



NATIONAL INSTITUTE OF TECHNOLOGY, TIRUCHIRAPPALLI
DEPARTMENT OF COMPUTER APPLICATIONS

COURSE PLAN – PART I			
Name of the programme and specialization	M.Tech Data Analytics		
Course Title	NATURAL LANGUAGE PROCESSING		
Course Code	CA6D1	No. of Credits	3
Course Code of Pre-requisite subject(s)			
Session	Jan 2023	Section	NA
Name of Faculty	Dr. S. Sangeetha	Department	Computer Applications
Official Email	sangeetha@nitt.edu	Telephone No.	0431-2503743
Name of PAC Chairman	Dr.B.Balaji		
Official E-mail	balaji@nitt.edu	Telephone No.	
Course Type (please tick appropriately)	<input type="checkbox"/> Core course <input checked="" type="checkbox"/> Elective course		
Syllabus (approved in BoS)			
<p>Natural Language Processing – Applications of NLP – Linguistic Background – NLP tasks – Ambiguities in NLP tasks – Finite state automata – Regular Expressions – Corpus – Text Normalization – Edit Distance – Boundary Determination – Tokenization – Stemming - Lemmatization.</p> <p>Morphological Analysis – Part of speech tagging – Shallow parsing – Dependency parsing – WordNet- Sematic similarity measures – Semantic representation – Coreference Resolution – Tools – Natural Language Toolkit – Stanford CoreNLP.</p> <p>Language model – n-gram language models – Hidden Markov Model – Conditional random Fields – Topic models – Graph Models – Machine Learning for NLP – Language Features – Maximum Entropy classifier – Phrase Based clustering.</p> <p>Deep Learning for NLP - Neural Networks – Vector Representations – Word Embeddings – Attention model – Encoder Decoder – Transformer based models – Tools: PyTorch.</p> <p>Applications and Case Studies – Question Answering – Machine Translation – Information retrieval – Information Extraction</p>			
REFERENCES:			
<ol style="list-style-type: none"> 1. Anders Søgaard, Ivan Vulić, Sebastian Ruder, Manaal Faruqui , Cross-Lingual Word Embeddings (Synthesis Lectures on Human Language Technologies), Morgan & Claypool Publishers, 2019 2. Delip Rao, Brian McMahan, Natural Language Processing with PyTorch: Build Intelligent Language Applications Using Deep Learning, O'Reilly, 2019 			

3. Daniel Jurfsky, James H. Martin Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition, Second Edition, Pearson 2013
4. Christopher D. Manning and Hinrich Schutze, "Foundations of Statistical Natural Language Processing", MIT Press, 1999.

COURSE OBJECTIVES

1. To introduce language processing techniques to enable Text data processing
2. To impart knowledge on text data processing using Statistical and Machine learning models
3. To describe the various Embeddings and Deep learning models for NLP
4. To introduce real world applications of Language processing

MAPPING OF COs with POs

Course Outcomes	Programme Outcomes (PO) (Enter Numbers only)
1. Identify the patterns in text and pre-process the large text corpus	1,2
2. Describe and work with basic NLP tasks	1,2
3. Analyze the text content to provide predictions related to a specific domain.	1,2,3,4
4. Use statistical and machine learning models with feature selection for NLP	1.,2,3,4
5. Adopt Deep learning models for NLP along with embeddings	1,2,3,4
6. Apply the concepts for solving NLP Applications.	1,2,3,5

COURSE PLAN – PART II

COURSE OVERVIEW

This course introduces the basics of language processing techniques including syntactic and semantic representation of text content. It also focuses on the usage of tools like Natural Language Toolkit and Stanford CoreNLP to computationally process the natural language text data. The course deals with language models, Graph models and the Machine learning techniques to model and handle text data. It also focuses on the vector representations of text data and discusses the transformer-based models for NLP in detail. The course ends with discussion of few applications.

COURSE TEACHING AND LEARNING ACTIVITIES

Week	Hour	Topic	Mode of Delivery
1	1	Natural Language Processing	Presentation
	2	Applications of NLP	-do-
	3	Linguistic Background, NLP tasks	-do-
2	1	Ambiguities in NLP tasks	Presentation
	2	Finite state automata	-do-
	3	Regular Expressions	-do-
3	1	Corpus, Text Normalization	Presentation
	2	Edit Distance	-do-
	3	Boundary Determination, Tokenization	-do-
4	1	Stemming, Lemmatization	Presentation
	2	Morphological Analysis	Presentation
	3	Part of speech tagging	-do-

	4	Shallow parsing, Dependency parsing	-do-
5	1	WordNet, Sematic similarity measures	Presentation
	2	Semantic representation	-do-
	3	Coreference Resolution	-do-
	4	Natural Language Toolkit	Presentation & Demo
6	1	Stanford CoreNLP	Presentation & Demo
	2	N-gram language model	Presentation
	3	Hidden Markov Model	-do-
	4	Conditional random Fields	-do-
7	1	Topic models	Presentation
	2	Graph Models	-do-
	3	Machine Learning for NLP – Language Features	-do-
8	1	Maximum Entropy classifier	-do-
	2	Phrase Based clustering.	-do-
	3	Deep Learning for NLP - Neural Networks	-do-
9	1	Vector Representations – Word Embeddings	-do-
	2	Attention model, Encoders Decoders	-do-
	3	Transformer based models	-do-
10	1	Transformer based models	-do-
	2	Transformer based models	-do-
	3	Transformer based models	-do-
11	1	PyTorch	Presentation & Demo
	2	PyTorch	Presentation & Demo
	3	Question Answering	Presentation
12	1	Machine Translation	Presentation
	2	Information retrieval	Presentation
	3	Information Extraction	Presentation

COURSE ASSESSMENT METHODS (shall range from 4 to 6)

S.No.	Mode of Assessment	Week/Date	Duration	% Weightage
1	Test 1	As per Academic schedule	1 Hr	15
2	Test 2	As per Academic schedule	1 Hr	15
3	Project	Week 7	4 Weeks	20
CPA	Compensation Assessment*	At the end of the course	1 Hr	15
4	Final Assessment *	At the end of the course	3 Hrs	50

COURSE EXIT SURVEY

- The students through the class representative may give their feedback at any time to the course faculty which will be duly addressed.
- The students may also give their feedback during Class Committee meeting.

COURSE POLICY (including compensation assessment to be specified)

ATTENDANCE POLICY (A uniform attendance policy as specified below shall be followed)

- 75% attendance in each course is mandatory.
- Students with less than 65% of attendance shall be prevented from writing the final assessment and shall be awarded 'V' grade.

ACADEMIC DISHONESTY & PLAGIARISM

- 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 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.
- The students are expected to come out with their original solution for problems given as assignment, and tests/examinations.

ADDITIONAL INFORMATION, IF ANY

The faculty is available for consultation in office from 4 pm to 5 pm on Monday every week.

FOR APPROVAL

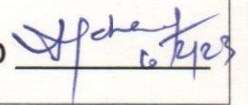
Course Faculty



CC- Chairperson



HOD



Head of the Department
Dept. of Computer Applications
National Institute of Technology
Tiruchirappalli - 620 015.
Tamilnadu, India.