



Faculty of Computer Studies (FoCS)

Course Name: Natural Language Processing

Course Code: T3568

(UG/PG): PG

Number of Credits: 3

Level: 4

Learning Objectives: The major focus of this course is to introduce Natural Language Processing so as to enable the students equip with office automation and to acquire the ability to process the historical documents familiarity with linear algebra is required.

Pedagogy:

Lectures
Class room discussion
Seminars

Course Outline:

Sr. No.	Topic	Hours
1	Introduction to Natural Language Processing: Structural features of texts in natural language, ambiguity on all levels of language, the main challenges of natural language processing, basic approaches to problem solving: manually written rules and machine learning.	6
2	Basic Text processing and Language Models: Preprocessing: tokenization and segmentation, normalization of words: stemming, lemmatization, morphological analyzers, regular expressions and edit distance. N-grams, perplexity, methods of smoothing, the use of language models, input prediction, error correction, speech recognition and text generation	10
3	Tagging problems and Hidden Markov Models: POS tagging, named entity recognition as a tagging problem, hidden Markov models and their advantages and disadvantages, the Viterbi algorithm	8
4	Parsing and Classification: Constituency and dependency trees, context-free grammar, probabilistic approach to parsing, lex-icalized PCFGs and CKY	10

	algorithm. Classification problems and naive Bayes classifier.	
5	Evaluation: Performance measures: accuracy, precision, recall and F-measure.	5
6	Case Study on Natural Language Processing Applications: Language Modeling, Text Summarization, Machine Translation and Computational Semantics.	6
	Total	45

Books Recommended:

1. Jurafsky D. Martin, J. "Speech and language processing: An introduction to speech recognition, computational linguistics and natural language processing", Prentice Hall, 2008.
2. Sarkar, D. Text Analytics with Python: A Practical Real-World Approach to gaining Actionable Insights from your Data. – Apress, 2016.

Suggested Evaluation Methods:

- 1) Multiple Choice Test
- 2) Test
- 3) Assignments

Parallel/Similar courses in the existing curriculum:

S.No.	Name of the course	Institute where it was offered
1	Speech and Language Processing	University of Edinburge
2	Natural Language Processing	IIT ,Mumbai
3	Natural Language Processing	Indian Institute of technology and Management

Name of Member	Dr. Dhandra	Dr. Gururaj			
Designation	Professor	Assistant Professor			
Org. / Inst.	SICSR	SICSR			
Signature					

Name of the Expert:

Signature:

Date: