

Choosing a Course Project (due April 14 at the latest)

As explained in class, you have a choice of several different projects to work on. You may work in groups of up to 4, if the project is appropriately challenging.

Some possible projects are already posted at <http://cis.poly.edu/cs912/projects.html>. You will receive email whenever additional projects are posted. Projects will be supervised either by the instructor (suel@poly.edu) or by two PhD students, Shuai Ding (dingshuai1985@gmail.com) or Jinru He (mikehe1117@gmail.com), as stated in the description. In any case, you need to tell me by 4/14 what project you plan to do.

Important: It is now up to you think about possible projects with your teammates, talk to me about details, and notify me about what you will work on. Deadline is 4/14, but earlier is better. Talk to me if there are questions! We are 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 may use the set of 3 million web pages from the New Zealand `nz` web domain that was made available for the previous homeworks, or other more interesting data sets. 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 suitable data sets is an important part of planning your project.

The project will be due at the end of the semester (around 5/17). In addition to the code itself, and an optional demo, every group should submit a paper about their project, with at least maybe 5 to 6 pages per person involved, properly typed in Word or LaTeX or some other software. (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.