Choosing a Course Project (due November 26)

As explained in class, you have a choice of several different course projects to work on. You may work in groups of one or two, but if you work together, more work will be expected. As a general guideline, projects should involve an amount of work slightly larger than that of the second assignment you have done, per person involved in the project.

Some possible projects will be posted on the project page of the course site. You will receive email whenever additional projects are posted. You are also encouraged to propose your own course projects on a relevant topic. Projects will be supervised either by the instructor or the TAs. In any case, you must submit by 11/26 a short description in pdf, stating what project you plan to do, what you plan to achieve, and what people will be involved. This should be about half to one page of text.

Important: It is now up to you think about possible projects, talk to me about details, and notify me about what you will work on. Talk to me if there are questions! I am happy to help, and you can ask for more hints at any point in time, by email or in person, but you have to take the initiative.

Most of the projects will require suitable data sets. If appropriate, you could use the data set from the homeworks, or larger web data sets we can provide, or data sets from the New York Times archive, Twitter data, data from sites such as Wikipedia, stackexchange, pinterest, kickstarter, github, amazon, or many others. You can acquire your own data, or request help if you need a particular type of data for your project (e.g., a larger web graph, or a crawl of a country). Identifying and if necessary retrieving and preprocessing suitable data sets is an important part of doing your project.

The project will be due at the end of the semester (around 12/20). In addition to the code itself, and an optional demo, every group should submit a paper about their project, with about 10 for one person, or 15 for two, properly typed in Word or LaTex or some other software using 11pt font size. (If the whole project is to write a paper, then of course more pages are expected!) In the paper, you should describe on a high level what you did, how you did it, maybe what other solutions there are that you decided not to use, and what the result of your work was, plus appropriate references to the research literature as needed.