

RE: VBA question: How to extract cell values in different language

Source:

<http://www.tech-archive.net/Archive/Excel/microsoft.public.excel.programming/2008-05/msg02757.html>

- *From:* Christopher <Christopher@xxxxxxxxxxxxxxxxxxxxxxxxxxxxx>
 - *Date:* Wed, 14 May 2008 10:59:00 -0700
-

I was really hoping to not have to work in binary mode, but I guess that's the only way then.

Unfortunately, I believe I'm still screwed if my OS is set to Japanese. The Range object I receive from the Cells method on the sheet seems to process the data before it gives it to me (modifies the data to fit ANSI with no extended characters).

Is there a way to get "raw" data from cells on a worksheet?

"Joel" wrote:

There are two ways you may correct the problem.

1) Open the file in binary Mode

from

Open TblName For Output As #1

to

Open TblName For Binary Access Write As #1

2) The problem could be with the Print #1 statement. Not sure. I know if you used Write #1 this function puts double quotes around the data. You may want to try the Scripting Object to write the data. Here is an example

```
Sub WriteFixed()
```

```
Const MyPath = "C:\temp\"
```

```
Const WriteFileName = "text.txt"
```

```
Const ForReading = 1, ForWriting = 2, ForAppending = 3
```

```
Const TristateUseDefault = -2, TristateTrue = -1, TristateFalse = 0
```

RE: VBA question: How to extract cell values in different language

```
Set fswrite = CreateObject("Scripting.FileSystemObject")

'open files
WritePathName = MyPath + WriteFileName
fswrite.CreateTextFile WritePathName
Set fwrite = fswrite.GetFile(WritePathName)
Set tswrite = fwrite.OpenAsTextStream(ForWriting, TristateUseDefault)

LastRow = Cells(Rows.Count, "A").End(xlUp).Row

For RowCount = 1 To LastRow
LastCol = Cells(RowCount, Columns.Count).End(xlToLeft).Column
OutputLine = ""
For ColCount = 1 To LastCol
Data = Cells(RowCount, ColCount).Text
If Len(Data) < 12 Then
Data = Data & WorksheetFunction.Rept(" ", 12 - Len(Data))
End If
OutputLine = OutputLine & Data
Next ColCount
tswrite.writeline OutputLine
Next RowCount

tswrite.Close

End Sub
```

"Christopher" wrote:

Hi Joel,

I understand that the same text data will appear differently between a simple application like Notepad and something more flexible, like Microsoft Word.

But the actual text data is being modified for some reason.

For example, if I look at my text data, I see my intended result is supposed to have 'a' with an accent (0xE0), but the code will spit out 'a' (0x61) instead. It's like excel is toying with me. It's smart enough to somehow know that 0xE0 is 0x61 with an accent on it and decides to "down-grade" it to 0x61. The application is obviously smart enough to know what type of language is entered, but it seems like all that data is lost when the VBA code starts running.

To me, it looks like there has to be some kind of global language value that must be set to the correct language before running the code. But for some reason, all the language related information appears to be read-only. I'm

RE: VBA question: How to extract cell values in different language

not sure what the point of these functions are if they are only "get" methods and no other method (like a "set" method) seems to use them as parameters.

I greatly appreciate the time you're taking with me Joel.

"Joel" wrote:

You haven't confused me. I completely understand the problem. Your comment about having binary data and not unicode data confirms my suspicions.

Let me explain what is happening. You are have 256 binary characters. Depending on the language settings the characters are displayed differently, the data doesn't change. How the characters are displayed I'm calling a font.

If you use Notepad which only displays binary data (ascii) and not the Unicode data on the same file with two different language setting gives you different result; then the data hasn't changed. Only the way the data is displayed has changed. This is a change of a Font.

Go to word a open the Insert – Symbol window. You will see most of the character (look at bottom of window) are characters from 0000–00FF. These are the ascii characters. When the value of a character is greater than FF then these are the unicones.

You will see that when you change Fonts the you will see the same numerical values for different characters. This is what happens when you change your language setting. You don't change the data, the way the data displays changes.

"Christopher" wrote:

I'm sorry, I may have confused you when I mentioned "visually".
I don't care about font.

RE: VBA question: How to extract cell values in different language

I just need the correct text data to appear in the output.

When I check the binary data on the Japanese file, I see that it basically forced the unicode data into single byte characters.

I'm used to C++ where I can just treat the text data binary files and run the data through functions to convert to UTF-8 and other encodings.

When I walk through the VBA code, it looks like the data is corrupted as soon as the data is accessed through the Cells function.

"Joel" wrote

You are writing text data which has 256 different characters. Each Byte is 8 bits (referred to as Ascii) of data (2 raised to the 8th). There is no font with text data. Depending on your language setting the text displays with the font as specified by the language setting.

If you import the text into Word which has multiple fonts you can change the font to any language you want without changing the language setting.

There is no problem with your excel program. The easiest thing to do is to open the text file and copy the data. then paste the data into a word document and select the proper Font for the language. Then save the word document. Word will

RE: VBA question: How to extract cell values in different language

preserve the correct font
which a text file cannot do.

"Christopher" wrote:

Hi,

I was
hoping if
someone
could help
me with a
language
problem I'm
facing.
The code
below is
suppose to
extract
strings from
a worksheet
and create
string table
files that
will be fed
into a
resrouce
file. I am
using Excel
2007
currently.

The style of
the
worksheet:
First
column has
the string
IDs
First row
contains the
name of the
file to create
(excluding
column 1,
which
is skipped)
Each
column
(file) has a

RE: VBA question: How to extract cell values in different language

different
language
(english,
french,
german,
italian,
portuguese,
spanish, and
japanese)

The style of
the output
file:
Has header
"STRINGTABLE"
and
"BEGIN"
Each row of
text has a
string ID
and string
value
Has footer
"END"

I'm not an
experienced
VB
programmer,
so I am
basically
depending
on Excel
to be smart
enough to
understand
what I want.
It "sort of"
works, but
it's
very
annoying
that I have
to keep
switching
languages
to get the
correct
output.

The

RE: VBA question: How to extract cell values in different language

problem:
If I set my
OS
language
settings to
English, the
non-Japanese
languages
are
extracted
properly,
but
Japanese
gets
corrupted. If
I set my OS
language to
Japanese,
languages
like French
loses all
accents on
characters.

What I'm
aiming for:
To get what
is exactly
(visually)
displayed
by the excel
spreadsheet
to be
outputted to
a file
regardless
of the OS
language
setting. I
don't
necessarily
care about
what format
it's in
(except not
in Unicode)
as long as
it's
consistent.
Granted the
work

RE: VBA question: How to extract cell values in different language

around is to
switch the
OS
language
settings, but
I might
have to add
more
languages
later on and
the work
around
itself makes
me cringe.

– Japanese
is using
Shift–JIS
– Other
languages
are using
ANSI
(extended
latin) or
UTF–8

I was
hoping
there is a
way to
change the
language
setting of
the
application
(on–the–fly)
to
something
that is
compatible
with the
language of
the strings
(column) I
am
currently
extracting.

I searched
the forum,
but couldn't

RE: VBA question: How to extract cell values in different language

find a
similar
issue
posted.
Sorry if
this issue
has already
been
addressed.

Any help
will be
greatly
appreciated.

The code:
Sub
Export()

```
Dim  
TblName  
As String  
Dim CurDir  
As String  
Dim Text  
As String  
Dim SID  
As String  
Dim  
WriteBuf  
As String  
Dim cCount  
As Integer  
Dim rCount  
As Integer  
Dim  
rowMax As  
Integer
```

```
'get  
directory  
base for tbl  
files  
CurDir =  
GetCurDir
```

```
'get  
maximum  
number of  
rows  
Dim wkst
```

RE: VBA question: How to extract cell values in different language

```
As  
Worksheet  
Set wkst =  
Worksheets("StringTables")  
rowMax =  
wkst.UsedRange.Rows.Count
```

```
'get first tbl  
column and  
initialize  
row and  
column  
count  
TblName =  
CurDir +  
wkst.Cells(1,  
2).Text  
rCount = 2  
cCount = 2
```

```
'keep  
looping as  
long as a tbl  
name exists  
Do While  
TblName  
<> CurDir  
Open  
TblName  
For Output  
As #1  
'print string  
table  
headers  
Print #1,  
"STRINGTABLE"  
Print #1,  
"BEGIN"
```

```
'get first  
string ID  
number and  
string value  
SID =  
wkst.Cells(rCount,  
1)  
Text =  
wkst.Cells(rCount,  
cCount)  
Do While  
rCount <=
```

RE: VBA question: How to extract cell values in different language

```
rowMax

'skip this
row if ID or
value is
missing
If SID <> ""
And Text
<> "" Then
'Chr(9) =
tab Chr(34)
= double
quote
WriteBuf =
Chr(9) +
SID +
Chr(9) +
Chr(34) +
Text +
Chr(34)
Print #1,
WriteBuf
End If

'increment
row count
rCount =
rCount + 1
'get next
pair of
string ID
and value
SID =
wkst.Cells(rCount,
1)
Text =
wkst.Cells(rCount,
cCount)
Loop

'print table
footer
Print #1,
"END"
Close #1

'reset row
count
rCount = 2
'increment
column
```

RE: VBA question: How to extract cell values in different language

```
count (new  
language)  
cCount =  
cCount + 1  
'setup for  
new tbl file  
TblName =  
CurDir +  
wkst.Cells(1,  
cCount).Text  
Loop  
  
End Sub
```