
Subject: Re: LoadFile problem with accented chars
Posted by [koldo](#) on Tue, 10 Feb 2009 08:23:50 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hello luzr

I have had problems. In reading an UTF16-Little Endian, ToUtf8 does not seem to do it well.

I have used this function:

```
String LoadFileBOM(const char *path)
{
    String s = LoadFile(path);
    if (((s[0]&0xFF) == 0xFF) && ((s[1]&0xFF) == 0xFE))    // UTF16 Little Endian
        s = ToUtf8(s.Mid(2).ToWString());
    else if (((s[0]&0xFF) == 0xEF) && ((s[1]&0xFF) == 0xBB) && ((s[2]&0xFF) == 0xBF)) // UTF8
        s = s.Mid(3);
    return s;
}
```

that is called from here:

```
String s = LoadFileBOM("demo_u_16le.txt");
String ss;
```

```
for (int i = 0; i < s.GetCount(); ++i)
    ss.Cat(Format("%d: %0x;\n", i, s[i]&0xFF));
```

```
ss.Cat(s);
TestLineEdit.SetData(ss);
TestEditString.SetData(ss);
TestDocEdit.SetData(ss);
```

```
Quote:0: 41;
1: 75;
2: 70;
3: c3;
4: a1;
```

but it is:

```
Quote:0: 41;
1: 0;
2: 75;
3: 0;
4: 70;
```

```
5: 0;  
6: e1;  
7: 0;
```

Another question. When loading an UTF8 file with the later code, LineEdit and DocEdit reads it right but EditString shows a strange char. Does EditString not show UTF8 or I am doing it wrong?.

Best regards
Koldo
