CPS 296.1 (Spring 2012): Project in Computational Journalism

Introduction
Crisis

- Traditional news media: fewer readers → lower ad revenue → fewer resources → less original investigative reporting
- Journalism’s watchdog function is in trouble

*Quis custodiet ipsos custodes?* (Who will guard the guardians?)

- Who will hold governments, corporations, and powerful individual accountable to society?

http://www.dbgallery.co.uk/historys-whos-who/195869_socrates.html
Opportunity

- **Democratizing data**: more data are becoming publicly available
- Computing has a proven track record with big data

- Computational journalism
  - Lower cost
  - Increase effectiveness
  - Broaden participation: democratizing data analysis

http://www.filetransit.com/images/screen/2f4df0324760b79935b80ea340398d82_Matrix_Code_Emulator.jpg
Course goal

- Explore how computing can help public interest journalism
- **Build something that people will use and learn from**
  - Or fail valiantly in our attempt
- Learn computer science/technology along the way

- **Main difference from most other CS courses: this course is completely project-driven**
  - Topics we cover depend on the project(s) we choose
So how does this work?

- At the beginning of the course, we will survey the landscape, hear from journalists, and discuss ideas.

- Around the sixth week of the class, we will decide on the project idea(s):
  - The number of projects will depend on how many of you remain committed after the add/drop deadline (Jan. 25).
  - I prefer fewer, more substantial projects providing end-to-end, polished solutions that can be put to real use.
  - Teamwork is vital:
    - Learn to maximize your contribution to/influence in a team.
So how does this work? (Cont’d)

- In the remainder of the course, we will learn whatever topics needed to build our project(s), and tackle technical challenges together
  - Schedule (and topics) and will be finalized once we have chosen the project(s)
  - To keep the project(s) on track, there will also be progress report presentations

- We will end the course by demonstrating our project(s) in front of a panel of judges
But where is the science/research?

• It is possible that a project may just involve applying existing technologies
  • There will be design and engineering challenges
  • You will learn teamwork and practical project skills
  • There may be good real-world impact
• For those looking for research and/or publications:
  • Seemingly simple tasks may require research to solve well and generally
  • I will use my experience to try to ensure there is potential for deeper research within the project(s)
Specific tasks

• Show up and participate
  • Most of our meetings will be more like discussions and project meetings than lectures

• Read and review
  • There will be (almost) weekly reading assignments; some require reviews (submitted through Sakai)

• Report, present, and lead discussion
  • You will be asked to report your progress, survey a topic, and lead discussion on papers in class

• And above all, get your hands dirty on the project!
Grading

• Class attendance: 10%
• Reading assignments (with reviews): 20%
• Topic presentations: 20%
• Project (including written/oral presentations): 50%

• A: [90%, 100%]
• B: [80%, 90%)
• C: [70%, 80%)
• D: [60%, 70%)
Course rules

- Be creative
- “Stay hungry. Stay foolish.”
- Question conventions
- Be responsible (to your fellow team members)
- Commit to Duke Community Standard

- You must be self-motivated to take this course
Introductions

• Your background
  • Information retrieval, databases, text/data mining, statistics, machine learning, web development, UI, data visualization, system admin, project management, journalism, public policy, civic engagement, startup, ...

• On a scale of 1 to 10, are you more interested in learning more CS (1) or building something real to advance public interest (10)?

• Do you have a pet project on computational journalism?
Course information

- Website: http://tinyurl.com/dukecjclass
  - Readings assignments
  - Tentative schedule
  - Slides for class meetings

- Office hours:
  - Fridays 1:15-4pm, or by appointment

- Mailing list
  - Announcements + things of general interest
Reading and reviewing

- Use the Sakai forum (link provided on the course website) to post paper reviews
  - Not graded, but visible to the whole class
  - Always due by 10pm the night before the paper is discussed in class
  - Discussion leaders don’t need to submit reviews
- For this Friday
Coming soon

• Berkman Center’s report on news and information in the digital age + Cohen’s piece on open government (next Wednesday)†

• Your survey/research of computational journalism websites/projects (the following week)
  • Everybody participates

• Long survey of information extraction†

• Short survey of crowdsourcing†

† Discussion leaders needed; reply to my sign-up email