SCHEME OF COURSE WORK

Department of Information Technology

Course Details:

COURSE TITLE	INTRODUCTION TO NATURAL LANGUAGE PROCESSING		
COURSE CODE	20ITH102	L T P C	3 1 0 4
PROGRAM	B.TECH		
SPECIALIZATION	CSE, IT		
SEMESTER	V		
PRE REQUISITES	Data Structures & Algorithms, Python Programming Lab, Numerical Methods, Probability and Statistics, Calculus and Linear Algebra.		
COURSES TO WHICH IT IS A PREREQUISITE	N/A		

Course Outcomes (COs):

1	Explain how NLP is applied in the real world. (L2)
2	Compare a traditional NLP pipeline and a DL-based NLP pipeline. (L2)
3	Select appropriate text representation from vectorization features, embeddings and handcrafted features. (L3)
4	Apply text classification to real world problems. (L3)
5	Build solutions for different Information Extraction tasks. (L3)

Course Outcome versus **Program Outcomes**

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	2	2	2		1										2
CO2	2	2	2		1										2
CO3	2	2	2		2										2
CO4	2	3	2		2										2
CO5	2	3	2		2										2

S - Strongly correlated, *M* - *Moderately* correlated, *Blank* - *No correlation*

Assessment Methods	Assignment / Quiz / Mid-Test /Seminar/viva

Teaching- Learning & Evaluation

Week	Topic/ Contents	Course Outcomes	Sample questions	Teaching learning strategy	Assessment method & schedule
1	NLP: A PRIMER: NLP in the Real World, NLP Tasks, What Is Language? Building Blocks of Language, Why Is NLP Challenging? Machine Learning, Deep Learning, and NLP: An Overview, Approaches to NLP,	CO1	1. A Explain various applications of NLP in the real world. 5M B Classify the building blocks of a language. 5M	Lecture	Assignment-1, Test- 1 Quiz-1
2	Heuristics-Based NLP, Machine Learning for NLP, Deep Learning for NLP, Why Deep Learning Is Not Yet the Silver Bullet for NLP, An NLP Walkthrough: Conversational Agents.	CO1	2. Differentiate between Machine Learning and Deep Learning 5M	Lecture	Assignment-1, Test- 1 Quiz-1
3	NLP PIPELINE: Data Acquisition, Text Extraction and Cleanup, HTML Parsing and Cleanup, Unicode Normalization, Spelling Correction, System-Specific Error Correction,	CO2	3. Explain various issues in Tokenization of text	Lecture	Assignment-1, Test- 1 Quiz-1
4	Pre-Processing, Preliminaries, Frequent Steps, Other Pre-Processing Steps, Advanced Processing Feature Engineering, Classical NLP/ML Pipeline, DL Pipeline, Modeling, Start with Simple Heuristics,	CO2	4. What are the various stages (phases) in a generic NLP pipeline? Explain each stage with examples.	Lecture	Assignment-1, Test- 1 Quiz-1
5	Building Your Model, Building THE Model, Evaluation, Intrinsic Evaluation Extrinsic	CO2	5. A. List and Explain Intrinsic evaluation methods. B. List and Explain Extrinsic evaluation	Lecture	Assignment-1, Test- 1 Quiz-1

	Evaluation, Post-Modeling Phases, Deployment, Monitoring, Model Updating Working with Other Languages, Case Study.		methods.		
6	TEXT REPRESENTATION: Vector Space Models, Basic Vectorization Approaches,	CO3	6. How Vector space model helps to find similarity between two documents.	Lecture	Assignment-1,2, Quiz-1, Test-1, 2
7	One-Hot Encoding, Bag of Words, Bag of N-Grams, TF-IDF, Distributed Representations,	CO3	7. Explain the process of TF-IDF calculation for a given term in a document with an example.	Lecture	Assignment-2, Test- 2, Quiz-2
8	TEST 1				
9	Word Embeddings, Going Beyond Words, Distributed Representations Beyond Words and Characters, Universal Text Representations, Visualizing Embeddings, Handcrafted Feature Representations.	CO3	8. Explain various challenged in word embedding.	Lecture	Assignment-2, Test- 2, Quiz-2
10	TEXT CLASSIFICATION: Applications, A Pipeline for Building Text Classification Systems, A Simple Classifier Without the Text Classification Pipeline, Using Existing Text Classification APIs, One Pipeline, Many Classifiers, Naive Bayes Classifier,	CO3, CO4	9. Draw the pipeline for text classification and explain the components.	Lecture	Assignment-2, Test- 2, Quiz-2
11	Logistic Regression, Support Vector Machine, Using Neural Embeddings in Text Classification. Word Embeddings, Subword Embeddings and fastText, Document Embeddings,	CO4	10. Illustrate logistic regression with an example.	Lecture	Assignment-2, Test- 2, Quiz-2
12	Deep Learning for Text Classification, CNNs for Text Classification, LSTMs for Text Classification, Text Classification with Large, Pre-Trained Language Models, Interpreting Text Classification	CO4	11. Draw the architecture of LSTM and explain its usage for text classification.	Lecture	Assignment-2, Test- 2, Quiz-2

	Models, Explaining Classifier Predictions with Lime,				
13	Learning with No or Less Data and Adapting to New Domains, No Training Data, Less Training Data: Active Learning and Domain Adaptation, Case Study: Corporate Ticketing.	CO4	12. List various types of learning based on the availability of data and explain.	Lecture	Assignment-2, Test- 2, Quiz-2
14	INFORMATION EXTRACTION: IE Applications, IE Tasks, The General Pipeline for IE, Keyphrase Extraction, Implementing KPE,	CO5	13. Describe the general process for Keyphrase extraction.	Lecture	Assignment-2, Test- 2, Quiz-2
15	Practical Advice, Named Entity Recognition, Building an NER System, NER Using an Existing Library, NER Using Active Learning, Practical Advice, Named Entity Disambiguation and Linking, NEL Using Azure API,	CO5	13. Define the following a)Named Entity Recognition. b) Named Entity Disambiguation c) Named Entity Linking	Lecture	Assignment-2, Test- 2, Quiz-2
16	Relationship Extraction, Approaches to RE, RE with the Watson API, Other Advanced IE Tasks, Temporal Information Extraction, Event Extraction, Template Filling, Case Study.	CO5	14. Write a short notes on Relationship extraction.	Lecture	Assignment-2, Test- 2, Quiz-2
17	TEST-2				