

Spreadsheet

Barbora Houfkova

The spreadsheet style is modeled like a spreadsheet consisting of columns of data and formulas. When the data changes, the dependent data changes as well.

The main idea behind this style is to spread data into columns and then use the other columns in order to get the requested results. Specifically, the words from the book will be placed in the first column (one per row), the stop words will be placed in the second column (one per row) and the other columns will contain data that resulted from operations on the input data.

Set of columns:

1st : all_words – all words from the book

2nd: stop_words – words from the stop word file

3rd: non_stop_words - all the words that are not stop words

4th: unique_words - all the words that are not stop words without duplicates

5th: counts - number of occurrences – related to the 4th column

6th: sorted_data - sorted data – in a form of tuples – combined 4th and 5th column

The code:

- each column includes a data element and formula – so for example by all_words[0] we mean the actual data – because all_words is a list including two elements
- the first 2 columns are the input data – therefore there are no formulas

In order to define functions, lambda/map/sorted is commonly used. Other interesting functions used: zip, itertools.repeat, operator.itemgetter. In line 34, the function **update** is defined. It is called only once after the data have been loaded in the first and second column. It means that if the second element of column variable is a formula (not None), then apply the function and save the result into the first element.

When importing the data into the first column, the command sys argv was used together with regular expressions. The last command is a loop: for each tuple in the column sorted_data, print word and count.

Discussion:

I like this style of programming because I am familiar with MS Excel. The code is not extensively long. In essence, we start with two main columns. By implying different functions, we get the last column including the requested information. We load the data, call the main function in order to apply the formulas to each column and then we loop over the tuples and print the results. Spreadsheet programming style is suitable for tasks related to data-modification when we need to start with raw data and apply different formulas on it. I chose this style of programming because they use a lot of useful methods such as regular expressions, map, sorted or lambda function.

Sources:

<https://docs.python.org/2/library/operator.html#operator.itemgetter> ITEMGETTER

<https://stackoverflow.com/questions/27930038/how-to-define-global-function-in-python> GLOBAL

<https://stackoverflow.com/questions/4117530/sys-argv1-meaning-in-script> SYS

<https://www.programiz.com/python-programming/methods/built-in/zip> ZIP

https://www.tutorialspoint.com/python/list_count.htm list count

<https://www.programiz.com/python-programming/methods/built-in/sorted> SORTED

<https://www.quora.com/How-do-I-put-multiple-arguments-into-a-map-function-in-Python> map and multiple