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](mailto: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’

**Ans: B) '[a-z]\*@[a-z]\*.com'**

1. Which of the following is a quantifier in regular expressions in python?

A) ‘\*’ B) ‘+’

C) ‘?’ D) ‘{‘

**Ans: B) ‘+’**

1. Which of the following captures a pattern having @ symbol followed by 4 alphabets? A) ‘@[/w]{4}’ B) ‘@.{4}’

C) ‘@[/w]{1,4}’ D) ‘@.{0,4}

**Ans**: **A) ‘@[/w]{4}’**

## url = [“h](http://www.telegraph.co.uk/formula-1/2017/10/28/mexican-grand-prix-2017-time-does-start-tv-)t[tp://www.telegraph.co.uk/formula-1/2017/10/28/mexican-grand-prix-2017-time-does-start-tv-](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? A) '/(\d{4})/(\d{1,2})/(\d{1,2})/' B) ‘^/[/d]{4)/[/d]{2}/[/d]{2}’

C) ‘/[0-9]{4}/[0-9]{2}/[/d]{2}’ D) None of the above

**Ans**:

1. Which of the following meta-sequence is to match all alphanumeric characters?
   1. /w B) /d

C) /s D) /m

**Ans**: **A) /w**

1. 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)

* 1. pattern="#\w+" B) pattern="#[A-z]\*"

C) pattern= '#[A-z0-9]+' D) None of them

**Ans**:

1. 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)**

* 1. pattern="@[A-z]\*" B) pattern="@[A-z]+"

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

**Ans**:

1. Which of the following operator is used to mark the start of the string in regular expressions?
   1. \* B) ^

C) & D) None of them

**Ans: B) ^**

1. Which of the following functions match the pattern only at the beginning of the string?
   1. re.match() B) re.search()

C) re.findall() D) All of the above

**Ans: A)re.match(), C)re.findall()**

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

A) {0,} B) {1,}

C) {0,2} D) {3,}

**Ans: A) {0,} C) {0,2}**

1. Which of the following meta-sequences represent the digits?
   1. \w B) \s

C) \d D) \D

**Ans: C) \d**

1. Which distribution do the frequency of the words in a large document follow?
   1. Normal Distribution B) Zipf Distribution

C) F-Distribution D) Chi-square

**Ans: B) Zipf Distribution D) Chi-square**

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

C) slept D) running



**Ans: B) worse**

1. 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?

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

C) re.sub(street, 'Rd') D) None of the above

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

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

## import re

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

* 1. <re.match object; span = (1, 5), match = 'abbb'>
  2. <re.match object; span = (1, 8), match = 'abbb'>
  3. <re.match object; span = (1, 3), match = 'abbb'>
  4. <re.match object; span = (1, 7), match = 'abbb'>

**Ans:**