Devi Ahilya University, Indore, India Institute of Engineering & Technology				IV Year B.E. (Computer Engineering) (Full Time)			
Subject Code & Name	**Instructions Hours per Week**			**Credits**			
8CERC1	**L**	**T**	**P**	**L**	**T**	**P**	**Total**
Information Retrieval & Extraction	**3**	**1**	**0**	**3**	**1**	**0**	**4**
Duration of Theory Paper:3 Hours							

Learning Objectives:

- To understand challenges, scale and approaches for Information Retrieval system.
- To study structure and components of Information Retrieval systems.
- To understand design of Information Retrieval system by study of different data structures and algorithms used in design..
- To study means of measuring performance and effectiveness of Information Retrieval system and techniques for improvement.
- To understand Information Extraction and inherent challenges.

Pre requisites: Understanding of Data Structures and Algorithms.

COURSE CONTENTS

UNIT-I

Introduction:Goals and history of IR. The impact of the web on IR., Boolean retrieval: Processing Boolean queries, The extended Boolean model versus ranked retrieval, The term vocabulary & postings lists: Document delineation and character sequence decoding, Determining the vocabulary of terms, Positional postings and phrase queries.

UNIT-II

Dictionaries and tolerant retrieval: Search structures for dictionaries, Wildcard queries, Spelling correction, Phonetic correction, Index Construction: Hardware basics, Blocked sort-based indexing, Single-pass in-memory indexing, Distributed indexing, Dynamic indexing, Index compression**:** Statistical properties of terms in information retrieval, Dictionary compression, Postings file compression.

UNIT-III

Scoring, term weighting & the vector space model: Parametric and zone indexes, Term frequency and weighting, The vector space model for scoring, Variant tf-idf functions.

Computing scores in a complete search system: Efficient scoring and ranking, Components of an information retrieval system, Vector space scoring and query operator interaction.

UNIT-IV

Evaluation in information retrieval: Information retrieval system evaluation, Standard test collections, Evaluation of unranked retrieval sets, Evaluation of ranked retrieval results, Assessing relevance, Results snippets**,** Relevance feedback and query expansion, XML retrieval.

UNIT-V

Information Extraction and Other Issues: Language models for information retrieval, Flat clustering, Hierarchical clustering, Web search basics, Web crawling, Information extraction: Task and evaluation.

Course Outcomes:

CO No.	**Course Objective**
CO1	To have a clear understanding of the design of Information Retrieval system.
CO2	To understand the build process of indexes and their efficient storage and usage for Information Retrieval tasks.
CO3	Design and code components of Information Retrieval system.
CO4	Understand evaluation of performance and effectiveness of Information Retrieval System.
CO5	To have an understanding of the working of Web Search system.
CO6	To be able to understand Information Extraction task.

CO-PO Relationship:

	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9	P10	P-11	P12
CO1	**3**	**-**	**2**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**
CO2	**3**	**3**	**3**	**2**	**2**	**-**	**-**	**-**	**-**	**-**	**-**	**-**
CO3	**3**	**3**	**3**	**3**	**3**	**-**	**-**	**-**	**-**	**-**	**-**	**-**
CO4	**3**	**3**	**-**	**-**	**2**	**-**	**-**	**-**	**-**	**-**	**-**	**-**
CO5	**3**	**-**	**3**	**3**	**2**	**2**	**-**	**-**	**-**	**-**	**-**	**-**
CO6	**3**	**2**	**2**	**3**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**

BOOKS RECOMMENDED:

[1] Christopher D. Manning, Prabhakar Raghavan and Hinrich Schutze, I*ntroduction to Information Retrieval*, Cambridge University Press Cambridge, 2014.
[2] Bruce Croft, Donald Metzler and Trevor Strohman, *Search Engines: Information Retrieval in Practice*. Addison Wesley, 2009.