**Natural Language Processing**

(International & Graduate students)

Homework 1 Regular Expressions and Word Segmentation idea

Due Jan. 30, 2024 via LEB2

The homework has two parts. Part 1 is programming using Python. Part 2 is to explain your idea how to segment Thai texts into words.

Part 1: Programming using Python. Import re in Python and write regex to do the following. Show your code, explanation and test run. (5 points)

1. Extract all email addresses from a string

Test string = "The email addresses are be-mansouri@maine.edu and behrooz.mansouri@cs.maine.edu"

1. Extract all phone numbers from a string

"The phone numbers are 207-587-3290, 5583446854, (207) 107-9293, 2075142279x0000"

Part 2: The text below is from ***zebra question*** by Shel Silverstein. First you should try to read and understand the text yourself. Answer by giving **your idea** how to teach the computer to segment this text into correct words **as complete as possible**. You can write your answer as pseudo-code or algorithm. (5 points)

IaskedthezebraAreyoublackwithwhitestripes?Orwhitewithblackstripes?Andthezebra askedme,Areyougoodwithbadhabits?Orareyoubadwithgoodhabits?Areyounoisywith quiettimes?Orareyouquietwithnoisytimes?Areyouhappywithsomesaddays?Orareyou sadwithsomehappydays?Areyouneatwithsomesloppyways?Orareyousloppywith someneatways?AndonandonandonandonAndonandonhewentI'llneveraskazebra  
Aboutstripesagain.