****

**BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI**

**WORK INTEGRATED LEARNING PROGRAMMES**

**COURSE HANDOUT**

**Part A: Content Design**

|  |  |
| --- | --- |
| **Course Title** | Natural Language Processing |
| **Course No(s)** |  |
| **Credit Units** | 3 units |
| **Course Author** | Prof. Vijayalakshmi and Dr. Chetana Gavankar |
| **Version No** | 4.0 |
| **Date** | August 2020 |

**Course Objectives**

|  |  |
| --- | --- |
| **No** | **Course Objective** |
| **CO1** | To learn the fundamental concepts and techniques of natural language processing (NLP) |
| **CO2** | To learn computational properties of natural languages and the commonly used algorithms for processing linguistic information |
| **CO3** | To apply NLP techniques in state of art applications |
| **CO4** | To learn implementation of NLP algorithms and techniques |

**Text Book(s)**

|  |  |
| --- | --- |
| T1 | Speech and Language processing: An introduction to Natural Language Processing, Computational Linguistics and speech Recognition by Daniel Jurafsky and James H. Martin[3rd edition] |
| T2 | Natural language understanding[2nd edition] by James Allen |

**Reference Book(s) & other resources**

|  |  |
| --- | --- |
| R1 | Handbook of Natural Language Processing, Second Edition—NitinIndurkhya, Fred J. Damerau, Fred J. Damerau |
| R2 | Natural Language Processing with Python by Steven Bird, Ewan Klein, Edward Lopper |

**Modular Content Structure**

1. **Introduction to Natural Language Understanding**

1.1 The Study of Language.

1.2 Applications of Natural Language Understanding.

1.3 Evaluating Language Understanding Systems.

1.4 The Different Levels of Language Analysis.

1.5 Representations and Understanding.

1.6 The Organization of Natural Language Understanding Systems.

**2. N-gram Language Models**

2.1 N-Grams

2.2 Evaluating Language Models

2.3 Generalization and Zeros

2.4 Smoothing

2.5 The Web and Stupid Backoff

**3.** **Part-of-Speech Tagging**

3.1 (Mostly) English Word Classes

3.2 The Penn Treebank Part-of-Speech Tag set

3.3 Part-of-Speech Tagging

3.4 Markov Chains

3.5 The Hidden Markov Model

3.6 HMM Part-of-Speech Tagging

3.7 Part-of-Speech Tagging for Morphological Rich Languages

**4**. **Hidden Markov Models**

4.1 The Hidden Markov Model

4.2 Likelihood Computation: The Forward Algorithm

4.3 Decoding: The Viterbi Algorithm

4.4 HMM Training: The Forward-Backward Algorithm

4.5 Maximum Entropy Markov Models   
 4.6 Bidirectionality

**5. Grammars and Parsing.**

5.1 Grammars and Sentence Structure.

5.2 What Makes a Good Grammar

5.3 A Top-Down Parser.

5.4 A Bottom-Up Chart Parser.

5.5 Top-Down Chart Parsing.

5.6 Finite State Models and Morphological Processing.

5.7 Grammars and Logic Programming.

5.8 Parsing

**6. Statistical Constituency Parsing**

6.1 Probabilistic Context-Free Grammars

6.2 Probabilistic CKY Parsing of PCFGs

6.3 Ways to Learn PCFG Rule Probabilities

6.4 Problems with PCFGs

6.5 Improving PCFGs by Splitting Non-Terminals

6.6 Probabilistic Lexicalized CFGs

6.7 Probabilistic CCG Parsing

6.8 Evaluating Parsers

**7. Dependency Parsing**

7.1 Dependency Relations

7.2 Dependency Formalisms

7.3 Dependency Treebanks

7.4 Transition-Based Dependency Parsing

7.5 Graph-Based Dependency Parsing

7.6 Evaluation

**8. Word sense and word net**

8.1 Word Senses

8.2 Relations between Senses

8.3 WordNet: A Database of Lexical Relations

8.4 Word Sense Disambiguation

8.5 Alternate WSD algorithms and Tasks

8.6 Using Thesauruses to Improve Embeddings

8.7 Word Sense Induction

**9. Statistical Machine translation**

9.1 Introduction

9.2 Approaches

9.3 Language Models

9.4 Parallel Corpora

9.5 Word Alignment

9.6 Phrase Library

9.7 Translation Models.

9.8 Search Strategies

**10. Semantic web ontology**

10.1 Introduction

10.2 Ontology and Ontologies

10.3 Ontology Engineering

10.4 Ontology Learning

10.5 State of the Art

**11. Question Answering**

11.1 IR-based Factoid Question answering   
 11.2 Knowledge-based Question Answering

11.3 Using multiple information sources: IBM’s Watson

11.4 Evaluation of Factoid Answers

**12 Dialogue Systems and Chatbots**

12.1 Properties of Human Conversation

12.2 Chatbots

12.3 GUS: Simple Frame-based Dialogue Systems

12.4 The Dialogue-State Architecture

12.5 Evaluating Dialogue Systems

12.6 Dialogue System Design

**13. Sentiment analysis**

13.1 The Problem of Sentiment Analysis

13.2 Sentiment and Subjectivity Classification

13.3 Document-Level Sentiment Classification

13.4 Feature-Based Sentiment Analysis

13.5 Sentiment Analysis of Comparative Sentences

**Learning Outcomes:**

|  |  |
| --- | --- |
| **No** | **Learning Outcomes** |
| LO1 | Should have a good understanding of the field of natural language processing. |
| LO2 | Should have an algorithms and techniques used in this field. |
| LO3 | Should also understand the how natural language processing is used in Machine translation and Information extraction. |

**Part B: Contact Session Plan**

|  |  |
| --- | --- |
| **Academic Term** | 2020 |
| **Course Title** | Natural Language processing |
| **Course No** | DSECLZG525 |
| **Lead Instructor** | Dr. Chetana Gavankar |

**Course Contents**

|  |  |  |  |
| --- | --- | --- | --- |
| **Contact session** | **List of Topic Title**  **(from content structure in Part A)** | **Topic #**  **(from content structure in Part A)** | **Text /**  **Ref Book / External resource** |
| 1 | **Introduction**   * The Study of Language. * Applications of Natural Language Understanding. * Evaluating Language Understanding Systems. * The Different Levels of Language Analysis. * Representations and Understanding. * The Organization of Natural Language Understanding Systems. | Chapter1 | T2 |
| 2 | **N-Grams Language models**   * Evaluating Language Models * Generalization and Zeros * Smoothing * The Web and Stupid Backoff | Chapter 3 | T1 |
| 3 | **Part-of-Speech Tagging**   * (Mostly) English Word Classes * The Penn Treebank Part-of-Speech Tag set * Part-of-Speech Tagging * Markov Chains * The Hidden Markov Model * HMM Part-of-Speech Tagging * Part-of-Speech Tagging for Morphological Rich Languages | Chapter8 | T1 and class notes |
| 4 | **Hidden Markov Model Algorithms**   * Likelihood Computation: The Forward Algorithm * Decoding: The Viterbi Algorithm * HMM Training: The Forward-Backward Algorithm * Maximum Entropy Markov Model * Bidirectionality | Appendix chapter A | T1 and class notes |
| 5 | **Grammars and Parsing**   * Grammars and Sentence Structure. * What Makes a Good Grammar * A Top-Down Parser. * A Bottom-Up Chart Parser. * Top-Down Chart Parsing. * Finite State Models and Morphological Processing. * Grammars and Logic Programming. * Parsing | Chapter3 | T2 |
| 6 | **Statistical Constituency Parsing**   * Probabilistic Context-Free Grammars * Probabilistic CKY Parsing of PCFGs * Ways to Learn PCFG Rule Probabilities * Problems with PCFGs * Improving PCFGs by Splitting Non-Terminals * Probabilistic Lexicalized CFGs * Probabilistic CCG Parsing * Evaluating Parsers | Chapter 14 | T1 |
| 7 | **Dependency Parsing**   * + Dependency Relations   + Dependency Formalisms   + Dependency Treebanks   + Transition-Based Dependency Parsing   + Graph-Based Dependency Parsing   + Evaluation | Chapter 19 | T1 and class notes |
| 8 | **Review of session 1 to session 7** |  |  |
| 9 | **Word sense and word net**   * Word Senses * Relations between Senses * WordNet: A Database of Lexical Relations * Word Sense Disambiguation * Alternate WSD algorithms and Tasks * Using Thesauruses to Improve Embeddings * Word Sense Induction | Chapter15 | T1 |
| 10 | **Statistical Machine translation**   * Introduction * Approaches * Language Models * Parallel Corpora * Word Alignment * Phrase Library * Translation Models * Search Strategies | Chapter 17 | R1 |
| 11 | **Semantic web ontology**   * Introduction * Ontology and Ontologies * Ontology Engineering * Ontology Learning | Chapter 24 | R1 and class notes |
| 12 | **Question Answering**   * IR-based Factoid Question answering * Knowledge-based Question Answering * Using multiple information sources: IBM’s Watson * Evaluation of Factoid Answers | Chapter 25 | T1 |
| 13 | **Dialogue Systems and Chatbots**   * Properties of Human Conversation * Chatbots * GUS: Simple Frame-based Dialogue Systems * The Dialogue-State Architecture * Evaluating Dialogue Systems * Dialogue System Design | Chapter 26 | T1 |
| 14 | **Sentiment analysis**   * The Problem of Sentiment Analysis * Sentiment and Subjectivity Classification * Document-Level Sentiment Classification * Feature-Based Sentiment Analysis * Sentiment Analysis of Comparative Sentences | Chapter 26 | R1 |
| 15 | **NLP Tools and State of Art Areas** |  |  |
| 16 | **Review of session 9 to session 15** |  |  |

**Evaluation Scheme**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Evaluation Component** | **Name**  (Quiz, Lab, Project, Midterm exam, End semester exam, etc) | **Type**  (Open book, Closed book, Online, etc.) | **Weight** | **Duration** | **Day, Date, Session, Time** |
| **EC – 1** | Quiz |  | 5% |  | To be announced |
| **EC – 2** | Assignment |  | 15% |  | To be announced |
| **EC – 3** | Mid-term Exam | Closed book | 30% | 2 hours | To be announced |
| **EC – 4** | End Semester Exam | Open book | 50% | 2.5 hours | To be announced |

***Note*** *- Evaluation components can be tailored depending on the proposed model.*

**Important Information**

Syllabus for Mid-Semester Test (Closed Book): Topics in Weeks 1-8 (1-18 Hours)

Syllabus for Comprehensive Exam (Open Book): All topics given in plan of study

Evaluation Guidelines:

1. EC-1 consists of either two Assignments or three Quizzes. Announcements regarding the same will be made in a timely manner.
2. For Closed Book tests: No books or reference material of any kind will be permitted. Laptops/Mobiles of any kind are not allowed. Exchange of any material is not allowed.
3. For Open Book exams: Use of prescribed and reference text books, in original (not photocopies) is permitted. Class notes/slides as reference material in filed or bound form is permitted. However, loose sheets of paper will not be allowed. Use of calculators is permitted in all exams. Laptops/Mobiles of any kind are not allowed. Exchange of any material is not allowed.
4. If a student is unable to appear for the Regular Test/Exam due to genuine exigencies, the student should follow the procedure to apply for the Make-Up Test/Exam. The genuineness of the reason for absence in the Regular Exam shall be assessed prior to giving permission to appear for the Make-up Exam. Make-Up Test/Exam will be conducted only at selected exam centres on the dates to be announced later.

It shall be the responsibility of the individual student to be regular in maintaining the self-study schedule as given in the course handout, attend the lectures, and take all the prescribed evaluation components such as Assignment/Quiz, Mid-Semester Test and Comprehensive Exam according to the evaluation scheme provided in the handout.

</DIV>