
Subject: Re: LoadFile problem with accented chars
Posted by [koldo](#) on Sun, 08 Feb 2009 21:11:52 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hello luzr

It seems it is a matter of Notepad itself. If the file has 7 bits chars there is no problem, but after

Using this test program:

```
CONSOLE_APP_MAIN
{
  String data = LoadFile("C:\\test.txt");
  for (int i = 0; i < data.GetCount(); ++i)
    puts(Format("%d: %d", i, data[i]));
  getchar();
}
```

```
0: 97
1: 45
2: -31
```

but after saving and opening the file some times, I get this:

```
0: -1
1: -2
2: 97
3: 0
4: 45
5: 0
6: -31
7: 0
```

and yesterday I got other output... The answer is that Notepad adds a "BOM" to the file if it thinks it requires a bigger encoding.

BOM (Byte Order Mark, http://unicode.org/faq/utf_bom.html#BOM) is a signature of letters in the beginning of files that shows its encoding. For example:

- EF BB BF means UTF-8
- FF FE means UTF-16, little-endian

So yesterday Notepad saved the file as UTF8 (beginning with -17 == EF) and today it saved it in UTF-16, little-endian (beginning with a -1 == FF)

Sorry, perhaps it is not easy but, do you know how to program a way to get a text file and converting it into utf8 to be properly viewed U++ programs?, as when entering these chars into

U++ controls I get strange symbols and errors. It will also be great for parsing them.

Best regards
Koldo
