

QR 48/CSCI E-2 Problem Set 2

Due Monday, March 2 at the beginning of class. Show how you got your answers. Use reasonable units (e.g., say "1 second" instead of "1,000,000 microseconds").

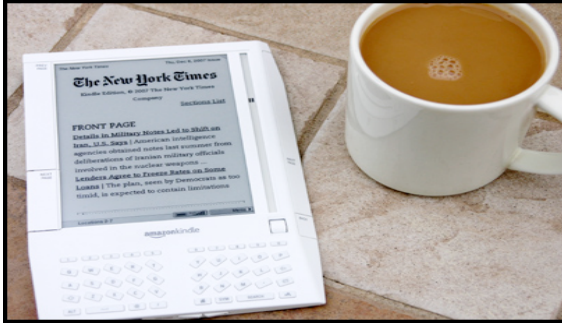
1) Linear Search and Binary Search

Imagine that we have created a dictionary for Harvard students, including names of buildings, classes, phrases like "illegitimum non carborundum," etc. A monumental effort produced a dictionary of 6905 terms, but left no time to put the entries in alphabetical order.

- A.** (3 points) If we use linear search and it takes only a second to check a single entry, what is the longest it would take to find a particular word?
[answer: $6905s/60 = 1hr\ 55m\ 5sec$]
- B.** (3 points) We want to find the word by using binary search. In order to be able to use this technique, what do we need to do to the list of entries in our dictionary?
[answer: order/sort the list – eg: alphabetically]
- C.** (3 points) If we are sure that we can use binary search and it takes one second to check an entry, what is the longest time it might take to find the word "Sever"?
[answer: $\lg(6905) = 12.75sec$]
- D.** (3 points) Our research team has added 8400 more words to the dictionary, slightly more than doubling its size. Now how does it take for a linear and binary search for a particular word (in the worst case, of course)?
[answer: $\lg(15305) = 13.9$, so 14sec]
- E.** (3 points) How big a dictionary could we search in an hour using binary search?
[2^{3600}]
- F.** (2 points) Give a sense of how big the number of part (E) is, by comparison with things in the physical universe.

2) Google Trends

Amazon Kindle is an e-book reader, a system for reading electronic books launched in U.S by Amazon.



- A. (3 points) Go to www.google.com/trends and search for kindle. You should see a graph. What does the top line show? What does the bottom line in the graph represent? By looking at the graph, Can you guess around what period the Kindle was launched? Note that you can focus on particular states and countries. Are there any places where interest in Kindles seems to be particularly high or low?
- B. (3 points) Check the last three months in 2008 for Xbox searches. When are people looking for Xboxes? Why, do you suppose? Do the answers vary by region?
- C. (3 points) Check all of 2008 and each individual month for searches for "bankruptcy" and "hangover." What patterns do you see? Can you explain them?
- D. (3 points) Obviously Google is collecting a lot of data about searches. Based on these experiments and your readings from *Blown to Bits*, succinctly and specifically list (1) ways in which this information is useful, and to whom, and (2) privacy concerns it might raise.

3) Cookies

- A. (4 points) Look through cookies on your own computer. Find at least 3 cookies that you are surprised that you have (e.g., a website you don't recognize, or one where you didn't expect to be tracked when visiting). Also, see if you can find a cookie that reveals personally identifiable information about yourself. For instance, look for your zip code, country of residence, or email address.
- B. (3 points) List the domains where these cookies come from and explain how their existence is surprising.

How to find your cookies:

Internet Explorer (Windows):

<http://kb.iu.edu/data/ajfh.html>

Internet Explorer (Mac):

<http://kb.iu.edu/data/ajfj.html>

Mozilla Firefox:

<http://kb.iu.edu/data/ajfi.html>

Safari:

<http://kb.iu.edu/data/amhi.html>

4) Traceroute

A tool called "traceroute" will show you every step of the path from a given computer on the Internet to another computer. System administrators often use this tool to

determine where network failures are occurring. You can also use it to find out who is responsible for delivering your packets.

- A. (3 points) Use traceroute to a few web sites using this page (which starts tracing from California-based company Dreamhost):

<http://www.yougetsignal.com/tools/visual-tracert/>

You might try a few foreign URLs to see how traffic goes across national boundaries. For example, look at the Chinese search engine baidu.com or the Dutch ISP xs4all.com. What do you observe about the nature of these "hops" and the connections between?

- B. (3 points) Find a URL that takes a relatively large number of hops to reach. What is the maximum number of hops for any of the sites you've tried? Are your packets going through any place where you might fear they are being monitored?

5) Primary and Secondary Storage



This is a picture of a hard drive similar to what you would find inside most personal computers. To retrieve a single piece of data from a new location on the disk, the head (at the point of the acute triangle pointing toward the center of the disk) has to be moved so it's on the right "track" (concentric circular band around the center of the disk) and the disk has to spin until the correct spot on that track is under the head. The time to move the head is called "seek time." (The soft clicking noise you may hear is the head moving.) The delay waiting for the disk to spin into position is called "latency."

- A. (2 points) Suppose the head has to move to a random spot on the disk to retrieve a record. Seek time is about 10ms. If the disk is spinning at 10,000 RPM (revolutions per minute), is seek time more or less than the average latency? (Show your work!)
[1minute = 1000ms * 60sec, so 60000ms/minute. 10000/60000 = 1/6 rotation per millisecond. Average latency takes .5 rotation, which equals 3ms. So, average latency is less than average seek.]
- B. (2 points) If the disk diameter is 3.5 inches and it can transfer 50 Mbyte/s from the outer track once the head is over the right spot on the disk, about how many bits are stored per inch of the disk track?
[3.5 * pi = ~11in. 50 Mbytes / 10,000 = 5.12Kilobytes per rotation. 5.12/11 = ~476.6 bytes = ~3813 bits.]
- C. (2 points) For a data transfer of 1KByte, what percentage of the total time (seek+latency+transfer) is just getting the head over the right spot on the disk?

**[Seek = 10ms. Latency = 3ms. Transfer = 50 Mbytes = 51200 kilobytes.
51200/1000 = 51.2Kbyte/ms, so transfer takes .0195ms.
(10+3+.0195)/.0195 = ~99.85% of the time is spent]**

- D.** (2 points) If it takes 100ns to retrieve a block of 32 bits from primary memory, how much slower is secondary storage than primary storage to retrieve 1KByte?
[32bits = 4bytes. 1024/4 = 256, so 100ns * 256 = 25600ns. 13.0195ms - .0256ms = 12.9939sec differences or 13.0195/.0256 = 508 times as slow]

6) CSCI E-2 GRADUATE CREDIT ONLY

Read Google's and Facebook's Terms of service:

<http://www.google.com/accounts/TOS>

<http://www.facebook.com/terms.php>

- A.** (2 points) Suppose you never signed up for a Gmail account but you use Google as a search engine. Did you ever agree to Google's Terms of Service?
- B.** (2 points) Does it say anywhere that Google has the right to use your search queries as part of the Trends data that it shows to anyone who asks?
- C.** (2 points) Suppose you put on your Facebook page that so-and-so is your boyfriend or girlfriend, but you don't change it when the relationship ends. Could Facebook legitimately claim that you are not adhering to your responsibilities under the terms to which you agreed? Why or why not?