

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI
HYDERABAD CAMPUS
FIRT SEMESTER 2016 – 2017
Course Handout (Part II)

01-08-2016

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 Reterival
Instructor-in-charge: Dr. Aruna Malapati (arunam@hyderabad.bits-pilani.ac.in)

1. Scope and Objectives

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), collaborative filtering, Language translation and Multimedia information retrieval.

2. Pre requisites: Programming in Java or C, and knowledge of core data structures and algorithms.

3.a. Text Book

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

3.b. Reference Books

- **R1:** Modern Information Retrieval, Ricardo Baeza-Yates and Berthier Ribeiro-Neto, Addison-Wesley, 2000. <http://people.ischool.berkeley.edu/~heerst/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. R ger Morgan & Claypool Publisher series 2010.
- **R5** Information Retrieval: Implementing and Evaluating Search Engines by S. Buttcher, C. Clarke and G. Cormack, MIT Press, 2010.
- **R6:** Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data by B. Liu, Springer, Second Edition, 2011.
- **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

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	To understand components of a recommender

	systems	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.

4.b. Lecture Schedule

SI No	Topic	Learning Objective	Lecture No	Module	Reading
1.	Course overview	Introduction to the course	1	M1	T1 Ch1
2.	Boolean retrieval	The term vocabulary postings lists	2,3	M1	T1 Ch 1 & 2,R1 Ch2 section 5
		Introduction to ad-hoc search	4		
3.	Dictionaries and tolerant retrieval	Wildcard queries Spelling correction	5	M1	T1 Ch 3
		Edit distances Phonetic correction	6		
4.	Index construction and Compression	Blocked sort-based indexing	7	M1	T1 Ch 3
		Single-pass in-memory indexing	8		
		Distributed indexing Dynamic indexing	9		
5.	Scoring, term weighting	Parametric and zone indexes Weighted zone scoring	10	M2	T1 Ch 6
		Learning weights Term frequency and weighting Tf-idf weighting	11		
6.	The vector space model for scoring	Dot products, Queries as vectors, Variant tf-idf functions,	12	M2	T1 Ch 6
		Document and query weighting schemes	13		
7.	Language Problems in IR	European Languages	14	M2	R3 Ch 1
		East Asian Languages			
		Other Languages			
8.	Cross Language Information Retrieval(CLIR)	Translation Approaches for CLIR	15	M2	R3 Ch2
		Handling many Languages	16		
		Using manually constructed Translation systems and resources for CLIR	17		
9.	Multimedia Information retrieval(MIR)	Basic Multimedia search technologies	18	M2	R4 Ch2,3
		Content based retrieval	19,20		
		Image and Audio data challenges	21,22		
		Multimedia IR Research	23		

10.	Recommender systems	Introduction to recommendation system	24	M3	R7 Ch1,2,3,4,5
		Collaborative , Content based recommendation	25		
		Hybrid recommendation systems	26		
11.	Web search basics	Search Engine Architecture	27,28	M4	T1 Ch 19 R1 Ch13, R2 Ch2
		Web characteristics	29		
		Advertising as the economic model	30		
		Index size and estimation	31		
		Crawler architecture	33,34		
		Distributing indexes	35,36		
12.	Link Analysis	The Web as a graph, Google’s Pagerank	37,38	M4	T1 Ch 21
		Hub and authorities (HITS), Web spam, SEO	38,40		
Total number of classes Planned			40		

5. Evaluation Scheme

6.a. Major Components

Component	Nature of Component	Duration	Date	Weightage
Project	Open Book / Take Home		To be announced in class	25%
Test 1	Closed Book	60 mins	9/9, 1.00--2.00 PM	20%
Test 2	Closed Book	60 mins	24/10, 1.00--2.00 PM	20%
Comprehensive	Closed Book	3 hours	05/12 AN	35%

5.b. Timeliness

- Assignments are to be completed in time with no postponements.
- Submissions 24 hours from deadline will have a penalty of 2 Marks per day.

6. Chamber Consultation: To be announced.

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

8. Make-up Policy: Make ups shall be granted on prior permission and only to genuine cases on submission of proofs from respective wardens.

Instructor-in-charge
CS F469