File Handling

# Before Class

1. Search the Internet for a definition of a computer file. List types of files you know.
2. Watch the video on how to deal with text files:

<https://youtube.com/playlist?list=PLi01XoE8jYohWFPpC17Z-wWhPOSuh8Er->

1. On the w3schools.com platform, do the lessons available in Python Tutorial - File Handling and PythonRegEx sections. Try to follow the examples in the tutorial on your computer.

<https://www.w3schools.com/python/default.asp>

1. Familiarise yourself with regular expression topic:
   1. Regex tutorial: <https://medium.com/factory-mind/regex-tutorial-a-simple-cheatsheet-by-examples-649dc1c3f285>
   2. Regular Expression HOWTO: <https://docs.python.org/3/howto/regex.html>
   3. Regular expressions online: <https://regex101.com/>
   4. Wikipedia: <https://en.wikipedia.org/wiki/Regular_expression>
   5. (In Polish) Wprowadzenie do wyrażeń regularnych:  
      <https://czterytygodnie.pl/wprowadzenie-wyrazen-regularnych-regex/>
2. Copy the following text to the website <https://regex101.com>

Forests cover about 30,5% of Poland's land area based on international standards. Its overall percentage is still increasing. Forests of Poland are managed by the national program of reforestation (KPZL), aiming at an increase of forest-cover to 33% in 2050. The richness of Polish forest (per SoEF 2011 statistics) is more than twice as high as European average (with Germany and France at the top), containing 2.304 billion cubic metres of trees. The largest forest complex in Poland is Lower Silesian Wilderness. More than 1% of Poland's territory, 3,145 square kilometres (1,214 sq mi), is protected within 23 Polish national parks. Three more national parks are projected for Masuria, the Polish Jura, and the eastern Beskids. In addition, wetlands along lakes and rivers in central Poland are legally protected, as are coastal areas in the north. There are over 120 areas designated as landscape parks, along with numerous nature reserves and other protected areas (e.g. Natura 2000).

Then create regular expressions they indicate in the text:

* 1. All words ‘Poland’ – (Poland)
  2. Country names (Poland, Germany and France) - (Poland|France|Germany)
  3. Punctuation marks (dots and commas) – (\.|\, )
  4. Numbers representing a year (four-digit numbers) - \d{4}
  5. Capital letters - [A-Z]
  6. Vowels - (a|e|i|o|u) , [aeiou]
  7. Words with at least five letters. - \w{5}
  8. Words starting with capital letters - \b[A-Z]\w\*

# During Class

## Reading from file

1. In any text editor (e.g. Windows Notepad), create a file countries.txt in which save, in separate lines, names of five countries. Then create a program that displays the file content.

file = open('countries.txt','r')  
file\_content = file.read()  
print( file\_content )  
file.close()

1. Create a program that displays the contents of the countries.txt text file. At the beginning of each line, display the line number. Tip: you have to read and display text file line by line.

# display text file, line by line  
...   
for line in file:  
 print(line, end="")  
...

1. Find any text file on the Internet and download it to your computer. Then write a program that displays its contents.
2. In any text editor, create a file numbers.txt in which save, in separate lines, integer numbers. Then write a program that reads numbers from the numbers.txt file and calculates their sum. Tip: Read the next line from the file and convert it into a numeric value.

## Writing to file

1. Create a program that saves, in separate lines, your name and surname, university name and field of study in a text file. Tip: open a file in writing mode and then use the write() method.
2. An array film\_titles contains any five film titles. Write a program that writes the film titles to a text file, each title on a separate line.
3. Create a program that allows you to add a name of next product you want to buy at the end of the text file shopping.txt. Enter the product name from the keyboard. Tip: open the file in appending mode.

# After Class

1. The following program displays the contents of a file, line by line:

f = open("filename.txt")  
for line in f:  
 print(line, end="")  
f.close()

Rewrite the program using the "with ..." as construct. Then check that the program is working properly.

1. Write a program that calculates the number of lines for any text file. The user enters the name of the file from the keyboard. Display the result of the calculation (the file name and the number of lines). Do not display the contents of the file. Sample result:

File name: countries.txt  
Number of lines: 14

1. Find any text file on the Internet that contains at least 30 lines of text and download that file to your computer. Then write a program that displays the first five lines from the file and then waits for the Enter key to be pressed. Then repeat displaying the next five lines from the file, waiting for the Enter key to be pressed each time.
2. Find any text file on the Internet and download it to your computer. Then write a program that copies the contents of this file to the copy.txt file. Copy the contents of the file as a whole. Finally, open both files in any text editor and check that their contents are the same.
3. Find any text file on the Internet and download it to your computer. Then write a program that copies the contents of this file to the copylines.txt file. Copy the contents of the file line by line. Finally, open both files in any text editor and check that their contents are the same.
4. Using any text editor, create the following two text files:

MeatAndFish.txt

Skinless white meat  
Tuna fish  
Luncheon meat  
Lean cuts of red meat

GrainsAndBread.txt

Bread  
Rice  
All purpose flour  
Breakfast cereal  
Pasta

Then write a program that creates a shoppinglist.txt file, in which save the contents of the MeatAndFish.txt and the GrainsAndBread.txt files.

1. Create a program that writes to a text file integer numbers in the range of <1,50>, every number in a separate line.
2. Create a program that writes 50 random integers between 100 and 999 to a text file, each number on a separate line.
3. Create a program that saves to a text file, numbers in the range of <1,10> with their second and third power. Sample result:

1,1,1  
2,4,8  
3,9,27  
4,16,64  
…

1. In any text editor, create a text file students.txt containing the following data in CSV format:

first\_name,last\_name,age,gender,email  
Decca,Blackstone,52,Male,dblackstone0@time.com  
Elena,Rechert,27,Female,erechert1@ucoz.com  
Bibbye,Norree,26,Female,bnorree2@reddit.com  
Alasdair,McCoole,53,Male,amccoole3@foxnews.com  
Hogan,Hatrey,26,Male,hhatrey4@zimbio.com

Then create a program that reads data from the CSV file and displays the first name, last name and email address of students under 30. Format the data as below. Sample result:

Elena Rechert erechert1@ucoz.com  
Bibbye Norree bnorree2@reddit.com  
Hogan Hatrey hhatrey4@zimbio.com

Tip: import and use csv module.

1. The announcement regarding the temperature forecast in degrees Celsius for the next three days was published on the Internet:

"Tuesday: 22C, Wednesday: 21C, Thursday: 26C "

Create a program that calculates the average temperature. Use regular expressions to extract the values of temperatures from the message.

import re  
  
message = "Tuesday: 22C, Wednesday: 21C, Thursday: 26C "  
temperatures = re.findall("\d{2}",message)  
# complete the program code  
# ...

1. Write a program that calculates the number of vowels in the text:

To be, or not to be, that is the question

Use regular expressions and the findall() method.

1. Write a program that computes the number of words in the following text. Use regular expressions.

To be, or not to be, that is the question

1. Find any text file on the Internet and download it to your computer. Then write a program that displays all words with at least six letters from the file. Display each word on a separate line. Use regular expressions.
2. The grades.txt file contains student’s grades. Create the file in any text editor.

Name: Peter  
Grades: 3.5, 4.0, 5.0, 4.5, 3.5, 3.0, 5.0

Then create a program that calculates the arithmetic mean of student’s grades.