

Re: Filename Encoding Help

Source:

<http://www.tech-archive.net/Archive/DotNet/microsoft.public.dotnet.framework/2008-06/msg00091.html>

- *From:* "Peter Duniho" <NpOeStPeAdM@xxxxxxxxxxxxxxxxxxxx>
 - *Date:* Sun, 08 Jun 2008 23:03:52 -0700
-

On Sun, 08 Jun 2008 16:29:39 -0700, Adhal <fakeemail@xxxxxxxxxxxx> wrote:

Thanks Göran,

You can use any unicode encoding, like UTF-7, UTF-8, UTF-16 or UTF-32.

I suggest UTF-8, it's the most efficient for regular text, and it's the default for all methods reading and writing text files in .NET.

Basically this program stores filenames & and other file details and it is going to be used only on windows XP and Vista. I want to support all languages that the filenames are capable of.

UTF-8 should be fine.

The problem is I am almost certain that Windows XP stores filenames in UTF-16 but I am not sure what Windows Vista does. I don't want to use UTF-32 if I don't need it, as increases the file size unnecessarily. :-?

That is exactly why Göran recommended UTF-8. It takes advantage of the fact that for certain Unicode characters, the first byte is sufficient. UTF-8 can represent the full Unicode spectrum, but many characters wind up encoded in just one or two bytes. It will produce the smallest files.

That depends on what characters the file names contains. The ASCII encoding only handles characters with character codes from 32 to 127. The ANSI character set will handle any characters in the ASCII character set.

It also depends on what you are going to use the file for. Is there any other program that will read the file?

Re: Filename Encoding Help

Ok this one is a bit more puzzling for me. Again it is a program that stores file names on Windows XP and Windows Vista into a text file, however the file can be opened by Windows 9x/me.

By what program?

Give you an example is best approach. I have Windows XP (Japanese) and I store the filenames in an ANSI text file. Now I take the file and open it up in Windows 98 (Japanese). I would expect the text file to open fine and all the characters appear fine. Am I right in my thinking?

It depends on what program opens it. Whether a Unicode format can be displayed depends on whether you're using a program that supports Unicode. I don't recall whether, for example, Notepad in Win98 supports Unicode. Win98 itself doesn't have a Unicode-enabled version of the Windows API, so it's possible Notepad also doesn't support Unicode on Win98.

A Unicode-enabled Windows API isn't required in order for a program to support Unicode, but it sure helps. :)

My understanding is this works if I store it as ANSI but not as ASCII. I haven't got Japanese Windows 98 to test this out. :(

The difference between ANSI and ASCII should be negligible with respect to dealing with MBCS or Unicode, since neither of the latter can be encoded in the former.

Pete

.