National Taiwan University, Department of Electronic Engineering

Data Structure, Spring 2020 Assignment #2

1 Problem Statement

In this homework, you’re going to implement regular expression matching. If we get a pattern *pattern*, we want to see if the pattern matches any input string *text*. The input string *text* could be empty and contains only lowercase letters $a - z$, and the pattern could be empty and contains lowercase letters $a - z$, and characters like ’.’ and ’*’. Rules for pattern are as below:

1. ’.’ Matches any single character.
2. ’*’ Matches zero or more of the preceding element.

It is worth noting that the pattern should match the **entire input text**. Below are some examples:

Example 1 :

text = ”aa”
pattern = ”a”
Output: false
Explanation: ”a” does not match the entire text ”aa”.

Example 2 :

text = ”aa”
pattern = ”a*”
Output: true
Explanation: ’*’ means zero or more of the preceding element, ’a’. Therefore, by repeating ’a’ once, it becomes ”aa”.

Example 3 :

text = ”ab”
pattern = ”.*”
Output: true
Explanation: ”.*” means ”zero or more (*) of any character (.)”.

Example 4 :

text = ”aab”
pattern = ”c*a*b”
Output: true
Explanation: c can be repeated 0 times, a can be repeated 1 time. Therefore, it matches ”aab”.

Example 5 :

text = ”mississippi”
pattern = ”mis*is*p*.”
Output: false

2 Input/Output Specification

Your should input two strings (text and pattern), and output a bool that returns **True** if the pattern matches the text given, and **False** if not.

3 Evaluation

We have provided a code file **main.py**. You have to fill in the function **pattern_match(***text***,** *pattern***)** where *text* is the input text and *pattern* is the pattern that we want to match. Write your codes in **TODO**.We will import your function from your code and evaluate it using our own datasets. Note that you should write your program **as efficient as possible!** Codes with time complexity **more than O(text size * pattern size) will get only get partial points!** We will test your codes with input string lengths to up to 3000 characters.

4 Submission

Please put your codes (including main.py or any other code files) into a directory named **studentID** and compress the directory into studentID.zip and upload studentID.zip to ceiba. The homework is due on **4/23**, at **4:00 am**.