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
Course Type	☐ Core course	■ Elective course	

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			
S.No.	**Week**	**Topic**	**Mode of Delivery**
1	1	Natural Language Processing Linguistic Background Mathematical Foundations	Presentation
2	2	Morphological Analysis Tokenization- Stemming Lemmatization - Boundary Determination.	Presentation
3	3	Reading unstructured data- - Representing text Part of speech tagging - Shallow parsing Syntactic representation	Presentation, Problem Solving
4	4	Text similarity WordNet based similarity Semantic representation.	Presentation, Problem Solving
5	5	Information retrieval, Information extraction Named Entity Recognition, Relation identification	Presentation, Problem Solving
6	6	Event Extraction, Language model - Probabilistic Models n-gram language models	Presentation, Problem Solving
7	7	Hidden Markov Model- Topic Modelling- Graph Models	Presentation, Problem Solving
8	8	Feature Selection and classifiers Rule-based Classifiers - Maximum entropy classifier	Presentation
9	9	Clustering-Word and Phrase-based Clustering. Tools – Natural Language Tool kit. Applications of Text Analytics	Presentation, Problem Solving, Demo
10	10	Applications in Social media Life science Legal and Tax regulatory	Demo, Presentation
11	11	Visualization Case studies.	Presentation, Discussion

COURSE ASSESSMENT METHODS

Sl.No.	Mode of Assessment	Week/Date	Duration	% Weightage
1	Test 1	Week 4	1 Hr	15%
2	Test 2	Week 8	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.)

- **Plagiarism**
  The students are expected to come out with their original solution for problems given as assignment, and tests/examinations.
- **Attendance**
  100% is a must. However, relaxation upto 25% will be given for leave on emergency requirements (medical, death, etc.) and representing institute events.

ADDITIONAL COURSE INFORMATION

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

FOR SENATE'S CONSIDERATION

Course Faculty _____ **CC-Chairperson** _____ **HOD** _____