Topics in AI HW 5 (Due May 1 midnight)

You may work with a partner on this homework; however, if you worked with a partner for a previous homework, I ask that you work with a *different* partner.

The goal of this assignment is for you to experience some of the problem solving necessary to extract information from natural language, and also some of the problems that you could encountered. The crux of this assignment is the web app at http://appoxy.herokuapp.com/info_ret/, which is a simple "programming interface" for manipulating text.

Your job, using that interface, is to extract out the prerequisites of the courses at Oxy, based only on their catalog description. In particular, you need to get the possible prereqsuites on a line by itself, as a department code and number (eg. "COGS 325"). You should know right off the bat that it is impossible to get this correct for all courses, if only because of the limited web interface. Part of the point is that even with text as structured as a catalog, and even with fairly powerful operations such as replacing text, there is a huge variation in phrasing that makes it difficult to do this automatically.

Note: if you want to write a actual program to attempt this task, you can download the course descriptions [from Github](#).

Once you've created a program that you're satisfied with, write a report describing your experience. In particular, answer the following questions:

- What courses could you successfully extract prerequisite information from? Why were you successful for these courses?
- What course could you almost extract prerequisite information from? What would have allowed you to succeed?
- What courses did you totally fail to extract prerequisite information from? Why did you fail? What would have allowed you to succeed?
- If you had to give your program a score out of 100, with 0 being an epic failure and 100 being a triumph, a huge success, what score would you give? Include lower and upper bounds on your score (ie. you think it would be unfair to go below/above those numbers), and justify your answer.
- Pick another corpus of text, and some information you might want to extract out of it. What would make that task difficult?
- What can you conclude about natural language processing/information retrieval from this exercise? Be specific in your answer.
- What can you conclude about our ability to understand language from this exercise? Be specific in your answer.