CENTRAL UNIVERSITY OF KERALA
DEPARTMENT OF COMPUTER SCIENCE
M.Sc. COMPUTER SCIENCE

ELECTIVE COURSE					
COURSE CODE	COURSE TITLE	CONTACT HRS/WEEK			CREDITS
		LEC	LAB	TUT	
CSC5001	Natural Language Processing	2	2	1	4

Lec = Lecture, Tut = Tutorial, Lab = Practical

This is a participatory, experimental and problem solving **skill development course**.

Course Objective:
The objective of the course is to provide theoretical and practical aspects of natural language processing.

By completing this course, students will obtain the following course/learning outcomes:
1. Knowledge gained:
   (i) State of art methods and algorithms for natural language processing
2. Skill gained:
   (ii) Skills in applying statistical approaches in natural language processing
   (iii) Skills in develop language modelling
3. Competency gained:
   (iv) Expertise in developing natural language processing algorithms for real world applications

Prerequisites: Basic knowledge in logical reasoning

Grading:

Lab implementation	– 12%
Participatory based group Project	– 10%
Assignment/Quiz/presentation	– 8%
Class Test	– 10%
Final Exam	– 60%

CSC5001 – Natural Language Processing

Module 1: Morphology and Finite-State Transducers
Survey of (Mostly) English Morphology, Finite-State Morphological Parsing, Combining FST Lexicon and Rules, Lexicon-free FSTs: The Porter Stemmer, Human Morphological Processing

Module 2: Probabilistic Models of Pronunciation and Spelling
Dealing with Spelling Errors, Spelling Error Patterns, Detecting Non-Word, Probabilistic Models, Applying the Bayesian method to spelling, Minimum Edit Distance, English Pronunciation Variation, The Bayesian method for pronunciation, Weighted Automata, Pronunciation in Humans

Module 3: N-grams
Counting Words in Corpora, Simple (Unsmoothed) N-grams, Smoothing, Backoff , Deleted Interpolation, N-grams for Spelling and Pronunciation, Entropy

Module 4: HMMs and Speech Recognition
Speech Recognition Architecture, Overview of Hidden Markov Models, The Viterbi Algorithm Revisited, Advanced Methods for Decoding, Acoustic Processing of Speech, Computing Acoustic Probabilities, Waveform Generation for Speech Synthesis, Human Speech Recognition

Text Book:
1. Daniel Jurafsky and James H. Martin, *Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition*, Pearson Education Series in Artificial Intell., 2008.

References:
2. Allen, James, *Natural Language Understanding*, Second Edition, Benjamin/Cumming, 1995.
3. Manning, Christopher and Heinrich, Schutze, *Foundations of Statistical Natural Language Proc.*, MIT Press, 1999.