CSS304: INFORMATION RETRIEVAL SYSTEMS

Hours/Week: 4	I.A. Marks: 30
Credits: 4	Exam. Marks: 70

Course Learning Objectives: Students will able to try,

Enable students to understand the various aspects of an Information retrieval system and its evaluation and to be able to design.

This module aims to give students an understanding of the fundamental techniques for hypermedia architectures, design and usability, document management and retrieval, metadata management, and searching the web.

Analyze the performance of information retrieval using advanced techniques such as classification, clustering, and filtering over multimedia.

Analyze ranked retrieval of a very large number of documents with hyperlinks between them.

_____

Course Outcomes: After completing the course, the students will be able to,

CO1:   Understanding the basics of Information Retrieval

CO2:   Realize the data structures like Inverted Indices used in Information retrieval systems.

CO3:   Realize the concepts of agile methods and software testing.

CO4:   Learn the different techniques for compression of an index including the dictionary and its posting list.

CO5: Developing the ability of develop a complete IR system from Scratch.

C06: Understanding the data structures like Inverted Indices used in Information retrieval systems.

CO7: Understanding the different techniques for compression of an index including the dictionary and its posting list.

_____

UNIT-I   12 Hrs.

Introduction: Definition, Objectives, Functional Overview, Relationship to DBMS, Digital libraries and Data Warehouses. Information Retrieval System Capabilities: Search, Browse, Miscellaneous

UNIT-II   12 Hrs.

Cataloging and Indexing: Objectives, Indexing Process, Automatic Indexing, Information Extraction. Data Structures: Introduction, Stemming Algorithms, Inverted file structures, N-gram data structure, PAT data structure, Signature file structure, Hypertext data structure.

UNIT-III   12 Hrs.

Automatic Indexing: Classes of automatic indexing, Statistical indexing, Natural language, Concept indexing, Hypertext linkages Document and Term Clustering: Introduction, Thesaurus generation, Item clustering, Hierarchy of clusters.

UNIT-IV   12 Hrs.

User Search Techniques: Search statements and binding, Similarity measures and ranking, Relevance feedback, Selective dissemination of information search, weighted searches of Boolean systems, Searching the Internet and hypertext. Information Visualization: Introduction, Cognition

and perception, Information visualization technologies. Text Search Algorithms: Introduction, Software text search algorithms, Hardware text search systems. Information System Evaluation: Introduction, Measures used in system evaluation, Measurement example –TREC results.

REFERENCE BOOKS:

1. Kowalski, Gerald, Mark T Maybury: Information Retrieval Systems: Theory and Implementation, Kluwer Academic Press, 1997.
2. Frakes, W.B., Ricardo Baeza-Yates: Information Retrieval Data Structures and Algorithms, Prentice Hall, 1992.
3. Yates, Modern Information Retrieval, Pearson Education, 1999.
4. Robert Korfhage, Information Storage &Retrieval, John Wiley & Sons, 1997.