CS4780: Information Retrieval
Course Policy

Hongning Wang
CS@Uva

http://www.cs.virginia.edu/~hw5x/Course/IR2020-Spring/_site/
Instructor

- Hongning Wang
  - Graduated from University of Illinois at Urbana-Champaign in 2014
Instructor

• Hongning Wang
  – Research area
    • Information retrieval
    • Data mining
    • Machine learning
  – Industry experience
    • Yahoo Labs
    • Microsoft Research
    • Snap
Goal of this course

- Discuss fundamental problems in information retrieval
  - Building blocks of search engine systems
  - Wide coverage of important IR techniques
    - Personalized recommendation
    - Online advertising
- Get hands-on experience by developing practical systems/components
- Prepare students for doing cutting-edge research in information retrieval and related fields
  - Open the door to the amazing job opportunities in IT industry
Outcomes

• Letters from former students

Dear professor,
Thank you so much for teaching me Information Retrieval which has been the most beneficial class this semester. I have got an internship position in Walmart Labs search team all because of the knowledge I learnt from your class. Although you are strict on the grade, but after all I think it's fair and still encourage me to learn better on IR.

Hi Professor Wang,
My name is XXX Zhang, and I just graduated from UVA in May. I will start working full-time at Google starting next Monday and I just got my team assignment today. I will be working at Google's search ranking team. I still remember the Information Retrieval class I took with you. That still remains one of my favorite CS classes at UVA!
I'm sending this email just to let you know that you have a former student working on search engines.

Hi Professor Wang,
I hope you are doing well! I am an alum of your Information Retrieval course and would like to say that you stood out as one of the best teachers I've had during my time at UVA.
I graduated in December of 2017 and since then I have been working as a Data Scientist here in Charlottesville. I've found a real passion in machine learning and plan on pursuing a career as a researcher in this field.
Character of this course

• Discussion oriented
  – This is how great ideas are created!
  – You are encouraged to express your thoughts, confusions, and suggestions
  – Focusing on why, rather than how
Prerequisites

• Programming skills – Important!
  – Basic data structures: CS 2150 or equivalent
  – **Java** is required for machine problems
    • Most open source packages are written in Java
  – Any language you choose for the rest of this course

• Math background
  – Probability
    • Discrete/continuous distributions, expectation, moments
  – Linear algebra
    • Vector, matrix, dot product
  – Optimization
    • Gradient-based methods
Pop-up quiz

1. Let $\mathbf{a}=(1,2,3)$ and $\mathbf{b}=(2,3,-2)$, the inner product between $\mathbf{a}$ and $\mathbf{b}$ is
   (a) 0          (b) 1         (c) 2             (d) 3

2. Let $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$, what is $\mathbf{A}^{-1}$,
   (a) $\begin{pmatrix} -1 & -2 \\ -2 & -1 \end{pmatrix}$ (b) $\begin{pmatrix} -\frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{1}{3} \end{pmatrix}$ (c) $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ (d) $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$
Pop-up quiz

3. What is the expectation of random variables drawn from Gaussian distribution N(0, 1),
   (a) 0          (b) 0.5         (c) 1             (d) 2

4. Complexity of merge sort,
   (a) $O(n)$   (b) $O(n^2)$   (c) $O(\log n)$   (d) $O(n \log n)$
Pop-up quiz

1. Let \( \mathbf{a} = (1, 2, 3) \) and \( \mathbf{b} = (2, 3, -2) \), the inner product between \( \mathbf{a} \) and \( \mathbf{b} \) is \( \text{ (c) } \) 
   (a) 0          (b) 1         (c) 2             (d) 3

2. Let \( \mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \), what is \( \mathbf{A}^{-1} \), \( \text{ (b) } \) 
   (a) \( \begin{pmatrix} -1 & -2 \\ -2 & -1 \end{pmatrix} \)  
   (b) \( \begin{pmatrix} -1 & 2 \\ 2 & 3 \end{pmatrix} \)  
   (c) \( \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \)  
   (d) \( \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \)
3. What is the expectation of random variables drawn from Gaussian distribution $N(0, 1)$, (a) 0 (b) 0.5 (c) 1 (d) 2

4. Complexity of merge sort, (c) (d) 
(a) $O(n)$ (b) $O(n^2)$ (c) $O(\log n)$ (d) $O(n \log n)$
Structure of this course

• Six major topics will be covered by lectures
  – E.g., Search engine architecture, retrieval models, search evaluation, relevance feedback, and link analysis

• Latest development will be covered by paper reading assignments and presentations
  – E.g., mobile search, recommendation, personalization, you name it!
Grading policy

• Reading assignments (10%)
  – Peer evaluation, after each chapter

• Homework (35%)
  – Machine problems (~3)

• Midterm Exam (20%)
  – Check points of key concepts (in class, 75 minutes)

• Course project (35%)
  – In the exam week

No curving will be applied!
Grading policy

- Reading assignments (10%)
  - Peer evaluation, after each chapter
- Homework (35%)
  - Machine problems (~3)
- Midterm Exam (20%)
  - Check points of key concepts (in class, 75 minutes)
- Course project (35%)
  - In the exam week
- Paper presentation (10%)
  - For graduate students only

No curving will be applied!

fairness will be guaranteed by the instructor
Reading assignments

• Read the instructor selected papers after each chapter
• Open-ended essay questions
• Peer evaluation on course forum
Paper presentation

• Choose to present the most recent works in the area of information retrieval
• Peer evaluation
• Choose from the instructor’s selected papers, which are beyond our course content, so as to increase our topic coverage
Midterm exam

• In the second half of semester
  – ~after spring break
• Covers all material we would have learnt by then
• In-class, 75 minutes
• Format
  – True/False question
  – Short answer questions
  – Short essay questions

Fact-based questions
Research-like open discussions
You design your midterm?

• After each chapter, based on your understanding, post one question related to the most important concept in that chapter on our course forum

• Read the others’ posted questions and vote on them

• The top voted questions will be included in the midterm

• The authors of those top voted questions should provide the answer, and will get bonus points
Course project

• Topics
  – Implement algorithms in assigned research papers
  – Self-selected topics with permission from the instructor

• Team work
  – 3-4 students per group

• Evaluation
  – Two-page proposal (25%)
  – 15-minutes in-class presentation (40%)
  – Written report (35%)
Late policy

• Homework
  – Submit via Collab (no extension)
  – Late penalty: 15%, two weeks after the due date; 30%, afterwards

• Course project
  – Final report is due before presentation (no extension)
Classroom participation

• HIGHLY APPRECIATED!
  – Helps me quickly remember your names
  – Reminds me what is still confusing
  – You can drive the lecture/discussion in this class!
Contact information

• Lecture
  – Instructor: Hongning Wang
  – Time: Tu/Th 2:00pm to 3:15pm
  – Location: Olsson Hall 011
  – Office hours
    • Time: Tu/Th 3:30pm to 4:30pm
    • Location: Rice Hall 408
    • Additional office hour can be requested by email
Contact information

• TA
  – Nan Wang (nw6a@virginia.edu)
  – Office hour
    • Time: Monday/Wednesday 10:30am to 11:30am
    • Location: Rice Hall 442
    • Additional office hour can be request by email
Thank you!

QUESTIONS?