Nirma University
Institute of Technology, School of Technology
MTech Computer Science and Engineering (Data Science)

Semester – II

L	T	P	C
3	0	2	4

Course Code	3CS42D102
Course Name	Information Retrieval

Course Learning Outcomes (CLOs):

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

1. comprehend 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: **Teaching Hours**

Unit I **4**

Introduction: Motivation and Applications of Information Retrieval (IR) systems, Architecture of an IR system, overview of search engine functionality and types of search queries

Unit II **6**

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

Unit III **10**

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)

Unit IV **5**

Web Information Retrieval: Link Analysis, Page Ranking, HITS. Introduction to Semantic Web.

Unit V **12**

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).

Unit VI **8**

Advanced Topics: Summarization, Personalization (Recommender Systems using content

based and collaborative filtering), Question Answering, Cross Language and Multimedia Information Retrieval.

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.

Laboratory Work:

Laboratory work will be based on above syllabus with minimum 5 experiments to be incorporated.

Suggested Readings^:

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

L=Lecture, T=Tutorial, P=Practical, C=Credit

^this is not an exhaustive list