

PRASAD V. POTLURI SIDDHARTHA INSTITUTE OF TECHNOLOGY

(Autonomous)

Kanuru, Vijayawada-520007

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING (Data Science)

IV B Tech – I Semester

Information Retrieval Systems

Course Code	20DS4702D	Year	IV	Semester	I
Course Category	PEC	Branch	CSE (Data Science)	Course Type	Theory
Credits	3	L-T-P	3-0-0	Prerequisites	Data Structures
Continuous Internal Evaluation	30	Semester End Evaluation	70	Total Marks	100

Course Outcomes		
Upon Successful completion of course, the student will be able to		
CO1	Describe the core concepts and architecture of information retrieval systems to differentiate them from other information systems.	L2
CO2	Apply appropriate data structures and indexing techniques to build efficient text retrieval components in information retrieval systems.	L3
CO3	Use algorithms for stemming and thesaurus construction to improve the accuracy and relevance of text analysis and retrieval processes.	L3
CO4	Analyze various string searching algorithms to assess their benefits and constraints in information retrieval applications.	L4

[illegible]

PRASAD V. POTLURI SIDDHARTHA INSTITUTE OF TECHNOLOGY

(Autonomous)

Kanuru, Vijayawada-520007

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING (Data Science)**IV B Tech – I Semester**

Syllabus		
Unit No	Contents	Mapped CO
I	Introduction to Information storage and retrieval systems: Domain Analysis of IR systems, IR and other types of Information Systems, IR System Evaluation, Information Retrieval System Capabilities, Basics of Data Structures and Algorithms related to Information Retrieval Systems	CO1
II	Inverted Files and Signature Files: Introduction, Structures used in Inverted Files, building an Inverted file using a sorted array, Modifications to the Basic Techniques. Signature Files: Concepts of Signature files, Compression, Vertical Partitioning, Horizontal Partitioning.	CO1, CO2, CO4
III	Cataloguing and Indexing: History and Objectives of Indexing, Indexing Process, Automatic Indexing, New Indices for Text, Lexical Analysis and Stoplists Data Structures: N-Gram Data Structure, Hypertext and XML Data Structure	CO1, CO2, CO4
IV	Stemming Algorithms and Thesaurus Construction: Types of Stemming algorithms, Experimental Evaluations of Stemming, stemming to Compress Inverted Files. Thesaurus Construction: Features of Thesauri, Thesaurus Construction, Thesaurus construction from Texts	CO1, CO3, CO4
V	String Searching Algorithms: Introduction, Preliminaries, The Naive Algorithm, The Knutt-Morris-Pratt Algorithm, The Shift-Or Algorithm, The Karp-Rabin Algorithm.	CO1, CO3, CO4

Learning Resources
Text Books
1. Modern Information Retrieval, Ricardo Baeza-Yates & Berthier Ribeiro-Neto, 2nd ed., 1999, Addison-Wesley
2. Information Storage and Retrieval Systems: Theory and Implementation, Gerald J. Kowalski & Mark T. Maybury, 2nd ed., September 30, 2000, Springer (Kluwer Academic)
References
1. Information Retrieval: Implementing and Evaluating Search Engines, Stefan Büttcher, Charles L. A. Clarke, Gordon V. Cormack, 1st Edition, 2010, The MIT Press.
2. Information Retrieval: A Health and Biomedical Perspective, William Hersh, 4th Edition, 2022, Springer.
E-Recourses and other Digital Material
1. https://www.analyticsvidhya.com/blog/2015/04/information-retrieval-system-explained/
2. https://www.analyticsvidhya.com/blog/2017/11/information-retrieval-using-kdtree/
3. https://medium.com/analytics-vidhya/information-retrieval-part-1-extracting-webpages-a9d0b715535d