Subject: Re: Making support for code pages unlimited
Posted by mirek on Sat, 13 Feb 2010 12:58:40 GMT

Mindtraveller wrote on Tue, 09 February 2010 17:45Recently I had some time to analyze encodings. And I've met a number of issues to be discussed.

First, I discovered that CHARSET_**** tables are not in unicode, but in UTF-8. If it is so, I should rebuild proposed tables according to this.

What makes you think that? Those tables are just UTF-16 codes for characters 128-255.

Quote:
Second issue is simple etude, which actually failed. Just look at this simple code:CONSOLE_APP_MAIN
{

}
 I executed this example as console app under Windows XP. My "native" console code page under XP is CP866. But on program run, instead of cyrillic letters, I've seen garbage symbols. My question: what is wrong in this example and why do you think it fails to convert symbols properly?

Well, it looks like things are more complicated. I believe our part is OK, but the console simply expcts the output to be in different charset. E.g.:

 http://illegalargumentexception.blogspot.com/2009/04/i18n-un
icode-at-windows-command-prompt.html

However, more thinking about it, I believe we should perhaps use Unicode variant of WriteFile and convert current app encoding (which is utf8 by default) to unicode.

In that case, however, you example would work without ToCharset...

Mirek