NIRMA UNIVERSITY

Institute:	Institute of Technology, School of Technology
Name of Programme:	MTech CSE, MTech CSE (Cyber Security), MTech CSE (Data Science)
Course Code:	6CS362ME25
Course Title:	Information Retrieval Systems
Course Type:	Department Elective-I
Year of Introduction:	2025-26

L	T	Practical Component				C
		LPW	PW	W	S	
3	0	2	-	-	-	4

Course Learning Outcomes (CLO):

At the end of the course, the students will be able to:

1. explain concepts, algorithms, data, and file structures to design an IR systems (BL2)
2. apply methodology for the design and evaluation of textual IR systems (BL3)
3. compare major types of IR systems based on theoretical foundations (BL4)
4. develop multimedia IR systems using relevant algorithms. (BL6)

Unit	Contents	Teaching Hours (Total 45)
Unit-I	**Introduction to IRS:** Motivation and Applications of Information Retrieval (IR) systems, Architecture of an IR system, overview of search engine functionality and types of search queries	04
Unit-II	**Document Representation and Indexing:** Text Processing (Tokenization, Stop Word Removal, Stemming), Building an inverted index, storage and compression, Zipf's Law, Information visualization using word cloud and histograms	06
Unit-III	**Retrieval Models:** Boolean, TF, TF-IDF models. Vector space formation of documents. Similarity measures and ranking (Euclidean, Cosine and Jaccard), Relevance Feedback (Rocchio method), Concept of Latent Semantic Indexing using Singular Value Decomposition, Evaluating performance of an IR system. Language Models (Basics of Bayes Theorem and Markov Models, n gram modeling)	10
Unit-IV	**Web Information Retrieval:** Link Analysis, Page Ranking, HITS. Introduction to Semantic Web.	05
Unit-V	**Machine Learning for Information Retrieval:** Naive Bayesian Classification for spam filtering and text classification. K nearest neighbours. Word2Vec and neural word embeddings. GloVe. Introduction to Recurrent Neural Networks. Clustering terms using documents (partitioning and hierarchical clustering).	12
Unit-VI	**Advanced Topics:** Summarization, Personalization (Recommender Systems using content based and collaborative filtering), Question Answering, Cross Language and Multimedia Information Retrieval, Retrieval Augmented Generation (RAG).	08

Self-Study:

The self-study contents will be declared at the commencement of the semester. Around 10% of the questions will be asked from self-study content.

Suggested Readings/ References:

1. C.D. Manning, P. Raghavan, H. Schütze, Introduction to Information Retrieval, Cambridge UP
2. R. Baeza-Yates, B. Ribeiro-Neto, Modern Information Retrieval, Addison-Wesley
3. B. Croft, D. Metzler, T. Strohman, Information Retrieval in Practice, Pearson Education
4. D.A. Grossman, O. Frieder, Information Retrieval: Algorithms and Heuristics, Springer
5. W.B. Croft, J. Lafferty, Language Modeling for Information Retrieval, Springer

Suggested List of Experiments:

Sr. No.	Name of Experiments/Exercises	Hours
1	Getting started with Kaggle and implementing MNIST Digits classification	02
2	Text Preprocessing using NLTK	02
3	Vector Space Model using CountVectorizer and TfIdf Vectorizer in Scikit-learn	02
4	Implementation of distance and similarity measures	02
5	Performance Evaluation of IR systems	02
6	Language Models using RNN and virtual labs on Natural Language Processing	04
7	Text Classification using naive Bayesian approach	04
8	Implementation of recommender systems	04
9	Multimedia Information Retrieval – Person Recognition using audio/video/image	04
10	Mini search engine using GUI.	04