Administrivia

(CPEG/ELEG 457/657: Search and Data Mining)

Hui Fang
Department of Electrical and Computer Engineering
University of Delaware

2

Welcome!

- **Instructor: Hui Fang**
  - **Research: Search, Information retrieval, Text Mining**
  - **Office Hour: by appointments**
  - **Email me at any time**
    - **hfang@udel.edu**
- **TA: Yue Wang and Hao Wu**
  - **Contact information: 209 Evans Hall**
  - **Office hours: 3-4pm on Mondays and Thursdays.**

3

Course Web site

- **Canvas**
  - **In-class announcements generally reflected on web page**
  - **Check it frequently for the updates (e.g., due dates)**
- **Basic information**
  - **https://www.eecis.udel.edu/~hfang/cpeg657.html**
  -

4

Course Goal

- **Introduce "Search", also known as "Information Retrieval" in academia**
- **Cover basic concepts of "Data Mining", in particular "Text Mining"**

- **Prepare students for doing research/development in information management and related fields**
- **Answer some questions you might have about search engines and other information systems**
  - **How does a search engine work?**
  - **What's the architecture of search engines?**
  - **How does Google News work?**
  - **How does an email spam filter work?**
  - **How does Amazon's recommendation system work?** 5

Prerequisites

- **Required:**
  - **Proficiency in programming (needed for machine problems and projects)**
  - **Ability and Willingness to learn new programing languages or tools**

- **Optional: Knowledge of basic probability & statistics (would be helpful for understanding algorithms deeply)**
- **Contact the instructor if you aren't sure** 6

Textbooks (recommended)

- *Text Data Management and Analysis: A Practical Introduction to Information Retrieval and Text Mining* , By C. Zhai and S. Massung.

- *(Optional) Introduction to Information Retrieval,*
  by C. D. Manning, P. Gaghavan, and H. Schuetze. (available free online)

7

1

Course Format

- Three 50-min lectures/week
- Assignments (written homework and machine problems)
- Midterm
- Course project
  - Real world data (Web, email, literature,...)
  - Group project

CPEG/ELEG 457/657

- CPEG vs. ELEG?
  - No difference

- 457 vs. 657
  - Different grading criteria
  - Different workload
    - E.g., students taking 657 are required to write a literature survey on the topic related to their projects.

Grading (CPEG/ELEG 457)

- Class participation and quizzes: 15%
- Assignments: 25%
- Midterm: 20%
- Course project: 40%
  - Pre-Proposal: 5%
  - Proposal: 10%
  - Final report and presentation: 25%

Grading (CPEG/ELEG 657)

- Class participation and quizzes: 10%
- Assignments: 25%
- Midterm: 20%
- Course project: 45%
  - Pre-Proposal: 5%
  - Proposal: 10%
  - Literature survey: 10%
  - Final report and presentation: 20%

Lectures

- Lecture slides will be posted AFTER the lecture
  - are to complement the lectures

- Many issues discussed in the lecture will be covered in the exams and assignments
  - Hence try to attend the lectures regularly

Quizzes

- Several (likely 3 to 5) in-class two-minute quizzes will be given at the end of some randomly chosen classes.
- The quiz questions should be very easy to answer if a student has really paid attention to what is discussed in the class.
- Besides being one factor contributing to grading, quizzes also serve as a way to keep track of class attendance.

Assignments

- Some paper-based, some may involve programming
  - Paper-based assignments will be collected at the beginning of class on the due date
  - Machine problems are due at 11:59pm on the due date.
- Late homework
  - Will be penalized on an hourly scheme
  - Details are available at course website.

14

Course Projects

- **Pre-proposal: Due on Feb. 24**
  - Submit one paragraph to describe what you would like to do
  - Specify who you would prefer to work with
- Proposal: around the mid-March
  - Written report (1 page)
  - Oral presentation (15-20 minutes)
  - Peer evaluation + instructor evaluation
- Literature survey (657 students only)
- Final report and presentation
  - Peer evaluation + instructor evaluation

15

Honesty

- Plagiarism and other forms of cheating will results in an F and/or other disciplinary action
  - Always cite your sources
    - Includes your fellow students
    - Include anything someone else wrote, whether published or not, whether on the Web or not.
  - Encouraged to discuss homework issues with each other
    - Not exams
  - Handed in work must be your own
    - Do not copy, do not let someone read your work

16

How to Get the Most?

- Attend every lecture and participate discussions
  - course syllabus + slides → road map
  - Many topics aren't covered well in any reading
- Use your peers in learning
  - Discuss in/out classes to enhance understanding
  - Do a project together
- Ask questions in class and after class!

17

Questions?

18

3