Birla Institute of Technology & Science, Pilani
Hyderabad Campus

INSTRUCTION DIVISION
SECOND SEMESTER 2015-2016
Course Handout Part II

Date: 11-05-2015

In addition to part-I (General Handout for all courses appended to the time table) this portion gives further specific details regarding the course.

Course No. : *CS F469*
Course Title : *Information Retrieval*
Instructor-in-Charge : Dr. Aruna Malapati (arunam@bits-hyderabad.ac.in)
Co-Instructor: : Dr. Lavika Goel (lavika.goel@pilani.bits-pilani.ac.in)

Scope and Objective of the Course:

Textbooks:

1. T1. C. D. Manning, P. Raghavan and H. Schutze. Introduction to Information Retrieval, Cambridge University Press, 2008.

Reference books

1. R1: Modern Information Retrieval, Ricardo Baeza-Yates and Berthier Ribeiro-Neto, Addison-Wesley, 2000. http://people.ischool.berkeley.edu/~hearst/irbook/
2. R2: Search Engines: Information Retrieval in Practice by Bruce Croft, Donald Metzler, and Trevor Strohman, Addison-Wesley, 2009.
3. R3: Cross-Language Information Retrieval by By Jian-Yun Nie Morgan & Claypool Publisher series 2010.
4. R4: Multimedia Information Retrieval by Stefan M. Rüger Morgan & Claypool Publisher series 2010.
5. R5 Information Retrieval: Implementing and Evaluating Search Engines by S. Buttcher, C. Clarke and G. Cormack, MIT Press, 2010.
6. R6: Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data by B. Liu, Springer, Second Edition, 2011.
7. R7: Ricci, F.; Rokach, L.; Shapira, B.; Kantor, P.B. (Eds.), Recommender Systems Handbook. 1st Edition., 2011, 845 p. 20 illus., Hardcover, ISBN: 978-0-387-85819-7

Course Plan:

Lecture No.	Learning objectives	Topics to be covered	Chapter in the Text Book
1	To understand the reason to study this course	Introduction and Motivation	
2-3	The term vocabulary postings lists and Introduction to ad-hoc search	Boolean retrieval	T1 Ch 1 & 2,R1 Ch2 section 5
4-5	Understand the importance of	Wildcard queries,	T1 Ch 3

	Dictionaries and tolerant retrieval	Spelling correction, Edit distances and Phonetic correction	
6-8	To be able to identify and implement suitable Index and Compression techniques	Blocked sort-based indexing Single-pass in-memory indexing Distributed indexing Dynamic indexing Parametric and zone indexes Weighted zone scoring	T1 Ch 4
9-11	Understand the importance of Scoring, term weighting	Parametric and zone indexes Weighted zone scoring Learning weights Term frequency and weighting Tf-idf weighting	T1 Ch 6
12-13		Dot products, Queries as vectors, Variant tf-idf functions, Document and query weighting schemes	T1 Ch 6
14-15		Probabilistic Information retrieval	T1 Ch 11
16	Open Challenges in text retrieval		
17	To be able to identify the issues while working with languages other than English	European Languages East Asian Languages Other Languages	R3 Ch 1
18-20	Ability to implement a CLIR using the IBM model	Translation Approaches for CLIR Handling many Languages Using manually constructed Translation systems and resources for CLIR	R3 Ch2
21	Research issues in CLIR		
22-25	To be able to identify the issues while working with multimedia like Image, Audio and video	Basic Multimedia search technologies Content based retrieval Image and Audio data challenges Multimedia IR Research	R4 Ch2,3
26	Research issues in MIR		
27-35	Differentiate between different recommender systems and suggest a suitable system based on the problem and data available.	Introduction to recommendation system Collaborative , Content based recommendation Hybrid recommendation systems	R7 Ch1,2,3,4,5
35-38	Understand the Architecture for a specific search engine.	Search Engine Architecture Web characteristics Advertising as the economic model Index size and estimation	T1 Ch 19
38-40	To be able to implement their own crawler depending on the problem.	Crawling Crawler architecture	T1 Ch 20

		Distributing indexes	
41-42	Under the famous algorithms for Pagerank etc.	The Web as a graph, Google's Pagerank Hub and authorities (HITS), Web spam, SEO	T1 Ch 21

Evaluation Scheme:

Component	Duration	Weightage (%)	Date & Time	Nature of Component
Programming Assignments	Take Home Open Book		To be announced	**30%**
Mid Sem exam	Closed Book	90min	6/10 2:00 - 3:30 PM	**30%**
Comprehensive	Partially Open book / Closed Book	3 hours	4/12 FN	**40%**

Chamber Consultation Hour: At Pilani Campus, every Monday 4:00 PM - 5:00 PM

Notices: All notices related to the course will be displayed on the **CSIS Notice Board**, and / or course website.

Make-up Policy:
Make ups for tests shall be granted by the I/C on prior permission and only to genuine cases with the permission of the warden concerned.
Make-up for comprehensive examination will be decided and scheduled by the Instruction Division.

INSTRUCTOR-IN-CHARGE