Extracting all text from a powerpoint file in VBA

Asked 9 years, 6 months ago   Active 8 years, 3 months ago   Viewed 12k times

▲

6

▼

🔖

1

🕒

I have a huge set of powerpoint files from which I want to extract all the text and just lump it all into one big text file. Each source (PPT) file has multiple pages (slides). I do not care about formatting - only the words.

I could do this manually with a file by just ^A ^C in PPT, followed by ^V in notepad; then page down in the PPT, and repeat for each slide in the powerpoint. (Too bad I can't just do a ^A that would grab EVERYTHING ... then I could use sendkey to copy / paste)

But there are many hundreds of these PPTs with different numbers of slides.

It seems like this would be a common thing to want to do, but I can't find an example anywhere.

Does anyone have sample code to do this?

vba   text   powerpoint   strip

<div align="right">

edited Jan 19 '11 at 6:51   asked Jan 12 '11 at 23:10

Todd Main   elbillaf

30.9k 10 73 136   **1,596** 7 34 59

</div>

2 Answers

<div align="right">

Active | Oldest | Votes

</div>

▲

3

▼

✓

🕒

Here's some code to get you started. This dumps all text in slides to the debug window. It doesn't try to format, group or do anything other than just dump.

```vba
Sub GetAllText()
Dim p As Presentation: Set p = ActivePresentation
Dim s As Slide
Dim sh As Shape
For Each s In p.Slides
    For Each sh In s.Shapes
        If sh.HasTextFrame Then
            If sh.TextFrame.HasText Then
                Debug.Print sh.TextFrame.TextRange.Text
            End If
        End If
```

```
    Next
End Sub
```

I'll give it a try and get back! – elbillaf Jan 13 '11 at 3:30

Cool. Note that the debug window has a limited amount of text that it can hold. You can kick out the results to a .txt or other file though. – Todd Main Jan 13 '11 at 5:44

Still have problem. Set PPT = CreateObject("PowerPoint.Application") PPT.Visible = True PPT.Presentations.Open filename:=src, ReadOnly:=True Set p = PPT.ActivePresentation For Each s In p.Slides For Each sh In s.Shapes If sh.HasTextFrame Then If sh.TextFrame.HasText Then Print #1, sh.TextFrame.TextRange.Text End If End If Next Next – elbillaf Jan 13 '11 at 23:57 ✎

Powerpoint seems to open, but I can't see the stuff...when I just copy / paste your code it works fine, but I'm processing hundreds (perhaps 1000s) of these files at a time. Not sure how to format this thing in follow-up comments. thanks,tff – elbillaf Jan 14 '11 at 0:00

To see the debug window, you need to open the VBE (Visual Basic Editor) in PowerPoint. Then open the Immediate window (I believe it is Ctrl+G). But if you want to see what is getting kicked out, instead of `Debug.Print` you could also use a `MsgBox`. I'd do that on a small PPT as it will quickly fill up much of the screen in the returned message box. Are you automating this from a different piece of software though? I ask because you're doing `Set PPT = CreateObject("PowerPoint.Application")`, which you wouldn't need to do if you were using the VBE in PowerPoint. – Todd Main Jan 16 '11 at 6:53 ✎

The following example shows code to loop through a list of files based on Otaku's code given above:

1

```
Sub test_click2()

Dim thePath As String
Dim src As String
Dim dst As String
Dim PPT As PowerPoint.Application
Dim p As PowerPoint.Presentation
Dim s As Slide
Dim sh As PowerPoint.Shape
Dim i As Integer
Dim f(10) As String

f(1) = "abc.pptx"
f(2) = "def.pptx"
f(3) = "ghi.pptx"

thePath = "C:\Work\Text parsing PPT\"
```

```
    On Error Resume Next
    Kill dst
    Open dst For Output As #1
      Set PPT = CreateObject("PowerPoint.Application")
      PPT.Activate
      PPT.Visible = True
      'PPT.WindowState = ppWindowMinimized
      PPT.Presentations.Open filename:=src, ReadOnly:=True
      For Each s In PPT.ActivePresentation.Slides
          For Each sh In s.Shapes
              If sh.HasTextFrame Then
                  If sh.TextFrame.HasText Then
                      Debug.Print sh.TextFrame.TextRange.Text
                  End If
              End If
          Next
      Next
      PPT.ActivePresentation.Close
    Close #1
Next i
Set PPT = Nothing

End Sub
```

Great effort. Thanks for sharing your final solution with the community, while still accepting otaku's answer. – froeschli Jan 18 '11 at 22:50