Birla Institute of Technology & Science, Pilani
Hyderabad Campus

SECOND SEMESTER 2025-2026
Course Handout Part II

Date: 30-12-2025

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. Hrishikesh Rajesh Terdalkar

1. Scope and Objectives of the Course

This course on Information Retrieval provides a comprehensive understanding of the principles, algorithms, and systems used to retrieve relevant information at scale. It spans the evolution of IR from classical lexical models and indexing techniques to modern neural retrieval and Retrieval-Augmented Generation (RAG) systems. Emphasis is placed on understanding retrieval as a pipeline, from text processing and indexing to ranking, evaluation, and system-level tradeoffs, while also engaging with contemporary challenges such as dense retrieval, hybrid systems, responsible retrieval, and multilingual access. Students are expected to develop both theoretical understanding and practical insight into building, analyzing, and evaluating modern search systems relevant to industry and research.

Upon successful completion of the course, the student should be able to:

- ✔ Explain the IR pipeline and its role in modern search systems.
- ✔ Build efficient indexing and retrieval mechanisms for large text collections.
- ✔ Apply exact-match, tolerant, and lexical ranking techniques for document retrieval.
- ✔ Evaluate IR systems using standard test collections, metrics, and significance tests.
- ✔ Use probabilistic, language-model, and learning-based approaches for ranking.
- ✔ Apply neural and dense retrieval methods, including hybrid lexical–dense systems.
- ✔ Design and analyze Retrieval-Augmented Generation (RAG) pipelines.
- ✔ Incorporate structured and graph-based signals (Knowledge Graphs, Web graphs) into retrieval systems.
- ✔ Address multilingual, cross-lingual, and domain-specific retrieval challenges.
- ✔ Recognize robustness, fairness, safety, and governance issues in retrieval systems

2. Desirable Prerequisites

- Programming proficiency in Python
- Data Structures and Algorithms
- Basics of Machine Learning

3. Resources

3.1. Textbook

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

3.2. Reference Books and Other Resources

- **R0.** Lecture Slides.
- **R1.** Speech and Language Processing (3rd ed. draft), Aug 2025. *D. Jurafsky and J. H. Martin*. https://web.stanford.edu/~jurafsky/slp3/
- **R2**. Search Engines: Information Retrieval in Practice, *W. B. Croft, D. Metzler, T. Strohman*. https://ciir.cs.umass.edu/irbook/
- **R3.** Modern Information Retrieval, *R. Baeza-Yates and B. Ribeiro-Neto*, Addison-Wesley. http://people.ischool.berkeley.edu/~hearst/irbook/
- **R4.** Domain-Specific Knowledge Graph Construction, *M. Kejriwal*, Springer. https://link.springer.com/book/10.1007/978-3-030-12375-8
- **R5.** An Introduction to Neural Information Retrieval, *B. Mitra and N. Craswell*, 2018. https://bits-hyderabad.short.gy/ms-neural-ir.pdf
- **RM.** Reference materials such as research papers shared in class.

4. Method of Conduct of the Course

The course will be conducted through a combination of lectures, hands-on programming assignments, a rigorous semester-long project, and student-led paper presentations. Lectures will emphasize core concepts, algorithms, and system-level design principles in the field of Information Retrieval, supplemented by discussion of contemporary research and real-world search systems. Programming assignments and the course project will focus on implementing and evaluating retrieval components and pipelines. Students are expected to engage with research literature, participate actively in discussions, and work independently or in approved groups for the project. Assessment will include both open-book and closed-book components, as specified, to evaluate practical skills as well as conceptual understanding. Appropriate use of external libraries and tools is permitted where specified; however, all submitted work must reflect the student's own understanding and effort.

5. Course Plan

Lecture No.	Learning outcomes	Topics to be covered	Chapter in the Textbook
1	Introduce IR; define key problems and scope.	☐ What is IR? IR vs DB vs NLP ☐ Retrieval pipeline ☐ Modern landscape (Lexical → Neural → RAG)	T1 Ch1 R2 Ch1 Ch2 R3 Ch1 Ch2
2-3	Build robust text preprocessing and statistics intuition	☐ Tokenization; normalization; stopwords; stemming ☐ Zipf's & Heaps' laws ☐ Term statistics	T1 Ch2 R1 Ch2 R2 Ch4
4-6	Implement exact-match retrieval end-to-end	☐ Boolean retrieval ☐ Inverted index; postings; phrase/proximity; skip pointers	T1 Ch1 Ch2 R2 Ch5
7-8	Handle noisy queries and tolerant matching	☐ Dictionary structures; ☐ Wildcard; edit distance; spelling correction	T1 Ch3 R2 Ch6
9-10	Construct scalable indexes	☐ BSBI/SPIMI ☐ Memory vs disk ☐ Incremental/dynamic indexing; distributed considerations	T1 Ch4 R2 Ch5 R3 Ch9
11-12	Optimize index storage & efficiency	☐ Dictionary compression ☐ Postings compression ☐ Time–space tradeoffs	T1 Ch5
13-14	Rank documents using lexical relevance models	☐ TF–IDF; cosine ☐ Length normalization ☐ Fielded scoring intuition	T1 Ch6 Ch7 R2 Ch7
15-16	Evaluate IR systems	☐ Test collections; ☐ Relevance; P/R; MAP; nDCG; significance basics & pitfalls	T1 Ch8 R2 Ch8 R3 Ch3
17-19	Apply probabilistic retrieval and LM-based IR	☐ Binary Independence Model ☐ BM25 ☐ Language models ☐ Query likelihood; smoothing	T1 Ch11 Ch12 R2 Ch7

20-21	Explain semantic mismatch and motivate neural retrieval	☐ Vocabulary mismatch;
		☐ Distributional hypothesis;
		☐ Why embeddings help
			R1 Ch5
			R5 Ch2

22-23	Use embeddings and contextual representations for retrieval	□ Artificial Neural Networks □ Word embeddings (e.g., GloVe, word2vec, fastText); contextual embeddings; pooling □ Sparse vs dense vs hybrid retrieval representations	R1 Ch5 Ch6 Ch8 Ch9 R5 Ch3 Ch4
24-26	Train neural rankers/retrievers	□ Representation vs interaction □ Ranking as supervised learning; Learning-to-Rank paradigms: pointwise, pairwise, listwise □ Feature-based vs neural LTR; reranking pipelines □ Losses; negatives; distillation	R2 Ch7 R5 Ch5 Ch6 Ch7 RM
27	Engineer dense retrieval at scale	□ Vector search pipeline □ Approximate Nearest Neighbour NN concepts □ Latency/recall tradeoffs	R0 RM
28-29	Cross-lingual IR and Indian IR	□ Multilingual embeddings □ Code-mixing □ Low-resource IR	R1 Ch5 Ch6 C9 R0 RM
30-32	Introduce Knowledge Graphs (KGs); Construct KGs from text; Use KGs to improve retrieval	□ What is a KG? □ Information extraction for triples; entity resolution □ Entity-centric retrieval □ KG-aware query expansion/reranking □ Hybrid KG+dense+lexical	R1 Ch20 Ch23 R4 Ch1 Ch2 Ch3 Ch4 Ch5 R0 RM
33-36	Introduce Retrieval Augmented Generation (RAG); Build RAG systems	□ RAG architectures □ Chunking; query rewriting; grounding/citations; reranking; □ Conversational retrieval □ GraphRAG as KG-aware RAG	R1 Ch11 R0 RM
37	Evaluate RAG systems	□ Retrieval metrics vs task metrics □ Attribution/faithfulness; error taxonomy; debugging	R1 Ch11 R0 RM
38-39	Model the Web as a graph; link-based ranking	□ PageRank □ Topic-sensitive PageRank; □ Link spam; HITS	T1 Ch21 R3 Ch13

40	Address responsible & robust retrieval	□ Responsible retrieval: fairness, harms, safety; □ Robustness; evaluation pitfalls; governance	R0 RM R1 Ch4 Ch7

6. Evaluation Scheme

Component	Duration	Weightage	Date & Time	Nature
Programming Assignment (1)	Take Home	**10%**	TBA (before mid-sem)	Open Book
Course Project	Take Home	**30%**	Continuous Evaluation (5% by mid-sem)	Open Book
Mid-Term Exam	90 mins	**20%**	09/03/2026 2:00PM to 03.30PM	Closed Book
Research Paper Presentation (1)	15 mins	**10%**	TBA (after mid-sem)	Open Book
Comprehensive Exam	3 hours	**30%**	04/05/2026 FN (9:30AM to 12:30PM)	Closed Book

7. Chamber Consultation Hour

4:00 to 5:00 PM on **Wednesdays** at H-108 (office chambers)

8. Notices

All notices related to the course will be displayed on **the course webpage on LMS**.

9. Make-up Policy

Make-up requests for mid-sem and comprehensive examinations may be approved for genuine cases with prior permission of the IC, and after rigorous scrutiny. Permission will be granted only if the candidate has applied makeup for all other registered courses.

10. Academic Honesty and Integrity Policy

Academic honesty and integrity are to be maintained by all the students throughout the semester, and no type of academic dishonesty is acceptable.

HRISHIKESH RAJESH TERDALKAR
Instructor-in-charge
CS F469