PVP 23

PRASAD V. POTLURI SIDDHARTHA INSTITUTE OF TECHNOLOGY

(Autonomous)

Kanuru, Vijayawada-520007

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING (AI&ML)

III B Tech – I Semester

Information Retrieval Lab

Course Code	**23AM3551**	**Year**	III	**Semester**	I
Course Category	PC	**Branch**	CSE (AI&ML)	**Course Type**	Practical
Credits	1.5	**L-T-P**	0-0-3	**Prerequisites**	Python Programming
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	Understand basic techniques in information retrieval and text preprocessing including document representation, stopword removal, and stemming.	**L2**
CO2	Apply core algorithms for text classification, clustering, and web crawling using standard datasets and real-world content using tools	**L3**
CO3	Analyze large-scale document collections and web content using indexing, graph-based ranking, and topic modeling techniques.	**L4**
CO4	Evaluate advanced techniques such as LSI, topic-specific PageRank, and Twitter mining to extract meaningful insights and trends.	**L5**

Contribution of Course Outcomes towards achievement of Program Outcome & Strength of correlation (3: High, 2: Medium, 1: Low)

	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PSO1**	**PSO2**
CO1	**2**								**2**				
CO2	**3**	**3**			**3**							**3**	
CO3	**3**	**3**										**3**	
CO4		**3**									**2**		

PVP 23

PRASAD V. POTLURI SIDDHARTHA INSTITUTE OF TECHNOLOGY

(Autonomous)
Kanuru, Vijayawada-520007

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING (AI&ML)

III B Tech – I Semester

Syllabus		
Exp. No.	**Contents**	**Mapped CO**
1	Representation of a Text Document in Vector Space Model and Computing Similarity between two documents.	CO1 to CO4
2	Pre-processing of a Text Document: stop word removal and stemming	CO1 to CO4
3	Construction of an Inverted Index for a given document collection comprising of at least 50 documents with a total vocabulary size of at least 1000 words.	CO1 to CO4
4	Classification of a set of Text Documents into known classes (You may use any of the Classification algorithms like Naive Bayes, Max Entropy, Rochio's, Support Vector Machine). Standard Datasets will have to be used to show the results.	CO1 to CO4
5	Text Document Clustering using K-means. Demonstrate with a standard dataset and compute performance measures- Purity, Precision, Recall and F-measure.	CO1 to CO4
6	Crawling/ Searching the Web to collect news stories on a specific topic (based on user input). The program should have an option to limit the crawling to certain selected websites only.	CO1 to CO4
7	To parse XML text, generate Web graph and compute topic specific page rank	CO1 to CO4
8	Implement Matrix Decomposition and LSI for a standard dataset.	CO1 to CO4
9	Mining Twitter to identify tweets for a specific period (and/or from a geographical location) and identify trends and named entities.	CO1 to CO4
10	Implementation of PageRank on Scholarly Citation Network	CO1 to CO4

Learning Resources
Text Books

PRASAD V. POTLURI SIDDHARTHA INSTITUTE OF TECHNOLOGY

(Autonomous)
Kanuru, Vijayawada-520007

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING (AI&ML)

III B Tech – I Semester

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