486: Information Retrieval and Web Search

Syllabus

Winter 2021

Schedule:

Lecture: T/Th 1:30-2:50pm, Rada Mihalcea
Discussions: M 4:00-4:50pm, Santiago Castro
M 5:00-5:50pm, Gauri Kambhatla
W 4:00-4:50pm, Ashkan Kazemi
W 5:00-5:50pm, Do June Min

Instructors

Rada Mihalcea, mihalcea@umich.edu
Office hours: Th 3-4pm

GSIs

Santiago Castro, sacastro@umich.edu
Gauri Kambhatla, gkambhat@umich.edu
Ashkan Kazemi ashkank@umich.edu
Do June Min, dojmin@umich.edu

GSI Office hours

M 11-1pm, Do June Min
T 9 - 11am, Gauri Kambhatla
W 1:30-3:30, Ashkan Kazemi
Th 4-6pm, Santiago Castro

Course description

This course will cover traditional material, as well as recent advances in Information Retrieval (IR), the study of indexing, processing, querying, and classifying data. Basic retrieval models, algorithms, and IR system implementations will be covered. While the course will primarily focus on IR techniques for textual data, it will also address IR for other media, including images/videos, music/audio files, and geospatial information. The course will also address topics in Web search, including Web crawling, link analysis, search engine development, social media, and crowdsourcing.

Website

Most course materials are made available on Canvas, and are considered required reading. A detailed schedule, including lecture topics, assignment due dates and exam dates, will also be available on Canvas.

Forum

We will be using Piazza to host a course forum. You are required to read this regularly; it is the venue we will use for important course announcements and assignment clarifications. In addition, it will be a significant source of help and hints on the assignments. We do not answer technical questions via email. In order to save everyone time, we want all students to have the

benefit of seeing each question and its answer, so please use the forum. Please do not post your own solutions, project code, test cases, or output to the forum. Also, please search the forum before posting to avoid questions that have already been asked and answered.

You can access Piazza on Canvas, or directly at:
https://piazza.com/umich/winter2021/eecs486wn2021

Contact

Please direct all technical questions to the Piazza forum. For questions of other nature, you can reach us by private message on Piazza or at the email addresses listed above.

Prerequisites

EECS 281 Data Structures and Algorithms

List of Topics

- Course Overview. Introduction to IR models and methods.
- Text properties. Text processing.
- Boolean model and extensions.
- Vector space model.
- Word embeddings.
- Web crawling.
- Link analysis. PageRank and HITS. Topic sensitive link analysis.
- Relevance feedback. Query expansion.
- Evaluation of IR systems. IR test collections.
- Text classification.
- Question answering.
- Keyword extraction. Text summarization.
- Social media. Social networks.
- Knowledge graphs.
- Image retrieval and video retrieval.
- Ethics of Web Search.
- Recent trends in information retrieval.

Required/recommended readings

(**required**) Introduction to Information Retrieval
Christopher D. Manning, Prabhakar Raghavan and Hinrich Schutze
Cambridge University Press, 2008.
Online version available: http://nlp.stanford.edu/IR-book/

(**recommended**) Readings in Information Retrieval
K.Sparck Jones and P. Willett
Morgan Kaufmann, 1997

(**recommended**) Modern Information Retrieval
Ricardo Baeza-Yates and Berthier Ribeiro-Neto
Addison-Wesley, 1999

Additional readings will be assigned throughout the semester.

Programming Environment

For this course, you must have a CAEN account. Anyone enrolled in EECS 486 is eligible for one, but it takes a little while to get it set up. You will have to write your programs in Python. We will grade your programs in the CAEN Linux environment and they must run correctly in this environment. You are free to develop your programs on any platform you like, but they must run on the CAEN platform.

Grading and Timeline

Programming assignments: 40%
Exam I (February 23, 6:30-8:00pm): 20%
Exam II (April 13, 6:30-8:00pm): 20%
Project (Multiple deadlines, Poster presentations April 16, 3:30-7:00pm): 20%

Assignments

There will be four programming assignments throughout the semester. The assignments are regarded as projects, not specs, and thus intentionally leave room for students to make design decisions. Assignments are due by 11:59pm on the due date. Assignments may be turned in up to 3 days late, with a penalty of 10% for each day late. No credit will be given after 3 days. No exceptions.

Exams

There will be two exams. The exams will be synchronous, on Zoom. Note that they will take place at a different time, to better accommodate for different time zones (6:30pm-8pm).

Project

The class includes a term project, where students will work in a team to address a problem related to information retrieval and web search. Each team will include 5 members who will have to equally contribute to the project. The project will include two checkpoints during the semester, and one final submission deadline and poster presentation at the end of the semester.

Honor Code

This class is run under the University of Michigan Honor Code. The Honor Code holds that students are honorable, trustworthy people and encourages them to behave with integrity in all phases of University life. By conforming to the Code, students do their work in an environment conducive to establishing high standards of personal integrity, professional ethic, and mutual respect. As a basic feature of the Code, students are placed upon their honor during all examinations, written quizzes, computer questions, homework, laboratory reports and any other work turned in for credit.

Accommodations for Students with Disabilities

If you think you need an accommodation for a disability, please let your instructor know at the beginning of the semester or at least 2 weeks in advance. As soon as you make us aware of your needs, we can work with the Services for Students with Disabilities (SSD) office to help us determine appropriate academic accommodations. SSD (734-763-3000; http://ssd.umich.edu) typically recommends accommodations through a Verified Individualized Services and Accommodations (VISA) form. Any information you provide is private and confidential and will be treated as such.