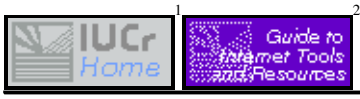

Browsing non–Latin Web pages

English is the *de–facto* standard language used on the World Wide Web. Therefore, if the author's language is not English, one makes pages in English for world–wide readers and at the same time pages for the local readers in the local mother tongue. Usually, both pages contain the same information.

So in many cases, to surf around the WWW, it is enough to prepare a WWW browser and a pc (and O.S.) that can display, at least, only English.

However, there are many interesting Web pages, that are not written in English. With the growth of the Net more and more such pages will be written. This is the reason why the *internationalization* (often abbreviated as **I18N** because there are eighteen characters *I* and *N*) of browsers is so important nowadays (*e.g.* see W3C[3] and Unicode[4]).

Major WWW browsers such as Netscape[5] and Microsoft Internet Explorer[6] have already adopted suitable character sets for many languages around the world. Especially in the pc world, recent versions of O.S.s have a facility so called *International support*. By using these browsers and facilities, a user can display the exotic pages without difficulties.

There is a vast number of combinations of O.S. and browsers, however, the common procedure to *How to browse exotic pages* is as follows:

1. obtain and install suitable fonts, and
2. make the browser recognize the fonts and language (encoding).

The degree of difficulty of these steps depends on your browsing environment.

Here let me limit the story to the Cyrillic character set used in Russian. I do so because many Web pages are written in Cyrillic and also it has a more complicated encoding than the others. In addition, some Non–Latin characters can be browsed in the same font as Cyrillic. If you are interested in browsing Cyrillic pages and would like to know more details, I suggest the links listed at the bottom of this page be visited.

Browsing Cyrillic pages

One of most difficult points is that there is more than one encoding method for Cyrillic in use. Actually the following four different encoding systems are known, however, only (1.) and (2.) are important for W3 browsing.

1. KOI8–R ––– Internet *de–facto* standard (RFC1489). Many of Russian web sites use this coding. KOI stands for *kod obmena informaticii* (code for information exchange) 8bit
2. CP1251 ––– Windows encoding used by some sites
3. CP866 ––– DOS encoding. Old DOS programs used this code. (A CP866 <––> KOI8–R converter is available.)

[1] http://www.iucr.ac.uk/iucr–top/welcome.html
[2] int.index.html
[3] http://www.w3.org/pub/WWW/International/
[4] http://www.standard.com/unicode/
[5] http://home.netscape.com
[6] http://www.microsoft.com/ie/

4. Macintosh encoding

In general as the encoding is not specified on the server the user should install both (1.) and (2.) coding fonts and switch between these two encodings until the Cyrillic text is displayed correctly. For Windows software, by using the latest version of Netscape[7] (above 3.0b2) and Microsoft Internet Explorer 3.0[8] with International Extensions for Microsoft Internet Explorer 3.0[9] and 3.0b2,[10] one can browse both (1.) and (2.) encoded sites using only the CP1251 font set. As for Macintosh software, the latest version of Netscape (above 3.0b2) can browse KOI8 encoded sites with Apple Standard Cyrillic Fonts.[11]

Obtaining Cryrillic fonts

The fonts for Windows can be downloaded as Cyrillic Web fonts (ForWWW.zip)[12] or as individual fonts of KOI8 encoding (Arial–Relcom[13] and ROL:KOI8/Courier)[14] and CP1251 encoding (ER Bukinist 1251[15] and ER Kurier 1251)[16]. They are installed by the command sequence: `ControlPanel` –> `Fonts` –> `File` –> `InstallNewFonts`.
Of course, many kinds of fonts (free or for payment) are found around the net.

For Macintoshes, KOI8–R fonts are also downloadable from the Russinification the Mac[17] page. Clearer fonts of Apple Standard Cyrillic Fonts[18] are also downloadable from Apple.

For X–windows, a resource package Xrus[19] including the Cronyx KOI8–R font set is available. I18N is progressing in many browsers based on X11R6 (Arena i18n, chimera, mosaic), and also in text based browsers (Lynx 2.5–FM, MULE+WWW–mode).

When using the browsers, set up the encoding and the fonts first. For example with the Windows and Netscape combination this is done with the commands: `Options` –> `GeneralPreference` –> `Fonts` and select `Arial Relcom` for a proportional font and `ROL KOI8-R Courier` for a fixed font with KOI8–R encoding. Set the `Encoding` to `Latin2(Central European)`. Also select `ER Bukinist 1251` for a proportional font and `ER Kurier 1251` for a fixed font with Windows encoding. Set the Encoding to Korea(?) or User defined. In the latest version of Netscape, Cyrillic(KOI8–R) and Cyrillic(Win1251) encoding is available. For more details, consult the Cyrillic for MS Windows Netscape[20] page.

East European Languages

Ukrainian
: Use the same fonts as Cyrillic. Visit Ohio Super Computer Center (OSC) Central and Eastern Europe (CEE) Ukrainian Server[21] for some interesting information.

Czech, Hungarian, Polish, Slovakian, Slovenian
: The webmasters usually prepare two kinds of page. One does not contain special characters and one does. To browse the latter, windows EE (Curier

7 ftp://ftp.netscape.com/pub/navigator/3.0/3.0b6a/windows/standard/
8 http://www.microsoft.com/ie/ie.html
9 http://www.microsoft.com/ie/download/
10 http://www.microsoft.com/ie/download/
11 ftp://ftp.apple.com/dts/mac/sys.soft.intl/russian.7.0.1/fonts–1.hqx
12 http://www.siber.com/sib/russify/ms–windows/fonts/ForWWW.zip
13 ftp://ftp.cs.umd.edu/pub/cyrillic/fonts_for_MS_windows/koi8_fin.zip
14 ftp://ftp.cs.umd.edu/pub/cyrillic/fonts_for_MS_windows/rolk8c.zip
15 ftp://ftp.cs.umd.edu/pub/cyrillic/fonts_for_MS_windows/bk1251.zip
16 ftp://ftp.cs.umd.edu/pub/cyrillic/fonts_for_MS_windows/co1251.zip
17 http://www.pitt.edu/~mapst57/rus/russian.html
18 ftp://ftp.apple.com/dts/mac/sys.soft.intl/russian.7.0.1/fonts–1.hqx
19 ftp://ftp.relcom.ru/misc/fonts/cyrillic/xwindows/xrus_bin.tgz
20 http://www.siber.com/sib/russify/ms–windows/netscape.html
21 http://www.osc.edu/ukraine.html

New CE, Times New Roman CE[22] CP1250) font and ISO Latin–2 (Times NR CE/Latin 2, Courier N CE/Latin 2, Arial CE/Latin2[23] ISO–8859–2) fonts are needed. Fonts for Macintosh (from Paragraph Corp.), Macintosh Central Europe (PT CP202 from Apple), Fonts from Polish Version of System 7.0.1[24] are also available.

Writing Cyrillic

To input Cyrillic in browsers or E–mail programs, the keyboard utilities,[25] keyboard script or keyboard resources are also needed in addition to suitable fonts.

In Windows95, `Multi language support` is prepared from the Windows95 CD. `Start` –> `add application` –> `multi language support` and select a suitable language. For example on using `Word Pad`, Cyrillic characters can be read and written after selecting a Cyrillic font in the application.

For MacOS, the Cyrillic Language Kit[26] is available from Apple computer.

Other topics and related sites

- SovInformBureau[27] – Russinify everything, Russian Humor, Travel, Art and more!
- KOI8–R Russian Net Character Set or (mirror)[28][29] – What is KOI8–R ? (including character table) and its usage on various platforms and in various applications
- Russinification Macintosh[30] – Russinification will be completed in 1–2–3–4 steps.
- Cyrillic for MS Windows Netscape[31] – Detailed installation instructions and links to the related sites.
- Fonts[32]
- Test pages:
  - KOI8–R[33]
  - CP–1251[34]
  - Third test[35]

  Test if your browser shows the correct characters.
- Browsing Japanese –– Japanese and Chinese use a *two–byte* code unlike the one–byte European language codes. So browsing Japanese pages is not easy and is further complicated by the existence of more than three different encodings. However, if you have enough bandwidth, visiting Shodouka Launchpad[36] is a nice idea. Because Japanese characters are converted into pictures when passing through this site, all browsers can display the picturized Japanese pages.

[Index][37] – 17th September 1996 – © Hidehiro Uekusa[38] – **Not to be copied or reproduced without permission**

[22] ftp://cech.cesnet.cz/pub/local/czwin3x.zip
[23] ftp://hopi.dtcc.edu/pub/berlin/fonts/latin2.zip
[24] ftp://ftp.apple.com/dts/mac/sys.soft.intl/polish.7.0.1/font.hqx
[25] http://www.siber.com/sib/#russify–ms–windows
[26] http://product.info.apple.com/productinfo/datasheets/as/cyrillic.html
[27] http://www.siber.com/sib/
[28] http://www.nagual.ru/~ache/koi8.html
[29] http://www.siber.com/sib/russify/ache/koi8.html
[30] http://www.pitt.edu/~mapst57/rus/russian.html
[31] http://www.siber.com/sib/russify/ms–windows/netscape.html
[32] http://babel.uoregon.edu/yamada/fonts/russian.html
[33] http://www.siber.com/sib/russify/test/testKOI8.html
[34] http://www.siber.com/sib/russify/test/test1251.html
[35] http://www.siber.com/sib/russify/ache/koi8.html#visual
[36] http://www.lfw.org/shodouka/
[37] int.index.html
[38] http://www.chem.titech.ac.jp/~uekusa/index–e.html