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

SECOND SEMESTER 2019-20
COURSE HANDOUT

Date: 16.01.2021

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	**:**	**Vinti Agarwal(vinti.agarwal@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:

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, which includes several retrieval concepts and techniques such as representation and indexing of data, text mining, websearch: basics and advances, multimedia retrieval, etc.

3. Text Books:

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

4. 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

5. Course Plan:

Module No.	Lecture Session	Reference	Learning outcomes
M1: Basic Information Retrieval Concepts	Lecture 1: Course Overview	T1 Ch1	Introduction to the course
	Lectures 2-4: Boolean retrieval	T1 Ch 1 & 2, R1 2.5	The term vocabulary postings lists and introduction to ad-hoc search
	Lectures 5-6: Dictionaries and tolerant retrieval	T1 Ch 3	Wildcard queries, Spelling correction, Edit distances, Phonetic correction
	Lectures 7-9: Index construction	T1 Ch 4	Blocked sort-based indexing, Single-pass in-memory indexing, Distributed indexing, Dynamic indexing
	Lectures 10-11: Scoring, term weighting	T1 Ch 6	Learning weights, Term frequency and weighting, tf-idf weighting
	Lecture 12: The vector space model for scoring	T1 Ch 6	Dot products, Queries as vectors, Variant tf-idf functions, Document and query weighting schemes
	Lecture 13: Evaluation of IR	T1 Ch 8	Evaluation of unranked retrieval sets Evaluation of ranked retrieval sets
	Lectures 14-15: Probabilistic Model(3 lec)	T1 Ch 11	Probabilistic Information retrieval
M2: Text Mining	Lectures 16-21: Text Mining (3.5 lec)	T1 Ch 13, 14, 16,17	Text Classification, Text Clustering
M3: Web Search and Link Analysis	Lectures 22-25: Web search basics (1.5)	T1 Ch 19 R1 Ch13, R2 Ch2	Search Engine Architecture Web characteristics The search user experience Index size and estimation
	Lectures 26-28: Web crawlers and indexes	T1 Ch 20 R2 Ch 3	Crawling, Crawler architecture, Distributing

			indexes
	Lectures 29-31: Link Analysis	T1 Ch 21	The Web as a graph,Google's Pagerank, Hub and authorities (HITS), Web spam
M4: Multimedia and Cross Lingual IR	Lectures 32-35: Cross Language Information Retrieval (CLIR)	R3 Ch2	Language Problems in IR, Translation Approaches for CLIR, Handling many Languages, Using manually constructed Translation systems and resources for CLIR, Research issues
	Lectures 36-40: Multimedia Information retrieval (MIR)	R4 Ch2,3	Basic Multimedia search technologies, Content based Retrieval, Research issues in MIR
M5: Recommender Systems	Lectures 40-42: Recommender systems	R5 Ch1,2,3,4,5	Introduction to recommendation systems, Collaborative, Content, Knowledge and Hybrid recommendation systems, and Applications.

6. Evaluation Scheme:

S. No.	Component	Mode	Duration	Date	Weight
1	**Mid Semester Test**	**Closed Book**	**90 min**	**TBA (visit AUGSD website)**	**30%**
2.	**Quiz**	**Open Book**	**30 min**	**(25 Feb- 05 March) TBA**	**10%**
3.	**Term paper/Assignment**	**Open Book**	**TBA**	**(15 April -29 April) TBA**	**25%**
4.	**Comprehensive Exam**	**Partially Open Book**	**2 hours**	**May 07th (FN)**	**35%**

7. Chamber Consultation Hour: Once in a week. Every Friday from 5:00 PM – 6:00 PM. Interested student(s) need to inform beforehand through email if consultation is required.

8. Notices:

All the notices/communication concerning this course will be through the CANVAS platform only. You are requested to check this periodically. E-mail will be used as and when required.

9. Make-up Policy:

To be granted only in case of serious illness or emergency. The student needs to apply for the make-up in advance with documentary proof and prior permission from AUGSD.

10. Note (if any):

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

Instructor-in-charge
Course No. CS F469