CPS 49S Google: The Computer Science Within and its Impact on Society - Spring 2008
Homework 1

- Due date: Thursday, Jan 31, 2008, 1.00 PM (before class). Late submissions will not be accepted (unless there are documented excuses from the dean).

- Submission: You should create a web page containing your answers and upload the page to your Duke personal web space as per the instructions at: http://oit.duke.edu/web-multimedia/web/webpage/index.html. Email the URL of this web page to shivnath@cs.duke.edu

- Don't forget to test using a browser that the URL you send is valid, and that your web page shows up properly.

- Email questions to shivnath@cs.duke.edu

- Total points = 100.

Question 1 [Points 10]

1. What is the URL of the official web page of Duke basketball coach Mike Krzyzewski?

2. What is the title of this web page?

3. Which keywords does the author of this web page give as the keywords under which the page should be indexed by a search engine?

Question 2 [Points 5] What is the IP address of Stanford University's web server? How did you find it?

Question 3 [Points 25]

A. Define the Database of Intentions in your own words.

B. List four different types of data that a search engine can collect, which can then be used to generate the Database of Intentions. Hint: One type of data is the time when each search is done.

C. For each type of data that you listed above, give an example that illustrates how this data is useful in the Database of Intentions. Your examples can be fictional. For example, if we record the time when each search is done, we may be able to say that more people search for Hillary Clinton than Barack Obama in the morning; while the opposite is true in the afternoon.

Question 4 [Points 10] On Page 12 of the textbook, the author claims that the following question can be answered from the Database of Intentions: "Which politician is popular in Iowa, New Hampshire, or California, and why?" Would you agree with the author's claim? Justify your answer.

Question 5 [Points 20] In this question you are asked to compare two search engines: Google (http://www.google.com/) and Yahoo! (http://www.yahoo.com/). List four different metrics based on which you will do the comparison. (For example, one metric is how fast the search engine returns search results.) For each of these metrics, report real results based on searches you do at the search engines.

Question 6 [Points 10] When the author talks about the graph on Page 28 of the textbook, he talks about a "massive tail".

A. What is the author referring to?

B. Do you think the "massive tail" is a good thing for Google? Justify your answer briefly.

Question 7 [Points 10] For each of the following queries, state whether the query is informational, transactional, navigational, or more than one of these.

A. USA Patriot Act

B. Mike Krzyzewski

C. Pirates Movie Tickets

D. Google Maps

E. Duke

Question 8 [Points 10] What does Chapter 2 of the textbook describe as the breakthrough that "led to a revolution in the field of information retrieval"?