Projects in CS 234

Overview

There will be three projects in CS 234 (Fall 2017). The first two projects are group projects, the third one is an individual project. You should consider the two first projects as "learning projects" in which you try out different things, explore techniques, and get practice getting better at all steps of the data science cycle by working together with your peers. Doing well in these two projects indicates to yourself, your peers, and me that you are taking your learning seriously and are always looking for opportunities to grow. To encourage experimentation and learning there will be no grade for these two projects, but only qualitative feedback about how close your output is to the standard we see in the research papers we'll be looking at (like the "Wikipedians are born"). Even if you don't learn everything in the first project, you should continue to learn and improve for the next ones. Then, in the individual project, you can showcase all your skills and knowledge and use that as your argument for how much you learned in the class, based on what you can produce.

Project 1 is about studying knowledge production and bias in Wikipedia by analyzing edits in certain articles and the editors that do them.

Project 2 is about studying Google searches: the searches that users do, the ones that Google suggests, and the different results it shows for such searches and how these might change over time.

Project 3 is broadly called "digital natives". In this project you will identify a source of data (start thinking now) that will allow us to study the generation (your generation) that are called "digital natives" because they grew up with technology. The data has to be "traces", that is content or interactions that are captured automatically by devices. These can be your own traces (for example, all your emails, your browsing history, all your text messages, your instagram interaction, Facebook history, google docs, online games, etc.), or you can collect the traces of your friends, in order to make broader statements. Ambitious projects will try to collect multiple sources of traces to paint a more nuanced picture of how the digital natives use technology across different platforms and how much technology is part of their daily lives.

In all projects I expect you to put your best efforts, be engaged, and take the initiative and pursue your own questions within the context of the project theme. I'll always be a resource where you can come to bounce off ideas and discuss their feasibility.

Timelines for these projects

Project 1 - Wikipedia [We have started this already on Sep 27]
Sep 27 - Nov 3

Project 2 - Google searches
Nov 3 - Dec 1

Project 3 - Digital Natives
From now until Dec 21 [last day of exams]

Publishing in International Conferences

There are dedicated international conferences, for example WebSci or ICWSM or CSCW, where the kind of investigative projects we're doing in this class can be submitted as research papers. I encourage you to consider this possibility. For example, WebSci in 2018 takes place in Amsterdam, and ICWSM in Silicon Valley. Especially if you are interested in going in grad school, publishing a paper as an undergraduate is seen as a major accomplishment.

Today

- We'll create groups of 3 students, one group will have 4 students.
- You'll get together as a group and decide what questions within the Wikipedia project you want to pursue.
- By the end of the day, one designated group member will send an email to Eni, cc-ing all group members to confirm the group creation and the questions you are interested to pursue.
- Next steps:
    - members do research to find papers or other online resources that have investigated these questions or similar ones.
    - Each of you reads 2 papers (or investigates two projects) and takes notes about what they did and how it can be useful to you (by replicating their work).
    - In your class portfolio, you create a new page, for wikipedia (like the one for the eating project), and start "blogging" about your progress and everything you learn, linking to notes, websites, ideas for visualization, inspiration for questions, etc. This is similar to lab journals (blue books) in science classes. Your notes are used to communicate with your group, me, but also with the world.