Course Title: Information Retrieval   **Credit: 3**
Course No: CSIT.425.1   **Number of period per week: 3+3**
Nature of the Course: Theory + Lab   **Total hours: 45+45**
Year: Fourth, Semester: Eighth
Level: B. Sc. CSIT

1. Course Introduction

This is a undergraduate-level introductory course for information retrieval. It will cover algorithms, design, and implementation of modern information retrieval systems. Topics include: retrieval system design and implementation, text analysis techniques, retrieval model, search evaluation, retrieval feedback, applications in web information management.

2. Objectives

The Student should be made to:
- Learn the information retrieval models
- Be familiar with Web Search Engine
- Be exposed to Link Analysis
- Understand Hadoop and Map Reduce
- Learn document text mining techniques

3. Specific Objectives and Contents

Specific Objectives	Contents
• Define Information Retrieval and discuss it components • Understand framework of search engines • Differentiate IR from Web Search	**Unit I: Introduction (6 hr)** 1.1. Introduction, History of IR, Components of IR, Issues ,Open source Search engine Frameworks 1.2. The impact of the web on IR, The role of artificial intelligence (AI) in IR, 1.3. IR Versus Web Search, Components of a Search engine, Characterizing the web
• Discuss Boolean and Vector Space Model • Demonstrate TF-IDF Weighting and cosine similarity • Exemplify probabilistic IR and LSI • Describe Relevance feedback and query expansion	**Unit II: Information Retrieval (12 hr)** 2.1. Boolean and vector-space retrieval models, Term weighting – TF-IDF weighting, cosine similarity 2.2. Preprocessing, Inverted indices, efficient processing with sparse vectors 2.3. Language Model based IR, Probabilistic IR, Latent Semantic Indexing 2.4. Relevance feedback, Pseudo-relevance feedback and query expansion

	Unit III: Web Search Engine-Crawling (8 Hrs)
• Understand structure of web and optimization ideas • Discuss architectures of we and crawling • Demonstrate web indexes and index compression	3.1. Web search overview, web structure, the user, paid placement, Search engine optimization/spam 3.2. Web size measurement, search engine optimization/spam 3.3. Web Search Architectures, crawling, meta-crawlers Focused Crawling, web indexes 3.4. Near-duplicate detection, Index Compression, XML retrieval.
	Unit IV: Web Search (10 Hrs)
• Demonstrate Link Analysis techniques and HITS algorithm • Discuss Searching and Ranking techniques • Exemplify relevance scoring and ranking of web search result • Demonstrate Recommendation generation algorithms	4.1. Link Analysis, hubs and authorities, Page Rank and HITS algorithms, Searching and Ranking 4.2. Relevance Scoring and ranking for Web, Similarity, Hadoop & Map Reduce, Evaluation 4.3. Personalized search, Collaborative filtering and content-based recommendation of documents and products, handling invisible Web 4.4. Snippet generation, Summarization, Question Answering, Cross- Lingual Retrieval.
	Unit V: Document Text Mining(9 Hrs)
• Understand basics of document text mining • Demonstrate text classification, clustering and categorization algorithms	5.1. Information filtering; organization and relevance feedback 5.2. Text Mining, Text classification and clustering, Categorization algorithms: naive Bayes; decision trees; and nearest neighbor – 5.3. Clustering algorithms: agglomerative clustering; k-means; expectation maximization (EM).

Evaluation System

Undergraduate Programs							
External Evaluation	**Marks**	**Internal Evaluation**	**Weight age**	**Marks**	**Practical**	**Weight age**	**Mark**
End semester examination		Assignments	20%		Practical Report copy	25%	
(Details are given in the separate table at the end)	60	Quizzes	10%	20	Viva	25%	20
		Attendance	20%		Practical Exam	50%	
		Internal Exams	50%				
Total External	60	Total Internal	100%	20		100%	20
Full Marks 60+20+20 = 100							

External evaluation

1. **End semester examination:**

   It is a written examination at the end of the semester. The questions will be asked covering all the units of the course. The question model, full marks, time and others will be as per the following grid.

2. **External Practical Evaluation:**

   After completing the end semester theoretical examination, practical examination will be held. External examiner will conduct the practical examination according to the above mentioned evaluation. There will be an internal examiner to assist the external examiner. Three hours time will be given for the practical examination. In this examination Students must demonstrate the knowledge of the subject matter.

Full Marks: 100, Pass Marks: 45, Time: 3 Hrs

Nature of question	Total questions to be asked	Total questions to be answered	Total marks	Weightage
Group A: multiple choice*	20	20	20×1 = 20	60%
Group B: Short answer type questions	7	6	6×8 = 48	60%
Group C: Long answer type questions	3	2	2×16 =32	60%
			100	100%

Each student must secure at least 50% marks in internal evaluation in order to appear in the end semester examination. Failed student will not be eligible to appear in the end semester examinations.

Internal evaluation

Assignment: Each student must submit the assignment individually. The stipulated time for submission of the assignment will be seriously taken.

Quizzes: Unannounced and announced quizzes/tests will be taken by the respective subject teachers. Such quizzes/tests will be conducted twice per semester. The students will be evaluated accordingly.

Attendance in class: Students should regularly attend and participate in class discussion. Eighty percent class attendance is mandatory for the students to enable them to appear in the end semester examination. Below 80% attendance in the class will signify NOT QUALIFIED (NQ) to attend the end semester examination.

Presentation: Students will be divided into groups and each group will be provided with a topic for presentation. It will be evaluated individually as well as group-wise. Individual students have to make presentations   on   the   given   topics.

Mid-term examination: It is a written examination and the questions will be asked covering all the topics in the session of the course.

Discussion and participation: Students will be evaluated on the basis of their active participation in the classroom discussions.

Instructional Techniques: All topics are discussed with emphasis on real-world application. List of instructional techniques is as follows:
- Lecture and Discussion
- Group work and Individual work
- Assignments
- Presentation by Students
- Quizzes
- Guest Lecture

Students are advised to attend all the classes and complete all the assignments within the specified time period. If a student does not attend the class(es), it is his/her sole responsibility to cover the topic(s) taught during that period. If a student fails to attend a formal exam/quiz/test, there won't be any provision for re-exam. Unless and until the student clears one semester he/she will not be allowed to study in the following semesters.

Laboratory Work

Student should implement IR algorithms discussed in the course by using weighting and similarity measures. Students also need to practice Web Search, Clustering, Classification, and Recommendation Generation Algorithms.

Prescribed Text

- *C. Manning, P. Raghavan, and H. Schütze*, Introduction to Information Retrieval, Cambridge University Press, 2008.
- *Ricardo Baeza, Yates and Berthier Ribeiro, Neto*, Modern Information Retrieval: The Concepts and Technology behind Search 2nd Edition, ACM Press Books 2011.
- *Bruce Croft, Donald Metzler and Trevor Strohman*, Search Engines: Information Retrieval in Practice, 1st Edition Addison Wesley, 2009.
- *Mark Levene*, An Introduction to Search Engines and Web Navigation, 2nd Edition Wiley, 2010.