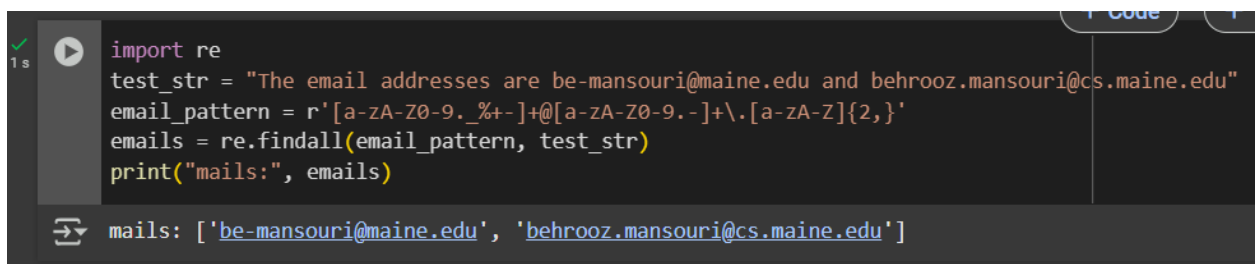


Part 1: Programming using Python. Import re in Python and write regex to do the following. Show your code, explanation and test run. (5 points)

1. Extract all email addresses from a string

```
import re
test_str = "The email addresses are be-mansouri@maine.edu and behrooz.mansouri@cs.maine.edu"
email_pattern = r'[a-zA-Z0-9._%+-]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,}'
emails = re.findall(email_pattern, test_str)
print("mails:", emails)
```

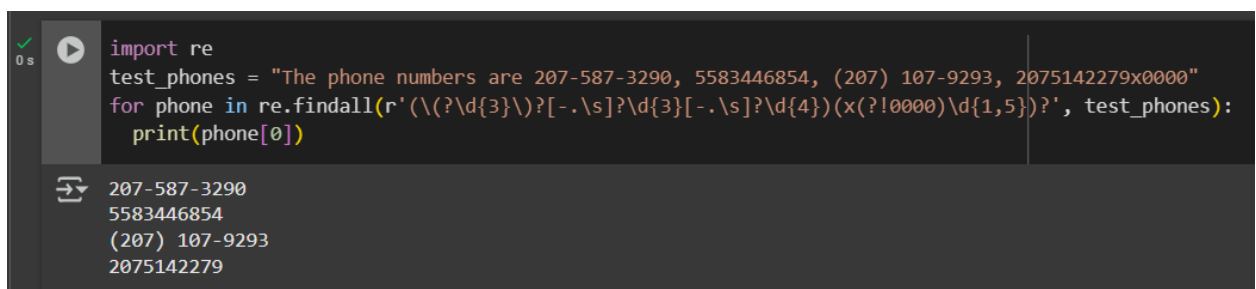


```
import re
test_str = "The email addresses are be-mansouri@maine.edu and behrooz.mansouri@cs.maine.edu"
email_pattern = r'[a-zA-Z0-9._%+-]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,}'
emails = re.findall(email_pattern, test_str)
print("mails:", emails)
```

mails: ['be-mansouri@maine.edu', 'behrooz.mansouri@cs.maine.edu']

2. Extract all phone numbers from a string

```
import re
test_phones = "The phone numbers are 207-587-3290, 5583446854, (207) 107-9293, 2075142279x0000"
for phone in re.findall(r'(\(?\d{3}\)?[-.\s]\d{3}[-.\s]\d{4})(x(?:0000)\d{1,5})?', test_phones):
    print(phone[0])
```



```
import re
test_phones = "The phone numbers are 207-587-3290, 5583446854, (207) 107-9293, 2075142279x0000"
for phone in re.findall(r'(\(?\d{3}\)?[-.\s]\d{3}[-.\s]\d{4})(x(?:0000)\d{1,5})?', test_phones):
    print(phone[0])
```

207-587-3290
5583446854
(207) 107-9293
2075142279

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)

laskedthezebraAreyoubblackwithwhitestripes?Orwhitewithblackstripes?Andthezebra
askedme,Areyougoudwithbadhabits?Orareyoubadwithgoodhabits?Areyounoisywith
quiettimes?Orareyouquietwithnoisytimes?Areyouhappywithsomesaddays?Orareyou
sadwithsomehappydays?Areyouneatwithsomesloppyways?Orareyousloppywith
someneatways?AndonandonandonandonAndonandonhewentI'llneveraskzebra
Aboutstripesagain.

= I asked the zebra Are you black with white stripes? Or white with black stripes? And the zebra asked me, Are you good with bad habits? Or are you bad with good habits? Are you noisy with quiet times? Or are you quiet with noisy times? Are you happy with some sad days? Or are you sad with some happy days? Are you neat with some sloppy ways? Or are you sloppy with some neat ways? And on and on and on and on And on and on he went I'll never ask a zebra About stripes again.

I think to learn a computer to do the job of separate words to get a readable one, we need to give a dictionary of knows words to it. It will do like I did: read each letter until I see a word that I know, add space and go to the next one.

In NLP approach, maybe we can use a model who predict words with the one who is the most accurate.