VELAGAPUDI RAMAKRISHNA

SIDDHARTHA ENGINEERING COLLEGE::VIJAYAWADA

(AUTONOMOUS)

DEPARTMENT OF IT

MICRO LEVEL SYLLABUS

Class	**B Tech**	**Regulation**	**VR20**
Subject Code	**20IT7404A**	**Year & Semester**	**IV YEAR, SEMESTER-I**
Title of the Subject	**NATURAL LANGUAGE PROCESSING**		

Unit No	Topics Covered (with Sub Topics)	Text Book	Chapter/ Section Number	Page Number
Unit I	**Regular Expressions, Text Normalization, Edit Distance**	[T1]	**2**	
	Regular Expressions		2.1	5-5
	Basic Regular Expression Patterns		2.1.1	5-8
	Disjunction, Grouping, and Precedence		2.1.2	8-9
	A Simple Example		2.1.3	9-10
	More Operators		2.1.4	10-11
	A More Complex Example		2.1.5	11-11
	Substitution, Capture Groups, and ELIZA		2.1.6	12-13
	Words		2.2	13-15
	Corpora		2.3	15-16
	Text Normalization		2.4	16-17
	Unix Tools for Crude Tokenization and Normalization		2.4.1	17-18
	Word Tokenization		2.4.2	18-20
	Byte-Pair Encoding for Tokenization		2.4.3	20-22
	Word Normalization, Lemmatization and Stemming		2.4.4	22-24
	Minimum Edit Distance		2.5	24-25
	The Minimum Edit Distance Algorithm		2.5.1	25-29
	Natural Language Toolkit Essentials (NLTK)			
	Introduction to NLP		1	
	Diving into NLTK		1.3	13-19
	Text Wrangling and Cleansing	[T2]	2	21-29
	What is Text wrangling		2.1	21-24
	Text cleansing		2.2	24-24
	Sentence splitter		2.3	24-25
	Tokenization		2.4	25-26
	Stemming,Lemmatization.		2.5,2.6	26-28
	Stop word removal, Rare word removal		2.7,2.8	28-29
Unit II	**N-gram Language Models**		3	31-32
	N-Grams		3.1	32-37
	Evaluating Language Models		3.2	37-38
	Perplexity		3.2.1	38-40
	Sampling sentences from a language model		3.3	40-40
	Smoothing		3.5	43-44
	Laplace Smoothing		3.5.1	44-46
	Add-k smoothing	[T1]	3.5.2	46-46
	Backoff and Interpolation		3.5.3	46-47
	Naive Bayes and Sentiment Classification		4	58-59
	Naive Bayes Classifiers		4.1	59-62
	Training the Naive Bayes Classifier		4.2	62-64
	Worked example		4.3	64-64
	Optimizing for Sentiment Analysis		4.4	64-66

	Naive Bayes for other text classification tasks		4.5	66-67
	Naive Bayes as a Language Model		4.6	67-67
Unit III	**Sequence Labeling for parts of Speech and Named Entities**	[T1]	**8**	**160-160**
	(Mostly) English Word Classes		8.1	161-163
	Part-of-Speech Tagging		8.2	163-165
	Named Entities and Named Entity Tagging		8.3	165-166
	HMM Part-of-Speech Tagging		8.4	167-167
	Markov Chains		8.4.1	167-168
	The Hidden Markov Model		8.4.2	168-168
	The components of an HMM tagger		8.4.3	169-169
	Context-Free Grammars and Constituency Parsing		**17**	**357-357**
	Constituency		17.1	358-358
	Context-Free Grammars		17.2	358-361
	Formal Definition of Context-Free Grammar		17.2.1	362-362
	Treebanks		17.3	362-364
	Grammar Equivalence and Normal Form		17.4	364-365
	Ambiguity		17.5	365-367
	CKY Parsing: A Dynamic Programming Approach		17.6	367-367
	Conversion to Chomsky Normal Form		17.6.1	367-368
	CKY Recognition		17.6.2	368-370
	CKY Parsing		17.6.3	370-371
	CKY in Practice		17.6.4	371-372
Unit IV	**Logical Representations of Sentence Meaning**	[T1]	**19**	405-406
	Computational Desiderata for Representations		19.1	406-408
	Model-Theoretic Semantics		19.2	408-411
	First-Order Logic		19.3	411-411
	Basic Elements of First-Order Logic		19.3.1	411-413
	Variables and Quantifiers		19.3.2	413-414
	Lambda Notation		19.3.3	415-415
	The Semantics of First-Order Logic		19.3.4	415-416
	Inference		19.3.5	416-418
	Event and State Representations		19.4	418-419
	Description Logics		19.5	419-422
	Word Senses and WordNet		**23**	457-458
	Word Senses		23.1	458-458
	Defining Word Senses		23.1.1	458-459
	How many senses do words have?		23.1.2	459-459
	Relations Between Senses		23.2	460-461
	WordNet: A Database of Lexical Relations		23.3	462-463
	Sense Relations in WordNet		23.3.1	463-464