

Unit 2 - Week 1

Course outline

How to access the portal?

Week 1

- Introduction
- Operations on a Corpus
- Probability and NLP
- Vector Space models
- Sequence Learning
- Machine Translation
- Preprocessing
- Statistical Properties of Words - Part 01
- Statistical Properties of Words - Part 02
- Statistical Properties of Words - Part 03
- Week 1 Lecture Materials
- Quiz : Assignment 1
- Week 1 Feedback : Applied Natural Language Processing
- Assignment 1 : Programming Exercise Self Assessment

Week 2

Week 3

Week 4

Week 5

Week 6

Week 7

Week 8

Week 9

Week 10

Week 11

Week 12

DOWNLOAD VIDEOS

Assignment 1

The due date for submitting this assignment has passed. As per our records you have not submitted this assignment.

Due on 2019-08-14, 23:59 IST.

Check all the right answers

1) Document preprocessing steps are

1 point

- Tokenization
- Substitution
- Normalization
- Feature Selection

No, the answer is incorrect.

Score: 0

Accepted Answers:

Tokenization

Substitution

Normalization

2) Pick the stemming actions

1 point

- was, am, are, is → be
- helped, helps → help
- troubled, troubling, trouble → trouble
- friend, friendship, friends, friendships → friend
- studied → studi
- All of the above

No, the answer is incorrect.

Score: 0

Accepted Answers:

helped, helps → help

studied → studi

3) Pick the lemmatization actions

1 point

- was, am, are, is → be
- study, studying, studied → study
- troubled, troubling, trouble → trouble
- has have, had → have
- All of the above

No, the answer is incorrect.

Score: 0

Accepted Answers:

All of the above

4) Case folding is used for _____

1 point

- Normalization
- Tokenization
- Stemming
- Lemmatization

No, the answer is incorrect.

Score: 0

Accepted Answers:

Normalization

5) Consider a corpus with 100000 documents. The word boon occurs in some documents (say, 200) with the following frequency:

1 point

$$TF_{d_1} = \frac{25}{127}, TF_{d_2} = \frac{3}{250}, TF_{d_3} = \frac{20}{650}, TF_{d_9} = \frac{15}{125} \text{ and } TF_{d_{1000}} = \frac{20}{800}$$

If the total number of words in the corpus is 100000, then arrange the documents according to the rank in the ascending order using TF*IDF for the word boon

- $[d_{1000}, d_1, d_2, d_9, d_3]$
- $[d_2, d_{1000}, d_3, d_9, d_1]$
- $[d_{1000}, d_2, d_3, d_9, d_1]$
- None of the above

No, the answer is incorrect.

Score: 0

Accepted Answers:

$[d_2, d_{1000}, d_3, d_9, d_1]$

6) The start of the sentence is matched by _____ and the end of the sentence is matched by _____

1 point

- ^ and \$
- \$ and ^
- \\$ and .
- \^ and \.

No, the answer is incorrect.

Score: 0

Accepted Answers:

^ and \$

7) Consider the following corpus of 4 documents:

1 point

Documents	Terms
D_1	NLP is an interesting subject
D_2	Many students are interested in learning NLP
D_3	ANN plays an important role in NLP applications
D_4	Do you play tennis?

The TF*IDF for the word NLP for D_1, D_2, D_3, D_4 is

- $[\frac{1}{5} \frac{1}{7} \frac{1}{8} 0] \log_{10}(\frac{3}{4})$
- $[\frac{1}{6} \frac{1}{7} \frac{1}{8} 0] \log_{10}(\frac{3}{4})$
- $[\frac{1}{5} \frac{1}{7} \frac{1}{8} 0] \log_{10}(\frac{4}{3})$
- None of the above

No, the answer is incorrect.

Score: 0

Accepted Answers:

$[\frac{1}{5} \frac{1}{7} \frac{1}{8} 0] \log_{10}(\frac{4}{3})$

8) A college student wants to toss a textbook to his roommate who is leaning out of a window directly above him. He throws the book upwards with an initial velocity of **8.0 m/s**. The roommate catches it while it is traveling at **3.0 m/s** [up]. (a) How long was the book in the air? (b) How far vertically did the book travel? A car accelerates in a straight line from rest at the rate of **2.3 m/s²**. What is its final velocity after 55 m? What is its time?

Check all the regular expressions that will find the terms highlighted in bold and underlined?

- $[d]? \setminus [d]^+ \setminus [a-z]^+ \setminus V[a-z]$
- $[d]? \setminus [d]^+ \setminus sm \setminus V$
- $[0-9]^+ \setminus [0-9]^+ \setminus [a-z]^+ \setminus V$
- $[d]? \setminus [d]^+ \setminus [a-z]^+ \setminus V[a-z]$

No, the answer is incorrect.

Score: 0

Accepted Answers:

$[d]? \setminus [d]^+ \setminus [a-z]^+ \setminus V[a-z]$

$[0-9]^+ \setminus [0-9]^+ \setminus [a-z]^+ \setminus V$

$[d]? \setminus [d]^+ \setminus [a-z]^+ \setminus V[a-z]$

9) Consider the following frequency rank table. Assuming the value of β as 3.0, what is the approximate value of α so that the Mandelbrot's rule satisfies the relationship of rank and frequency for the terms "Mr" and "very"

1 point

Word	Frequency	Rank
for	1321	24
have	1301	25
is	1220	26
with	1187	27
Mr	1153	28
very	1151	29

- 0.045
- 0.055
- 0.55
- None of the above

No, the answer is incorrect.

Score: 0

Accepted Answers:

0.055