YA-NewsWatcher and Character Sets

Introduction
Yet Another NewsWatcher adds support for character transliterations other than Mac <-> Latin1.

Not everyone in the world uses the same alphabet, and not all computers have the same character sets mapped to the 256 different possible 8-bit character codes. The use of standardized character sets and transliterations helps to work around some of the problems this causes.

The earliest versions of NewsWatcher performed no character transliteration at all. That was fine, as long as both the writer of the message and the reader of the article were using Macs. When this was not true, the results would be funny looking text. For example, the sender on a Mac would use the Apple character (shift-option-K), and the receiver on an IBM PC might see some graphic character. Starting with one of then 2.0b releases, John Norstad included support for the standard and by far most commonly used character set mappings. When sending a message, the Mac to Latin1 (also known as ISO-8859-1) transliteration would be used. This maps the extended Mac characters to the closest matching standard Latin1 characters, so that when a Mac poster uses the é character it is translated to the appropriate matching character code in the Latin1 character set. Similarly, when receiving messages, NewsWatcher would assume that the article was written using the Latin1 character set, and would transliterate characters back to the Mac characters set using an inverse transliteration.

This works fine most of the time. It doesn't work well when the article being received wasn't in the Latin1 character set, which may happen in non-English language newsgroups.

So, like the excellent email program Eudora, and many other internet applications, YA-NewsWatcher adds support for additional transliteration tables. This makes it easy for third parties to customize the program for non-English languages (such as Cyrillic) without having to modify and recompile the program code, and without losing the ability to effectively post to and read English language newsgroups.

How it Works
When YA-NewsWatcher is first launched, it finds (and creates if necessary) a folder named "YA-NewsWatcher Tables" in the same folder as the active preferences folder. It scans this folder for files that have the necessary 'taBL' resources, loads the transliteration tables and adds the names of the tables to some transliteration menus. These appear as popup menus in a number of dialogs throughout the program, and allow the user to specify what kind of character transliteration will be performed. For those who want to create new transliteration tables, the format of these files is described the end of this document.

There is a new Preferences dialog panel that lets the user set the default character transliteration that will be performed when displaying articles, when posting a message, and when sending a message. By default these are set to the fixed values used by older versions of YA-NewsWatcher (Mac to Latin1 when sending; Latin1 to Mac when receiving). The default transliteration done on articles can also be set for individual newsgroups by using the "Newsgroup Settings" dialog. The transliteration can also be changed when sending a message by clicking on the "Message Settings" button in message windows. When sending a message, YA-NewsWatcher now also inserts proper MIME headers, indicating what character set was used.

One other common transliteration that may be useful is a Mac to US ASCII conversion. This attempts to replace non-ASCII characters in the Mac character set with the nearest matching ASCII character. For example, all the forms of accented e characters (é, è, ê, ë) are replaced with a plain e. This can be useful for emailing a message to a user whose email system doesn't handle 8-bit

characters. Included with the YA-NewsWatcher distribution is a sample Mac to US ACSII table, located in the "Tables" folder. Just place the file "Macintosh -> US ASCII" in the "YA-NewsWatcher Tables" folder and you'll be able to specify this transliteration when sending messages. Also included are four more sets of tables, thanks to Andreas Prilop (Cyrillic KOI-8, Cyrillic Windows, Greek, and Latin-2)

Finally, please note that character set support in YA-NewsWatcher is still incomplete. There are still two obvious features that need to be added. Both these features should be included in a future release of YA-NewsWatcher.

One is a way to change the character set and font for an article after it has been opened. What currently happens is the article is opened using the specified (either program-wide default or newsgroup-specific) character transliteration, and is both displayed that way and stored in the RAM article cache. If you try to close the window, change the character transliteration, and reopen the article, the already transliterated text in the RAM cache will be redisplayed, and the new transliteration settings will be ignored.

The second thing that YA-NewsWatcher should do and still does not is to deduce the proper character set for articles which have the proper MIME headers. This would automatically detect the correct transliteration table to use, and would override any default value (which would still be used for the numerous posts that don't have MIME headers).

Table File Format
YA-NewsWatcher now has a new icon for files of type 'TABL' (creator 'NNTP', of course) that can be used for transliteration table files. However, any file in the "YA-NewsWatcher Tables" will be scanned for tables, no matter what its creator or type.

The table files are very similar but not identical in format to those used by Eudora. Table resources (type 'taBL') are classified into two kinds: those used when sending, and those used when receiving. Those used for sending must have even resource ID numbers; those used for receiving must have odd resource ID numbers. This is the same convention used by Eudora.

The name of the 'taBL' resource is shown in YA-NewsWatcher's popup menus, and tables are remembered by resource name, not resource number. So resource ID numbers need only be unique within a given file.

For each 'taBL' resource there MUST be a matching 'STR ' resource with the same ID number. The contents of the string should be the (often standardized) name of the character set it is used with. For example, both the Macintosh -> ISO-8859-1 and ISO-8859-1 -> Macintosh tables have 'STR ' resources that contain the string "ISO-8859-1". When the matching table is used to transliterate a message being sent, this string is used in the MIME charset= header. In a future version of YA-NewsWatcher that performs automatic transliteration table selection based on an article's MIME headers, this string will be used to select which input transliteration table to use.

I encourage people creating additional table files to put just one pair of tables in each file. This will make it easier for users to select which tables they do or do not want to use without having lots of duplicate or unwanted tables causing confusion.

Special thanks to Jean-Pierre Kuypers for all his help in testing the new character set transliteration features.