# UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
## End Semester Examination, May 2019

**Course: Natural Language Processing**
**Semester:  II**
**Program: MTech in Computer Science**                          Time 03 hrs.
**Course Code: CSAI7006**                                        Max. Marks: 100

## SECTION A

| | | | |
|---|---|---|---|
| Q 1 | Discuss four applications of Natural Language Processing in brief? | **4** | **CO1** |
| Q 2 | Explain Lexical Diversity. | **4** | **CO2** |
| Q 3 | Explain frequency distribution. | **4** | **CO2** |
| Q 4 | Explain collocations and bigrams. | **4** | **CO3** |
| Q 5 | Discuss in short about<br>　i)　　Dispersion plot<br>　ii)　　Hapaxes | **4** | **CO1** |

## SECTION B

| | | | |
|---|---|---|---|
| Q 6 | Explain the basic stages involved in Natural Language Processing? | **10** | **CO2** |
| Q 7 | Analyze and explain the concept of transformation based tagging? Illustrate how is it different with other automatic tagging techniques? | **10** | **CO5** |
| Q 8 | With the help of regular expression in Python, write the code to filter out all the numbers from the given text? | **10** | **CO3** |
| Q 9 | Describe any three automatic tagger available in NLTK? Also write their usage example.<br><br>OR<br><br>Demonstrate how morphological and syntactic clues can be useful in determining the category of the given word. | **10** | **CO5** |

## SECTION-C

| | | | |
|---|---|---|---|
| Q 10 | Given the dataset of 'n' documents on news. Apply the concept of NLP to classify them into different tags. Discuss the features and classifier used. Also elaborate the method for evaluating your proposed method.<br><br>OR | **20** | **CO5** |

| | | | |
|---|---|---|---|
| | Given the dataset of 'n' documents on inaugural speeches by different prime ministers of India.Apply the concept of NLP to classify any given unknown document (generated from speech) into its respective speaker. Discuss the features and classifier used. Also describe the method for evaluating your proposed method. | | |
| Q 11 | Create a pattern for automatically tagging given sentence into at least five tags. Also write the Python program for automatically tagging given document using the created pattern. | **20** | **CO4** |

## UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
### End Semester Examination, May 2019

**Course: Natural Language Processing**
**Semester:  II**
**Program: M Tech in Computer Science**                    **Time 03 hrs.**
**Course Code: CSAI7006**                              **Max. Marks: 100**

<table>
<tr><td colspan="3" align="center"><b>SECTION A</b></td></tr>
<tr><td></td><td></td><td></td><td></td></tr>
<tr><td>Q 1</td><td>Mention four challenges to Natural Language Processing in brief.</td><td>4</td><td>CO1</td></tr>
<tr><td>Q 2</td><td>Code a program in Python to find the lexical diversity of the given text.</td><td>4</td><td>CO2</td></tr>
<tr><td>Q 3</td><td>Show the Python code for displaying 10 most frequent words from any of the NTLK corpus.</td><td>4</td><td>CO2</td></tr>
<tr><td>Q 4</td><td>Show Python code for finding collocations and creating bigrams?</td><td>4</td><td>CO3</td></tr>
<tr><td>Q 5</td><td>Explain in brief:-<br>   i)      textonyms<br>   ii)     Hapaxes</td><td>4</td><td>CO1</td></tr>
<tr><td colspan="4" align="center"><b>SECTION B</b></td></tr>
<tr><td>Q 6</td><td>Explain Stopwords from the context of NLP. Write the Python code for filtering out all the stopwords from the given text.</td><td>10</td><td>CO2</td></tr>
<tr><td>Q 7</td><td>Analyze the concept of POS tagging or simply tagging in Natural Language Processing. Illustrate any such five tags along with their examples.</td><td>10</td><td>CO5</td></tr>
<tr><td>Q 8</td><td>With the help of regular expression in Python, write the code to find out all the 8 letter word with 'j' at its 6<sup>th</sup> place and 't' at its 8<sup>th</sup> place from the given text.</td><td>10</td><td>CO3</td></tr>
<tr><td>Q 9</td><td>Analyze the concept of n-gram tagger in difference to Unigram tagger. Illustrate how to combine the two taggers to get the best results.<br><div align="center">OR</div>Demonstrate how semantic and syntactic clues can be useful in determining the category of the given word.</td><td>10</td><td>CO5</td></tr>
<tr><td colspan="4" align="center"><b>SECTION-C</b></td></tr>
<tr><td>Q 10</td><td>Apply the concept of natural language processing for the problem of summarizing the content of given document. The document should have atleast 8 paragraphs.<br><br><div align="center">OR</div><br><br>Apply the concept of Natural Language Processing for evaluating the answer given by a student for the given question. Consider the answer to be single sentence.</td><td>20</td><td>CO5</td></tr>
</table>

| | | | |
|------|------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|-------|
| Q 11 | Analyze how the web content can be useful from the perspective of Natural Language Processing. How to extract information from the web using Python? | **20** | **CO4** |