

Assignment 1
CS 410/510: Information Retrieval on the Internet
Due Tuesday January 16, 2007
8 points

Goal: The goal of this assignment is not only to think about possible ways that a search engine might rank documents in response to a query but also to think about how you might find information about search engines and how you might test a hypothesis. This will also help you prepare for future assignments in which you will actually test some hypotheses about how search engines work.

Details:

This assignment is to be done individually. You may discuss the assignment with your classmates, but the work you turn in should be your own. See the *Policies* and *Academic Integrity* sections in the course syllabus.

Find aspects, beyond those explored in the current Google Garden, of web pages that someone purports to influence rank order on Google (or another search engine). Undergraduates must find two aspects, graduate students must find three. For each aspect:

- a) Describe the aspect.
- b) Give a reference to it (for example, the document or web page in which it appears) and briefly describe how you found it (for example what search engine you used and what search terms you entered, or from what site you started browsing).
- c) Give an example of two web pages (you may create your own example web pages; they do not have to be real) that would be ranked differently on account of this aspect.
- d) Describe a test that could determine if this aspect really is influencing rank order.

Turn in your answers as hard copy at the beginning of class on January 16, 2007. Make sure your answers are legible. Word processed is best.