



**BITS Pilani**  
Pilani | Dubai | Goa | Hyderabad

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

### **COURSE HANDOUT**

#### **Part A: Content Design**

<b>Course Title</b>	Information Retrieval
<b>Course No(s)</b>	
<b>Credit Units</b>	4 (Unit split between Class Hours + Lab/Webinar)
<b>Course Author</b>	Dr. Chetana Gavankar
<b>Version No</b>	1.0
<b>Date</b>	June 2022

#### **Course Objectives**

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

#### **Text Book(s)**

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

#### **Reference Book(s) & other resources**

<b>R1</b>	Modern Information Retrieval, Ricardo Baeza-Yates and Berthier Ribeiro-Neto, Addison-Wesley, 2000. <a href="http://people.ischool.berkeley.edu/~hearst/irbook/">http://people.ischool.berkeley.edu/~hearst/irbook/</a>
<b>R2</b>	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
<b>R3</b>	Cross-Language Information Retrieval by By Jian-Yun Nie Morgan & Claypool Publisher series 2010
<b>R4</b>	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.

### **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. Evaluation in Information Retrieval
  - 5.1. Evaluation in unranked retrieval sets
  - 5.2. Evaluation in ranked retrieval sets
6. Web Search
  - 6.1. Web Search Basics
  - 6.2. Web Crawlers and Indexes
  - 6.3. Link Analysis
7. Cross Lingual Retrieval
  - 7.1. Language Problems in IR
  - 7.2. Approaches for CLIR
  - 7.3. Neural
8. Multimedia Information Retrieval
  - 8.1. Multimedia Search Technologies
  - 8.2. Content Based Retrieval
9. Recommender Systems
  - 9.1. Collaborative and Content Based RS
10. Neural IR
  - 10.1 Introduction to Deep neural network
  - 10.2 Deep neural network for IR

**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**

<b>Academic Term</b>	Mtech DSE June 2021
<b>Course Title</b>	Information Retrieval
<b>Course No</b>	DSECLZG537
<b>Lead Instructor</b>	Dr. Chetana Gavankar

Contact Session	List of Topic Title (from content structure in Part A)	Topic # (from content structure in Part A)	Text/Ref Book/external resource
1	<ul style="list-style-type: none"> <li>● Introduction <ul style="list-style-type: none"> <li>○ Information Vs Data Retrieval</li> <li>○ Basic Concepts</li> <li>○ The retrieval process</li> <li>○ Taxonomy of IR</li> <li>○ Classic IR and Alternative models</li> </ul> </li> </ul>	1.1, 1.2	R1 Ch1, Ch2
2	<ul style="list-style-type: none"> <li>● Boolean Retrieval <ul style="list-style-type: none"> <li>○ Inverted index</li> <li>○ Processing Boolean queries</li> <li>○ Boolean Vs Ranked retrieval</li> <li>○ Term vocabulary and postings lists</li> <li>○ Phrase queries</li> </ul> </li> </ul>	2.1	T1 Ch 1, 2
3	<ul style="list-style-type: none"> <li>● Dictionary and Tolerant Retrieval <ul style="list-style-type: none"> <li>○ Search Structures for dictionaries</li> <li>○ Wildcard queries</li> <li>○ Phonetic Correction</li> </ul> </li> </ul>	2.2	T1 Ch3

4	<ul style="list-style-type: none"> <li>● Index Construction and Compression <ul style="list-style-type: none"> <li>○ Blocked sort-based Indexing</li> <li>○ Single pass in-memory indexing</li> <li>○ Distributed and dynamic indexing</li> <li>○ Dictionary comparison</li> <li>○ Postings file compression</li> </ul> </li> </ul>	2.3	T1 Ch4,5
5	<ul style="list-style-type: none"> <li>● Vector Space Model <ul style="list-style-type: none"> <li>○ Term frequency and weighting</li> <li>○ The vector space model for scoring</li> <li>○ Tf-idf functions</li> </ul> </li> </ul>	3.1, 3.2	T1 Ch6
6	<ul style="list-style-type: none"> <li>● Classification &amp; Clustering for IR <ul style="list-style-type: none"> <li>○ Feature Selection</li> <li>○ Vector space classification</li> <li>○ Document Representation</li> <li>○ Rocchio classification</li> <li>○ Evaluating Classification</li> </ul> </li> </ul>	4.1	T1 Ch13, 14 T1 Ch16, 17
7	<ul style="list-style-type: none"> <li>● Evaluation in Information Retrieval <ul style="list-style-type: none"> <li>○ Evaluation in unranked retrieval sets</li> <li>○ Evaluation in ranked retrieval sets</li> </ul> </li> </ul>		T1 Chp 8
8	<ul style="list-style-type: none"> <li>● Review</li> </ul>		
9	<ul style="list-style-type: none"> <li>● Web Search <ul style="list-style-type: none"> <li>○ Web characteristics</li> <li>○ The search user experience</li> <li>○ Index size and estimation</li> </ul> </li> </ul>	5.1	T1 Ch19
10	<ul style="list-style-type: none"> <li>● Web Crawling and Indexes <ul style="list-style-type: none"> <li>○ Crawling</li> <li>○ Crawler Architecture</li> <li>○ Distributed Indexes</li> </ul> </li> </ul>	5.2	T1 Ch20
11	<ul style="list-style-type: none"> <li>● Link Analysis <ul style="list-style-type: none"> <li>○ The web as a graph</li> <li>○ Google's page rank</li> <li>○ Hub and Authorities (HITS)</li> </ul> </li> </ul>	5.3	T1 Ch21
12	<ul style="list-style-type: none"> <li>● Cross Lingual IR (CLIR) <ul style="list-style-type: none"> <li>○ Language problems in IR</li> <li>○ Translation Approaches</li> <li>○ Handling Many Languages</li> <li>○ Resources for CLIR</li> </ul> </li> </ul>	6.1, 6.2	R3 Ch2
13	<ul style="list-style-type: none"> <li>● Multimedia IR <ul style="list-style-type: none"> <li>○ Basic Multimedia search technologies</li> <li>○ Content Based Retrieval</li> </ul> </li> </ul>	7.1,7.2	R4 Ch2,3
14	<ul style="list-style-type: none"> <li>● Recommender System <ul style="list-style-type: none"> <li>○ Collaborative recommendation</li> </ul> </li> </ul>	8.1	R2 Ch1-5

	<ul style="list-style-type: none"> <li>o Content based recommendation</li> <li>o Other type &amp; hybrid recommendations</li> </ul>		
15	<b>NEURAL IR</b> 15.1.Introduction to Deep neural networks 15.1.1Input text representation 15.1.2 Standard architectures 15.2 Deep neural networks for IR 15.2.1 Document auto encoders 15.2.2 Siamese networks 15.2.3 Interaction based network 15.2.4 Lexical and semantic matching 15.2.5 Matching with multiple document fields	6&7	An Introduction to Neural Information Retrieval by Bhaskar Mitra, Nick Craswell Link: <a href="https://www.microsoft.com/en-us/research/uploads/prod/2017/06/fntir2018-neuralir-mitra.pdf">https://www.microsoft.com/en-us/research/uploads/prod/2017/06/fntir2018-neuralir-mitra.pdf</a>
16	<ul style="list-style-type: none"> <li>• Review</li> </ul>		

#### **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 1	Online	5%		

	Quiz2	Online	5%		
	Assignment	Take home	20%		
<b>EC – 2</b>	Mid-Semester Test	Closed Book	30%		
<b>EC – 3</b>	Comprehensive Exam	Open Book	40%		

**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.