

Problem 1. (20 points)

Navigate around the Course Web-page, fill the online form and record your transaction number.

Problem 2. (10 points)

Estimate the number of web-pages stored by Google. Justify your answer.

Problem 3. (30 points)

You have two options: Either pursue option 1 or alternatively, option 2. Doing both won't gain you more points; only knowledge.

Option 1.

Download to AFS or your own PC/laptop one Web crawler such as WebSphinx (there is a version of it available in Java as a .jar file). Google to find out where it is. Read the online available manual and run the program. If you have never run a java program that's fine; under afs do a `java -help | more` for information.

When you manage to have it running, go and grab first all the text files at

<http://www.cs.njit.edu/~alexg/courses/cis435/handouts.html>

If you feel more confident, try

<http://www.cs.njit.edu/~alexg/courses/cis435/index.html>

Be considerate, careful, and restrain yourself. Do such testing late in the afternoon or early in the morning.

Report in one page (handwritten or typewritten, the choice is yours) all your actions to run/setup WebSphinx, and give a summary listing of all text files (alphabetical order, multiple names say, 4-10 per line) you grabbed from the former URL (...handouts.html). Be brief and to the point.

Option 2.

Alternatively, read the manuals/paper introduction/documentation of 3 web crawlers and do a comparison of their capabilities. Be brief, tabulate your results and restrict yourselves to an one page summary. ■