DEPARTMENT OF COMPUTER APPLICATIONS

NATIONAL INSTITUTE OF TECHNOLOGY, TIRUCHIRAPPALLI

Course Title	NATURAL LANGUAGE COMPUTING		
Course Code	CA612	No. of Credits	3
Department	Computer Applications	Faculty	Dr. S.Sangeetha
e-learning platform	http://egov.nitt.edu/moodle/login/index.php		
Course Teacher(s)/Tutor(s) E-mail	sangeetha@nitt.edu	Telephone No.	0431-2503743
Class committee Chairperson	Dr.R.Eswari		
Course Type	☐ Core course	■	Elective course

SYLLABUS

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 Tool Kit, Apache OpenNLP Applications of Text Analytics – Applications in Social Media - Life Science - Legal Text–Visualization -Case Studies.

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 Natural Language Toolkit to computationally process the text data. The course ends with few text analytics applications and its visualization as case studies.

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.*

Course Outcomes

- *Process the text data at syntax and semantic level.*
- *Extract Information from Text data*
- *Analyze the text content to provide predictions related to a specific domain.*

COURSE TEACHING AND LEARNING ACTIVITIES

Week	class	Topic	Mode of Delivery
1	1	Natural Language Processing	Presentation
	2	Linguistic Background	
	3	Mathematical Foundations	
2	1	Morphological Analysis.	Presentation
	2	Tokenization- Stemming	
	3	Lemmatization - Boundary Determination	
3	1	Reading unstructured data- - Representing text	Presentation Problem Solving
	2	Part of speech tagging - Shallow parsing	
	3	Syntactic representation	
4	1	Text similarity	Presentation, Problem Solving
	2	WordNet based similarity	
	3	Semantic representation.	
5	1	Information retrieval	Presentation, Problem Solving
	2	Information extraction -Named Entity Recognition,	
	3	Relation identification	
6	1	Event Extraction	Presentation, Problem Solving
	2	Language model - Probabilistic Models	
	3	n-gram language models	
7	1	Hidden Markov Model	Presentation,
	2	Topic Modelling	
	3	Graph Models	
8	1	Feature Selection and classifiers	Presentation
	2	Rule-based Classifiers	
	3	Maximum entropy classifier	
9	1	Clustering-Word and Phrase-based Clustering.	Presentation, Problem Solving, Demo
	2	Tools – Natural Language Tool kit.	
	3	Applications of Text Analytics	
10	1	Applications in Social media	Demo, Presentation
	2	Applications in Life science	
	3	Applications in Legal and Tax regulatory	
11	1	Visualization	Presentation, Discussion
	2	Case studies.	
	3	Assignment Presentations	
	4	Assignment Presentations	

COURSE ASSESSMENT METHODS

Sl.No.	Mode of Assessment	Week/Date	Duration	% Weightage
1	Test 1	After completion of 2 units	1 Hr	15%
2	Test 2	After completion of 4 units	1 Hr	15%
3	Assignment	Refer Instructions given in Moodle	Refer Instructions given in Moodle	20%
4	Semester	At the end of course	3 hrs	50%

ESSENTIAL READINGS

1. C. Manning and H. Schutze, "Foundations of Statistical Natural Language Processing", MIT press, 1999.
2. Steven Struhl, "Practical Text Analytics: Interpreting Text and Unstructured Data for Business Intelligence", Kogan Page, 2015.
3. Matthew A. Russell "Mining the Social Web", O'Reilly Media, 2013.

COURSE EXIT SURVEY (mention the ways in which the feedback about the course is assessed and indicate the attainment also)

- 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 plagiarism, academic honesty, attendance, etc.)

MODE OF CORRESPONDENCE

Through email: sangeetha@nitt.edu

PLAGIARISM

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

COMPENSATION ASSESSMENT POLICY

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. The prior permission and required document must be submitted for absence
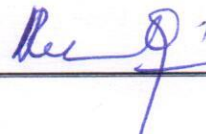
Attendance

100% is a must. However, relaxation upto 25% will be given for leave on emergency requirements (medical, death, etc.) and representing institute events.

For Approval

Course Faculty _S.SANGEETHA_ CC-Chairperson _____ HOD _S.R.Balasundaram_