

## CONFIDENTIAL

<b>Name of Examination</b> <small>(Please tick, symbol is given)</small>	:	MID		END	✓	SUPPLE	
<b>Name of the School</b> <small>(Please tick, symbol is given)</small>	:	SOE		SOCS	✓	SOP	
<b>Programme</b>	:	B.Tech.					
<b>Semester</b>	:	7 <sup>th</sup> semester					
<b>Name of the Course</b>	:	Natural Language Processing					
<b>Course Code</b>	:	CSEG-415					
<b>Name of Question Paper Setter</b>	:	Bikram Pratim Bhuyan					
<b>Employee Code</b>	:	40001825					
<b>Mobile &amp; Extension</b>	:	9854350562					
<b>Note: Please mention additional Stationery to be provided, during examination such as Table/Graph Sheet etc. else mention “NOT APPLICABLE”: NOT APPLICABLE</b>							
<b>FOR SRE DEPARTMENT</b>							
<b>Date of Examination</b>	:						
<b>Time of Examination</b>	:						
<b>No. of Copies (for Print)</b>	:						

Name:

Enrolment No:



**UNIVERSITY OF PETROLEUM AND ENERGY STUDIES**

**End Semester Examination, December 2018**

**Course: Natural Language Processing (CSEG-415)**

**Semester: 7<sup>th</sup>**

**Programme: B.Tech.**

**Time: 03 hrs.**

**Max. Marks: 100**

**SECTION A**

**All questions in SECTION A are compulsory**

S. No.		Marks	CO
Q 1	What do you mean by Natural Language Processing? Mention some areas where NLP is applied.	4	CO1
Q 2	What are the challenges of NLP? What do you mean by supervised learning?	4	CO1
Q 3	What do you mean by function and content words? Give examples. If we read Tom Sawyer, who dominated as the most frequent words?	4	CO3
Q 4	If first corpus has $TTR1 = 0.013$ and second corpus has $TTR2 = 0.13$ , where $TTR1$ and $TTR2$ represents type/token ratio in first and second corpus respectively; then what can you say about both of the corpus? Explain your suggestion.	4	CO3
Q 5	In the sentence, "In Dehradun I took my hat off. But I can't put it back on."; compute the total number of word tokens and word types. Bigram models are what ordered Markov Models?	4	CO3

**SECTION B**

**(Q 6, 7, 8 are compulsory. Attempt Q9A or Q9B)**

Q 6	Given the following sentences: "I want to eat. I want to sing. I eat Chinese." If you are following the bigram model; what is the probability of the following sentence: "I want to eat Chinese"? Also compute the probability of the following sentence: "I want to sing and eat "?	10	CO3
Q 7	In Vector Space Model, suppose we have two sentences bear the words; S1:<man, eat, eat>; S2:<man, eat, chicken, chicken>; S3:<man, eat, chicken>. Find the cosine and Jaccard similarity between S1 and S3.	10	CO3
Q 8	How is the sigmoid model related to probability? What is the range of the sigmoid function $S(X)$ ? Simulate the 'OR' function using a basic neural network without weights. What should be the threshold?	10	CO2
Q 9A	"I made her duck". What are the possible interpretations that you can make out from the statement? If some indices are inserted in a max-heap. What is the complexity of finding the minimum element? Explain the Hidden Markov Model related to NLP with examples.	10	CO3
Q 9B	For text compression in NLP we use the Huffman coding technique. Given the following sentences: "I want to eat. I want to sing. I eat Chinese. He too want to eat Chinese. I want to sing and eat." Give the Huffman tree. Compute in ratio how much text was compressed using the technique.	10	CO3

<p style="text-align: center;"><b>SECTION-C</b>  <b>(Q 10 is compulsory. Attempt Q11A or Q11B)</b></p>			
Q 10	<p>Consider the following context-free grammar:</p> <p style="margin-left: 40px;"> S → NP VP  NP → Det N  VP → V  VP → V NP  N → dog  N → cat  N → mouse  Det → the  V → sees  V → hates  V → sneezes </p> <p>Which of the following sentences are recognized by this grammar and why?</p> <p style="margin-left: 40px;"> i) The dog sneezes the cat  ii) The mouse hates  iii) The cat the mouse hates  iv) The mouse hates the mouse  v) The mouse sees the mouse </p>	20	CO3
Q 11A	<p>Consider a simple three-state Markov model of the weather. Any given day, the weather can be described as being</p> <ul style="list-style-type: none"> <li>• State 1: precipitation (rain or snow) • State 2: cloudy • State 3: sunny</li> </ul> <p>Transitions between states are described by the transition matrix</p> $A = \{a_{ij}\} = \begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$ <p style="margin-left: 40px;"> a) Draw the state transition graph.  b) Given that the weather on day t=1 is sunny, what is the probability that the weather for the next 7 days will be “sun, sun, rain, rain, sun, clouds, sun”?  c) What is the probability that the weather stays in the same known state Si for exactly T consecutive days? </p>	20	CO4
Q 11B	<p>We seek to classify documents as being about sports or not. Each document is associated with a pair (x, y), where x is a feature vector of word counts of the document and y is the label for whether it is about sports (y = 1 if yes, y = 0 if false). The vocabulary is size 3, so feature vectors look like (0, 1, 5), (1, 1, 1), etc.</p>	20	CO4

Consider a naive Bayes model with the following conditional probability table:

word type	1	2	2
$P(w \mid y = 1)$	1/10	2/10	7/10
$P(w \mid y = 0)$	5/10	2/10	3/10

and the following prior probabilities over classes:

$P(y = 1)$	$P(y = 0)$
4/10	6/10


Consider the document with counts  $x = (1, 0, 1)$ .

- Which class has highest posterior probability?
- What is the posterior probability that the document is about sports?

## CONFIDENTIAL

<b>Name of Examination</b> <small>(Please tick, symbol is given)</small>	:	MID		END	✓	SUPPLE	
<b>Name of the School</b> <small>(Please tick, symbol is given)</small>	:	SOE		SOCS	✓	SOP	
<b>Programme</b>	:	B.Tech.					
<b>Semester</b>	:	7 <sup>th</sup> semester					
<b>Name of the Course</b>	:	Natural Language Processing					
<b>Course Code</b>	:	CSEG-415					
<b>Name of Question Paper Setter</b>	:	Bikram Pratim Bhuyan					
<b>Employee Code</b>	:	40001825					
<b>Mobile &amp; Extension</b>	:	9854350562					
<b>Note: Please mention additional Stationery to be provided, during examination such as Table/Graph Sheet etc. else mention “NOT APPLICABLE”: NOT APPLICABLE</b>							
<b>FOR SRE DEPARTMENT</b>							
<b>Date of Examination</b>	:						
<b>Time of Examination</b>	:						
<b>No. of Copies (for Print)</b>	:						

**SET: 2**

Name:			
Enrolment No:			
<p align="center"><b>UNIVERSITY OF PETROLEUM AND ENERGY STUDIES</b></p> <p align="center"><b>End Semester Examination, December 2018</b></p> <p> <b>Course: Natural Language Processing (CSEG-415)</b> <span style="float: right;"><b>Semester: 7<sup>th</sup></b></span> </p> <p> <b>Programme: B.Tech.</b> <span style="float: right;"><b>Max. Marks: 100</b></span> </p> <p><b>Time: 03 hrs.</b></p>			
<p align="center"><b>SECTION A</b></p> <p align="center"><b>All questions in SECTION A are compulsory</b></p>			
S. No.		Marks	CO
Q 1	Mention some areas where NLP is applied. What are the challenges of NLP?	4	CO1
Q 2	I made her duck. What are the possible interpretations that you can make out from the statement? If some indices are inserted in a max-heap. What is the complexity of finding the minimum element?	4	CO1
Q 3	What do you mean by function and content words? Give examples. If we read Tom Sawyer, who dominated as the most frequent words?	4	CO3
Q 4	If first corpus has $TTR1 = 0.059$ and second corpus has $TTR2 = 0.59$ , where $TTR1$ and $TTR2$ represents type/token ratio in first and second corpus respectively; then what can you say about both of the corpus? Explain your suggestion.	4	CO3
Q 5	In the sentence, "India is my homeland. I will not give up on it." compute the total number of word tokens and word types. Bigram models are what ordered Markov Models?	4	CO3
<p align="center"><b>SECTION B</b></p> <p align="center"><b>(Q 6, 7, 8 are compulsory. Attempt Q9A or Q9B)</b></p>			
Q 6	Given the following sentences: "I want to eat. I want to sing. I eat Chinese." If you are following the bigram model; what is the probability of the following sentence: "I want to eat Chinese"? Also compute the probability of the following sentence: "I want to sing and eat "?	10	CO3
Q 7	For text compression in NLP we use the Huffman coding technique. Given the following sentences: "I want to eat. I want to sing. I eat Chinese. He too want to eat Chinese. I want to sing and eat." Give the Huffman tree. Compute in ratio how much text was compressed using the technique.	10	CO3
Q 8	How is the sigmoid model related to probability? What is the range of the sigmoid function $S(X)$ ? Simulate the 'OR' function using a basic neural network without weights. What should be the threshold?	10	CO2
Q 9A	Explain Naive Bayes and Hidden Markov Models related to NLP.	10	CO2
Q 9B	What are the different types of learning? Explain each type with examples.	10	CO2

**SECTION-C**  
**(Q 10 is compulsory. Attempt Q11A or Q11B)**

Q 10	<p>Consider the following productions:</p> <p><math>S \rightarrow NP VP</math>  <math>NP \rightarrow NP PP</math>  <math>NP \rightarrow \text{sushi}</math>  <math>NP \rightarrow I</math>  <math>NP \rightarrow \text{chopsticks}</math>  <math>NP \rightarrow \text{you}</math>  <math>VP \rightarrow VP PP</math>  <math>VP \rightarrow \text{Verb NP}</math>  <math>\text{Verb} \rightarrow \text{eat}</math>  <math>PP \rightarrow \text{Prep NP}</math>  <math>\text{Prep} \rightarrow \text{with}</math>  Where;  NP – noun phrase  VP –verb phrase  PP -preposition phrase.</p> <p>a) Use the CYK parsing algorithm to find if the sentence "I eat sushi with chopsticks with you" belongs to the above grammar.  b) Explain the CYK algorithm.</p>	20	CO3
Q 11A	<p>Consider a simple three-state Markov model of the weather. Any given day, the weather can be described as being</p> <ul style="list-style-type: none"> <li>• State 1: precipitation (rain or snow)</li> <li>• State 2: cloudy</li> <li>• State 3: sunny</li> </ul> <p>Transitions between states are described by the transition matrix</p> $A = \{a_{ij}\} = \begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$ <p>d) Draw the state transition graph.  e) Given that the weather on day t=1 is sunny, what is the probability that the weather for the next 7 days will be “sun, sun, rain, rain, sun, clouds, sun”?  f) What is the probability that the weather stays in the same known state <math>S_i</math> for exactly T consecutive days?</p>	20	CO4
Q 11B	<p>We seek to classify documents as being about sports or not. Each document is associated with a pair (x, y), where x is a feature vector of word counts of the document and y is the label for whether it is about sports (y = 1 if yes, y = 0 if false). The vocabulary is size 3, so feature vectors look like (0, 1, 5), (1, 1, 1), etc. Consider a naive Bayes model with the following conditional probability table:</p>	20	CO4

word type	1	2	2
$P(w \mid y = 1)$	1/10	2/10	7/10
$P(w \mid y = 0)$	5/10	2/10	3/10

and the following prior probabilities over classes:

$P(y = 1)$	$P(y = 0)$
4/10	6/10

Consider the document with counts  $x = (1, 0, 1)$ .

- c) Which class has highest posterior probability?
- d) What is the posterior probability that the document is about sports?