

NATURAL LANGUAGE PROCESSING - WORKSHEET 2

All the questions in this worksheet have one or more than one correct answers. Choose all the correct options to answer the questions:

1	Consider the below string:	
1.	"please mail me at nitin12@gmail.com"	
	Which of the following patterns can capture the mail id in above string?	
	A) '.*@[a-z]*.com '	B) '[a-z]*@[a-z]*.com'
	C) '[/w]*@[/w]*.[/w]*'	D) '[/w]+com'
2.		/ E =
۷٠	A) '*'	B) '+'
	C) '?'	D) '{'
2	Which of the following captures a pattern having	<i>/</i> (
3.		
	A) '@[/w]{4}'	B) '@.{4}'
4	C) '@[/w]{1,4}'	D) '@. {0,4}
4.		2017/10/28/mexican-grand-prix-2017-time-does-start-tv-
	channel-odds-lewisl/2017/05/12" Which of the following regexp patterns can be used to extract date from the above url?	
		B) '^/[/d]{4)/[/d]{2}/[/d]{2}'
	, , , , , , , , , , , , , , , , , , , ,	D) None of the above
5.	Which of the following meta-sequence is to match all alphanumeric characters?	
	A) /w	B) /d
	C) /s	D) /m
6.	Which of the following regexp pattern which would extract all the hashtags from the below string?	
	String = "sachin will love to play cricket at #lords in #ICCcricketworldcup #2k15"	
	Import re	
	re.findall(pattern, String)	
	A) pattern="#\w+"	B) pattern="#[A-z]*"
	C) pattern= '#[A-z0-9]+'	D) None of them
7.	Which of the following regexp pattern which wo	ould extract all the mentions (for example @aakash,
	@nk_154) from the below string?	
	String = "I would like to thank @akshay_154,	@nitin12, @asthaMishra "
	Import re re.findall(pattern, String)	
	A) pattern="@[A-z]*"	B) pattern="@[A-z]+"
	C) pattern= '@[A-z0-9]+'	D) pattern= '@\w+'
8.	Which of the following operator is used to mark	
٠.	A) *	B) ^
	C) &	D) None of them
9.	Which of the following functions match the pattern only at the beginning of the string?	
	A) re.match()	B) re.search()
	C) re.findall()	D) All of the above
10	Which of the following is same as "*" operator?	D) All of the above
10.	A) {0,}	B) {1,}
		D) {3,}
11	C) {0,2} Which of the following meta seguences represent	
11.	Which of the following meta-sequences represer	
	A) \w	B) \s
	C) \d	D) \D



12. Which distribution do the frequency of the words in a large document follow?

A) Normal Distribution

B) Zipf Distribution

C) F-Distribution

D) Chi-square

13. Which of the following words cannot be reduced to their base words by stemming (PorterStemmer, Lancaster etc.) correctly?

A) eating

B) worse

C) slept

D) running

14. Suppose we want to Replace Road with rd.

street = '21 Ramakrishna Road'

Which of the following statements can be used in python to do the task?

A) re.sub('Road', 'Rd', street)

B) re.sub('Rd', 'Road', street))

C) re.sub(street, 'Rd')

D) None of the above

15. What will be the output of the following lines of code?

import re

re.search("aabbbbbb", "ab{3,5}?")

- A) < re.match object; span = (1, 5), match = 'abbb'>
- B) < re.match object; span = (1, 8), match = 'abbb'>
- C) < re. match object; span = (1, 3), match = 'abbb'>
- D) < re.match object; span = (1, 7), match = 'abbb'>

