Worksheet 2

NLP

1. Consider the below string:

“please mail me at nitin12@gmail.com”

Which of the following patterns can capture the mail id in above string?

Answer: A) ‘.\*@[a-z]\*.com ‘ C) '[/w][\*@[/w]\*.[/w]\*](mailto:*@[/w%5d*.%5b/w%5d*)'

2. Which of the following is an quatifier in regular expressions in python?

Answer: A) ‘\*’ B) ‘+’

3. Which of the following captures a pattern having @ symbol followed by 4 alphabets?

Answer: A) ‘@[/w]{4}’ B) ‘@.{4}’

4. url = **“http://www.telegraph.co.uk/formula-1/2017/10/28/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?

Answer: A) '/(\d{4})/(\d{1,2})/(\d{1,2})/'

5. Which of the following meta-sequence is to match all alphanumeric characters?

Answer: A) /w

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)**

Answer: A) pattern="#\w+" C) pattern= '#[A-z0-9]+'

7. Which of the following regexp pattern which would 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)**

Answer: C) pattern= '@[A-z0-9]+' D) pattern= ‘@\w+’

8. Which of the following operator is used to mark the start of the string in regular expressions?

Answer: B)^

9. Which of the following functions match the pattern only at the beginning of the string?

Answer: A) re.match()

10. Which of the following is same as “\*” operator?

Answer: A) {0,}

11. Which of the following meta-sequences represent the digits?

Answer: C) \d

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

Answer: B) Zipf Distribution

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

Answer: C) slept

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?

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

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

**import re**

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

Answer**:** C) <re.match object; span = (1, 3), match = 'abbb'>