BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI
WORK INTEGRATED LEARNING PROGRAMMES

COURSE HANDOUT

Part A: Content Design

Course Title	Information Retrieval
Course No(s)	SS ZG537
Credit Units	4 (Unit split between Class Hours + Lab/Design/Fieldwork + Student preparation) Ex. 1-1-2, (total 4 units or credits) ie 1 unit for class room hours, 1 unit for lab hours, 2 units for student preparation. Typically 1 unit translates to 32 hours
Course Author	Poonam Goyal
Version No	1.0
Date	

Course Objectives

No	Course Objective
CO1	To understand structure and organization of various components of an IR system
CO2	To understand information representation models, term scoring mechanisms, etc. in the complete search system
CO3	To understand architecture of search engines, crawlers and the web search
CO4	To understand cross lingual retrieval and multimedia information retrieval

Text Book(s)

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

Reference Book(s) & other resources

R1	Modern Information Retrieval, Ricardo Baeza-Yates and Berthier Ribeiro-Neto, Addison-Wesley, 2000. http://people.ischool.berkeley.edu/~hearst/irbook/
R2	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
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	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.

Modular Content Structure

1. Introduction
    - 1.1. Information Retrieval
    - 1.2. Basic Search Model
2. Basic Information Retrieval Concepts
    - 2.1. Boolean Retrieval
    - 2.2. Dictionaries and Tolerant Retrieval
    - 2.3. Index Construction and Compression
3. Vector Space Model
    - 3.1. Scoring, Term Weighting
    - 3.2. The Vector Space Model for Scoring
4. Text Mining
    - 4.1. Text Classification
    - 4.2. Vector Space Classification
    - 4.3. Text Clustering
5. Web Search
    - 5.1. Web Search Basics
    - 5.2. Web Crawlers and Indexes
    - 5.3. Link Analysis
6. Cross Lingual Retrieval
    - 6.1. Language Problems in IR
    - 6.2. Approaches for CLIR
7. Multimedia Information Retrieval
    - 7.1. Multimedia Search Technologies
    - 7.2. Content Based Retrieval
8. Recommender Systems
    - 8.1. Collaborative and Content Based RS

Learning Outcomes:

No	Learning Outcomes
LO1	Students will gain understanding about an information retrieval system as a whole and about its components.
LO2	Students will have knowledge about the design issues and their solutions of different type of models including Boolean, vector space etc.
LO3	Students will have detailed understanding about text indexing, mining, weighting schemes etc.
LO4	Students will acquire knowledge about cross lingual and multimedia information retrieval.
LO5	With the acquired knowledge students will be able to design and build different kind of information retrieval systems.

Part B: Contact Session Plan

Academic Term	
Course Title	Information Retrieval
Course No	
Lead Instructor	

Contact Hour	List of Topic Title (from content structure in Part A)	Topic # (from content structure in Part A)	Text/Ref Book/external resource
1-2	● Introduction o Information Vs Data Retrieval o Basic Concepts o The retrieval process o Taxonomy of IR o Classic IR and Alternative models	1.1, 1.2	R1 Ch1, Ch2
3-5	● Boolean Retrieval o Inverted index o Processing Boolean queries o Boolean Vs Ranked retrieval o Term vocabulary and postings lists o Phrase queries	2.1	T Ch2
6-7	● Dictionary and Tolerant Retrieval o Search Structures for dictionaries o Wildcard queries o Phonetic Correction	2.2	T Ch3
8-10	● Index Construction and Compression o Blocked sort-based Indexing o Single pass in-memory indexing	2.3	T Ch4

	o Distributed and dynamic indexing o Dictionary comparison o Postings file compression		
11-12	● Vector Space Model o Term frequency and weighting o The vector space model for scoring o Tf-idf functions	3.1, 3.2	T Ch6
13-15	● Text Mining o Classification ● Naïve Bayes ● Vector space classification ● Evaluating Classification	4.1	T Ch13, 14
16-18	● Text Mining o Clustering ● Flat clustering ● Hierarchical clustering	4.2	T Ch16, 17
19	● Web Search o Web characteristics o The search user experience o Index size and estimation	5.1	T Ch19
20-22	● Web Crawling and Indexes o Crawling o Crawler Architecture o Distributed Indexes	5.2	T Ch20
23-24	● Link Analysis o The web as a graph o Google's page rank o Hub and Authorities (HITS)	5.3	T Ch21
25-27	● Cross Lingual IR (CLIR) o Language problems in IR o Translation Approaches o Handling Many Languages o Resources for CLIR	6.1, 6.2	R3 Ch2
28-29	● Multimedia IR o Basic Multimedia search technologies o Content Based Retrieval	7.1,7.2	R4 Ch2,3
30-31	● Recommender System o Collaborative recommendation o Content based recommendation o Other type & hybrid recommendations	8.1	R2 Ch1-5
32	● Review		

Work integration: Detailed plan

No	Activity description (Examples are given below)
1	Apply Domain modelling concept to the work you are doing in the work place
2	Present the architecture of the software you are working on
3	Analyse the test plan of the software project you are working on and identify areas where it can be further improved
4	Seminar / talk by Project manager in the company on a topic of relevance to the course

Evaluation Scheme

Evaluation Component	Name (Quiz, Lab, Project, Mid term exam, End semester exam, etc)	Type (Open book, Closed book, Online, etc.)	Weight	Duration	Day, Date, Session, Time
EC - 1	Quiz	Online	10%		To be announced
	Assignment	Take home	10%		To be announced
EC - 2	Mid-Semester Test	Closed Book	30%	2 hrs	To be announced
EC - 3	Comprehensive Exam	Open Book	50%	3 hrs	To be announced

Note *- Evaluation components can be tailored depending on the proposed model.*

Syllabus for Mid-Semester Test (Closed Book): Topics in Weeks 1-7
Syllabus for Comprehensive Exam (Open Book): All topics given in plan of study

Evaluation Guidelines:

1. EC-1 consists of either two Assignments or three Quizzes. Announcements regarding the same will be made in a timely manner.
2. For Closed Book tests: No books or reference material of any kind will be permitted. Laptops/Mobiles of any kind are not allowed. Exchange of any material is not allowed.
3. For Open Book exams: Use of prescribed and reference text books, in original (not photocopies) is permitted. Class notes/slides as reference material in filed or bound form is permitted. However, loose sheets of paper will not be allowed. Use of calculators is permitted in all exams. Laptops/Mobiles of any kind are not allowed. Exchange of any material is not allowed.
4. If a student is unable to appear for the Regular Test/Exam due to genuine exigencies, the student should follow the procedure to apply for the Make-Up Test/Exam. The genuineness of the reason for absence in the Regular Exam shall be assessed prior to giving permission to appear for the Make-up Exam. Make-Up Test/Exam will be conducted only at selected exam centres on the dates to be announced later.

It shall be the responsibility of the individual student to be regular in maintaining the self-study schedule as given in the course handout, attend the lectures, and take all the prescribed evaluation components such as Assignment/Quiz, Mid-Semester Test and Comprehensive Exam according to the evaluation scheme provided in the handout.