmodmap` utility. This is also tricky. Generally, what you should do is:

- Assign the `Mode_switch` keysym to some key, and
- Add `Mode_switch` to some spare modifier map

Now the key to which the `ModeShift` is assigned will act as a mode switch. This means that while it is held down, the keyboard is in alternative mode.

Moreover, if you add a *lockable* key to that modifier's map, this key will lock the alternative mode.

Note: There are some problems however. Serge Vakulenko (`vak@cronyx.com`) pointed out that the different X Server implementations may have different rules of assignments the mode switches (like, for example, some servers restrict the set of the keys which may work in toggle mode to, say, `CapsLock`, `NumLock`, and `ScrollLock`). Hopefully, this is a subject to change in the next release of the X Window System. For more details, see the X Protocol specification.

Unfortunately, I didn't manage to make the CapsLock key have the same functionality in the alternative mode, namely, to lock the upper case. It seems to me, it is impossible to do it, because of the idiotic X input translation design. If I am wrong, please correct me.

Let's see an example. Suppose, one wants to use the right Alt as a mode switch and the ScrollLock as as a mode lock. First of all, one should check the default modifiers' map. This is accomplished by running the `xmodmap` without arguments:

```
$ xmodmap

xmodmap:  up to 2 keys per modifier, (keycodes in parentheses):

shift       Shift_L (0x32),  Shift_R (0x3e)
```

```
lock        Caps_Lock (0x42)
control     Control_L (0x25)
mod1        Alt_L (0x40),  Alt_R (0x71)
mod2        Num_Lock (0x4d)
mod3
mod4
mod5
```

According to the above, the plan of attack is the following:

1. remove the `Alt_R` key from the `mod1` map
2. assign the `Mode_switch` keysym to the `Alt_R` key
3. assign the `Scroll_Lock` keysym to the keycode 78 (the code of the *actual* ScrollLock)
4. add the `Mode_switch` to the spare (`mod3`) map, and
5. add the `Scroll_Lock` keysym to the `mod3` map

Thus, here is the solution:

```
remove mod1 = Alt_R
keysym Alt_R = Mode_switch
keycode 78 = Scroll_Lock
add mod3 = Mode_switch
add mod3 = Scroll_Lock
```

If you use the latter solution, you may combine both the table and the mode directives in your `~/.Xmodmap` file.

Such files are generally supplied with the various X Cyrillic stuff packages. The good example is the tables in the perfect package by Serge Vakulenko described above.

Once you have such file containing the table, you should run the command:

```
xmodmap filename
```

...every time you start X. Modify your X startup file to do it. The system-wide file is `/usr/lib/X11/xinit/xinitrc`; the personal one is either `~/.xinitrc`, or `~/.Xclients`, or `~/.xsession`, depending on waht you have.

NOTE: If `xmodmap` complains on your table, try to load the default table first. The default one is usually located in `/usr/lib/X11/etc/xmodmap.std`.

5 Cyrillic support in TeX and LaTeX

In this section I'll describe several ways to make TeX and LaTeX typeset Cyrillic texts. There are several ways, which differ in setup sophistication and usage convenience. For example, one possibility is to start without any preliminary setup and use the *Washington AMSTeX Cyrillic fonts*. On the other hand, you may install a LaTeX package, providing a very high degree of Cyrillic setup. I have an experience with two such packages. One is the `cmcyralt` package by Vadim V. Zhytnikov (`vvzhy@phy.ncu.edu.tw`) and Alexander Harin (`harin@lourie.und.ac.za`), and the other one is the `LH` package by the *CyrTUG* group with styles and hyphenation for LaTeX2e by Sergei O. Naoumov (`serge@astro.unc.edu`). I'll describe both.

Note, that there are two versions of LaTeX available - 2.09 is the old one, while 2e is a new pre-3.0 release. If you are using LaTeX 2.09, then switch quickly to the 2e. The latter retains compatibility with the old one, but has much more features. Hopefully, version 3 will be released soon. I describe a LaTeX 2e setup.

Also, both of these packages require the Cyrillic text to be typeset using the *Alt* codeset, not *KOI-8*! This is caused by historical reasons, since the creators of these packages used to work with `EmTeX` - the MS-DOG version of TeX (they didn't know about Linux yet :-). Switching to the *KOI-8* requires some effort and is being expected to be done soon. So far, use some utility to convert your russian text from *KOI-8* to *Alt*. See section 8.1.

5.1 Using the Washington Cyrillic

This package was created for the American Mathematic Society to provide documents with Russian references. Therefore, the authors were not very careful and the fonts look quite clumsy. This package is usually referred to as a `"really bad cyrillic package for TeX"`.

Nevertheless, we'll discuss it, because it is very easy to use and doesn't require any setup - this collection is supplied with most of TeX distributions.

Of course, you won't be able to use such luxury as automatic hyphenation, but anyway...

1. Prepend your document with the following directives:

```
\input cyracc.def
\font\tencyr=wncyr10
\def\cyr{\tencyr\cyracc}
```

2. Now to type a cyrillic letter, you enter

```
\cyr
```

and use a corresponding latin letter or a TeX command. Thus, the lower case of the Russian alphabet is expressed by the following codes:

```
a b v g d e \"e zh z i {\u i} k l m n o p r s t u f kh c ch sh shch
{\cprime} y {\cdprime} \`e yu ya
```

It is extremely inconvenient to convert your Russian texts to such encoding, but you can automate the process. The translit program *XREF* supports a TeX output option.

5.2 Using the cmcyralt package for LaTeX

The `cmcyralt` package can be found on any CTAN (Comprehensive TeX Archive Network) site like `ftp.dante.de`. You should get two collections: the fonts collection from `fonts/cmcyralt` and the styles and hyphenation rules from `macros/latex/contrib/others/cmcyralt`.

Note: Make sure you have the `Sauter` package installed, since `cmcyralt` requires some fonts from it. You can get this package from CTAN site as well.

Now you should do the following:

1. Put the new fonts to the TeX fonts tree. On my system (Slackware 2.2) I created a `cmcyralt` directory in the `/usr/lib/texmf/fonts/cm/`. Create the `src`, `tfm`, and `vf` subdirectories in it. Put there `.mf`, `.tfm`, and `vf` files respectively.
2. Put the font driver files (`*.fd`) from the styles archive to the appropriate place (in my case it was `/usr/lib/texmf/tex/latex/fd`).
3. Put the style files (`*.sty`) to the appropriate LaTeX styles directory (in my case `/usr/lib/texmf/tex/latex/sty`).

Now the hyphenation setup. This requires to remake the LaTeX base file.

1. The file `hyphen.cfg` contains the directives for both English and Russian hyphenation. Extract the one for Russian and place it to the LaTeX hyphenation config file `lthyphen.ltx`. In my case, that file was in `/usr/lib/texmf/tex/latex/latex-base`.
2. Put the `rhyphen.tex` to the same directory. It is needed for making the new base file. Later, you can remove it.
3. Do '`make`' in that directory. Don't for get to make a link from `Makefile` to `Makefile.unx`. During the make process check the output. There should be a message:

   ```
   Loading hyphenation patterns for Russian.
   ```

   If everything goes OK, you will get the new `latex.fmt` in that directory. Put it to the appropriate place, where the previous one was (like `/usr/lib/texmf/ini/`). **Don't forget to save the previous one!**.

This is it. The installation is complete. Try processing the examples found in the styles archive. If you are to create the PostScript files without any problems, then everything is OK. Now, to use Cyrillic in LaTeX, prepend your document with the following directive:

```
\usepackage{cmcyralt}
```

For more details, see the `README` file in the `cmcyralt` styles archive.

Note: if you do have problems with the examples, provided you have installed the things right, then probably your TeX system hasn't been installed correctly. For example, during my first try, every attempt to create the `.pk` files for the russian fonts failed (`MakeTeXPK` stage). A substantial investigation discovered some implicit conflict between the *localfont* and *ljfour* `METAFONT` configurations. It used to work before, but kept crashing after the `cmcyralt` installation. Contact your local TeX guru - TeX is very (sometimes too much) complicated to reconfigure it without any prior knowledge.

5.3 Using the CyrTUG package

You can obtain the CyrTUG package from the *SunSite archive* `<ftp://sunsite.unc.edu/pub/academic/russian-studies/Software>`. Get the files `CyrTUGfonts.tar.gz`, `CyrTUGmacro.tar.gz`, and `hyphen.tar.Z`.

The process of installation doesn't differ from the previous one.

6 Miscellaneous utilities setup

Generally, to set the certain utility up to handle the Cyrillic requires just to allow the 8 bit input. In some cases it is required to tell the application to show the extended ASCII characters in their "native" form.

6.1 bash

Three variables should be set on order to make `bash` understand the 8-bit characters. The best place is `~/.inputrc` file. The following should be set:

```
set meta-flag on
set convert-meta off
set output-meta on
```

6.2 csh/tcsh

The following should be set in `.cshrc`:

```
setenv LC_CTYPE iso_8859_5
stty pass8
```

If you don't have the POSIX `stty` (impossible for Linux), then replace the last call to the following:

```
stty -istrip cs8
```

6.3 emacs

The minimal cyrillic support in `emacs` is done by adding the following calls to one's `.emacs` (provided that the Cyrillic character set support is installed for console or X respectively):

```
(standard-display-european t)

(set-input-mode (car (current-input-mode))
   (nth 1 (current-input-mode))
   0)
```

This allows the user to view and input documents in Russian.

However, such mode is not of a big convenience because `emacs` doesn't recognize the usual keyboard commands while set in Cyrillic input mode. There are a number of packages which use the different approach. They don't rely on the input mode stuff established by the environment (either X or console. Instead, they allow the user to switch the input mode by the special `emacs` command and `emacs` itself is responsible for re-mapping the character set. The author took a chance to look at three of them. The `russian.el` package by Valery Alexeev (`ava@math.jhu.edu`) allows the user to switch between cyrillic and regular input mode and to translate the contents of a buffer from one Cyrillic coding standard to another (which is especially useful while reading the texts imported from MS-DOG).

The only inconvenience is that emacs is still treating the russian characters as special ones, so it doesn't recognize russian words' bounds and case changes. To fix it, you have to modify the syntax and case tables of emacs:

```
;; there is a garbage in the variables below, since SGML doesn't like
;; cyrillic characters. You have to put the uppercase and lowercase
;; parts of the Russian alphabet respectively (see the actual files)

(setq *russian-abc-ucase* "*** SGML SUCKS ***")
(setq *russian-abc-lcase* "*** SGML SUCKS ***")


(let ((i 0)
      (len (length *russian-abc-ucase*)))

     (while (< i len)
       (modify-syntax-entry (elt *russian-abc-ucase* i) "w  ")
       (modify-syntax-entry (elt *russian-abc-lcase* i) "w  ")
       (set-case-syntax-pair (elt *russian-abc-ucase* i)
                             (elt *russian-abc-lcase* i)
                             (standard-case-table))
       (setq i (+ i 1))))
```

For this purpose I created a `rusup.el` file which does this, as well as a couple handy functions. You have to load it in your `~/.emacs`.

Another alternative is the package `remap` which tries to make such support more generic. This package is written by Per Abrahamsen (`abraham@iesd.auc.dk`) and is accessible at `ftp.iesd.auc.dk`.

As for the author's opinion, I would suggest to start using the `russian.el` package because it is very easy to setup and use.

6.4 ispell

There is an ***rspell*** package created by Neal Dalton (`nrd@cray.com`) for `ispell`, but I experienced some problems making it work right away. Try it - maybe you will be luckier.

6.5 ksh

As for the public domain `ksh` implementation - `pdksh 5.1.3`, you can input 8 bit characters only in `vi` input mode. Use:

```
set -o vi
```

6.6 less

So far, `less` doesn't support the KOI-8 character set, but the following environment variable will do the job:

```
LESSCHARSET=latin1
```

6.7 mc (The Midnight Commander)

As of version 3.1.2, select the *full 8 bits* item in the **Options/Display** menu.

6.8 netscape

Set the following resource in your `.Xdefaults` file:

```
Netscape*documentFonts*registry:        koi8
```

Instead, if you want to set it up globally, add the following line to the application defaults file for the `Netscape` (usually `/usr/lib/X11/app-defaults/Netscape`):

```
*documentFonts*registry:                koi8
```

6.9 rlogin

Make sure that the shell on the destination site is properly set up. Then, if your `rlogin` doesn't work by default, use '`rlogin -8`'.

6.10 sendmail (aka "The Doom of a Sysadmin")

As of version 8, `sendmail` handles 8-bit data correctly by default. If it doesn't do it for you, check the `EightBitMode` option and option `7` given to mailers in your `/etc/sendmail.cf`. See *"Sendmail. Operation and Installation Guide"* for details.

6.11 zsh

Use the same way as with `csh` (see section 6.2). The startup files in this case are `.zshrc` or `/etc/zshrc`.

7 Printing

To print the text files containing the Russian characters using on PostScript printers, you need two things: the fonts and the 8bit-aware software able to print the texts using those fonts. The best package to print the text files in PostScript is `a2ps` by Evan Kirshenbaum (`evan@csli`) and Miguel Santana (`miguel@imag.imag.fr`). The last version is 8bit-aware. You can get it from `imag.imag.fr:/archive/postscript/`.

Check the `sunsite.unc.edu:/pub/academic/russian-studies/Software/` in order to obtain the Cyrillic PostScript fonts. Also, there is a lot of fonts in the *SimTel* `<ftp://oak.oakland.edu/pub/SimTel/>` collection.

8 Useful Tools

8.1 Conversion Utilities

There are a number of programs able to convert from *KOI-8* to *Alt* and back. Look at http*** or http*** for a list of handy little utilities. You can even use the special mode for `emacs` (see section 6.3).

However, I would especially recommend a translit package. It supports many popular codesets and is even able to produce a TeX***XREF file from text in Russian. Also, RedHat users will enjoy an ***RPM package of translit*** available at the RedHat ***FTP*** site.

8.2 Programmer's tools

So far, I explained the ways to make the programs accept and display the Cyrillic codeset. However the full *localization* of the system comprises much more. All discussed above is not enough. The system should be

friendly for a user who doesn't necessarily speak English. In my own opinion, it is not a big deal to become familiar with English at the level of the programs' messages. However, it is not quite fair to require it.

Thus, the next level of localization requires the programs to be customizable to the requirements of different languages and data representation habits.

Before, that was done by developing some abstraction of the messages to output from the program's code. Now, such mechanism is (more or less) standardized. And, of course, there are free implementations of it!

The good news is that GNU finally adopted the way of making the internationalized applications. Ulrich Drepper (`drepper@ipd.info.uni-karlsruhe.de`) developed a package `gettext`. This package is available at all GNU sites like *prep.ai.mit.edu* `<ftp://prep.ai.mit.edu/pub/gnu/>`. It allows you to develop programs in the way that you can easily make them support more languages. I don't intend to describe the programming techniques, especially because the `gettext` package is delivered with excellent manual.

So, if you are developing programs which output messages (have you ever developed any program which didn't?), then don't be lazy to put a little (yes, really little) effort to make your program locale-aware.

Request for collaboration: If you want to learn the `gettext` package and to contribute to the GNU project simultaneously; or even if you just want to contribute, then you can do it! GNU goes international, so all the utilities are being made locale-aware. The problem is to translate the messages from English to Russian (and other languages if you'd like). Basically, what one has to do is to get the special `.po` file consisting of the English messages for a certain utility and to append each message with it's equivalent in Russian. Ultimately, this will make the system speak Russian if the user wants it! For more details and further directions contact Ulrich Drepper (*drepper@ipd.info.uni-karlsruhe.de* `<mailto:drepper@ipd.info.uni-karlsruhe.de>`).

9 Summary of the various useful resources

My stuff related to the Cyrillic setup `<ftp://ftp.netvision.net.il/pub/home/b/belikoff/cyrillic/>`

RFC 1489 `<file://ds.internic.net/rfc/rfc1489.txt>`

The remap package for Emacs `<ftp://ftp.iesd.auc.dk/>`

Many fonts collections for X `<ftp://ftp.funet.fi/pub/culture/russian/comp/xwin>`

Information on Cyrillic Software `<http://sunsite.oit.unc.edu/sergei/Software/Software.html>`

Useful Cyrillic packages `<ftp://sunsite.unc.edu/pub/academic/russian-studies/Software>`

The kbd package for Linux `<ftp://sunsite.unc.edu/pub/Linux/system/Keyboards/>`

X fonts collections `<ftp://ftp.switch.ch/mirror/linux/X11/fonts/>`

rspell `<ftp://sunsite.unc.edu/pub/academic/russian-studies/Software/rspell.tar.gz>`