Syllabus

Com S 435/535

Fall 18

Modern data sets are really BIG, and these data sets give rise to new computational problems as wells as new algorithmic and implementation challenges. In this course we study design of algorithms for a few computational problems that involve massive data sets. The course places equal emphasis on theoretical (algorithmic and mathematical ideas) and practical aspects (working with real world data sets). Tentative topics:

Modern data sets are really BIG, and these data sets give rise to new computational problems as wells as new algorithmic and implementation challenges. In this course we study design of algorithms for a few computational problems that involve massive data sets. The course places equal emphasis on theoretical (algorithmic and mathematical ideas) and practical aspects (working with real world data sets). Tentative topics. (We may not have time to cover all of these topics).

1. Basics of Probability Theory
   - Random Variables,
   - Expectation, Variance.
   - Markov, Chebyshev and Chernoff Bounds
2. Probabilistic Data Structures
   - Theory behind hashing
   - Memory efficient data structures: Bloom Filters, CountMin Sketch
3. Similar Items
   - Measuring Similarity among items
   - Document Similarity
   - Dimensionality Reduction
   - Locality Sensitive Hashing
   - Latent Semantic Analysis
4. Components of Search Engine:
   - Crawling and Ranking
     - Web Crawlers
     - Google's page rank algorithm.

  - Hubs and Authorities in Web.
- Information Retrieval
  - Boolean Retrieval
  - Ranked retrieval
  - Index Building and Index Compression

5. Social Network Analysis.
   - Substructures and Communities
   - Influence Propagation
6. Matrix Decompositions
7. Online Advertising and Recommendation Systems
   - Advertising on Web, Placing ads along with search results, Ad words Problem
   - Predicting items (movies/songs/news articles) an individual user might like. Content Based Recommendation, Collaborative Filtering Recommendation

Pre-requisites: Com S 311
Catalogue Description: Challenges involved in solving computational problems on massive data sets. Discussion of computational problems that arise in the context of web search, social network analysis, recommendation systems, and online advertising etc. Theoretical aspects include modeling the computational problems using graphs, study of similarity measures and hash functions, and design of efficient algorithms for graphs. Practical aspects include implementation and performance evaluation of the algorithms on real world data sets. Graduate credit requires a written report on current research.

Staff

- Instructor: Pavan Aduri, 112 Atanasoff Hall. `pavan@cs.iastate.edu`. Office Hours: TBA
- Teaching Assistant. Madhavan Rajagopalan. `madhavrp@iastate.edu`

Text Book. There is no required text book. Students are expected to take notes during the lectures. In addition, I will provide lecture notes. Lecture notes are typically posted within one week of lecture. You may use Mining of Massive Data Sets by Leskovec, Rajaraman and Ullman as a reference book.

Grading. Grading is based on homework's, programming projects (in Java) and one mid-term exam, and a final exam. Dates for the exams are yet to be decided. Graduate students are expected to do a *research component* that involves reading research papers and producing a report. Graduate students may be assigned more homework problems than undergraduate students.

For Undergraduate students, Homeworks are worth 15 %, Programming projects are worth 35 %, midterm exam is worth 20% and final exam is worth 30%. For graduate students, Homeworks are worth 10%, programming projects are worth 30%, midterm exam is worth 20%, final exam is worth 30% and the research component is worth 10%.

A Programming Assignment will be due during the dead week.

Outcomes.

- Ability to apply mathematical ideas and algorithmic principles in modeling computational problems that arise in the context of massive data sets.
- Ability to use theoretical tools in a practical environment to design programs to solve computational problems that arise in the context of massive data sets.

Academic Dishonesty Policy. The class will follow Iowa State Universities policy on academic dishonesty. Anyone suspected of academic dishonesty will be reported to the Dean of Students Office. Please see
`http://www.dso.iastate.edu/ja/academic/misconduct.html`

Disabilities. Iowa State University complies with the Americans with Disabilities Act and Sect 504 of the Rehabilitation Act. If you have a disability and anticipate needing accommodations in this course, please contact your instructor to set up a meeting within the first two weeks of the semester or as soon as you become aware of your need. Before meeting with instructor, you will need to obtain a SAAR form with recommendations for accommodations from the Disability Resources Office, located in Room 1076 on the main floor of the Student Services Building. Their telephone number is 515-294-7220 or email `disabilityresources@iastate.edu.` Retroactive requests for accommodations will not be honored.

DeadWeek. This class follows the Iowa State University DeadWeek policy as noted in section 10.6.4 of the Faculty Handbook.

Harassment and Discrimination. Iowa State University strives to maintain our cam- pus as a place of work and study for faculty, staff, and students that is free of all forms of prohibited discrimination and harassment based upon race, ethnicity, sex (including sexual assault), pregnancy, color, religion, national origin, physical or mental disability, age, marital status, sexual orientation, gender identity, genetic information, or status as a U.S. veteran. Any student who has concerns about such behavior should contact his/her instructor, Student Assistance at 515-294-1020 or email dso-sas@iastate.edu, or the Office of Equal Opportunity and Compliance at 515-294-7612.

Religious Accommodation. If an academic or work requirement conflicts with your religious practices and/or observances, you may request reasonable accommodations. Your request must be in writing, and your instructor or supervisor will review the request. You or your instructor may also seek assistance from the Dean of Students Office or the Office of Equal Opportunity and Compliance.

Contact for Information On University Policies. If you are experiencing, or have experienced, a problem with any of the university policies (on academic dishonesty, dead week, harassment and discrimination, ,religious accommodation) email academicissues@iastate.edu.