

Midterm

LBSC 690
October 24, 2011

Name: _____
by writing my name I swear by the honor code

Read all of the following information before starting the exam:

- Show all work, clearly and in order, if you want to get full credit. I reserve the right to take off points if I cannot see how you arrived at your answer (even if your final answer is correct).
- You have two and a half hours to complete this exam (but you shouldn't need it).
- This exam is closed book, but open notes. You may use a calculator, but you shouldn't need to.
- Justify your answers algebraically whenever possible to ensure full credit. Be sure to have units for all answers that call for them.
- Circle or otherwise indicate your final answers.
- Please keep your written answers brief; be clear and to the point. I will take points off for incorrect or irrelevant statements.
- This test has seven problems, but you only need to complete **five** questions. If you answer more than five questions, I will grade the first five completed questions. (If you change your mind on which questions you want graded, be sure to clearly identify which question shouldn't be graded.)
- Good luck!

Scenario

All of the questions will be about the same scenario. You're working at a major urban library (the Metropolis library system) that just got a grant to use technology to promote the social reading experience. Part of that plan is to promote the most popular books at kiosks in every library entrance.

Your job is to help get it off the ground; these questions will talk about creating a backend database, storing the data, deploying the data, and examining logs.

1. Your database administrator has started on a design to store the information. Currently, she has two tables ("authors" and "books") with the following information:

AuthorName	Nationality
Margaret Atwood	Canadian
Abelard	French
⋮	⋮

BookTitle	BookYear
The Blind Assassin	2000
Sic et Non	1120
Surfacing	1972
⋮	⋮

This is a *relational database*, so she needs to connect authors to the books they've written. Help her out!

- (a) What columns would you need to add to which tables?
 - i. To the author table, add an id column
 - ii. To the book table, add an id column
 - iii. To the book table, add an author column
- (b) What kind of keys would they be?
 - i. "author" would be a foreign key
 - ii. Both "id" columns would be primary keys
- (c) What values might work for these tables? (Hint: Atwood wrote *The Blind Assassin* and *Surfacing*, Abelard wrote *Sic et Non*)

ID	AuthorName	Nationality
1	Margaret Atwood	Canadian
2	Abelard	French
⋮	⋮	⋮

ID	Author	BookTitle	BookYear
1	1	The Blind Assassin	2000
2	2	Sic et Non	1120
3	1	Surfacing	1972
⋮	⋮	⋮	⋮

2. Part of the plan is to also include a small snippet of the book that people can read at the kiosk for each book (a teaser). Your database administrator tells you that there is a 60 KB limit on how much raw text can be stored. About how many words (in the conventional sense, not the computer sense) of ASCII-encoded text is this equivalent to? Is this a reasonable limit for a snippet? You may find some of the following factoids useful:

- The entropy of English is 1.1 bits
- The average English NYT newspaper article has 157.3 unique words
- The average English word has 5.1 characters
- The average English sentence has 14.3 words

Let's assume that there are 6 characters per English word (including whitespace). Because we're using ASCII, that's one byte per character.

$$\frac{60 \text{ KB}}{1} \frac{1 \text{ word}}{6 \text{ characters}} \frac{1 \text{ character}}{1 \text{ byte}} \frac{1000 \text{ byte}}{1 \text{ KB}} = 10,000 \text{ words} \quad (1)$$

3. In addition to the 60KB snippet of each book, this is a multimedia display, so they want to include a high definition picture of the author (800 KB) and an image of each book cover (340 KB).

(a) How much space would this take up in GB if there were eight thousand books and one thousand authors?

$$\begin{aligned} \frac{60 \text{ KB} + 340 \text{ KB}}{\text{book}} \frac{8 \cdot 10^3 \text{ books}}{1} + \frac{800 \text{ KB}}{\text{author}} \frac{10^3 \text{ authors}}{1} \\ = 8 \times 400 \cdot 10^3 \text{ KB} + 800 \cdot 10^3 \text{ KB} \\ = 4000 \cdot 10^3 \text{ KB} = 4 \cdot 10^6 \text{ KB} \end{aligned}$$

$$\frac{4 \cdot 10^6 \text{ KB}}{1} \frac{1 \text{ GB}}{10^6 \text{ KB}} = 4 \text{ GB}$$

(b) How long (in hours) would it take to transfer the whole dataset on a 3.2 mbs network connection?

$$\begin{aligned} \frac{4 \text{ GB}}{1} \frac{8 \text{ bit}}{\text{byte}} \frac{10^3 \text{ Mb}}{\text{Gb}} = 32 \cdot 10^3 \text{ Mb} \\ 32 \text{ Mb} \frac{1 \text{ s}}{3.2 \text{ Mb}} = 10^4 \text{ s} \frac{1 \text{ hour}}{3600 \text{ s}} = 2.8 \text{ hour} \end{aligned}$$

4. The transfers are taking longer than expected. Looking at the logs, it seems that a computer with the IP 129.302.236 is connecting to the machine and uploading files. Your boss worries that it could be a malicious outsider because the primary server for distributing new information to the kiosks is “kioskserver.metropolis.us.” The database administrator disagrees, saying that it is the kioskserver.

(a) Do you trust the logs? Why might you not?

An IP address is composed of 4 “octets” (8-bit numbers), this has five numbers, and one of them is greater than 256. So the IP address is not valid, making the logs suspect.

(b) On further investigation, you discover the IP address was actually 156.32.12.62. Given this information, how could both your administrator and your boss be right? How could you easily find out?

A domain name like “kioskserver.metropolis.us” is associated with an IP address. It could be that the kiosk server actually has that IP address. There are a variety of methods to find out:

- Ping “kioskserver” and see what IP address is used
- Look at “kioskserver” network configuration
- Connect to the IP address and see if it looks like the kioskserver (e.g. if it’s running a webserver)

(c) To placate your boss (and employ good practices), what could you suggest to ensure that malicious outsiders couldn’t connect to your kiosks? Assume you can change how machines are addressed and that the kiosks only have to communicate with “kioskserver.”

First, make sure that all of the kiosks are not visible externally. Make sure the gateway has a secure firewall. Place a firewall on the kiosks to only allow connects from “kioskserver.”

5. The interface uses a logging mechanism to keep track of what's going on when users search for popular books or browse the collection. The developer said that it generates valid XML, but you're not so sure. Find at least five problems with the XML snippet below and explain why they're incorrect.

```
<session>
<query language="spanish>
<text>Where is the bath room?</TEXT>
<time>14:01:23</time>
</queryend>
<response size="3">
<result book_id="1701" title="Tender is the night">
<result book_id="32" title="Where angels fear to tread" />
<result book_id="234" title="The Wife of Bath" />
<time>14:02:58</time>
</session>
```

- (a) Missing closing quote around spanish
 - (b) "TEXT" should be lowercase
 - (c) "query" should be ended by query, not queryend
 - (d) "result" is a singleton, and should have a trailing slash
 - (e) Missing namespace / definition tag
 - (f) There is no end tag to the "response"
6. Given the logging information from the previous question (forgetting the sloppy XML, assume everything else is correct), what interface issues can you identify with the system in terms of the following aspects and how might you (briefly) correct them?
- (a) Affordance / Constraints
 - (b) Causality / Responsiveness
 - (c) Displaying state / Individual differences

7. Finally, a question that has nothing to do with the kiosk task. This question is about representing colors in HTML and hexadecimal.

(a) What color (use words) is represented by the color #00FF00? How about #0000FF?

The first is “green,” the second is “blue.”

(b) What color do you get if you add the two colors together (show your work in **any** base you’re comfortable with).

$$\begin{array}{r} 00 \text{ FF } 00 \\ + 00 \text{ 00 FF} \\ \hline 00 \text{ FF FF} \end{array}$$

This is “cyan.”

(c) What color is represented by #00FF01? Would most people be able to discern it from #00FF00?

It’s “green” with the tiniest bit of blue. No, most people would not be able to discern it from #00FF01.

(d) What color is represented by #00FF01 + #0000FF (again, show your work). Why is this such a different answer from part (b)?

$$\begin{array}{r} 1 \text{ 11 } 1 \\ 00 \text{ FF } 01 \\ + 00 \text{ 00 FF} \\ \hline 01 \text{ 00 00} \end{array}$$

This is “black” with the very slightest bit of red.