Network Update

Chris Higgins, Steven Sharp, Ruthanne Dubois
University of Maryland—College Park

Non-Roman-Based Character Scripts and the WWW

Use of the WWW for language classes continues to increase. Teachers of all languages encourage their students to explore the WWW for authentic materials, current news and more. This past year more and more of our faculty at the University of Maryland were accessing sites which were encoded for non-Roman-based character scripts. In researching what we needed to view these sites, I was amazed at the number of different encoding systems. For instance Arabic has at least two encoding systems, Japanese three and Cyrillic no less than seven. Russian alone has five encoding systems. Both Netscape and Microsoft Internet Explorer provide support for at least 15 encoding systems as well as a way to add user-defined systems.

As with most information on the web, there is an overabundance of sites wholly or partially dedicated to fonts and multilingual web browsing. Discussion of fonts on the WWW tends to focus on writing fonts for word processing, desktop publishing, etc. A small percentage of the font-related sites deal with browsing the WWW. Below is a short list of sites dealing with fonts for six relatively common non-Roman-based character scripts. There are many more character sets available via the WWW. Of these a large percentage deal primarily with Roman-based character sets with special diacritic marks or characters such as non-Cyrillic Eastern European languages like Polish. We have also listed a number of sites where both writing fonts and web fonts and character sets are available for multiple languages. Finally, there are numerous software products available to help simplify viewing non-Roman-based character scripts via the WWW. We have listed just a few that deal with multiple languages.

"This past year more and more of our faculty at the University of Maryland were accessing sites which were encoded for non-Roman-based character scripts."

Fonts and Character Sets via the WWW

Arabic

There seems to be no consistent way to read Arabic from the web. ISO 8859-6 (Arabic) is the standard encoding adopted by the Arab League and CP 1256 is the code page used by Microsoft for Arabic Windows. These two seem to be the most common but may not work with every page.

ISO 8859-6—To get more information and download these font sets: **http://www.hike.te.chiba-u.ac.jp/ikeda/ISO/iso8859.html** or **http://wwwwbs.cs.tu-berlin.de/~czyborra/charsets/**

LangBox International Arabic Support—Fonts for Arabic for Unix boxes. There is also a multilingual page which shows other scripts for Unix: **http://sparta.spartacus.com/langbox/arabic.html**.

The HotTea applet developed by Global Publishing Group represents a new technology allowing users to view Persian/Arabic without any effort. It is supported on the following platforms: Win 95, NT with Netscape 2.X or later, HotJava, Microsoft Internet Explorer 3.X with Java upgrade; Unix with Netscape 2.X or later, HotJava; Power PC Macintosh with Netscape 2.X or later: **http://www.gpg.com/hottea/**

Chinese

Chinese doesn't yet have the ease of adding on a font that will allow you to read, at least some, pages that are written in Chinese. They require some sort of add-on (helper) or the use of Chinese Windows, or the Chinese Language Kit (Mac) The following pages will help you with information about how effective each of these methods are.

How to read Chinese on the Internet—look at: **http://www.hknet.com/HKNet/chinfaq.html**

Viewing and Listening to Chinese on the WWW—Gives a list of resources for how to configure your computer so that you can surf the web and read Chinese characters: **http://www.webcom.com/~bamboo/chinese/www.html**

Read Chinese in Net Applications—A useful site which contains information about how Chinese is handled by Windows and Mac computers, and gives links to add-ons and helper programs that will help you get Chinese up and running: **http://www.biol.uregina.ca/liu/pub/read-chn.shtml**

Chinese Character Pronunciations: **http://www.webcom.com/ocrat/reaf/**

Greek

With Netscape or Microsoft Internet Explorer Greek fonts work fairly easily on the WWW.

Greek Fonts on the Net—Fonts for Windows, Mac, and Dos/Unix: **http://www.hellenic.net/fonts/**

Obtaining and installing Greek fonts — This site has directions for installing fonts for Windows, Mac, DOS, and others with directions: **http://www.hyper.gr/makthes/fonts.html**

Greek fonts—Info for Greek fonts on Unix systems, including Lynx, and X-Windows and X-Mosaic: **http://www.softlab.ntua.gr/local/greek.html**

Different font configurations are required for viewing various Hebrew sites. Hebrew fonts for Windows Internet sys-

Hebrew tems are available from a number of sources. The two most popular Internet font types are ElroNet and Web Hebrew fonts. They are available as both "regular" and "Euro" fonts. If the text appears reversed, switch to or from "Euro." Also the Hebrew characters can be stored in a variety of formats. Sometimes you might find it necessary to switch between ElroNet and Web Hebrew in order to read the characters. In addition we sometimes use a technique that works well on Hebrew Windows with regular Hebrew fonts (particularly Narkisim). (Courtesy Reading Hebrew... Internet Hebrew.)

Directions for viewing Hebrew—Has information for Windows users of Microsoft Internet Explorer 3.0+, Netscape Navigator 3.0+, 2.0, and other browsers:
http://www.hebhtml.co.il/HebHTML/2heb.htm

How to view Hebrew on the Internet—Has information on Hebrew fonts, for PC, Mac and Unix, as well as information on installation and printing:
http://ietn.snunit.k12.il/hebrew.htm

Reading Hebrew—Contains information for installation for fonts to view Hebrew on Unix, PC and Mac computer systems:
http://www.macom.co.il/Hebrew/poetry/how.to.read.hebrew.html

Japanese There are three major Japanese character set encodings: JIS (ISO-2022-JP), SJIS (Shift-JIS) and EUC-JP (Extended Unix Code for Japanese). ASCII Character set for Japanese:
http://staff.navisoft.com/gww/fonts/Charsets.html#JIS-201

Encodings of 94x94-character sets:
http://www.ifcss.org/ftp-pub/software/info/cjk-codes/94x94-encodings.html

A Japanese extension—used with MS Internet Explorer to view Japanese on the WWW:
http://www.panix.com/~tn/j-pc-os.html#msie

Russian Cyrillic is much easier to decode then some language scripts, mainly because it is supported. Part of the problem though, is that it is perhaps overly supported. There are at least four different encoding systems which are supported generally throughout the Russian sphere of influence. These are KOI8-R, WIN 1256, a DOS version, and Mac Russian encoding. Unfortunately, depending on the computer you have, you may or may not be able to read some of these. Fortunately, most web pages are encoded in at least two of these. I have found it best to work with KOI8-R for Mac computers and WIN 1256 for Windows based environments, because there are pages which auto-detect what your system is and place you in an encoded document based on your system, leaving you no choice.

Dazhdbog's Grandchildren Software—is a resource page for many different aspects of Russia. It includes information about KOI8-R:
http://sunsite.oit.unc.edu/sergei/Software/Software.html

Home of KOI8-R —Russian Net Character Set—You can find KOI8-R, a Cyrillic font, usable on Mac, and sometimes in Windows, for reading common Cyrillic pages. Also contains info about making a Cyrillic page:
http://www.nagual.pp.ru/~ache/koi8.html

SovInfoBureau Russify Anything—A resource on the different types of encodings (KOI8-R, unicode, Rosetta, CP1251, and others). Gives instructions and fonts for WIN, XWIN, DOS, OS/2, TeX. Links to Mac, Commadore, and VT200, 320 terminals pages.
http://www.siber.com/sib/russify/Russification
http://www.as.ua.edu/gnrn/russification.html

Russian Audience Instructions:
http://www.civ.com/realtors/RUSSDIR.HTM

Cyrillic for MS Windows Netscape—Directions for the cyrillicification of Windows, with different versions of Netscape, and how to write and read email:
http://www.siber.com/sib/russify/ms-windows/netscape.html

Multiple Language Resources

This lists a variety of sources of fonts via the WWW. All of the fonts available at these sites may not function with WWW browsers but can be downloaded and used for word processing and more.

Unicode—A universal language font, created by committee.
http://www.unicode.org/

Rosetta—A universal language font by a person:
http://www.pluris.com/rosetta/

The ISO 8859 Alphabet Soup— ISO 8859 is a standardized series of 8-bit character sets for writing in Western alphabetic languages. It was designed by the European Computer Manufacturer's Association (ECMA). The series is included in the Internet charset register for use with MIME. ISO 8859 is not even remotely as complete as Unicode, but it is already a major improvement over the 7-bit US-ASCII:
http://wwwwbs.cs.tu-berlin.de/user/czyborra/charsets/

ASCII Encoding:
http://staff.navisoft.com/gww/fonts/Charsets.html

Yamada Font Archive:
http://babel.uoregon.edu/Yamada/fonts.html

Internet resources for Language Teachers—Contains links to fonts for a variety of languages:
http://www.hull.ac.uk/cti/langsite.htm/fonts.htm

Using Fonts in Netscape and Microsoft Internet Explorer

To use the fonts below you must first install them on your hard drive. (Macs and Windows machines obviously deal with fonts in different ways, so install them as you normally would.) Once the fonts are available on your computer, in Netscape under general preferences choose fonts. Here you will need to set up the encoding as well as the fonts* style for each encoding. Finally, under the options menu, click on document encoding and select the appropriate encoding system. Each system has its quirks and will work with some pages and not with others.

Microsoft Internet Explorer works similarly to Netscape except you change the fonts* for each encoding from the preferences menu under edit. To change the encoding to read different pages, choose font encoding under the view menu and select the appropriate encoding system.

(*Note that in your browser setup you should specify both proportional and fixed fonts.)

Software Solutions

In our continued search for a simple solution to the script problem, we discovered that the Macintosh operating system and the various MS Windows operating systems are available in many other languages providing easy access to the necessary elements for viewing that languageís character script on the WWW. There also exist a plethora of language-specific and multilingual web browsers as well as a variety of language versions for the more popular web browsers. However, to view a variety of character scripts in the same system, we found a number of software products for both Mac and Windows machines which make choosing the appropriate character script as easy as clicking a button.

Apple Language Kits For Macs has Language Kits which are available for Arabic and Persian, Chinese, Cyrillic, Hebrew, Indian, Japanese and Korean:
http://www.macos.apple.com/multilingual/

On the Windows side the programs range from a single language to more than two dozen. Two programs offer a great deal of flexibility as well as a large number of languages.

GlobalSurf from DynaLab Inc.—This program supports more than twenty font encoding systems including Arabic, Chinese, Cyrillic, Greek, Hebrew, Japanese, Korean, Turkish and a variety of other Eastern European and Scandinavian languages: **http://www.dynalab.com/**

Internet with an Accent by Accent Software—With a browser, HTML editor, navigator plug-in and more, this program supports over 20 character script encoding systems for

viewing and creating multilingual sites on the WWW.
http://www.accentsoft.com/ ■

Chris Higgins is Coordinator for Foreign Language Instructional Technology at The Language Center, University of Maryland—College Park.