

Information Storage and Retrieval

Fall, 2026

Lecture:	Wednesday 13:50-16:25	Instructor:	Yong Zhao
Classroom:	S508	Office:	N523
Office Hours:	Wednesday 9:30-12:30	Email:	yong.zhao@scai-jl.cn

Teaching Assistant: Yu Fu (fuyu05@stu.scu.edu.cn)

Recitation:(optional)

Textbook:

Introduction to Information Retrieval Illustrated Edition, Christopher Manning, Cambridge University Press, 2012.

Prerequisite:

Data Structure

Description:

This course focuses on the foundations of modern search engine and fundamental and advanced techniques of text-based information systems, including indexing, query understanding, (large) language models, learning to rank, interactive search, evaluation and user models, question answering, conversational system, and neural models for IR..

Course Objectives:

The course is designed to give students a solid background in information storage and retrieval systems, particularly for them to understand:

- How do search engines work?
- Effectiveness and efficiency
- How do users behave with them?
- How do users determine usefulness of information?
- How can a search engine mimic this process?
- Why do search engines fail?
- How can they be evaluated (off-line)?
- How can they be monitored and tuned (on-line)?

Learning Outcomes:

At the end of the class, the students should be able to understand the key concepts and applications related to information retrieval.

The students should have clear understanding of:

- What is information retrieval
- The working mechanism of search engines
- Efficient text indexing
- Boolean and vector-space retrieval models
- Evaluation and interface issues
- IR techniques for the web, including crawling, link-based algorithms, and metadata usage
- Document clustering and classification
- Traditional and machine learning-based ranking approaches
- RAG for large language models

Assessments:

Assessment Type	Percentage	Classification
Class activities	15	Open-ended coursework
Laboratory		Performance / open-ended
Homework	15	
Group project	30	
Midterm Exam		
Final Exam	40	Fixed-answer / standard key
...		

(Note: The percentage of assessment items with fixed answers or a standard key should not exceed 70%.)

Grade Composition:

Class letter grade is given within the following score ranges:

90.00 – 100.00 A 85.00 – 89.99 A- 80.00 – 84.99 B+ 76.00 – 79.99 B 73.00 – 75.99 B-
70.00 – 72.99 C+ 66.00 – 69.99 C 63.00 – 65.99 C- 60.00 – 62.99 D 0.00 – 59.99 F

GPA grades will be further determined based on the given letter grades to show on your transcripts.

Make-up Policy:

Students are expected to be present, on time, and prepared for all class meetings. If a student is late or absent, that student should notify the instructor via email before the start time of the class.

MATERIAL COVERED: The sequence of the sections covered in this class is:

Lecture slides will be available for copying or posted on blackboard.

1. Introduction and background

2. Information retrieval
3. Ad-hoc retrieval
4. Controlled vocabulary
5. Full-text indexing
6. Boolean retrieval
7. Indexing and query processing
8. Statistical Properties of Text
9. Document Representation
10. Retrieval Model
 - vector space model
 - language modeling
 - RAG
11. Evaluation
 - test-collection construction
 - evaluation metrics
 - experimentation
 - user studies
12. Federated Search
13. Clustering
14. Text Classification

COLLABORATION AND ACADEMIC HONESTY POLICY

Individual work on all homework and examinations is required, Cheating and copying other students' homework/exam are strictly prohibited. Any violation of this policy will be treated severely.

Collaboration amongst students to understand the course material and to work on course projects is strongly encouraged, however each student should take on different/distinguishable responsibilities in the course projects.

COURSE READING MATERIAL

Textbook – Database Management Systems (3rd edition) - by Raghu Ramakrishnan and Johannes Gehrke, McGraw Hill, 2003.