Determine what character encoding is used by a file

file -bi [filename]

Example output:

steph@localhost ~ $ file -bi test.txt
text/plain; charset=us-ascii

Use vim to change a file's encoding

If you use the vim text editor, you can configure it to save files as utf-8. Place the following in your /etc/vim/vimrc or ~/.vimrc file:

set encoding=utf-8
set fileencoding=utf-8

You will only notice a difference in the encoding if you edit the file and add unicode (utf-8) characters (most character keys on the keyboard will create a unicode equivalent if you hold down the alt key). Start vim, edit the file and add some unicode characters. If you create a test file containing the following...

steph@localhost ~ $ cat utf8test.txt
abcdefghijklmnopqrstuvwxyz
á ãä ç éêëìíîïðñò õö øùú

...then the file command should tell you the file is utf-8:

steph@localhost ~ $ file -bi utf8test.txt
text/plain; charset=utf-8

If you then remove the UTF-8 characters and save the file, it will be us-ascii again.

Change a file's encoding from the command line

To convert the file contents to from ASCII to UTF-8:

iconv -f ascii -t utf8 [filename] > [newfilename]

Or

recode UTF-8 [filename]

To convert the file contents from UTF-8 to ASCII:

iconv -f utf8 -t ascii [filename]

Because UTF-8 can contain characters that can't be encoded with ASCII, this command will generate an error unless you tell it to strip non-ASCII characters using the -c flags:

steph@localhost ~ $ iconv -f utf-8 -t ascii utf8test.txt
abcdefghijklmnopqrstuvwxyz
iconv: illegal input sequence at position 27
steph@localhost ~ $ iconv -c -f utf-8 -t ascii utf8test.txt
abcdefghijklmnopqrstuvwxyz

A similar thing can be achieved using the -f flag with the recode command.

steph@localhost ~ $ recode ascii utf8test.txt
recode: utf8test.txt failed: Invalid input in step `ANSI_X3.4-1968..CHAR'
steph@localhost ~ $ recode -f ascii utf8test.txt
steph@localhost ~ $ cat utf8test.txt
abcdefghijklmnopqrstuvwxyz

Warning: If you use the iconv -c flag or the recode -f flag, you could loose characters.

Change the filename encoding

To convert the filename from ascii to UTF-8:

Warning: Run this without the --notest option first, to make sure there will be no problems.

convmv -f ascii -t utf8 --notest [filename] > [newfilename]

References

- [Saving files as utf-8](#)
- [Tip of the Trade: GNU Recode](#)
- [How to create UTF-8 files?](#)

Last modified: 25/09/2008 Tags: (none)