

NIRMA UNIVERSITY

Institute:	Institute of Technology
Name of Programme:	BTech All (Other than CSE)
Course Code:	4CS106IE25
Course Title:	Information Retrieval Systems
Course Type:	Interdisciplinary Minor-Elective
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, students will be able to –

1. relate the concepts and various components of information retrieval systems (BL2)
2. apply theoretical foundations for the development of information retrieval systems (BL3)
3. choose design and evaluation parameters for information retrieval systems (BL5)
4. develop practical skills to handle and design information retrieval systems. (BL6)

Unit	Contents	Teaching Hours (Total 45)
Unit-I	Introduction to Information Retrieval Systems: Concept and architecture of Information Retrieval systems, Boolean Retrieval, the term vocabulary and posting lists, text processing - tokenization, stop words removal, stemming, lemmatization, posting lists intersection via skip pointers, positional posting lists, and phrase queries.	08
Unit-II	Data Structures: Dictionaries and tolerant retrieval - wildcard queries, spelling correction, phonetic correction. Inverted indexing, index construction - types of indices, Index compression - Heap's law and Zipf's law, dictionary compression.	08
Unit-III	Scoring and Vector Space Models and Language Models: Scoring, term weighting and vector space model, term frequency and weighting, TF-IDF weighting, Dot products, similarity measures for computing score, Evaluating IR systems, Relevance feedback - the Rocchio method, pseudo and indirect relevance feedback, query expansion, types of language models for information retrieval, the query likelihood model.	14
Unit-IV	Document Classification and Clustering: Text classification using naive Bayesian method, Bernoulli and multinomial models, evaluation of text classification, other classification techniques	08
Unit-V	Web-based IR: Web crawling, web search, and link analysis, Meta search engines, Multimedia IR - retrieving information from audio, video, and images	07

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 contents

Suggested Readings/ References:

1. Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze, An Introduction to Information Retrieval, Cambridge University Press

2. D.A. Grossman, O. Frieder, Information Retrieval: Algorithms and Heuristics, Springer
3. W.B. Croft, J. Lafferty, Language Modeling for Information Retrieval, Springer
4. G. Kowalski, M.T. Maybury, Information Storage and Retrieval Systems, Springer
5. Grigoris Antoniou and Frank van Harmelen, A Semantic Web Primer, The MIT Press
6. Croft, D. Metzler, T. Strohman, Information Retrieval in Practice, Pearson Education

Suggested List of Experiments:

Sr. No.	Title	Hours
1	Getting started with Kaggle and implementing MNIST Digits classification	04
2	Text Preprocessing using NLTK	04
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 and virtual labs on Natural Language Processing	04
7	Text Classification using a naive Bayesian approach	04
8	Implementation of meta-search algorithms	02
9	Multimedia Information Retrieval – Person Recognition using audio/video/image.	04
10	Mini search engine using GUI	02