CS 525/DS 595: Information Retrieval & Social Web
Syllabus, Worcester Polytechnic Institute, Spring 2020
http://web.cs.wpi.edu/~kmlee/cs525

1. **Credits**: 3
    a. Class Meets: **Wednesday 6:00pm - 8:50pm, Zoom**

2. **Instructor**: Kyumin Lee, (435) 797-8420, kmlee@wpi.edu
    Office Hours: **Tuesday 9:30-10:30am** and **Wednesday 4-5pm via Zoom**

3. **Primary Textbook**: Introduction to Information Retrieval, Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schutze, Cambridge University Press.

4. **Additional Texbooks**:
    a. Mining of Massive Datasets. Jure Leskovec, Anand Rajarman and Jeffrey D. Ullman. Cambridge University Press.
    b. 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 components and operation of search engines providing search services. Components include web crawlers, indexers, link-based ranking algorithms, and recommender systems. Students will conduct a team-based project.
    b. Prerequisites: I expect all students to have had some previous exposure to basic probability, statistics, algorithms, and data structures. 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 storage and retrieval, including topics such as adversarial information retrieval, and social information management.

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. Recommeders

      g. Distributed word representations for IR

Communication:

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

Grading Policy:

The course grading policy is as follows:

2.5%	Attendance and in-class discussion
2.5%	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

Attendance and Quizzes:

We will check your attendance twice during the semester to encourage you to show up and engage in discussion. 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. You will present and may demonstrate your project in April 29.

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 and Cheating:
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) 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:

If you need course adaptations or accommodations because of a disability, or if you have medical information to share with me, please make an appointment with me as soon as possible. If you have not already done so, students with disabilities who believe that they may need accommodations in this class are encouraged to contact the Office of Disability Services (ODS) as soon as possible to ensure that such accommodations are implemented in a timely fashion. This office in Daniels Hall has phone (508-831-4908) and can be reached by email at disabilityservices@wpi.edu.