

## NATURAL LANGUAGE PROCESSING WITH PYTHON

Course Code 12IS5B6

L:T:P:S 3:0:0:0

Credits 3

CIE Marks 100

SEE Marks 100

SEE Duration 3 Hrs

### Unit-I

#### Introduction to Language Processing and Python

Computing with Language: Texts and Words, A Closer Look at Python: Texts as Lists of Words Computing with Language: Simple Statistics, Back to Python: Making Decisions and Taking Control, Automatic Natural Language Understanding

6 Hrs

#### Accessing Text Corpora and Lexical Resources

Accessing Text Corpora, Conditional Frequency Distributions, More Python: Reusing Code, Lexical Resources, WordNet

### Unit-II

#### Processing Raw Text

Accessing Text from the Web and from Disk, Strings: Text Processing at the Lowest Level

Text Processing with Unicode, Regular Expressions for Detecting Word Patterns, Useful Applications of Regular Expressions, Normalizing Text ,Regular Expressions for Tokenizing Text, Segmentation, Formatting: From Lists to Strings

7 Hrs

#### Categorizing and Tagging Words

Using a Tagger, Tagged Corpora, Mapping Words to Properties Using Python Dictionaries Automatic Tagging, N-Gram Tagging, Transformation-Based Tagging, How to Determine the Category of a Word

### Unit-III

**Learning to Classify Text** Supervised Classification, Further Examples of Supervised Classification, Evaluation, Decision Trees, Naive Bayes Classifiers, Maximum Entropy Classifiers, Modeling Linguistic Patterns

6 Hrs

#### Extracting Information from Text

Information Extraction, Chunking, Developing and Evaluating Chunkers, Recursion in Linguistic Structure ,Named Entity Recognition, Relation Extraction

### Unit-IV

#### Analyzing Sentence Structure

Some Grammatical Dilemmas, What's the Use of Syntax?, Context-Free Grammar, Parsing with Context-Free Grammar, Dependencies and Dependency Grammar, Grammar Development

7 Hrs

#### Building Feature-Based Grammars

Grammatical Features, Processing Feature Structures, Extending a Feature-Based Grammar

## **Unit-V**

### **Analyzing the Meaning of Sentences**

Natural Language Understanding, Propositional Logic, First-Order Logic, The Semantics of English Sentences, Discourse Semantics

**7 Hrs**

### **Managing Linguistic Data**

Corpus Structure: A Case Study, The Life Cycle of a Corpus ,Acquiring Data, Working with XML

### **Course Outcome**

- 1 Define the terms used in NLP and scope of NLP
- 2 Explain the need and working of different components of a working NLP tools
- 3 Apply different techniques available in the development of the components in NLP
- 4 Analyze the behavior of each component based on the parameters
- 5 Develop the different components using appropriate structures and techniques

### **References**

- 1 Steven Bird, Ewan Klein and Edward Loper, "Natural Language Processing with Python", O'Reilly Publication, First Edition, ISBN 978-0-596-51649-9
- 2 James Allen – Natural Language Understanding, Pearson Education, 2<sup>nd</sup> Edition, ISBN: 978-81-317-0895-8, 1995
- 3 Christopher D. Manning Foundations of Statistical Natural Language Processing, The MIT Press; 1<sup>st</sup> edition, ISBN: 0-262-13360-1, 1999
- 4 KaviNarayana Murthy - "Natural Language Processing - An Information Access Perspective", EssEss Publications, 1<sup>st</sup> Edition, ISBN: 81-7000-485-3, 2006

### **Scheme of Continuous Internal Evaluation for Theory**

CIE consists of Three Tests each for 45 marks (15 marks for Quiz + 30 marks for descriptive) out of which best two will be considered. In addition 10 marks to be earned through assignment or seminar on emerging topics.

### **Scheme of Semester End Evaluation—Theory**

1. Question No. 1 consisting of objective type/short type questions, it is compulsory and it carries 20 marks, covering the entire syllabus.
2. There are five units. Each unit will have two questions of 16 marks each, students have to answer one question from each unit.