KOMMURI PRATAP REDDY INSTITUTE OF TECHNOLOGY

R18 JNTUH CSE III-I SEMESTER

CS523PE: INFORMATION RETRIEVAL SYSTEMS (Professional Elective - II)

ACADEMIC YEAR 2021-22 COURSE FILE

Prepared by

B.BUVANESH GOUD

Asst.Professor

Department of CSE

Vision of the Institute

To emerge as a premier institute for high quality professional graduates who can contribute to economic and social developments of the Nation.

Mission of the Institute

Mission	Statement
IM_1	To have holistic approach in curriculum and pedagogy through industry interface to meet the needs of Global Competency.
IM_2	To develop students with knowledge, attitude, employability skills, entrepreneurship, research potential and professionally ethical citizens.
IM_3	To contribute to advancement of Engineering & Technology that would help to satisfy the societal needs.
IM_4	To preserve, promote cultural heritage, humanistic values and spiritual values thus helping in peace and harmony in the society.

Vision of the Department

To Provide Quality Education in Computer Science for the innovative professionals to work for the development of the nation.

Mission of the Department

Mission	Statement
DM_1	Laying the path for rich skills in Computer Science through the basic knowledge of mathematics and fundamentals of engineering
DM_2	Provide latest tools and technology to the students as a part of learning infrastructure
DM_3	Training the students towards employability and entrepreneurship to meet the societal needs.
DM_4	Grooming the students with professional and social ethics.

Program Educational Objectives (PEOs)

PEO's	Statement
PEO1	The graduates of Computer Science and Engineering will have successful career in technology.
PEO2	The graduates of the program will have solid technical and professional foundation to continue higher studies.
PEO3	The graduate of the program will have skills to develop products, offer services and innovation.
PEO4	The graduates of the program will have fundamental awareness of industry process, tools and technologies.

Program Outcomes

PO1	**Engineering Knowledge:** Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
PO2	**Problem Analysis:** Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.
PO3	**Design/development of Solutions:** Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
PO4	**Conduct investigations of complex problems:** Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
PO5	**Modern tool usage:** Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modelling to

	complex engineering activities with an understanding of the limitations.
PO6	**The engineer and society:** Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.
PO7	**Environment and sustainability:** Understand the impact of the professional engineering solutions in societal and environmental context, and demonstrate the knowledge of, and need for sustainable development.
PO8	**Ethics:** Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
PO9	**Individual and team network:** Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
PO10	**Communication:** Communicate effectively on complex engineering activities with the engineering community and with society at large, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
PO11	**Project management and finance:** Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.
PO12	**Life-Long learning:** Recognize the need for, and have the preparation and able to engage in independent and life-long learning in the broadest context of technological change.

PROGRAM SPECIFIC OUTCOME'S:

PSO1	**Foundation of mathematical concepts**: To use mathematical methodologies to crack problem using suitable mathematical analysis, data structure and suitable algorithm.
PSO2	**Foundation of Computer Science:** The ability to interpret the fundamental concepts and methodology of computer systems. Students can understand the functionality of hardware and software aspects of computer systems.
PSO3	**Foundation of Software development:** The ability to grasp the software development lifecycle and methodologies of software systems. Possess competent skills and knowledge of software design process.

SYLLABUS

Prerequisites:

1.   Data Structures

UNIT - I

Introduction to Information Retrieval Systems: Definition of Information Retrieval System, Objectives of Information Retrieval Systems, Functional Overview, Relationship to Database Management Systems, Digital Libraries and Data Warehouses

Information Retrieval System Capabilities: Search Capabilities, Browse Capabilities, Miscellaneous Capabilities

UNIT - II

Cataloging and Indexing: History and Objectives of Indexing, Indexing Process, Automatic Indexing, Information Extraction

Data Structure: Introduction to Data Structure, Stemming Algorithms, Inverted File Structure, N-Gram Data Structures, PAT Data Structure, Signature File Structure, Hypertext and XML Data Structures, Hidden Markov Models

UNIT - III

Automatic Indexing: Classes of Automatic Indexing, Statistical Indexing, Natural Language, Concept Indexing, Hypertext Linkages

Document and Term Clustering: Introduction to Clustering, Thesaurus Generation, Item Clustering, Hierarchy of Clusters

UNIT - IV

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 to Information Visualization, Cognition and Perception, Information Visualization Technologies

UNIT - V

Text Search Algorithms: Introduction to Text Search Techniques, Software Text Search Algorithms, Hardware Text Search Systems

Multimedia Information Retrieval: Spoken Language Audio Retrieval, Non-Speech Audio Retrieval, Graph Retrieval, Imagery Retrieval, Video Retrieval

TEXT BOOK

1. Information Storage and Retrieval Systems – Theory and Implementation, Second Edition, Gerald J. Kowalski, Mark T. Maybury, Springer

REFERENCES

1. Frakes, W.B., Ricardo Baeza-Yates: Information Retrieval Data Structures and Algorithms, Prentice Hall, 1992.

2. Information Storage & Retrieval By Robert Korfhage – John Wiley & Sons.

3. Modern Information Retrieval By Yates and Neto Pearson Education.

COURSE OBJECTIVES

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

1. To learn the important concepts and algorithms in IRS

2. To understand the data/file structures those are necessary to design, and implement information retrieval (IR) systems.

COURSE OUTCOMES

Students will be able to:

   I. **C316.1:** Define the Information retrieval system and its objectives along with various capabilities. [Remember]

  II. **C316.2:** Understand to apply IR principles to locate relevant information from large collections of data using various indexing process and data structures. [Understand]

 III. **C316.3:** Implement to design different document clustering algorithms. [Apply]

 IV. **C316.4:** Analyze different retrieval systems for web search tasks. [Analyze]

  V. **C316.5:** Investigate various information retrieval systems which falls under Multimedia retrieval systems. [Create]

Mapping of Course Outcomes with PO's and PSO's:

High -3   Medium -2   Low-1

Course Outcomes	Program Outcomes												Program Specific Outcomes		
CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
C316.1	3	-	-	3	-	-	-	2	2	2	-	3	2	-	-
C316.2	-	3	-	-	-	-	-	1	-	2	-	-	-	3	-
C316.3	-	3	3	-	-	2	-	-	2	2	-	1	3	2	2
C316.4	2	2	2	-	-	-	1	2	1	-	-	-	-	-	2
C316.5	3	3	-	3	1	-	-	2	-	-	-	1	2	3	-

CO-PO Mapping Justification

C316.1: Define the Information retrieval system and its objectives along with various capabilities.

[Remember]

	Justification
PO1	Students will be able to know the concept of Mathematics.
PO4	Students will be able to solve complex problems and interpret the data.
PO8	Students will be able to define the responsibilities and ethics.
PO9	Students will be able to know how to work in a team.
PO10	Students will be able to communicate with the people effectively.
PO12	Students will be able to know how to learn the evolving techonologies.

C316.2: Understand to apply IR principles to locate relevant information from large collections of

data using various indexing process and data structures. [Understand]

	Justification
PO2	Students will able to identify different problems and they can analyse.
PO8	Students will be able know the principles of indexing and data structures.
PO10	Students will be able to communicate the different principles involved in indexing process & data structures.

C316.3: Implement to design different document clustering algorithms. [Apply]

	Justification
PO2	Students will be able to analyse the clustering algorithms.
PO3	Students will be able to develop the new algorithms as per the requirement.
PO6	Students will be able to assess the consequences while using the algorithms.
PO9	Students will be able to develop the different algorithms used in retrieval system.
PO10	Students will be able to communicate effectively in analyzing the algorithms.
PO12	Students will be able to learn the analysis of writing an algorithm in their own way.

C316.4: Analyze different retrieval systems for web search tasks. [Analyze]

	Justification
PO1	Students will be able to examine the web tasks.
PO2	Students will be able to identify the problem while search retrieval.
PO3	Students will be able to design new search techniques in retrieving information.
PO7	Students will be able to examine the professional knowledge regarding information retrieval.
PO8	Students will be able to develop the responsibilities in maintaining the searching techniques.
PO9	Students will be able to work in a team and analyze the retrieval problems.

C316.5: Investigate various information retrieval systems which fall under Multimedia retrieval systems. [Create]

	Justification
PO1	Students will be able to apply the knowledge of information retrieval by using different search based techniques.
PO2	Students will be able to analyze the different complex problems multimedia.
PO4	Students will be able to conduct and interpret data for gathering information.
PO5	Students will be able to create, select and apply the complex activities for retrieving the information.
PO8	Students will be able to apply the Multimedia design principles for information retrieval.
PO12	Students will be able to recognize the need of learning and can investigate the different retrieval technologies.

Lesson Plan - Information Retrieval Systems (CS523PE)

Faculty Name: B.Buvanesh Goud	Year/Sem: III/I	Academic Year: 2021-2022

w.e.f: 06-09-2021

L.No.	Name of the Topic	Reference Books	Delivery Method
1	**Unit-1: Introduction to Information Retrieval System**: Definition of Information Retrieval System	T1(1-3)	Chalk & Talk
2	Objectives of Information Retrieval Systems	T1(4-9)	Chalk & Talk
3	Functional Overview Item Normalization	T1(10-15)	Chalk & Talk
4	Selective Dissemination of Information	T1(16-17)	Chalk & Talk
5	Document Database Search, Index Database Search	T1(18-19)	Chalk & Talk
6	Multimedia Database Search, Relationship to DBMS	T1(20-21)	Chalk & Talk
7	Digital Libraries and Data Warehouses	T1(22-23)	Chalk & Talk
8	Information Retrieval System Capabilities: Introduction	T1(27-28)	Chalk & Talk
9	Search Capabilities	T1(28-37)	Chalk & Talk
10	Browse Capabilities	T1(38-41)	Chalk & Talk
11	Miscellaneous Capabilities	T1(41-48)	Chalk & Talk
12	Revision	T1(1-48)	Chalk & Talk
13	**Unit-2: Cataloging and Indexing** History of Indexing & Cataloging	T1(51-53)	Chalk & Talk
14	Objectives and Process of Indexing	T1(54-58)	Chalk & Talk
15	Automatic Indexing	T1(58-63)	Chalk & Talk
16	Indexing by Concept and Information Extraction	T1(63-68)	Chalk & Talk
17	Data Structure: Introduction to Data Structure	T1(71-73)	Chalk & Talk
18	Stemming Process and Algorithms, Inverted File Structure	T1(74-84)	Chalk & Talk

19	N-Gram Data Structures	T1(85-88)	Chalk & Talk
20	PAT Data Structure	T2(88-92)	Chalk & Talk
21	Signature File Structure	T1(93-94)	Chalk & Talk
22	Hypertext and XML Data Structures,	T1(95-99)	Chalk & Talk
23	Hidden Markov Models	T1(99-102)	Chalk & Talk
24	**Unit-3: Automatic Indexing** Introduction & Classes of Automatic Indexing	T1(105-107)	Chalk & Talk
25	Statistical Indexing, Probabilistic Weighting	T1(108-110)	Chalk & Talk
26	Vector Weighting, Bayesian model	T1(111-122)	Chalk & Talk
27	Natural Language, Index phrase generation	T1(123-127)	Chalk & Talk
28	Natural Language processing	T1(128-129)	Chalk & Talk
29	Concept Indexing	T1(130-131)	Chalk & Talk
30	Hypertext Linkages	T1(132-134)	Chalk & Talk
31	Summary and Revision	T1(105-135)	Chalk & Talk
32	Document and Term Clustering _Introduction	T1(139-142)	Chalk & Talk
33	Thesaurus Generation	T1(143)	Chalk & Talk
34	Manual and Automatic term Clustering	T1(144-153)	Chalk & Talk
35	Item Clustering	T1(154-156)	Chalk & Talk
36	Hierarchy of Clusters	T1(156-159)	Chalk & Talk
37	Summary and Revision	T1(139-160)	Chalk & Talk
38	**UNIT_IV :** User Search Techniques	T1(165)	Chalk & Talk
39	Search Statements and Binding	T1(166-167)	Chalk & Talk
40	Similarity Measures	T1(167-173)	Chalk & Talk
41	Ranking & Ranking Algorithms	T1(174-175)	Chalk & Talk
42	Relevance Feedback	T1(175-179)	Chalk & Talk
43	Selective Dissemination of Information Search	T1(179-185)	Chalk & Talk

44	Weighted Searches of Boolean Systems	T1(186-190)	Chalk & Talk
45	Searching the INTERNET and Hypertext	T1(191-193)	Chalk & Talk
46	Summary	T1(139-195)	Chalk & Talk
47	Information Visualization: Introduction	T1(199-202)	Chalk & Talk
48	Cognition and Perception	T1(203-204)	Chalk & Talk
49	Aspects of the Visualization process	T1(204-207)	Chalk & Talk
50	Information Visualization Technologies	T1(208-217)	Chalk & Talk
51	Summary and Revision	T1(191-218)	Chalk & Talk
52	**UNIT-V Text Search Algorithms _Introduction**	T1(221)	Chalk & Talk
53	Text search techniques	T1(221-224)	Chalk & Talk
54	Software Text Search Algorithms	T1(225-226)	Chalk & Talk
55	Boyer-Moore Algorithm	T1(226-229)	Chalk & Talk
56	Aho-Corasick Algorithm	T1(230-233)	Chalk & Talk
57	Hardware Text Search Systems	T1(233-235)	Chalk & Talk
58	Fast Data Finder Architecture	T1(236-238)	Chalk & Talk
59	Summary	T1(221-239)	Chalk & Talk
60	**Multimedia Information Retrieval**	T1(241-242)	Chalk & Talk
61	Spoken Language Audio Retrieval	T1(242-243)	Chalk & Talk
62	Non-Speech Audio Retrieval	T1(244-245)	Chalk & Talk
63	Graph Retrieval	T1(245-246)	Chalk & Talk
64	Imagery Retrieval	T1(246-249)	Chalk & Talk
65	Video Retrieval	T1(249-255)	Chalk & Talk
66	Summary	T1(241-255)	Chalk & Talk
67	Review		Talk

Web Reference:

https://en.wikipedia.org/wiki/Information_retrieval

https://www.sciencedirect.com/topics/computer-science/information-retrieval-systems

https://nlp.stanford.edu/IR-book/

https://nios.ac.in/media/documents/SrSecLibrary/LCh-015B.pdf

TIME TABLE

Class: III-B.Tech I Sem   A.Y: 2021-22   W.E.F- 06-09-2021   LH:- B-202

	I 9:30-10:20	II 10:20 - 11:10	11:10-11:20	III 11:20-12:10	IV 12:10-1:00	1:00-1:40	V 1:40-2:30	VI 2:30-3:20	VII 3:20-4:10
MON	SE	FLAT		IPR	PPL		CN	WT	LIB/INT
TUE	PPL	IRS	**B**	SEED CLASSES		**L**	FLAT	SE	WT
WED	SEED CLASSES		**R**	AI	IRS	**U**	SE – CN/WT LAB		
THU	SEED CLASSES		**E**	CN	SE	**N**	IPR	PPL	SPORTS
FRI	WT	CN	**A**	FLAT	IRS	**C**	CN/WT– SE LAB		
SAT	CN	SE	**K**	WT	FLAT	**H**	AI	ACS LAB	

UNIVERSITY CALENDAR

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD

Academic Calendar 2021-22

B. TECH./B.PHARM. III & IV YEARS I & II SEMESTERS

I SEM

S. No	Description	Duration	
		From	To
1	Commencement of I Semester classwork	06.092021	
2	1st Spell of Instructions (including Dussehra Recess)	06.09.2021	06.11.2021 (9 Weeks)
3	Dussehra Recess	11.10.2021	16.10.2021 (1 Week)
4	First Mid Term Examinations	08.11.2021	13.11.2021 (1 Week)
5	Submission of First Mid Term Exam Marks to the University on or before	20.11.2021	
6	2nd Spell of Instructions	15.11.2021	08.01.2022 (8 Weeks)
7	Second Mid Term Examinations	10.01.2022	18.01.2022 (1 Week)
8	Preparation Holidays and Practical Examinations	19.01.2022	25.01.2022 (1 Week)
9	Submission of Second Mid Term Exam Marks to the University on or before	25.01.2022	
10	End Semester Examinations	27.01.2022	09.02.2022

The following Table contains other resources like Question Bank

Name of the Content	Resource URL
Assignment -1	https://drive.google.com/file/d/1bY5crHp36zYWTCPF_F84Jqe3aSmdpJpY/view?usp=sharing
Assignment-2	https://drive.google.com/file/d/1qKHTLvRDcs-zlmt_-qP3p-3hM7u9tSuV/view?usp=sharing
Question Bank	https://drive.google.com/file/d/1a-PhkU3Eahrl5zNdRernOZvSkHnotOtq/view?usp=sharing
Digital Notes Unit-I_Part-1	https://drive.google.com/file/d/10ZILTegkYn6YFuDOwQQXcyqrGWwwbw6G/view?usp=sharing
Digital Notes Unit-I_Part-2	https://drive.google.com/file/d/1Ow7BHx1AUBDcGDcdPuSXiGpCIIclLpNy/view?usp=sharing
Digital Notes Unit-II_Part-1	https://drive.google.com/file/d/1wBLEkOahNibeIXwKxq_oJGlmzqv5JAJR/view?usp=sharing
Digital Notes Unit-II_Part-2	https://drive.google.com/file/d/1s8GeqmSLFNVUoXsGB3bAL2I51iIms2Uq/view?usp=sharing
Digital Notes Unit-III_Part-1	https://drive.google.com/file/d/1Y1tbSe8Wk26OF3v6fuJbnjuddB1QObUk/view?usp=sharing

Digital Notes Unit-III_Part-2	**https://drive.google.com/file/d/1e6bBIy3EtHAdpG_Emk4fUp4i RQZD2BRl/view?usp=sharing**
Digital Notes Unit-IV	**https://drive.google.com/file/d/1osUrRZwRU6zxP5CUk0EpPZ MBJ_GvV4VG/view?usp=sharing**
Digital Notes Unit-V	**https://drive.google.com/file/d/1_b8bRwT-EK8lpFoPHX7DzFBO9ZkWPxUx/view?usp=sharing**