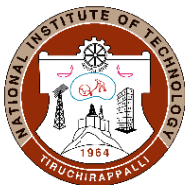




COURSE PLAN – PART I			
Name of the programme and specialization	M.TECH / CSE		
Course Title	Statistical Natural Language Processing		
Course Code	CS619	No. of Credits	3
Course Code of Pre-requisite subject(s)			
Session	July/ January 2020	Section (if, applicable)	A / B
Name of Faculty	Dr. Rajeswari Sridhar	Department	CSE
Official Email	srajeswari@nitt.edu	Telephone No.	
Name of Course Coordinator(s) (if, applicable)	Batch – 2019 – 2021		
Official E-mail		Telephone No.	
Course Type (please tick appropriately)	<input type="checkbox"/> Elective course		
Syllabus (approved in Senate)			
UNIT I Morphology And Part-Of-Speech Processing Introduction –Regular Expressions and Automata-Non-Deterministic FSAs. Transducers –English Morphology -Finite-State Morphological Parsing -Porter Stemmer -Tokenization-Detection and Correction of Spelling Errors. N-grams –Perplexity -Smoothing -Interpolation -Backoff . Part-of-Speech Tagging – English Word Classes -Tagsets -Rule-Based -HMM -Transformation-Based Tagging - Evaluation and Error Analysis. Hidden Markov and Maximum Entropy Models UNIT II Speech Processing Phonetics –Articulatory Phonetics -Phonological Categories - Acoustic Phonetics and Signals -Speech Synthesis –Text Normalization –Phonetic and Acoustic Analysis -Diphone Waveform synthesis –Evaluation-Automatic Speech Recognition –Architecture -Hidden Markov Model to Speech -MFCC vectors -Acoustic Likelihood Computation -Evaluation. Triphones –Discriminative Training -Modeling Variation. Computational Phonology-Finite-State Phonology –Computational Optimality Theory -Syllabification -Learning Phonology and Morphology UNIT III Syntax Analysis Finite-State and Context-Free Grammars -Dependency Grammars. Syntactic Parsing – Ambiguity -Dynamic Programming Parsing Methods –CKY-Earley and Chart Parsing-Partial Parsing-Evaluation. Statistical Parsing –Probabilistic Context-Free Grammars –Probabilistic CKY Parsing of PCFGs –Probabilistic Lexicalized CFGs –Collins Parser – Shallow parsers – Dependency parsing			



UNIT IV Semantic and Pragmatic Interpretation Representation of Meaning –Desirable Properties -Computational Semantics -Word Senses -Relations Between Senses –WordNet -Event Participants-Proposition Bank -Frame Net –Metaphor. Computational Lexical Semantics –Word Sense Disambiguation-Supervised Word Sense Disambiguation - Dictionary and Thesaurus Methods-Word Similarity -Minimally Supervised WSD - Hyponymy and Other Word Relations -Semantic Role Labeling -Unsupervised Sense Disambiguation. Computational Discourse -Discourse Segmentation -Unsupervised Discourse -Segmentation -Text Coherence -Reference Resolution –Phenomena –Features and algorithms -Pronominal Anaphora Resolution

UNIT V Applications Information Extraction –Named Entity Recognition -Relation Detection and Classification –Temporal and Event Processing -Template-Filling -Biomedical Information Extraction. Question Answering and Summarization -Information Retrieval - Factoid Question Answering -Summarization -Single and Multi-Document Summarization - Focused Summarization -Evaluation. Dialog and Conversational Agents –Properties of Human Conversations -Basic Dialogue Systems

*Programming Assignments are mandatory

Text books

1. Jurafsky and Martin, “Speech and Language Processing”, Pearson Prentice Hall, Second Edition, 2008.
2. Christopher D. Manning and Hinrich Schütze, “Foundations of Statistical Natural Language Processing”, MIT Press, 1999.
3. Stevan Bird, “Natural Language Processing with Python”, Shroff, 2009.
4. James Allen, “Natural Language Understanding”, Addison Wesley, Second Edition, 2007.
5. Nitin Indurkha, Fred J. Damerau, “Handbook of Natural Language Processing”, (Chapman & Hall/CRC Machine Learning & Pattern Recognition), Second Edition, 2010.
6. Alexander Clark, Chris Fox, Shalom Lappin, “The Handbook of Computational Linguistics and Natural Language Processing”, Wiley-Blackwell, 2012

COURSE OBJECTIVES

- To understand the need for morphological processing and their representation
- To know about the various techniques used for speech synthesis and recognition
- To appreciate the syntax analysis and parsing that is essential for natural language processing
- To learn about the various representations of semantics and discourse
- To have knowledge about the applications of natural language processing

MAPPING OF COs with POs

Course Outcomes	Programme Outcomes (PO) (Enter Numbers only)
1. Identify the different linguistic components of natural language	1, 2,



2. Design a morphological analyzer for a given natural language	1, 3, 4
3. Decide on the appropriate parsing techniques necessary for a given language and application	1, 2, 3, 5, 6
4. Design new tag set and a tagger for a given natural language	3, 4, 5
5. Design applications involving natural language	3, 4, 5

COURSE PLAN – PART II			
COURSE OVERVIEW			
<p>Natural language processing deals with designing and development of algorithms involved in making the machine understand information similar to how humans understand information given as audio, text, video, etc. Making meaningful sense of the contents of the information is essential for humans. Natural language processing deals with algorithms that essentially makes meaningful sense of any audio, video, text that are given as input.</p>			
COURSE TEACHING AND LEARNING ACTIVITIES			(Add more rows)
S.No.	Week/Contact Hours	Topic	Mode of Delivery
1	06/01/2020 to 10/01/2020 2 hours	Unit 1: Introduction, Need and applications of NLP	Powerpoint presentation
2	06/01/2020 to 10/01/2020 1 hour	Biology of Speech processing, place and manner of articulation	Powerpoint presentation
3	13/01/2020 to 17/01/2020 2 hours	Word Boundary detection, issues in speech processing, Block Diagram of a speech processing system	Powerpoint presentation
4	13/01/2020 to 17/01/2020 1 hour	N-gram models for language representation	Powerpoint presentation
5	20/01/2020 to 24/01/2020 2 hours	N-gram models for language representation	Powerpoint presentation
6	20/01/2020 to 24/01/2020 1 hour	Regular expression, Finite State machines	Powerpoint presentation
7	27/01/2020 to 31/01/2020 2 hours	Finite State Machines, Morphology learning, HMM	Powerpoint presentation



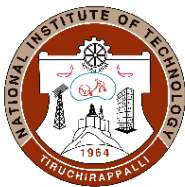
8	27/01/2020 to 31/01/2020 1 hour	HMM - Decoder	Powerpoint presentation
9	03/02/2020 to 07/02/2020 2 hours	Text to Speech – Pronunciation, methods, reasoning, grapheme to phoneme conversion	Powerpoint presentation
10	03/02/2020 to 07/02/2020 1 hour	Named Entities, Named Entity recognition	Powerpoint presentation
11	10/02/2020 to 14/02/2020 2 hours	Parts of Speech – tagging, tagger	Powerpoint presentation
12	10/02/2020 to 14/02/2020 1 hour	Parsing – Top down, Bottom Up, Earley parser	Powerpoint presentation
13	17/02/2020 to 21/02/2020 2 hours	Earley Parser, Problems	Powerpoint presentation
14	17/02/2020 to 21/02/2020 1 hour	Cycle Test 1	
15	24/02/2020 to 28/02/2020 2 hours	CYK Parser, Probabilistic CYK parser	Powerpoint presentation
16	24/02/2020 to 28/02/2020 1 hour	Rule Based Parsing, Hybrid Parsing	Powerpoint presentation
17	02/03/2020 to 06/03/2020 1 hour	Maximum Entropy Models, Generalized vs Specialized computation	Powerpoint presentation
18	02/03/2020 to 06/03/2020 2 hours	Shallow Parsing	Powerpoint presentation
19	09/03/2020 to 13/03/2020 1 hour	Lexical knowledge networks, Wordnets	Powerpoint presentation
20	16/03/2020 to 20/03/2020 2 hours	Semantic Roles, Semantic analysis, WSD	Powerpoint presentation



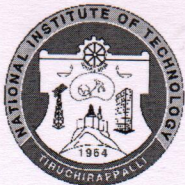
21	16/03/2020 to 20/03/2020 1 hour	WSD, Multilinguality	Powerpoint presentation
22	23/03/2020 to 27/03/2020 2 hours	Metaphors, Co-reference resolution, anaphora resolution	Powerpoint presentation
23	23/03/2020 to 27/03/2020 1 hour	Sentiment analysis, Text entailment	Powerpoint presentation
24	30/03/2020 to 03/04/2020 1 hour	Question Answering system	Powerpoint presentation
25	06/04/2020 to 10/04/2020 1 hour	Cycle Test 2	
26	06/04/2020 to 10/04/2020 2 hours	Information Retrieval, Cross-lingual information retrieval	Powerpoint presentation
27	13/04/2020 to 17/04/2020 2 hours	Single, Multi-document Summarization	Powerpoint presentation
28	13/04/2020 to 17/04/2020 1 hour	Multi-document summarization	Powerpoint presentation

COURSE ASSESSMENT METHODS (shall range from 4 to 6)

S.No.	Mode of Assessment	Week/Date	Duration	% Weightage
1	Cyle Test 1	17/02/2020 to 21/02/2020 1 hour	1 hour	15
2	Cycle Test 2	06/04/2020 to 10/04/2020 1 hour	1 hour	15
3	Assignment 1	17/02/2020 to 21/02/2020 1 hour	2 hours	10
4	Project	13/04/2020 to 17/04/2020 1 hour	20 hours	20



CPA	Compensation Assessment*			
5	Final Assessment *	As per academic schedule	3 hours	40
*mandatory; refer to guidelines on page 4				
COURSE EXIT SURVEY (mention the ways in which the feedback about the course shall be assessed)				
<ol style="list-style-type: none"> 1. Students' feedback through PAC meetings 2. Feedbacks are collected before final examination through MIS or any other standard format followed by the institute 3. Students, through their Class Representatives, may give their feedback at any time to the course faculty which will be duly addressed. 				
COURSE POLICY (including compensation assessment to be specified)				
<u>MODE OF CORRESPONDENCE (email/ phone etc)</u>				
Email, in-person – after 4.00 pm.				
<u>COMPENSATION ASSESSMENT POLICY</u>				
<ol style="list-style-type: none"> 1. One compensation assessment will be given after completion of Cycle Test 1 and 2 for the students those who are absent for any assessment due to genuine reason. 2. Compensatory assessments would cover the syllabus of Cycle tests 1 & 2 3. The prior permission and required documents must be submitted for absence signed by HoD/CSE. 				
<u>ATTENDANCE POLICY</u> (A uniform attendance policy as specified below shall be followed)				
<ul style="list-style-type: none"> ➤ At least 75% attendance in each course is mandatory. ➤ A maximum of 10% shall be allowed under On Duty (OD) category. ➤ Students with less than 65% of attendance shall be prevented from writing the final assessment and shall be awarded 'V' grade. 				
<u>ACADEMIC DISHONESTY & PLAGIARISM</u>				
<ul style="list-style-type: none"> ➤ Possessing a mobile phone, carrying bits of paper, talking to other students, copying from others during an assessment will be treated as punishable dishonesty. ➤ Zero mark to be awarded for the offenders. For copying from another student, both 				



NATIONAL INSTITUTE OF TECHNOLOGY, TIRUCHIRAPPALLI

students get the same penalty of zero mark.

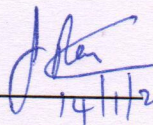
- The departmental disciplinary committee including the course faculty member, PAC chairperson and the HoD, as members shall verify the facts of the malpractice and award the punishment if the student is found guilty. The report shall be submitted to the Academic office.
- The above policy against academic dishonesty shall be applicable for all the programmes.

ADDITIONAL INFORMATION, IF ANY

1. The Course Coordinator is available for consultation during the time intimated to the students then and there.
2. Relative grading adhering to the instructions from the office of the Dean (Academic) will be adopted for the course.

FOR APPROVAL

Course Faculty


14/11/2020

CC- Chairperson


17/11/2020

HOD