

Natural Language Processing B.tech VI Semester Section-1

u19cs039@coed.svnit.ac.in [Switch account](#)



Draft saved

Your email will be recorded when you submit this form

* Required

MCQ and Short Answers

*

spelling rules are used to model the changes that occur in a word, usually when two morphemes combine is known as

- a. Morphotactics
- b. Orthographic rules
- c. Lexicon rules
- d. Finite-state transducer

☐ a

☐ b

☐ c

☒ d



*

How many trigrams phrases can be generated from the following sentence, after performing following text cleaning steps:

- Stopword Removal
- Replacing punctuations by a single space

" <S> #The grass is always-greener on the other side of the @fence </S>"

A) 3

B) 4

C) 5

D) 6

E) 7

☐ A

☐ B

☐ C

☒ D

☐ E



For the below question, In <your roll number> consider your own roll number.
Calculate Bigram Probabability of "NLP </s>".

Consider a mini-corpus:

<s> <your roll number> is a student of SVNIT </s>

<s> Has <your roll number> opted for NLP</s>

<s> Machine learning is must for NLP </s>

1/3

*

Which of the below is true for stemming:

- a. map from foxes to fox; but might need to know that foxes is plural.
- b. map from foxes to foxs;
- c. map from foxes to fox; but might not need to also know that foxes is plural.
- d. map from foxes to foxes only;

☒ a

☐ b

☐ c

☐ d



For the below question, In <your roll number> consider your own roll number.
Calculate Bigram Probabability of "<s> <your roll number>".

Consider a mini-corpus:

<s> <your roll number> is a student of SVNIT </s>

<s> Has <your roll number> opted for NLP</s>

<s> Machine learning is must for NLP </s>

|

2/3



*

Tag	Description	Example	Tag	Description	Example
CC	Coordin. Conjunction	<i>and, but, or</i>	SYM	Symbol	<i>+, %, &</i>
CD	Cardinal number	<i>one, two, three</i>	TO	"to"	<i>to</i>
DT	Determiner	<i>a, the</i>	UH	Interjection	<i>ah, oops</i>
EX	Existential 'there'	<i>there</i>	VB	Verb, base form	<i>eat</i>
FW	Foreign word	<i>mea culpa</i>	VBD	Verb, past tense	<i>ate</i>
IN	Preposition/sub-conj	<i>of, in, by</i>	VBG	Verb, gerund	<i>eating</i>
JJ	Adjective	<i>yellow</i>	VCN	Verb, past participle	<i>eaten</i>
JJR	Adj., comparative	<i>bigger</i>	VBP	Verb, non-3sg pres	<i>eat</i>
JJS	Adj., superlative	<i>wildest</i>	VBZ	Verb, 3sg pres	<i>eats</i>
LS	List item marker	<i>1, 2, One</i>	WDT	Wh-determiner	<i>which, that</i>
MD	Modal	<i>can, should</i>	WP	Wh-pronoun	<i>what, who</i>
NN	Noun, sing. or mass	<i>llama</i>	WP\$	Possessive wh-	<i>whose</i>
NNS	Noun, plural	<i>llamas</i>	WRB	Wh-adverb	<i>how, where</i>
NNP	Proper noun, singular	<i>IBM</i>	\$	Dollar sign	<i>\$</i>
NNPS	Proper noun, plural	<i>Carolinas</i>	#	Pound sign	<i>#</i>
PDT	Predeterminer	<i>all, both</i>	"	Left quote	<i>' or "</i>
POS	Possessive ending	<i>'s</i>	"	Right quote	<i>' or "</i>
PRP	Personal pronoun	<i>I, you, he</i>	(Left parenthesis	<i>[, (, {, <</i>
PRP\$	Possessive pronoun	<i>your, one's</i>)	Right parenthesis	<i>],), }, ></i>
RB	Adverb	<i>quickly, never</i>	,	Comma	<i>,</i>
RBR	Adverb, comparative	<i>faster</i>	.	Sentence-final punc	<i>. ! ?</i>
RBS	Adverb, superlative	<i>fastest</i>	:	Mid-sentence punc	<i>: ; ... --</i>
RP	Particle	<i>up, off</i>			

Use the Penn Treebank tagset to tag each word in the following sentence.

He is a short, skinny guy with a long, happy, mean-looking face.

- He/PRP is/VBZ a/DT short/JJ ,/, skinny/JJ guy/NN with/IN a/DT long/JJ ,/, happy/JJ ,/, mean-looking/JJ face/NN
- He/PRP is/VBZ a/DT short/JJ ,/, skinny/JJ guy/NN with/IN a/DT long/JJ ,/, happy/JJ ,/, mean-looking/JJ face/JJ
- He/PRP is/VBZ a/DT short/JJ ,/, skinny/JJ guy/PRP with/IN a/DT long/JJ ,/, happy/JJ ,/, mean-looking/JJ face/NN
- He/PRP is/VBZ a/DT short/JJ ,/, skinny/JJ guy/PRP with/IN a/DT long/JJ ,/, happy/JJ ,/, mean-looking/JJ face/JJ

- ☐ a
- ☐ b
- ☐ c
- ☒ d



*

Write regular expressions for the following language:
the set of all strings with two consecutive repeated words

- a. $([a-zA-Z1-9]^+ " ^+[a-zA-Z1-9])$
- b. $([a-zA-Z]^+)\backslash s^+ " ^+([a-zA-Z]^+)\backslash s$
- c. $([a-zA-Z]^+)\backslash s^+\backslash 1$
- d. $([a-zA-Z]^+ " ^+[a-zA-Z])$

- ☐ a
- ☐ b
- ☒ c
- ☐ d



*

Which of the below is an example of stemming operation

1. was, am, are, is → be
2. helped, helps → help
3. troubled, troubling, trouble → trouble
4. friend, friendship, friends, friendships → friend
5. studied → studi
6. All of the above

A. 2, 3 & 4

B. 3 & 5

C. 6

D. 2, 3, 4 & 5

E. 2 & 5

☐ A

☐ B

☐ C

☒ D

☐ E



*

Which of the given models can be used for the purpose of document similarity?

- A) Training a word 2 vector model on the corpus that learns context present in the document
- B) Training a bag of words model that learns occurrence of words in the document
- C) Creating a document-term matrix and using cosine similarity for each document
- D) All of the above

- ☐ A
- ☐ B
- ☐ C
- ☒ D



*

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

A. ^ and \$

B. \$ and ^

C. \ \$ and .

D. \ ^ and \ .

☐ A

☒ B

☐ C

☐ D



*

In a HMM, the possible state transitions are from state JJ to states NN, VB, JJ and RB. Following are the known state transitions probabilities;

$$P(NN|JJ) = 1/4, P(VB|JJ) = 1/6, \text{ and } P(JJ|JJ) = 1/4.$$

What is the transition probability value of $P(RB|JJ)$?

a) $1/4$

b) $1/2$

c) $1/5$

d) $1/3$

☐ a

☐ b

☐ c

☒ d



*

The edit distance between two strings can be zero In which of the following cases?

a) b) When the lengths of the two strings are equal

b) When the two strings are equal

c) The edit distance can never be zero

d) When one string is a substring of another

☐ a

☒ b

☐ c

☐ d



*

Which of the below statements are true

1. The raw Co-occurrence matrices are usually sparse
 2. one hot-vector a dense vector.
- A. Only 1
- B. Only 2
- C. Both 1 & 2
- D. None of the them

- ☐ A
- ☐ B
- ☒ C
- ☐ D

What will be the cost of converting the word strength to trend using the minimum edit distance technique *

- ☐ 5
- ☐ 6
- ☐ 7
- ☒ 4
- ☐ 3



N-grams are defined as the combination of N keywords together. How many bi-grams can be generated from a given sentence:

" <S> The grass is always greener on the other side of the fence </S> "

A) 10

B) 13

C) 11

D) 12

☐ A

☒ B

☐ C

☐ D

Clear selection

For the below question, In <your roll number> consider your own roll number. Calculate Unigram Probabability of your roll number.

Consider a mini-corpus:

<s> <your roll number> is a student of SVNIT </s>

<s> Has <your roll number> opted for NLP </s>

<s> Machine learning is must for NLP </s>

1/3



*

Which of the below is/are the examples of inflectional morphemes

☐ ation

☒ s

☒ ed

☒ ing

*

“ I saw bats ” is which type of ambiguity ?

A. Semantics

B. Lexical

C. Syntactic

D. None of the above

☐ A

☒ B

☐ C

☐ D



For the below question, In <your roll number> consider your own roll number.
Calculate Unigram Probabability of word "NLP".

Consider a mini-corpus:

<s> <your roll number> is a student of SVNIT </s>

<s> Has <your roll number> opted for NLP</s>

<s> Machine learning is must for NLP </s>

|

2/17

What will be the cost of converting the word strength to trend using the minimum edit distance technique *

- ☐ 5
- ☐ 6
- ☐ 7
- ☒ 4
- ☐ 3



*

You have created a document term matrix of the data, treating every tweet as one document. Which of the following is correct, in regards to document term matrix?

1. Converting all the words in lowercase will not affect the dimensionality of the data
2. Normalization of words in the data will reduce the dimensionality of data
3. Removal of stopwords from the data will affect the dimensionality of data

A) Only 1

B) Only 2

C) Only 3

D) 1 and 2

E) 2 and 3

F) 1, 2 and 3

☐ A

☐ B

☐ C

☐ D

☒ E

☐ F



*

Tag	Description	Example	Tag	Description	Example
CC	Coordin. Conjunction	<i>and, but, or</i>	SYM	Symbol	<i>+, %, &</i>
CD	Cardinal number	<i>one, two, three</i>	TO	"to"	<i>to</i>
DT	Determiner	<i>a, the</i>	UH	Interjection	<i>ah, oops</i>
EX	Existential 'there'	<i>there</i>	VB	Verb, base form	<i>eat</i>
FW	Foreign word	<i>mea culpa</i>	VBD	Verb, past tense	<i>ate</i>
IN	Preposition/sub-conj	<i>of, in, by</i>	VBG	Verb, gerund	<i>eating</i>
JJ	Adjective	<i>yellow</i>	VCN	Verb, past participle	<i>eaten</i>
JJR	Adj., comparative	<i>bigger</i>	VBP	Verb, non-3sg pres	<i>eat</i>
JJS	Adj., superlative	<i>wildest</i>	VBZ	Verb, 3sg pres	<i>eats</i>
LS	List item marker	<i>1, 2, One</i>	WDT	Wh-determiner	<i>which, that</i>
MD	Modal	<i>can, should</i>	WP	Wh-pronoun	<i>what, who</i>
NN	Noun, sing. or mass	<i>llama</i>	WP\$	Possessive wh-	<i>whose</i>
NNS	Noun, plural	<i>llamas</i>	WRB	Wh-adverb	<i>how, where</i>
NNP	Proper noun, singular	<i>IBM</i>	\$	Dollar sign	<i>\$</i>
NNPS	Proper noun, plural	<i>Carolinas</i>	#	Pound sign	<i>#</i>
PDT	Predeterminer	<i>all, both</i>	"	Left quote	<i>' or "</i>
POS	Possessive ending	<i>'s</i>	"	Right quote	<i>' or "</i>
PRP	Personal pronoun	<i>I, you, he</i>	(Left parenthesis	<i>[, (, {, <</i>
PRP\$	Possessive pronoun	<i>your, one's</i>)	Right parenthesis	<i>],), }, ></i>
RB	Adverb	<i>quickly, never</i>	,	Comma	<i>,</i>
RBR	Adverb, comparative	<i>faster</i>	.	Sentence-final punc	<i>. ! ?</i>
RBS	Adverb, superlative	<i>fastest</i>	:	Mid-sentence punc	<i>: ; ... --</i>
RP	Particle	<i>up, off</i>			

Using the above Penn Treebank tagset, which of the following Word - POS TAG pair is incorrect.

. . . I/PRP am/VBP sitting/VBG in/IN Mindy/NNP 's/POS restaurant/NN putting/VBG on/RP the/DT gefillte/NN fish/NN ,/, which/WDT is/VBZ a/DT dish/NN I/PRP am/VBP very/RB fond/VB of/RP ,/

- a. gefillte/NN
- b. fond/VB
- c. is/VBZ
- d. on/RP

- ☐ a
- ☒ b
- ☐ c
- ☐ d



*

Which is/are adj-root1 that would include adjectives that can occur with un- and -ly

- ☐ Big
- ☒ Clear
- ☒ Happy
- ☐ Small

*

“ Priya loves her mother, and Sita does too ” is which type of ambiguity ?

- A. Semantics
 - B. Lexical
 - C. Syntactic
 - D. None of the above
-
- ☒ A
 - ☐ B
 - ☐ C
 - ☐ D



*

What language would below regular expression match:

`\bRam\b.*\bKrishna\b|\bKrishna\b.*\bRam\b`

- a. all strings that have both the word Ram and the word Krishna in them
- b. all strings that have both the word Ram and the word Krishna in them but Not Ram that merely contain the word Ram
- c. all strings that have both the word Ram and the word Krishna in them including Ram that merely contain the word Ram
- d. all strings that have both the word Ram and the word Krishna in them but Not Ram that merely contain the word Krishna

☒ a☐ b☐ c☐ d[Back](#)[Submit](#)[Clear form](#)

Never submit passwords through Google Forms.

This form was created inside of Sardar Vallabhbhai National Institute of Technology, Surat. [Report Abuse](#)

Google Forms

