
Subject: Re: Running Asian Scripts in Radicore
Posted by [nnonnes](#) on Thu, 25 Jan 2007 16:46:50 GMT
[View Forum Message](#) <> [Reply to Message](#)

To the best of my knowledge, WordPad is not able to save files in UTF-8 format. It does have a "Unicode Text Document" option for saving files but I think this is 2-byte (UCS-2) Unicode, not multi-byte (UTF-8). It is therefore inappropriate for Thai characters, which occupy 3 bytes in UTF-8 format. It probably wouldn't work properly for other character sets, either, even those that take up 2 bytes in UTF-8.

If you change the Radicore code to eliminate the conversion from latin1, then I suggest that you amend your article to make UTF-8 encoding mandatory for all `language_text.inc` and `sys.language_text.inc` files that contain non-ASCII characters. This is because the `convertEncoding` function would no longer convert non-UTF-8 files accurately -- for example, it would no longer handle latin1 files having characters outside the ASCII character set.

The article should also recommend using UTF-8 encoding everywhere as a best practice, for it produces exactly the same output as ASCII encoding (for ASCII characters) whilst handling virtually every other language on Earth.

I will report what I learn as I test the various `mbstring.func_overload` options.
