



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 Computing		
Course Code	CA612	No. of Credits	3
Course Code of Pre-requisite subject(s)			
Session	July 2019	Section	---
Name of Faculty	Dr. S. Sangeetha	Department	Computer Applications
Official Email	sangeetha@nitt.edu	Telephone No.	0431-2503743
Name of Course Coordinator(s)	Dr.P.J.A.Alphonse		
Official E-mail	alphonse@nitt.edu	Telephone No.	0431-2503742
Course Type (please tick appropriately)	<input type="checkbox"/> Core course <input checked="" type="checkbox"/> Elective course		
Syllabus (approved in BoS)			
Natural Language Processing—Linguistic Background—Mathematical Foundations—Morphological Analysis—Tokenization—Stemming—Lemmatization—Boundary Determination. Reading unstructured data—Representing text data—Part of speech tagging—Syntactic representation—Text similarity—WordNet based similarity—Shallow parsing—Semantic representation. Information retrieval and Information extraction—Named Entity Recognition—Relation Identification—Template filling. Language model—Probabilistic Models—n-gram language models—Hidden Markov Model—Topic Modelling—Graph Models—Feature Selection and classifiers—Rule-based Classifiers—Maximum entropy classifier—Clustering—Word and Phrase-based Clustering. Tools—Natural Language Toolkit, Apache Open NLP.Applications of Text Analytics—Applications in Social media—Life science—Legal Text—Visualization—Case studies.			
References:			
1. Christopher D.Manning and Hinrich Schutze, "Foundations of Statistical Natural Language Processing", MIT Press, 1999. 2. Steven Struhl, "Practical Text Analytics: Interpreting Text and Unstructured Data for Business Intelligence", KoganPage, 2015. 3. Matthew A.Russell, "Mining the Social Web", O'Reilly Media, 2013. 4. Steven Bird, Ewan Klein and Edward Loper, "Natural Language Processing with Python", 1 st Edition, O'Reilly Media,2009			

COURSE OBJECTIVES

- To get introduced to language processing technologies to process the text data.
- To understand role of Information Retrieval and Information Extraction for Text analytics
- To acquire knowledge on text data analytics using language models.

MAPPING OF COs with POs

Course Outcomes	Programme Outcomes (PO) (Enter Numbers only)
1. Process the text data at syntax and semantic level.	1,2,4
2. Extract Information from Text data	2,3,4
3. Analyze the text content to provide predictions related to a specific domain.	2,3,4

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 introduces Information retrieval and Information extraction techniques, the base for Text Analytics. The course deals with language models, Graph models and the Machine learning techniques to model and handle text data. It then focuses on the usage of tools like Natural Language Toolkit to computationally process the natural language text data. The course ends with few text analytics applications and its visualization as case studies.

COURSE TEACHING AND LEARNING ACTIVITIES

Week	Hour	Topic	Mode of Delivery
1	1	Natural Language Processing	Presentation
	2	Linguistic Background	-do-
	3	Mathematical Foundations	-do-
2	1	Morphological Analysis.	Presentation
	2	Tokenization- Stemming	-do-
	3	Lemmatization - Boundary Determination	-do-
3	1	Reading unstructured data- - Representing text	Presentation, Problem Solving
	2	Part of speech tagging - Shallow parsing	-do-
	3	Syntactic representation	-do-
4	1	Text similarity	Presentation, Problem Solving
	2	WordNet based similarity	-do-
	3	Semantic representation.	-do-
5	1	Information retrieval	Presentation, Problem Solving
	2	Information extraction: Named Entity Recognition,	-do-
	3	Relation identification	-do-
6	1	Event Extraction	Presentation, Problem Solving
	2	Language model - Probabilistic Models	-do-
	3	n-gram language models	-do-
7	1	Hidden Markov Model	Presentation, Problem Solving
	2	Topic Modelling	Presentation

	3	Graph Models	Presentation
8	1	Feature Selection and classifiers	Presentation
	2	Rule-based Classifiers	Presentation
	3	Maximum entropy classifier	Presentation
9	1	Clustering-Word and Phrase-based Clustering.	Presentation, Problem Solving
	2	Tools – Natural Language Tool kit.	Demo
	3	Applications of Text Analytics	Presentation
10	1	Applications in Social media	Demo, Presentation
	2	Applications in Life science	Demo, Presentation
	3	Applications in Legal and Tax regulatory	Demo, Presentation
11	1	Visualization & Case studies.	Presentation, Discussion
	2	Word embedding	Presentation
	3	Word embedding	Presentation

COURSE ASSESSMENT METHODS (shall range from 4 to 6)

S.No.	Mode of Assessment	Week/Date	Duration	% Weightage
1	Test 1	Week 4	1 Hr	15
2	Test 1	Week 8	1 Hr	15
3	Programming Assignments (Submission through Moodle)	Week 6	3 Weeks	30
CPA	Compensation Assessment*	At the end of the course	1 Hr	15
4	Final Assessment *	At the end of the course	3 Hrs	40

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)

- 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.

ACADEMIC DISHONESTY & PLAGIARISM

- 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 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 Course Coordinator is available for consultation office from 4 pm to 5 pm on Monday every week.

FOR APPROVAL

Course Faculty  CC- Chairperson  HOD 