SRI VENKATESWARA COLLEGE OF ENGINEERING

COURSE DELIVERY PLAN - THEORY

Department of Information Technology	LP: Sub Code Rev. No: 00 Date: 20.01.2025
~~B.E/~~B.Tech~~/M.E/M.Tech~~ : Information Technology Regulation: R2022 PG Specialisation : NA Sub. Code / Sub. Name : IT22602-Natural Language Processing and Understanding Unit : 1	

Unit Syllabus: FINITE AUTOMATA & LEXICAL ANALYSIS
Introduction- Basic Mathematical Notation and techniques- Finite Automaton – DFA & NDFA – Finite Automaton with €- moves – Regular Languages- RegularExpression – Equivalence of NFA and DFA – Lexical Analysis-language processors -The Phases of Compiler-Errors Encountered in Different Phases-compiler Construction Tools

Objective: Students will learn the automata theory and tools of the compiler

Session No *	Topics to be covered	Ref	Teaching Aids
1	Introduction- Basic Mathematical Notation and techniques	1- Ch 1; pg 28-31	BB/PPT
2	Finite Automaton – DFA	1- Ch 2; pg 45-54	BB/PPT
3	Finite Automaton – NDFA	1- Ch 2; pg 55- 60	BB/PPT
4	Finite Automaton with €- moves	1- Ch 2; pg 72- 80	BB/PPT
5	Equivalence of NFA and DFA	1- Ch 2; pg 60- 65	BB/PPT
6	Regular Languages- Regular Expression	1- Ch 3; pg 83-108	BB/PPT
7	Lexical Analysis-language processors	R1- Ch 1; pg 1 -4	BB/PPT
8	The Phases of Compiler	R1- Ch 1; pg 4 -11	BB/PPT
9	Errors Encountered in Different Phases-compiler Construction Tools	R1- Ch 1; pg 12	BB/PPT
Content beyond syllabus covered (if any):			

* Session duration: 50 minutes

Sub. Code / Sub. Name: IT22602-Natural Language Processing and Understanding

Unit : III

Unit Syllabus: WORD LEVEL ANALYSIS

Context-Free Grammars, Grammar rules for English, Treebanks, Normal Forms for grammar – Dependency Grammar – Syntactic Parsing - Unsmoothed N-grams, Evaluating N-grams, Smoothing, Interpolation and Backoff – Word Classes, Part-of-Speech Tagging, Rule-based, Stochastic and Transformation-based tagging, Issues in PoS tagging – Hidden Markov and Maximum Entropy models.

Objective: Students will learn how to represent sentence using grammar

Session No *	Topics to be covered	Ref	Teaching Aids
10	Context-Free Grammars, Grammar rules for English	1- Ch 5; pg:169-180 2- Ch 12; pg:389-408	BB/PPT
11	Parse Trees, Treebanks	1- Ch 5; pg 181-190 2- Ch 12; pg:408-415	BB/PPT
12	Normal Forms for grammar, Dependency Grammar	2- Ch 12; pg:415-416 2- Ch 12; pg:418-421	BB/PPT
13	Syntactic Parsing	2- Ch 12; pg:431-454	BB/PPT
14	Unsmoothed N-grams, Evaluating N-grams	2- Ch 4; pg:83-91	BB/PPT
15	Interpolation and Backoff	2- Ch 4; pg:103-107	BB/PPT
16	Word Classes, Part-of-Speech Tagging	2- Ch 5; pg:124-136	BB/PPT
17	Rule-based, Stochastic and Transformation-based tagging, Issues in PoS tagging	2- Ch 5; pg:136-153	BB/PPT
18	Hidden Markov and Maximum Entropy models.	2- Ch 6; pg:173-207	BB/PPT

Content beyond syllabus covered (if any):

* Session duration: 50 mins

Sub. Code / Sub. Name: IT22602-Natural Language Processing and Understanding

Unit: IV

Unit Syllabus: SEMANTICS AND PRAGMATICS

Semantic Analysis: Meaning Representation-Lexical Semantics- Ambiguity-Word Sense Disambiguation. Requirements for representation, First-Order Logic, Description Logics – Semantic analysis, Word Senses, Relations between Senses, Thematic Roles, Selectional restrictions – Word Sense Disambiguation, WSD using Supervised, Dictionary & Thesaurus, Bootstrapping methods – Word Similarity using Thesaurus and Distributional methods.

Objective: Students will learn word syntax and semantic analysis

Session No *	Topics to be covered	Ref	Teaching Aids
19	Semantic Analysis: Meaning Representation	2- Ch 17; pg:553-560	BB/PPT
20	Lexical Semantics - Word Sense Disambiguation	2- Ch 20; pg:650	BB/PPT
21	Requirements for representation, First-Order Logic, Description	2- Ch 17; pg:563-588	BB/PPT
22	Logics – Semantic analysis	2- Ch 18; pg:593-602	BB/PPT
23	Word Senses, Relations between Senses	2- Ch 19; pg:627-633	BB/PPT
24	Thematic Roles Selectional restrictions	2- Ch 19; pg:635-645	BB/PPT
25	WSD using Supervised methods	2- Ch 20; pg:655-660	BB/PPT
26	WSD using Dictionary & Thesaurus, Bootstrapping methods	2- Ch 20; pg:662-666	BB/PPT
27	Word Similarity using Thesaurus and Distributional methods.	2- Ch 20; pg:668-684	BB/PPT
Content beyond syllabus covered (if any):			

* Session duration: 50 mins

Sub. Code / Sub. Name: IT22602-Natural Language Processing and Understanding

Unit: V

Unit Syllabus: DISCOURSE ANALYSIS

Discourse segmentation, Coherence – Reference Phenomena, Anaphora Resolution using Hobbs and Centering Algorithm – Coreference Resolution –applications

Objective: Students will learn pragmatics and discourse analysis.

Session No *	Topics to be covered	Ref	Teaching Aids
28	Discourse segmentation	2- Ch 21; pg:697-705	PPT
29	Text Coherence	2- Ch 21; pg: 705-711	PPT
30	Reference Phenomena	2- Ch 21; pg: 714-720	PPT
31,32	Anaphora Resolution using Hobbs algorithm	2- Ch 21; pg: 720-722	PPT
33	Centering Algorithm	2- Ch 21; pg: 722-724	PPT
34	Co-reference Resolution	2- Ch 21; pg: 724-728	PPT
35,36	Applications	2- Ch 22; pg: 743-750 2- Ch 23; pg: 805-819	PPT

Content beyond syllabus covered (if any):

* Session duration: 50 mins

Sub. Code / Sub. Name: IT22602-Natural Language Processing and Understanding

Unit: II

Unit Syllabus: NATURAL LANGUAGE GENERATION

Natural Language Generation: Architecture of NLG Systems- Generation Tasks and Representations - Application of NLG. Machine Translation: Problems in Machine Translation- Characteristics of Indian Languages- Machine Translation Approaches Translation involving Indian Languages, Porter Stemmer, Lemmatizer

Objective: Students will learn the architecture of NLP systems and applications of natural language processing techniques

Session No *	Topics to be covered	Ref	Teaching Aids
37	Natural Language Generation: Architecture of NLG Systems	Internet: https://wisconsin.pressbooks.pub/naturallanguage/chapter/natural-language-generation/	PPT
38	Generation Tasks and Representations	R4- Ch 6; pg: 126-140	PPT
39	Application of NLG	2- Ch 22; pg: 737-740	PPT
40	Machine Translation: Problems in Machine Translation	2- Ch 25; pg: 879-885	PPT
41	Characteristics of Indian Languages	Internet: https://www.w3.org/2006/10/SSML/papers/CHARACTERISTICS_OF_INDIAN_LANGUAGES.pdf	PPT
42,43	Machine Translation Approaches	2- Ch 25; pg: 887-900	PPT
44	Translation involving Indian Languages	Internet: https://anoopkunchukuttan.gitlab.io/publications/presentations/kiit_indicnlp_talk_oct2020.pdf	PPT
45	Porter Stemmer, Lemmatizer	2- Ch 3; pg: 68 3-Ch 3; pg: 107-108	PPT
Content beyond syllabus covered (if any):			

* Session duration: 50 mins

TEXT BOOKS:

1. John. E. Hopcroft, Rajiv Motwani and Jeffrey D Ullman, —Introduction to Automata Theory, Languages and Computation‖, Third Edition, Pearson Education, 2014
2. Daniel Jurafsky, James H. Martin,Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech, Pearson Publication, 2014.
3. Steven Bird, Ewan Klein and Edward Loper, Natural Language Processing with Python‖, First Edition, O_Reilly Media, 2009.

REFERENCES:

1. K. C. Louden, Compiler Construction Principles and Practice, Thomson Learning Inc. 1st Ed, 2007.
2. L. Breck Baldwin, —Language Processing with Java and LingPipe Cookbook, Atlantic Publisher, 2015.
3. M. Richard M Reese, —Natural Language Processing with Java, OReilly Media, 2015.
4. Nitin Indurkhya and Fred J. Damerau, —Handbook of Natural Language Processing, Second Edition, Chapman and Hall/CRC Press, 2010.
5. Tanveer Siddiqui, U.S. Tiwary, —Natural Language Processing and Information Retrieval, Oxford University Press, 2008.

	Prepared by	Approved by
Signature		
Name	Dr. V Vidhya / Mr. E Sivakumar	Dr V Vidhya
Designation	Professor / AP	Professor & HoD
Date	20.01.2025	
Remarks *:		
Remarks *:		

* If the same lesson plan is followed in the subsequent semester/year it should be mentioned and signed by the Faculty and the HOD