



**FIRST SEMESTER 2019-20**  
**COURSE HANDOUT**

**Date: 02.08.2019**

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. Lavika Goel (lavika.goel@pilani.bits-pilani.ac.in)

**1. Course description:**

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 and multimedia retrieval and filtering systems.

**2. Scope and Objective of the Course:**

1	To understand structure and organization of various components of an IR system
2	To understand information representation models, term scoring mechanisms, etc. in the complete search system
3	To understand architecture of search engines, crawlers and the web search
4	To understand cross lingual retrieval and multimedia information retrieval
5	To understand the concepts of Recommender Systems.

**3. Reference books:**

R1	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>
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
R8	Koehn P., "Statistical Machine Translation", Cambridge University Press, 2010.
R9	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> )

#### 4. Course Plan:

Module no.	Lecture Session		Learning Outcomes	Reference
M1: Basic Information Retrieval concepts	1	Reason to study this course	Introduction and Motivation	R1 Ch1, Ch2
	2-3	The term vocabulary postings lists and Introduction to ad-hoc search	Boolean retrieval	R9 Ch. 1 & 2, R1 Ch2 section 5
	4-5	Dictionaries and tolerant retrieval	Wildcard queries, Spelling correction, Edit distances and Phonetic correction	R9 Ch. 3
	6-8	Index and Compression techniques	Blocked sort-based indexing Single-pass in-memory indexing Distributed indexing Dynamic indexing Parametric and zone indexes Weighted zone scoring	R9 Ch. 4
	9-11	Scoring, term weighting	Parametric and zone indexes Weighted zone scoring Learning weights Term frequency and weighting Tf-idf weighting	R9 Ch. 6





**BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, Pilani**  
**Pilani Campus**  
**AUGS/ AGSR Division**

	12-13		Dot products, Queries as vectors, Variant tf-idf functions, Document and query weighting schemes	R9 Ch. 6
M2: Text mining	14-17	Text mining	Text Mining: Classification, Naïve Bayes, Vector space classification, Evaluating Classification, Clustering, Flat clustering, Hierarchical Clustering	R9 Ch13, 14, 16, 17
M3: Cross language Information Retrieval	18	Open Challenges in text retrieval Identification of the issues while working with languages other than English	European Languages East Asian Languages Other Languages	R3 Ch. 1
	19-23	Implementing a CLIR using the IBM model Research issues in CLIR	Translation Approaches for CLIR Handling many Languages Using manually constructed Translation systems and resources for CLIR	R3 Ch.2 R8: Ch. 4, 5, 6
M4: Recommender Systems	24-30	Recommender systems and suggesting a suitable system based on the problem and data available.	Introduction to recommendation system Collaborative, Content based recommendation, Hybrid recommendation systems.	R7 Ch. 1,2,3,4,5
M5: Multimedia Information Retrieval	31-37	Working with multimedia like Image, Audio and video Research issues in MIR	Basic Multimedia search technologies Content based retrieval Image and Audio data challenges Multimedia IR Research	R4 Ch. 1,2,3



**Save Paper.  
Save Trees.  
Save the World.**





**BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, Pilani**  
**Pilani Campus**  
**AUGS/ AGSR Division**

M6: Web search and Link Analysis	38-40	Web Searching	Web Search Basics, Web Crawlers and Indexes Link Analysis: The web as a graph, Google's page rank, hubs and authorities.	R9 Ch. 19, 20, 21
----------------------------------	-------	---------------	---	-------------------

**5. Evaluation Scheme:**

Component	Nature of component	Duration	Date & Time	Weightage (%)
Quizzes/ Assignments	To be announced	To be announced	To be announced	30%
Midsem exam	Closed Book	90min	<TEST_1>	30%
Comprehensive exam	Partly Open book	3 hours	<TEST_C>	40%

**6. Chamber Consultation Hour:** Monday 4-5 PM.

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

**8. Make-up Policy:**

Provision for make ups of assignments is not there. Make-up for midsem and compre tests shall be granted by the I/C on prior permission only before the date of the exam and also only for genuine cases/medical reasons where the student is medically unfit to sit on the day of the exam with the permission of the warden concerned. No make-up requests will be entertained after the exam is over. Make-up for comprehensive examination will be decided and scheduled by the AUGSD/AGSR Division.

**INSTRUCTOR-IN-CHARGE (CS F469)**

