**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:**

**“please mail me at nitin12@gmail.com”**

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

C) '[/w]\*@[/w]\*.[/w]\*'

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

A) ‘\*’ B) ‘+’ C) ‘?’

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

B) ‘@.{4}’ D) ‘@.{0,4}

**4. url = “http://www.telegraph.co.uk/formula-1/2017/10/28/mexican-grand-prix-2017-time-does-start-tvchannel-odds-lewisl/2017/05/12”**

**Which of the following regexp patterns can be used to extract date from the above url?**

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

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

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)**

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)**

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?**

B) ^

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

A) re.match()

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

A) {0,}

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

C) \d

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

B) Zipf Distribution

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

B) worse 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?**

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}?")**

A) <re.match object; span = (1, 5), match = 'abbb'>