COMSATS University Islamabad
Department of Computer Science
Course Description Form (CDF)

Course Information

Course Code: **AIC365** | Course Title: **Natural Language Processing**
Credit Hours: **3(2,1)** | Lecture Hours/Week: **2**
Lab Hours/Week: **3** | Pre-Requisites: **None**

Course Objectives

- To discuss the basic concepts of Natural Language Processing (NLP);
- To discuss the different algorithms and techniques in NLP;
- To present NLP problem, students should be able to analyse, assess and justify which algorithms are most appropriate for solving the problem;
- To identify the suitable evaluation measures for a given problem.

Course Content

Overview; Approaches to NLP; NLP Pipeline; Zipf'law; Collocations ;Text Representation; Accessing Text Corpora & Lexical Resources; Processing Raw Text; Categorizing & Tagging Words; Text Classification;; N-grams Models; Word Sense Disambiguation; Hidden Markov Model & Word Guessing; Maximum Entropy; Part-of-Speech Tagging; Natural Language Parsing; Building Feature-Based Grammars; Natural Language Generation; Statistical Approaches; Text Alignment & Machine Translation; Information Retrieval & Information Extraction; Deep Learning for NLP; Applications of NLP.

Unit wise Major Topics:

Unit	Topic	No. of Teaching Hours
1.	Overview; Approaches to NLP; NLP Pipeline; Collocations; Concept of Linguistic Resources; Corpora & Lexical Resources; Vocabulary; and Zipf'law.	3
2	Text Representation; Accessing Text; Processing Raw Text; Categorizing & Tagging Words; Text Normalization & its Importance; Tokenization Issues: Regular Expressions, Basic Patterns in Regular Expressions, Writing a Tokenizer using Regular Expressions; Minimum Edit Distance Algorithm; Stemming; Sentence Segmentation; and Word Sense Disambiguation.	3
3	N-gram Language Models; Evaluation of Language Models; Concept of Data Sparseness; Smoothing: Add-one Smoothing, Add-k Smoothing Back-off & Interpolation, and Kneaser-Ney Smoothing.	3
4.	Text Classification: Text classification using Naïve Bayes Classifier, Evaluation of a Classifier by using Precision, Recall & F1- measure; Sigmoid Function & Logistic Regression for Text Classification; Sentiment analysis; Part of Speech Tagging: HMM Tagger & Viterbi Algorithm, and Named Entity Recognition.	6

5.	Natural Language Parsing; Building Feature-Based Grammars; Context Free Grammar; Grammar Equivalence & Normal Form; and CKY Parsing.	3
6.	Lexical semantics, vector semantics, vector representation of words, Cosine similarity, TF-IDF, PMI, WordNet, Word2Vec, learning classifier, skip-gram embedding, and Visualization of Embedding.	3
7.	Deep Learning for NLP; RNNs as Language Models: Managing Context in RNN using LSTM, The Encoder-Decoder Model with RNNs, and Attention Mechanism.	4.5
8.	Language Generation; Statistical Approaches; Self-Attention Networks; Transformers as Language Models; BERT; Fine-tuning of Large Language Models; Natural Text Alignment & Machine Translation; and Information Retrieval & Information Extraction.	4.5
Total Contact Hours		**30**

Mapping of CLOs and GAs

Sr.#	Unit #	Course Learning Outcomes	Blooms Taxonomy Learning Level	GA
CLO's for Theory				
CLO-1	1-2	Describe fundamental concepts of Natural Language Processing.	*Understanding*	2
CLO-2	3-4	Apply the concept of language modelling to solve a given problem.	*Applying*	3-4
CLO-3	5-6	Apply different Natural Language Processing techniques to evaluate syntax and semantics of the text.	*Applying*	3-4
CLO-4	7-8	Analyze the working of large language models.	*Analyzing*	2-4
CLO's for Lab				
CLO-5	1-8	Build an NLP application using a large language model.	*Creating*	2-5

CLO Assessment Mechanism

Assessment Tools	CLO-1	CLO-2	CLO-3	CLO-4	CLO-5
Quizzes	Quiz 1	Quiz 2	Quiz 3	Quiz 4	-
Assignments	Assignment 1	Assignment 2	Assignment 3	Assignment 4	-
Mid Term Exam	Mid Term Exam	Mid Term Exam	-	-	Lab Mid Term Exam
Final Term Exam	Final Term Exam				Project/ Lab Final Term Exam

Text and Reference Books

Textbook:

1. Speech and Language Processing, Jurafsky & Martin, 2023.

Reference Books:

2. Natural Language Processing in Action: Understanding, analyzing, and generating text with Python, Lane H, Hapke H, Howard C, Manning Publications, 2019.
3. Natural Language Processing with Python, Steven Bird, Ewan Klein, Edward Loper – Analyzing Text with the Natural Language Toolkit (O'Reilly 2009, website 2018).