

CS 547/DS 547: Information Retrieval
Syllabus, Worcester Polytechnic Institute, Fall 2025
Salisbury Labs 104

1. **Credits:** 3
 - a. Class Meets: **Thursday 6:00pm - 8:50pm**
2. **Instructor:** Kyumin Lee, kmlee@wpi.edu
Pronouns: he/him/his
Office Hours: Thursday 4:00 - 5:00 pm
Office: Unity Hall 363
3. **TA:** Daniel Sardak, dsardak@wpi.edu
Office Hours: Mondays, Tuesdays and Fridays 2:00 - 3:00 pm
Office: Unity Hall 341
4. **Textbooks/Course Readings:**
 - a. Introduction to Information Retrieval, Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schutze, Cambridge University Press.
 - b. Mining of Massive Datasets. Jure Leskovec, Anand Rajarman and Jeffrey D. Ullman. Cambridge University Press.
 - c. An Introduction to Neural Information Retrieval. Bhaskar Mitra and Nick Craswell. Now Publishers Inc.
 - d. Research papers selected by the instructor.
5. **Specific Course Information:**
 - a. Course Description: Introduce theory, design, and implementation of text-based and Web-based information retrieval systems. Students learn the key concepts and models relevant to information retrieval and natural language processing on the Web and social systems. Topics include vector space model, crawling, indexing, web search, ranking, recommender systems, embedding and language model. Students will conduct a team-based project. All programming assignments will be in Python.
 - b. Prerequisites: I expect all students to have had some previous exposure to statistical learning at the level of DS 502/MA 543 and programming skills at the level of CS 5007. You should be able to design and develop large programs and learn new software libraries on your own.
6. **Specific goals for the course**
 - a. Course Objectives
By the end of the semester, you will be able to:

- i. Understand the key concepts and models relevant to information storage and retrieval, including efficient text indexing, vector space model, Web search.
- ii. Design, implement, and evaluate the core algorithms underlying a fully functional IR system, including the indexing, retrieval, and ranking components.
- iii. Identify the salient features and apply recent research results in information retrieval.

7. Brief list of topics to be covered

- a. Vector space Model
- b. Crawling
- c. Indexing
- d. Web search
- e. Link-based algorithms like PageRank
- f. Recommenders
- g. Distributed word representations and Large Language Models for IR
- h. Retrieval-augmented generation

Communication:

All course announcements will be posted via the Canvas course mailing list.

Grading Policy:

The course grading policy is as follows:

5%	In-Class Quizzes
24%	Assignments
20%	Midterm
20%	Final
31%	Project

The grading scale for graduate students is A:100-90, B:89.9-80, C:79.9-70, D:69.9-60, F:59.9-0

The grading scale for undergraduate students is A:100-90, B:89.9-80, C: 79.9-70, NR:69.9-0

Quizzes:

We will have 2 in-class quick quizzes (5 to 10 minutes each) spread across the semester to keep up with the readings.

Assignments:

There will be four assignments. Each assignment is proportion to 6% of your grade. All programming assignments will be in Python. You will have total **4 late days** during the semester. You can use **up to 2 late days for each assignment** without penalty. After you consume the total 4 late days or two late days for an assignment (whichever comes first), then you will get penalty proportion to extra late days (e.g., 10% off for the next late day, 20% for the next two late days and so on).

For example, you submitted your first assignment 2 days late. You will not get any penalty, but use 2 out of 4 late days. Or if you submit your first assignment 3 days later than due date, you will use 2 late days (again up to 2 late days for an assignment), and get 10% off penalty because of the third late day. For each assignment, we will **NOT accept** your solution **more than 3 days late**.

You may discuss an assignment with your colleague, but you should write a program and a report by yourself and should NOT copy and paste your colleague's solution. If you discussed an assignment with your colleague, explicitly report the colleague's name and what you discussed in your submission.

Exam:

The midterm and final exams are closed book and will be held in class. You may bring one standard 8.5" by 11" piece of paper with any notes you think appropriate or significant (front and back).

Project:

In the final project, you will apply algorithms, methods and techniques that you learned from this course to your project. The detailed information regarding the final project will be announced in class.

Add policy:

A student can add/drop through the 10th day of the semester without a fee. After day 10 of the semester, students can add courses (with instructor approval) with a \$100 late fee. Students must be officially registered for this course. No assignments or tests of any kind will be graded for students whose names do not appear on the class list. Refer to <https://www.wpi.edu/offices/registrar/course-registration/add-drop>

Drop policy:

No drops are allowed after the 10th day of the semester.

Learning Aids:

Lecture notes and schedule will be available in the course web page. Your homework solution and project report is supposed to be submitted via Canvas.

Plagiarism, Cheating and AI Policy:

Unless explicitly noted, all work you submit must be your own work. You are encouraged to discuss with others about ideas and material in the course, in preparing for exams, in understanding homework problems, project statements, etc. However, all homework solutions, exams are to be written individually, and the solutions should be your own, unless otherwise specified. Projects encourage teamwork, that is, in that case you are expected to work closely with your partner/(s) to solve problems and prepare a common agreed-upon solution.

Note in particular that copying of any material, may it be a single sentence or a figure, from any location (including the internet and generative AI including ChatGPT) without proper acknowledgement of the source constitutes plagiarism. If in doubt, please ask for clarification. Any violation of the WPI's

guidelines for academic integrity will result in no credit for the course and referral to the Student Affairs Office for disciplinary action. [You should be familiar with the WPI Policy on Academic Honesty](#).

Students with Disabilities:

Students with approved academic accommodations should plan to submit their accommodation letters through the [Office of Accessibility Services Student Portal](#). Should you have any questions about how accommodations can be implemented in this particular course, please contact me as soon as possible. Students who are not currently registered with the Office of Accessibility Services (OAS) but who would like to find out more information about requesting accommodations, documentation guidelines, and what the accommodated interactive process entails should plan to contact OAS either by email: AccessibilityServices@wpi.edu, by phone (508) 831-4908, or by stopping by the office on the 5th floor of Unity Hall.

Writing Center:

Would you like feedback on a paper that you're writing--or even a presentation or poster that you're creating? The Writing Center is here for you! Located in Salisbury Labs 233, the Writing Center offers free one-on-one consultations (in-person and over Zoom) to help you improve your communication skills. Tutors will read/listen to your work, give you feedback about your composition's strengths and weaknesses, and help you (and your team) chart a path forward as you revise. And since tutors are WPI students themselves, you can count on a friendly and supportive atmosphere. Consultations are open to all WPI students, undergraduate and graduate, for all classes and projects (even projects you're doing outside of class), and tutors will happily work with you at any stage of your process—early brainstorming, revising a draft, polishing a final draft. To see appointment options for both in-person and Zoom meetings, go to wpi.mywconline.com, or learn more about our services by visiting our homepage: wpi.edu/+writing

Snow happens in New England:

Should the instructor be unable to attend class due to weather, illness or some other circumstance then the course will be held online via Zoom at the regular course time. Should this situation arise, a notification will be shared on the course Canvas site along with details on how to access the Zoom session.