

---

Subject: UTF8 and accented characters

Posted by [forlano](#) on Thu, 26 Mar 2009 20:51:01 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

Hello,

in the past I have signaled a problem in misaligned characters in a simple text file in which there were accented characters. It looked like this below:

In principle there is a shift toward the left for each accented character. It seems there is a "eated" space for each accented character.

Today I have a simple test case that produces another problem that maybe related with the previous one.

```
#include <Core/Core.h>

using namespace Upp;

CONSOLE_APP_MAIN
{
    String a = "ultimatepp ultimatepp ultimatepp\n";

    SaveFile("a.txt", a);
}
```

This is what I see if in Project/Package organizer I set ENCODING=Default

If instead I set ENCODING=UTF8 I get the problem with accented characters:

Does not help to have as first line the command

```
SetDefaultCharset(CHARSET_UTF8);
```

Is it normal? Of course I am missing something... but what?

Thanks,

Luigi

---

## File Attachments

- 1) [text2.jpg](#), downloaded 926 times
- 2) [text1.jpg](#), downloaded 876 times

3) [text3.jpg](#), downloaded 529 times

---

---

Subject: Re: UTF8 and accented characters  
Posted by [mirek](#) on Fri, 27 Mar 2009 07:56:00 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

forlano wrote on Thu, 26 March 2009 16:51  
Is it normal? Of course I am missing something... but what?  
Thanks,  
Luigi

IMO:

You must always consider that there are two encodings involved:

Encoding that TheIDE uses to edit the file (that is in defined in Project menu).

Encoding that your code thinks the file is in. (This is set by SetDefaultCharset mainly).

If there is no match, you can see those weird things

Also, if you switch the encoding for some file, TheIDE does not reencode it. That is why you see weird things after switching to utf-8 - the file that is NOT in utf8 gets displayed as if it was...

But you can still convert the file to another encoding, using Project/Convert to encoding.

Mirek

---