

Pytuto (Python Tutorial) 8th assignment.

Hyungwon Yang
Hyung8758@gmail.com
EMCS Lab

This task was designed for those who just started python. I wish them try to complete the task and improve their python skill. Please search as many solutions as you can through the internet and ask questions to me and your teammates to tackle down the tasks. However, DO NOT share codes. Good luck.

Introduction

Regular Expression

Reading materials such as books or magazines that you read everyday contain a lot of information. Thanks to the advanced technology, accessing to those informative repositories is getting easier and this implies that we are able to obtain any information we want. However, in the midst of the flood of knowledge, spotting the information that we truly want is getting harder and this tendency tells us that we need to learn a special skill: finding and reorganizing the data adequately.

Regular expression is one of the powerful tools that supports the users to search specific words or symbols from the enormous reading materials. No matter what reading materials you are dealing with, regular expression will help you out to extract desirable information so that the users can improve their researches by adding more essential data. Most of the computer languages support regular expression and python provides the users with the module named 're'.

Let's look at the simple examples of regular expression.

```
>>> import re
>>> sentence = 'My dog is cute so my mother strokes the dog everyday'
>>> re.findall('dog', sentence)
>>> ['dog', 'dog']
>>>
>>> box = re.search('dog', sentence)
>>> box.group()
>>> 'dog'
```

If you search 're' module on the Internet, you can find useful guides that instruct you how to use regular expression module. Before you search the Internet, check the fundamental rules of 're' module in python.

Recommended websites:

1. <https://docs.python.org/2/library/re.html?highlight=re#module-re>
2. <http://devanix.tistory.com/296>
3. <http://egloos.zum.com/sweeper/v/3065126>

Task

A) Simple Regular Expression Usage

A-1. Text Cleaning

- Import 'YouRaiseMeUP.txt'
- Erase all the commas, periods, dashes and quotation marks. You have to remove all the marks except English letters.
- Save it named 'cleaned_YouRaiseMeUp.txt'

B) Advanced Regular Expression Usage

B-1. Text handling (4 small tasks)

- Import 10 text files. ('context1.txt' to 'context10.txt')
- Combine all the sentences from 10 text files.
- Remove the words that starts with a,b,c such as 'apple' and save it as 'combined_context.txt'
- Reimport all 10 text files again and find words that contain capitalized characters such as 'Grandmother', and 'More' and save them as a text file named 'SelectedWordsList.txt'. This text file lists all capitalized letters.

```
Grandmother
I
When
More
Color
```

List all the capitalized words:

- Count how many number words such as '15' or '16' are in the texts. Print the answer.
- Find the words consisted of 5 letters such as 'paint' and change those letters to '[removed]'
- Ex) 'I never realized that she was there' > 'I [removed] realized that she was [removed]'
- Save it named 'removed_context.txt'.

If you finish your assignment, please send the code to 'hyung8758@gmail.com'. The script name should be '?th_Assignment_YourName.py'. In the code script, the assignment number (e.g, 1st assignment), your name and email address should be written in the first line. Ask questions and give comments to me. It is always welcome.