

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI
INSTRUCTION DIVISION
First Semester 2016-2017
COURSE HANDOUT (PART II)

In addition to Part-I (general handout for all courses appended to this time table) this portion gives further details pertaining to the course.

Course No.: CS F469

Course Title: **Information Retrieval**

Instructor-in-charge: **POONAM GOYAL** (poonam@pilani.bits-pilani.ac.in)

1. Objective and Scope

This course studies the theory, design, and implementation of text-based information systems. The Information Retrieval core components of the course include statistical characteristics of text, representation of information needs and documents, several important retrieval models (Boolean, vector space, probabilistic, inference net, language modeling, link analysis), clustering algorithms, collaborative filtering, automatic text categorization, and experimental evaluation. The software architecture components include design and implementation of high-capacity text retrieval and text filtering systems. The course is designed to provide students with a broad understanding in the design and use of information retrieval techniques. The course also aims at providing a holistic view of information retrieval.

2. Text Book

- **T1:** C. D. Manning, P. Raghavan and H. Schutze. Introduction to Information Retrieval, Cambridge University Press, 2008. <http://nlp.stanford.edu/IR-book/>

3. Reference Books

- **R1:** Modern Information Retrieval, Ricardo Baeza-Yates and Berthier Ribeiro-Neto, Addison-Wesley, 2000. <http://people.ischool.berkeley.edu/~hearst/irbook/>
- **R2:** Search Engines: Information Retrieval in Practice by Bruce Croft, Donald Metzler, and Trevor Strohman, Addison-Wesley, 2009.
- **R3:** Cross-Language Information Retrieval by By Jian-Yun Nie Morgan & Claypool Publisher series 2010
- **R4:** Multimedia Information Retrieval by Stefan M. Rüger Morgan & Claypool Publisher series 2010.
- **R5** 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

4. Course Plan

4a. Modules and Learning Objectives

Module	Title	Learning Objective(s)
M1	Basic information retrieval concepts	To understand what Information retrieval is and how to represent data in Boolean form and index the data.
M2	Text and vector space classification	To understand Scoring, Term Weighting, the Vector Space Model and scoring in the complete search system.
M3	Recommender systems	To understand components of a recommender system, types of recommender systems and evaluation measures in recommender systems.
M4	Web search and link analysis	To understand architecture of search engines, crawlers and the web graph.

4b. Lecture Schedule

L. No	Topic	Learning Objective	Module	Reading
1	Course overview	Introduction to the course	M1	T2 Ch1
2-4	Boolean retrieval	The term vocabulary postings lists and introduction to ad-hoc search	M1	T1 Ch 1 & 2, R1 2.5
5-6	Dictionaries and tolerant retrieval	Wildcard queries Spelling correction Edit distances Phonetic correction	M1	T1 Ch 3
7-9	Index construction and Compression	Blocked sort-based indexing Single-pass in-memory indexing Distributed indexing Dynamic indexing	M2	T1 Ch 4
10-11	Scoring, term weighting	Parametric and zone indexes Weighted zone scoring	M2	T1 Ch 6
		Learning weights Term frequency and weighting Tf-idf weighting		
12-13	The vector space model for scoring	Dot products, Queries as vectors, Variant tf-idf functions, Document and query weighting schemes	M2	T1 Ch 6
14-18	Text Mining	Text Classification Text Clustering	M2	T1 Ch 13, 14, 16,17
19-21	Cross Language Information Retrieval (CLIR)	Language Problems in IR Translation Approaches for CLIR Handling many Languages Using manually constructed Translation systems and resources for CLIR Research issues	M2	R3 Ch2
22-26	Multimedia Information retrieval(MIR)	Basic Multimedia search technologies Content based Retrieval Research issues in MIR	M2	R4 Ch2,3
27-29	Recommender systems	Introduction to recommendation system Collaborative, Content, Knowledge and Hybrid recommendation systems	M3	R5 Ch1,2,3,4,5
30-34	Web search basics	Search Engine Architecture Web characteristics	M4	T1 Ch 19 R1 Ch13,

		Advertising as the economic model The search user experience Index size and estimation		R2 Ch2
35-37	Web crawlers and indexes	Crawling Crawler architecture Distributing indexes	M4	T1 Ch 20 R2 Ch 3
39-41	Link Analysis	The Web as a graph, Google's Pagerank Hub and authorities (HITS) Web spam, SEO	M4	T1 Ch 21

5. Evaluation Schedule

Component	Duration	Weightage(%)	Date & Time	Venue	Remarks
Mid Sem Exam	90 Mins.	30			Closed Book
Labs/Assignments		30			To be announced
Comprehensive	3 Hours	40			Partly open

6. Assignments

Assignment(s) (programming/reading) will be given to the students. This will immensely help the students in gaining a better understanding of the subject.

7. Chamber Consultation Hours

To be announced in the class.

8. Make-up Policy

Prior Permission is must and Make-up shall be granted only in genuine cases based on individual's need and circumstances.

8. Notices

All the notices concerning this course will be displayed on the **CSIS notice board** or **course website**.

Instructor-in-charge