# Midterm 2

To answer this question, you need to use Python **functions**, **file handlers,** and **modules**. You need to split the codes related to different functions into different files and import them as modules to a **main.py** file. So, the only file that I run to see the result is the **main.py** and that file should call other modules as they fit. In the following paragraphs, I have suggested how the modules could look like. Those are only suggestions and you can follow different names or a different number of modules, but your code must be 1) modularized and 2) **at least should have 4 modules**, not including the **main.py. (so you need to submit at least 5 files).**

I will run the code from the command line (not from IDE). The main.py file gets a **text file** as an **argument**. Let's assume the name of the file is file.txt. Your code processes the content of the file.txt so I will run your code as it's shown in the following command line:

$ python main.py file.txt

The content of file.txt is shown below (I have uploaded this file to blackboard). Please consider that you are not supposed to change the format or fix the typos in the file manually, rather if you want to do so, you have to do it programmatically (by Python code). Those are there intentionally put in the file to be fixed by your code.

> Python is easy to learn and enjoyable. Like other languages, Python provides a complete set of control flow elements, including while and for loops, and conditionals. .
> In python,  the indent is important !  Things can easily go wrong if you do not follow indentation.
> PythoN uses the level of indentation to group blocks of code with control elements.

The code gives a report that shows:

1) the number of words in the file.txt (1 Mark)

2) the number of times each word in the file.txt occurs without considering the case, so Python and python are the same in your code (I suggest you create the **counter** module to do this job for you) (1 Mark)

3) which word(s) are the most common in the document (suggest to create the **most_common** module to do this job) (1 Mark)

4) which words are the least common (the **least_common** module). (1 Mark)

5) Please note that you may need to clean up the content of the file programmatically (**not manually**) before you start. For example, you may want to convert everything to lower case (the **lower_case**

module) and delete all the punctuations (the **punch_remover** module). You may even want to break the text into a series of words (if you choose to do that, call it the **splitter** module).

> Python is easy to learn and enjoyable. Like other languages, Python provides a complete set of control flow elements, including while and for loops, and conditionals. .
> In python, the indent is important ! Things can easily go wrong if you do not follow indentation.
> PythoN uses the level of indentation to group blocks of code with control elements.

Be careful about extra whitespaces. For example, in the second paragraph I deliberately put two spaces between "**python,"** and "**the indent**". So just relying on the whitespace to find the words is not a reliable solution and it gives the wrong output. Also "do not" count on the punctuations. For example, I intentionally put two periods at the end of the first paragraph (conditionals. .) or I used ! in the second paragraph. (4 Marks)

6) At the end, your code needs to create a folder and saves the report inside a file (you pick the name of the file). This can be done in the report module. (2 Marks)