# Homework Assignment 3: for loop, string processing

|  |  |
| --- | --- |
| Assigned date | 2024-09-20 |
| Due date | 2024-10-04 |
| Estimate required time | 120 minutes |

This is an individual assignment.

* You may consult with professor and TA about any aspect of the assignment.
* You may consult with other students only in a general way, e.g., about debugging or Python issues, or questions about wording on the assignment.
* You cannot actively work with someone unless the assignment specifically grants permission to work together with another student.

Remember to follow the CMPT140/166 Coding Standard.

## Purpose

Nowadays, coding skills are not reserved for programmers or software engineers. Perhaps our traditional view of someone who does coding is someone who builds software for other people to use repeatedly such as someone who develops operating systems or Word processing software or even mobile applications. However, with coding environment becoming increasingly accessible, coding (or more accurately, scripting) skills have become a skill set similar to the ability to operate spreadsheet software. In this scenario, the programmer writes scripts to help him/her achieve tasks required by his/her work such as looking through Excel files to find various data records or renaming files in a systemic way. In these use-cases, the end goal/product of the software development is not the script itself but the task the script is designed to accomplish. The scripts the programmer writes may only be executed once. However, it is still worth the effort as tasks accomplished by scripts avoid random human errors. And if the volume of the input is substantial (e.g. if we need to go through hundreds of files), the task may not be doable manually in a practical and timely manner.

In this homework assignment, you will get a taste of what these custom scripts look like.

You are given a list of ID’s and are asked to change its format. The task you are doing here is sometimes called “data cleaning”. In our perspective, the details or the background knowledge (e.g. what does these ID’s represent) does not matter that much. However, having some information about the background often would help us appreciate the task and also may help us with software design decisions.

The ID’s you are given are ID’s of human specimens for cancer studies. These are pieces of tumor tissue taken out of patients when they had surgery to remove their tumors. These pieces of tumor tissue are not used for treatment of disease or diagnosis and would otherwise be thrown away. However, they are invaluable for medical research as they allow scientists to gain understanding on the nature/behavior of these cancers. We need to change the ID format so that it would be compatible with a new database system. It would be error-prone and time consuming if we do this by hand. The logic is a bit too complicated for this to be done in Excel (via Excel formula). So, you decided to use Python to help you with this task.

## Specification

ID format change needed is shown by the following example:

"VOA1392a" -> "VOA1392@A"

"Voa1392A" -> "VOA1392#A"

We can assume the last character is always an alphabet.

If the last character is lower case, replace it with @[upper case of last character]

If the last character is upper case, replace it with #[upper case of last character]

Make sure “VOA” is always all upper case.

In your script, please do the following

1. Reformat the ID’s according to the description above.
2. Store the newly formatted ID’s in a list.
3. Iterate through this list and print the content of the list to console i.e. using print() statements. In real life scenario, you would write the content of the list to a file. However, we did not cover writing to files using Python yet. So, this will be for another day!

Please download the template Python script from course website (learn.twu.ca): “Homework Assignment 3”. The template has more details regarding specific aspects of this assignment.

If you are using online/cloud-based Python IDE, you will need a way to save the specimen id file to the cloud. Please see the file “pythonanywhere upload and read files.docx” for instructions on how to do so for pythonanywhere.

## Deliverables

Create one python file named hw1-LastnameFirstname.py (e.g. hw1-LeungSamuel.py) where Lastname and Firstname are student’s last and first name respectively. Please ...

Please submit your source code file hw1-LastnameFirstname.py at <http://learn.twu.ca>

Grading scheme for codes:

* Documentation (20%)
* Correct execution (60%)
* Correct style and structure (20%)

## Hints

1. We will be using the Python library Pandas to read in an Excel file. The codes to load the library, call the corresponding function (read\_excel) to read the Excel file as well as the for-loop to iterate (i.e. go through) all lines of an Excel file is given in the template file already. However, you will have to do the following
   1. Make sure (and possibly replace) the file name “specimen\_ids.xlsx” (this is the input to the function read\_excel) so that it refers to the file you are pointing to. If your file is in the same folder as your script and your file is named exactly “specimen\_ids.xlsx”, you do not need to change anything.
   2. Add in your codes inside the for loop (for i in range(len(row\_content)) ... ) to do the ID formatting.
2. The following code extract the last character of a string:

A black screen with white text

Description automatically generated

1. The following code removes the last character of a string:

A black screen with white text

Description automatically generated

1. The following code check if ALL letters in a string is upper case

A screen shot of a computer code

Description automatically generated

1. The following code generates a string that is all upper case

A black background with white text

Description automatically generated