Course code	Course Name	L-T-P - Credits	Year of Introduction
IT362	Information Retrieval	3-0-0-3	2017

Pre-requisites: CS205 Data structures

Course Objectives

- To provide with foundation knowledge in information retrieval.
- To equip with sound skills to solve computational search problems.

Syllabus

Introduction to the Concepts of Information Retrieval, Retrieval models, Searching the web and Parallel and Distributed Information Retrieval systems.

Expected outcome .

The students will be able to

i. use different information retrieval techniques in various application areas
ii. apply IR principles to locate relevant information large collections of data and analyse performance of retrieval systems when dealing with unmanaged data sources
iii. implement retrieval systems for web search tasks.

Text Books:

1. C. Manning, P. Raghavan, and H. Schütze, "Introduction to Information Retrieval", Cambridge University Press, 2008.
2. C.J. Van Rijsbergen , Information Retrieval:, http://www.dcs.gla.ac.uk/Keith/Preface.html
3. Ricardo Baexa-Yates and Berthier Ribeiro-Neto, "Modern Information Retrieval", Addison Wesley Longman, 1999.

References:

1. Bruce Croft, Donald Metzler and Trevor Strohman, "Search Engines: Information Retrieval in Practice", 1st Edition Addison Wesley, 2009.
2. Manu Konchady, "Building Search Applications: Lucene, Ling Pipe", First Edition, Gate Mustru Publishing, 2008.
3. Mark Levene, "An Introduction to Search Engines and Web Navigation", 2nd Edition Wiley, 2010.
4. Ophir Frieder, "Information Retrieval: Algorithms and Heuristics: The Information Retrieval Series", 2nd Edition, Springer, 2004.
5. Stefan Buettcher, Charles L. A. Clarke, Gordon V. Cormack, "Information Retrieval: Implementing and Evaluating Search Engines", The MIT Press, 2010.

Course Plan

Module	Contents	Hours	Sem. Exam Marks
I	Introduction – Information versus Data Retrieval. Modeling of Information retrieval. Boolean Model, Vector Model, Probabilistic Model, Set Theoretical Models, Structured Text Retrieval Models.	7	15%
II	Classification, Measures of Association, Cluster Hypothesis, Single Link Clusters, File Structures, Inverted Files, Index Sequential Files, Ring Structures, Doubly Chained Trees, Hash Addressing.	7	15%
	FIRST INTERNAL EXAMINATION		
III	Evaluation, Relevance, Precision and Recall, Interpolation, Averaging techniques, The Swets Model.	7	15%
IV	Search Engines, Boolean Search, Matching Functions, Serial Search, Cluster Representatives, Cluster based retrieval.	7	15%
	SECOND INTERNAL EXAMINATION		
V	Web search basics – Web characteristics - crawling and indexes – Features of a crawler – Crawler architecture – DNS	7	20%

	resolution – The URL frontier – Distributing indexes – Connectivity servers.		
VI	Link Analysis – The Web as a graph – Anchor text and the web graph, PageRank – Markov chains, Page Rank computation, Topic-specific Page Rank, Hubs and authorities.	7	20%
END SEMESTER EXAM			

QUESTION PAPER PATTERN

Maximum Marks: 100 Exam Duration: 3 hours

The question paper shall consist of Part A, Part B and Part C.

Part A shall consist of three questions of 15 marks each uniformly covering Modules I and II. The student has to answer any two questions (15×2=30 marks).

Part B shall consist of three questions of 15 marks each uniformly covering Modules III and IV. The student has to answer any two questions (15×2=30 marks).

Part C shall consist of three questions of 20 marks each uniformly covering Modules V and VI. The student has to answer any two questions (20×2=40 marks).

Note : Each question can have a maximum of 4 subparts, if needed