

## Guide - VBA data parsing for Tweets and ANTA

For other guides and resources, visit [ethos.itu.dk](http://ethos.itu.dk). For questions, write [ethos@itu.dk](mailto:ethos@itu.dk).

It may sometimes be the case when using tools such as TCAT or Kimono that one has a large amount of data in an excel sheet, yet wants to be able to analyze it in a program such as ANTA.

If for example you have been running a TCAT query and have thousands of tweets, it may be interesting to run a semantic analysis on the text in the tweets. TCAT does not allow for this by itself, and ANTA cannot analyze texts helpfully if they are all in the same document. With a row per tweet it would take a very long time to extract the individual tweets into .txt files in order to upload them to ANTA.

This guide details a work-around for this problem. It will only work on a machine running Windows, due to a limitation in Mac OS<sup>1</sup>. It works by using a small Visual Basic for Applications (VBA) script, supplied here, and requires that you have Microsoft Excel. Your data must be arrayed in two columns, one with the file name, another with the file content.

### **Step 1: Save your file as a .xlsm file**

VBA scripts only work in .xlsm files, and so you must first save your file in this format. Choose file -> save as. Then from the format drop-down menu, select Excel Macro-Enabled Workbook (.xlsm).

### **Step 2: Load the VBA script**

Choose the 'Developer' tab from the ribbon (Home, Layout, Tables, etc). It may not be visible, in which case you need to make it so by way of your settings/preferences.

In the 'Developer' tab, press 'Editor'. You should now see a pop-up window. Right click on the sheet in which your data is (usually Sheet1), mouse to 'Insert' and choose 'Module'. A new pop-up window should open.

Copy-paste the following text into it and save it:

---

<sup>1</sup> The VBA technically does work on a mac, but will return an error message for all filenames over 32 characters, including the .txt extension. This appears to be a legacy limitation in Mac OS's basic text handling application. In theory you could reduce all file names to the limit and run the script.

18-05-2015

```
“Sub CreateFile()  
Do While Not IsEmpty(ActiveCell.Offset(0, 1))  
    MyFile = ActiveCell.Value & ".txt"  
    'set and open file for output  
    fnum = FreeFile()  
    Open MyFile For Output As fnum  
    'use Print when you want the string without quotation marks  
    Print #fnum, ActiveCell.Offset(0, 1) & " " & ActiveCell.Offset(0, 2)  
Close #fnum  
ActiveCell.Offset(1, 0).Select  
Loop  
End Sub”  
(Remember to remove the quotation marks)
```

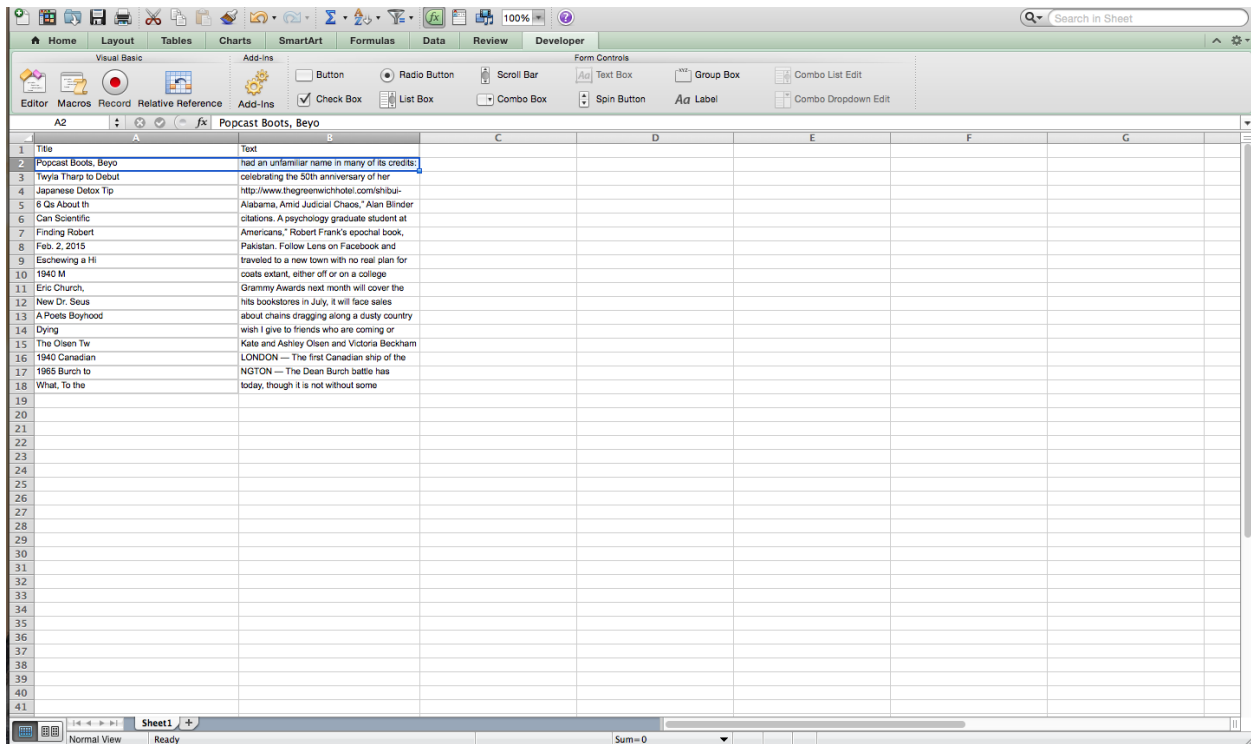
### **Step 3: Clean the data, run the script**

You are almost ready to run the script. Before you do so, check that the titles of your files contain no special characters such as semi-commas, question-marks, slashes, etc. The VBA code cannot handle these.

The data should be aligned next to one another, with the title column to the left of the content column.

In order to run the script, select your two columns (do not select the column headers, just the data). You can just select the two first fields in the top row. Then, still under the 'Developer' ribbon, press 'Macros', select the 'CreateFiles' macro and press 'Run'.

18-05-2015



The script will now automatically create a .txt file for each of the rows and output them to your computer. It should do so to the folder in which the excel sheet is, or a folder in which you have recently worked in excel.

If you encounter issues or errors make sure to check the column containing the file names.

## Step 4: Uploading to ANTA

Depending on your computer settings, you may run into trouble uploading your newly created .txt files to ANTA. ANTA only accepts UTF-8 encoded text files, and the above script may not necessarily produce this, depending on your computer's regional settings. If so use the following program to change the encoding of the files:

<http://sourceforge.net/projects/cp-converter/>

## Conclusion

The above is a work-around for creating text-files. It is admittedly quite clumsy, but it gets the job done. For questions, comments and ideas for improvements, please write [ethos@itu.dk](mailto:ethos@itu.dk).