

Re: How to read UTF-8 chars using VBA

Source:

<http://www.tech-archive.net/Archive/VB/microsoft.public.vb.general.discussion/2006-12/msg00485.html>

- *From:* "Tony Proctor" <tony_proctor@xx>
 - *Date:* Tue, 5 Dec 2006 16:56:02 -0000
-

As I stressed above, you have to read it from the file in binary mode (not with Line Input), and then convert it from UTF-8 to Unicode.

The following code assumes the utf.txt file contains a properly-flagged UTF-8 content (e.g. as saved by Notepad). Note that this means there will be a magic 3-byte UTF-8 marker at the start of the file. The UTF8->Unicode conversion will convert this to the Unicode marker &HFEFF.

```
'===== Start Form1 =====
Private Const CP_UTF8 = 65001

Private Declare Function MultiByteToWideChar Lib "kernel32" ( _
ByVal CodePage As Long, ByVal dwFlags As Long, _
ByVal lpMultiByteStr As Long, ByVal cchMultiByte As Long, _
ByVal lpWideCharStr As Long, ByVal cchWideChar As Long) As Long

Public Function sUTF8ToUni(bySrc() As Byte) As String
' Converts a UTF-8 byte array to a Unicode string
Dim lBytes As Long, lNC As Long, lRet As Long

lBytes = UBound(bySrc) - LBound(bySrc) + 1
lNC = lBytes
sUTF8ToUni = String$(lNC, Chr(0))
lRet = MultiByteToWideChar(CP_UTF8, 0, VarPtr(bySrc(LBound(bySrc))),
lBytes, StrPtr(sUTF8ToUni), lNC)
sUTF8ToUni = Left$(sUTF8ToUni, lRet)
End Function

Private Function ConvertUTF8File(sUTF8File As String) As String
Dim iFile As Integer, bData() As Byte, sData As String, lSize As Long

' Get the incoming data size
lSize = FileLen(sUTF8File)
If lSize > 0 Then
ReDim bData(0 To lSize - 1)

' Read the existing UTF-8 file
iFile = FreeFile()
```

Re: How to read UTF-8 chars using VBA

```
Open sUTF8File For Binary As #iFile
Get #iFile, , bData
Close #iFile

' Convert all the data to Unicode (all VB Strings are Unicode)
sData = sUTF8ToUni(bData)
Else
sData = ""
End If
ConvertUTF8File = sData
End Function

Private Sub Form_Load()
Dim vLine As Variant, sFileBody As String

' Load the UTF-8 file body into a Unicode string variable
sFileBody = ConvertUTF8File("utf.txt")
' Remove the leading Unicode marker from the body (i.e. the &HFEFF
sequence)
sFileBody = Mid$(sFileBody, 2)
Debug.Print sFileBody
' Now add the separate lines to out list box
For Each vLine In Split(sFileBody, vbCrLf)
List1.AddItem CStr(vLine)
Next vLine
End Sub
'===== End Form1 =====
```

Tony Proctor

"MSK" <mannaikarthik@xxxxxxxx> wrote in message
news:1165332779.981591.277710@xx

How did you load ? Can you please pass me the code.. ? Did you use any
builtin methods to convert ?

I tried to load like the following

```
open <filename> for input as #1
do while end of file
line input #1 ,ABC
loop
close #1
```

using the ABC (string) variable I filled the combo box

Can we get the y,w (circumflex) chars also as it is?

MSK

Re: How to read UTF-8 chars using VBA