

**Assignment 2 [10 points]:**

*Instructions:* This homework is an individual-base. Each of you is expected to submit programs that solve each of the following problems. **Due date of submissions is on Saturday, 13/11/2021.**

Write python program code that help scientists in identifying DNA pattern for different species, define the similarities between different DNA strands for stored species DNA. Identify whether the DNA of interest is a new strand or not. Your program keeps DNA file names for each specie in a dictionary in the form:

SpeciesDict= {"specieA": "DNA1.txt", "specieB": "DNA2.txt", ..}.

Each file contains a DNA strand which consists of series of A, T, C, G nucleotides.

*Example:*

DNA1.txt contains: AAAAGCCTGACTTTGCTAAATGCATTTTGCCAGGAATCG

DNA2.txt contains: AACCGGTTACGATTCGGGAAAAGCCGCACCAGGCTTGA

Your program should contain the following functions:

- A) Write a function *getFile(sName)*: that accepts specie name and return its file path, the function should return not found if the specie name does not exist.
- B) Write a function *retDNA(DNAfile)*: that accepts file path and print on screen the DNA strand inside it. This function should use Exceptions to handle wrong paths.
- C) Write a function *addDNA(specieName)*: that accepts new spice name, read its DNA sequence from the user and save it in a new DNA file. Then this function should call the *addSpecie* function to add the spice name and its file path to species dictionary.

- D) Write a function *addSpecie (specieName)*: that accepts specie name to add new specie name and its file path to species dictionary.
- E) Write function *compare(specie1, specie2)*: accepts two species to find DNA sequence similarities in their DNA files. A match is considered for patterns with sizes greater than or equal to 4 letters (size>=4), and they are located at the same positions in both files.

*Example:*

DNA1.txt contains: AAAAGCCTGACTTTGCTAAATGCATTTGCCAGGAATCG

DNA2.txt contains: AACCGTTACGATTCGGGAAAAGCCGCACAGGCTTGA

- As you can see from the example the first pattern (AAAAGCC) is not counted because they are not at the same locations. But the next pattern (CCAGG) is counted because they are at the same positions at position 29. Note: this function should call *getFile* function to retrieve the file path of a certain specie.
- F) Write *main()* function that display a proper menu for users on screen. And read their choices, and call the proper functions:
1. Print DNA for a specie
  2. Add new specie
  3. Compare two species

## Submission Instructions & Grading

Be sure that you have submitted the following two item:

- Submit the following sheet on MyCourses' respective assignment before the deadline. it includes the rubrics that will be used for grading the assignment.

GCIS123 - Software Dev. & Problem Solving I		
Assignment 2		
CRITERIA	MAXIMUM GRADE	STUDENT GRADE
GETFILE	1	
RETDNA	2	
ADDDNA	1	
ADDSPECIE	1	
COMPARE	3	
MAIN	1	
USING COMMENTS	1	
<b>TOTAL</b>	<b>10</b>	