CPS 49S Google: The Computer Science Within and its Impact on Society - Spring 2008
Homework 2

- Due date: Friday, Feb 29, 2008, 5.00 PM. 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**]

Page 41 of the textbook refers to a "breadth algorithm". In your own words, explain this concept and the problem that it solves.

Question 2 [**Points 5**]

The Lycos search engine used techniques to determine the meaning of a page (Page 53 in the textbook). Explain using one or more examples how knowing "the meaning of a page" helps a search engine.

Question 3 [**Points 5**]

List three features that were pioneered by the Excite search engine.

Question 4 [**Points 10**]

Page 105 of the textbook states: "Gross studied his IdeaLab companies' traffic acquisition numbers and computed the costs of each company's campaigns down to the single visitor."

1. What type of information was Gross looking for?
2. How does knowing this information help?

Question 5 [**Points 10**]

Page 108 of the textbook states: "... it pays to be a supply-side sugar daddy in the middle of a high-demand transaction with clear market imbalances." Relate this statement to search engines, users who search using these search engines, and advertisers who advertise on these search engines.

Question 6 [**Points 10**]

In your opinion, what were the main reasons why GoTo.com/Overture did not become as successful as Google.

Question 7 [**Points 15**]

1. Give the PageRank equations for all fives pages A, B, C, D, and E in Figure 1. Assume $d = 0.85$.
2. Compute the PageRanks of these five pages.

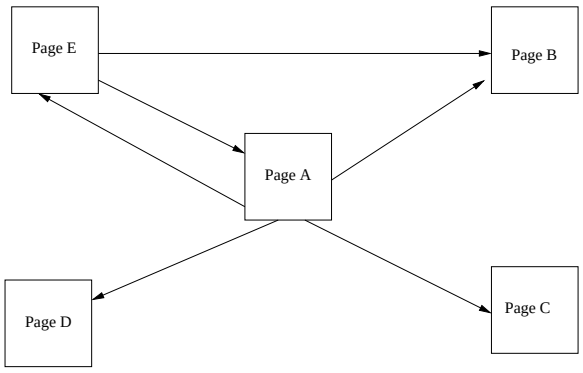

Figure 1: Link graph for Question 7

Question 8 [**Points 10**]

Assume that you are the owner of the Web site dukeathl.com. (Note: this Web site is for the purposes of this question only, and does not really exist.) You are considering the following two options:

A. Getting a link to your Web site from a popular portal like Yahoo.com.

B. Adding 10 more links between pages within your Web site.

1. Which option is better if you want to increase the average PageRank of your Web site?
2. Which option is better if you want to spread out the PageRanks more evenly among pages on your Web site?
3. Which option will be more expensive to implement?

Question 9 [**Points 15**]

In the reading "The Anatomy of a Large-Scale Hypertextual Web Search Engine," the authors list several challenges that a search engine has to deal with. These challenges include:

1. The number of pages on the Web is very large, and is growing rapidly.
2. The number of searches done per day has increased a lot over the last few years.
3. The Web has become more heterogeneous.

In your own words, state the implications of each of the above three challenges on the Crawler, Indexer, and Searcher components of a search engine.

Question 10 [**Points 10**]

1. Using examples illustrate why synonymy and polysemy in the English language make a search engine's task difficult.
2. What is a semantic network, and how does it help a search engine?