

NIRMA UNIVERSITY

Institute:	Institute of Technology, School of Technology
Name of Programme:	BTech CSE
Course Code:	4CS102ME25
Course Title:	Information Retrieval Systems
Course Type:	Department Elective- III
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. correlate the concepts and various components of information retrieval systems (BL2)
2. demonstrate the usage of different data/file structures in building computational search engines (BL2)
3. apply theoretical foundations for the development of information retrieval systems (BL3)
4. identify design and evaluation parameters for information retrieval systems. (BL4)

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.	09
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.	09
Unit-IV	Document Classification and Clustering: Text classification using naive Bayesian method, Bernoulli and multinomial models, evaluation of text classification, other classification techniques	05
Unit-V	Web-based IR: Web crawling, web search and link analysis, Meta search engines, Multimedia IR - retrieving information from audio, video, and images, Cross Lingual IR- Language Problems in IR, Translation Approaches for CLIR, handling many Languages Using manually constructed Translation systems and resources for CLIR, Research issues	07

Unit-VI	Recommender Systems: Components of a recommender system, types of recommender systems, and evaluation measures in recommender system	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. Wani, Zahid Ashraf, and Huma Shafiq, *Online Information Retrieval Systems Trending from Evolutionary to Revolutionary Approach.*, Encyclopedia of Information Science and Technology, IGI Global.
2. Manning, C. D., Raghavan, P., & Schütze, H., *Introduction to information retrieval*, Cambridge: Cambridge University Press.
3. Grossman, D. A., *Information retrieval: Algorithms and heuristics*, Springer.
4. Chakrabarti, S., *Mining the Web: Discovering k knowledge from hypertext data*, Morgan Kaufmann.
5. Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman, *Mining of Massive Datasets*, Cambridge University Press.

Suggested List of Experiments:

Sr. No.	Name of Experiments/Exercises	Hours
1	Study various existing information retrieval systems and prepare a writeup by comparing various information retrieval systems based on different characteristics. Prepare a report(writeup) with the following details for each information retrieval system. Identify the domain of the given information retrieval system The data source Objective Characteristics Observation and Conclusion	02
2	Consider a static corpus of 10 documents from a political news domain. Implement a Boolean retrieval model by considering the term document incidence matrix. Your implemented Boolean model should retrieve the required document/documents by processing AND & OR queries. Display sparsity value of term-document incidence matrix.	02
3	Perform text preprocessing using NLTK.	02
4	Extract features and opinions from the product review to enhance the performance of the traditional recommender system. To extract features, consider frequent nouns, and to extract opinions, consider nearer words from a frequent noun. Use Part of Speech tagging as a preprocessing technique before extracting features and opinions from product reviews.	04
5	Consider a corpus of N documents. Implement Vector Space model (TFIDF considers normalized term frequency). Your implemented vector space model should rank the relevant retrieved documents by processing query.	04
6	Implement a program for document classification using Naive Bayes and Bernoulli. Compare the performance of both algorithms using suitable accuracy parameters.	02

- | | | |
|----|---|----|
| 7 | Implement a Web crawler using a simple queue data structure. | 04 |
| 8 | Implement a Traditional Recommender System using the Collaborative Filtering technique. The recommender system predicts a rating for an unknown item for a specific user. | 04 |
| 9 | Evaluation of Recommender system using RMSE (Root mean square error). | 02 |
| 10 | Implement the Soundex algorithm. Your code should generate four-character codes based on the pronunciation of English words. | 04 |