

NIRMA UNIVERSITY

Institute:	Institute of Technology
Name of Programme:	Master of Computer Application (2-Years Programme)
Course Code:	3MCAD359
Course Title:	Information Retrieval
Course Type:	Departmental Elective
Year of Introduction:	2021-22

Credit Scheme

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. summarize the concepts, algorithms, data/file structures necessary to design, and implement IR systems
2. apply methodology for the design and evaluation of IR systems
3. compare major types of IR systems, the different theoretical foundations underlying these systems
4. develop the practical skills for IR systems design

Syllabus:

Total Teaching hours: 45

Unit	Syllabus	Teaching hours
Unit-I	Introduction: 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 & 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 modelling)	10
Unit-IV	Web Information Retrieval: Link Analysis, Page Ranking, HITS. Introduction to Semantic Web, introduction to information retrieval from social websites	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

DS

Unit-VI **Advanced Topics:** Summarization, Personalization (Recommender Systems using content based and collaborative filtering), Question Answering, Cross Language and Multimedia Information Retrieval 08

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

- 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
 - 6.G. Kowalski, M.T. Maybury, Information Storage and Retrieval Systems, Springer

Suggested List of Experiments:	Sr. No.	Title	Hours
	1	To study and do the comparison of search engines	02
	2	To study the Text Pre-processing and it's tools exploration	02
	3	To study and understand Python's library for Text Mining	02
	4	To study and Implement different retrieval models and comparison of their accuracy by testing on certain documents	04
	5	To study and implement Dimensionality Reduction using Singular Mode Decomposition	02
	6	To classify a given set of labelled email documents, as spam or non-spam using Naïve Bayesian classifier	04
	7	To study and Implement collaborative filtering based Recommender System for a specific domain	04
	8	To study and implement Document clustering on webpages	02
	9	To study and implement Multimedia Information Retrieval – Person Recognition using audio, video or image	06
	10	To study and implement one of the practicals from Virtual Labs	02

Suggested Case List: -NA-