Subject: Re: LoadFile problem with accented chars
Posted by cbpporter on Wed, 11 Feb 2009 14:25:08 GMT

View Forum Message <> Reply to Message

s = ToUtf8(s.Mid(2).ToWString());
Well by looking at this code from the old version, I believe you should be getting the results you were getting (the wrong ones).

When calling ToWString, the input will be considered to be in Utf8, but you have an Utf16 stuffed inside an String (Utf8). That explains the extra zeros.

On the other hand, the new versions may work, but they are a little heavy on conversions and allocations so I wouldn't use them on large files. Especially for Utf8 BOM, I believe the solution would be to create a LoadFile which detects the BOM and allocates and fills a buffer/String without the BOM directly. I'll have to check out LoadFile before I can elaborate on this.