

Assignments 2 & 7
CS 410/510 Information Retrieval on the Internet
Assignment 2 due Tuesday, 23 January
Assignment 7 due Tuesday, 27 February
Assignment 2: 10 points; Assignment 7: 20 points

Goal: In this project you will investigate empirically the behavior of a search engine (or several search engines). Which aspect to investigate is up to you: ranking, retrieval, stemming, stopwords, etc. You can investigate a web-based search engine or a desktop tool.

Details: There are two submissions connected with this project, due on 23 January and 27 February, at the beginning of class. This project can be done in groups of 2 or 3 people. Turn in a single paper from each group for each submission, with the names of all group members on it.

The main steps in the project are:

- Decide *what* system(s) to investigate, and *which* aspect(s) to study, perhaps performing some initial steps to understand the system or aspect better.
- Formulate a hypothesis about the behavior of the system on that aspect.
- Devise an experiment you can carry out to test your hypothesis.
- Carry out the experiment.
- Analyze the results of your experiment to see if your hypothesis is confirmed or not.

Notes:

1. You will not lose points if it turns out your hypothesis is incorrect.
2. If you are going to “plant” something like the Google Garden for a search engine to crawl and index, you need to put your material on line soon. It may take a week or more to be discovered by a crawler, and indexing will happen even later. Pages nested below other pages can take even longer.

Your first submission (Assignment 2) should include the following items:

- A. What system(s) and which aspect(s) you are studying.
- B. A statement of your hypothesis, and an explanation of what kind of observable behavior it should cause.
- C. A description of the experiment you plan to carry out to test your hypothesis.
- D. An explanation of what kind of experimental results would confirm your hypothesis, and what kind of results would falsify it.

Your second submission (Assignment 7) should include the following items:

- E. A detailed description of the experimental procedure you followed, explaining any changes you needed to make from your original plan (in Item C.).
- F. A presentation of the results of your experiment.
- G. An analysis of those results, including a conclusion on whether your hypothesis is confirmed or not (or possibly partially confirmed).

H. (For CS 510 students) A proposed explanation of what mechanisms the search engine could be using that would explain the behavior you observe.

We will provide feedback between submissions about the scope and experimental procedure for your project.