
Subject: Running Asian Scripts in Radicore
Posted by [nnonnes](#) on Thu, 25 Jan 2007 11:26:17 GMT
[View Forum Message](#) <> [Reply to Message](#)

We are building an application using Radicore that needs to run in various East Asian languages as well as English; i.e. Thai and Simplified Chinese. I've run into a couple of issues and devised successful workarounds that I thought I'd share.

In function `getLanguageText`, the statement:

```
$string = convertEncoding($string, 'latin1', 'UTF-8');
```

does not work when the `language_text.inc` or `sys.language_text.inc` files are already encoded in another character set such as UTF-8. This would always be the case when the translator is keying in text in a language like Thai.

I studied the `convertEncoding` function and discovered that it automatically inserts 'UTF-8' when the second argument ('latin1' as called by the `getLanguageText` function) is null. So, the above statement should be:

```
$string = convertEncoding($string, "", 'UTF-8');
```

This seems to solve the problem. Of course it will only work if the text files are really saved in UTF-8 and not some other character set that's not latin1. To prevent the problem from recurring we have decided to encode all text files in UTF-8, except the ones that we are absolutely certain contain nothing but ASCII characters.

I have also discovered that when you are saving a text file using UTF-8 encoding, using the standard Notepad editor that ships with Windows, it saves the file as "UTF-8 with signature." The signature is a special 2-byte character HEX(FE FF) that the Unicode standard recognizes as the "Byte Order Mark" (BOM). When Notepad saves the file, the BOM is inserted right at the beginning of the document. The resulting HTML pages that the Radicore XSLTs create, always contain the BOM unless the text file is saved in ASCII.

This wouldn't be a big deal except that Internet Explorer 7 has a bug that causes it to misbehave when it sees a BOM.

The solution is to save all the files containing non-ASCII text as plain old UTF-8, without the signature. But Notepad won't let you do this. The "Notepad2" freeware text editor seems to work fine. It's actually much nicer than Notepad so I plan to use it for all our work.

The third issue is really a question for Tony. Following the "Internationalisation and the Radicore Development Infrastructure" article, we enabled the multibyte string functions in our PHP (5.*) configuration. Not knowing exactly what to do, we decided to go all the way and implement all the overload functions in `php.ini`, as in:

```
mbstring.func_overload = 7
```

After this, I discovered that Radicore no longer processes multiple records when you select multiple rows from a screen such as LIST1. I have not investigated this problem any further but would like an opinion on whether function overloading is the culprit. I'd also appreciate a recommendation on what this setting should be.
