LBSC 690 Midterm Exam **Fall 2007**

Name: __Jimmy Lin_____

You have 150 minutes to complete this exam. Time will begin as soon as you start reading the first question.

- You may use any material, including the text book, lecture slides, and notes. You may also use anything found on the Internet that existed before the exam started.
- You may NOT communicate with any other person during this exam, either in person or using electronic means.

As strategies for completing the exam, keep the following in mind:

- If you find a question to be ambiguous, make reasonable assumptions as you see fit, but write down your assumptions.
- You are more likely to get partial credit for a wrong answer if you show your work.
- Be careful not to get carried away and run over the time limit by spending too much time on one question. Plan ahead, and don't devote more time to a question than it is worth.

Please write your answers in the space provided.

Score Summary (for use by grader)

Question	Possible points	Actual points
1	15	15
2	20	20
3	30	30
4	20	20
5	15	15
TOTAL	100	100

1. [15 points total] Moore's Law. Give three separate reasons why librarians, archivists, or anyone that works in an information-rich environment should pay attention to Moore's Law. No more than a couple of sentences for each reason. Organize as a bulleted list.

Loosely interpreted, Moore's Law says that computer speed doubles every 18 or months; a similar prediction is made about storage capacity. It is important for several reasons:

- Planning. When planning for a long-term project, the technology available tomorrow will be very different than the technology available today.
- Increased storage capacity means that it's easier to store everything rather than figure out what to keep. However, archival theory tells us this is bad practice, and only pushes problems off into the future. This puts tremendous pressure on appraisal, indexing, management, etc.
- Computers are getting faster, able to store more, etc., while human cognitive capacity remains essentially unchanged. This puts increasing pressure on interfaces to cope with information overload.
- Obsolescence of retained records and software. Advances in processing power usually means that software and file formats are rapidly falling out of date. Organizations need to implement plans to ensure that old data remains accessible.

[Grading note: I did not give full credit if you gave generic answers that don't actually have anything to do with Moore's Law]

2. [20 points total] HCI. Do you think Washington D.C. Craig's List is a well-designed site? Explain why or why not. See the following URL: http://washingtondc.craigslist.org/

I would like you to answer this question from two perspectives: static design (e.g., site organization) and interaction design (e.g., supporting user navigation). For each, give at least three reasons. In all cases, illustrate with specific design principles and concepts discussed in class. Organize your discussion with succinct bullet points.

NOTE: You do not need to argue exclusively one way or another: in other words, you can say "these aspects represent good design, those aspects represent poor design".

A. [10 points] Static design

- **Good:** organization based on function, based on what you want to do.
- **Good:** organization based on demographics, based on who you are.
- **Good:** leverages transfer effects, like newspaper classified ads.
- **Good:** consistent page layout
- **Bad:** poor contrast to separate headings from content
- **Bad:** poor use of spacing
- **Bad:** unclear what many of the labels refer to

B. [10 points] Interaction design

- **Good:** quick access to content, able to get to actual ads quickly
- **Good:** static categories breeds familiarity, especially across different locations
- **Bad:** limited support for lateral navigation, difficult to access related content
- **Bad:** limited support for browsing (ads in reverse chronological order only)
- **Bad:** limited support for browsing (difficult to remember which ads you've seen)
- **Bad:** poor search capabilities

3. [30 points total] Jimmy's Data Pigeon Service. Since I collaborate a lot with colleagues in computer science, I often need to move data between Hornbake and A.V. Williams (where my colleagues are located).

For all these questions, you can assume that 1MB = 1000KB, 1GB = 1000MB, etc.

Typically, I get a transfer rate of 200 KB/second on FTP between Hornbake and A.V. Williams. The latency of the connection between these two locations is 100ms.

A. [4 points] In total, how long will it take to FTP 100 KB from Hornbake to A.V. Williams?

0.5 seconds transfer time + 0.1 seconds latency = total 0.6 seconds

B. [4 points] In total, how long will it take to FTP 200 KB from Hornbake to A.V. Williams?

1.0 seconds transfer time + 0.1 seconds latency = total 1.1 seconds

C. [4 points] In total, how long will it take to FTP 1 MB from Hornbake to A.V. Williams?

5.0 seconds transfer time + 0.1 seconds latency = total 5.1 seconds

D. [4 points] In total, how long will it take to FTP 1 GB from Hornbake to A.V. Williams?

5000 seconds transfer time + 0.1 seconds latency = total 5000.1 seconds

E. [5 points] Generalize the effect that latency has on total transfer time as the amount of data you are transferring goes up.

Latency becomes increasingly negligible as the amount of data goes up.

F. [5 points] Say I train a pigeon to carry USB thumb drives from Hornbake to A.V. Williams; let's examine how fast data can be transferred using this method. Assume that I am using a USB drive that holds 0.5 GB. Here are the steps:

1. I copy data over to the USB drive. It takes 30 seconds to fill up the entire drive.
2. I attach the USB drive to the pigeon and send the pigeon over to A.V. Williams. This takes 30 seconds.
3. My colleague detaches the USB drive and copies out the data. It takes 30 seconds to copy everything off the drive.
4. My colleague sends the carrier pigeon back to me. This takes 30 seconds.
5. Process repeats.

Given this system for moving data around, how long does it take to move 2 GB over to my colleague in A.V. Williams from Hornbake?

8 minutes.

G. [4 points] Compared to FTP, characterize using pigeons to move data around:

Bandwidth: **higher**
Latency: **higher**

4. [20 Points total] XML as an enabler. We spent a significant amount of time in class talking about XML as an enabling technology, i.e., XML enables new capabilities that weren't possible before. For example:

- **MathML** allows users to display complex math equations in Web browsers.
- **RSS** allows users to control what they read, through RSS readers that perform aggregation, filtering etc.
- **XML** is used by large companies to manage their supply chains: to lower cost, ensure customers are better served, etc.
- **XML** is used in interlibrary loan to facilitate faster transactions, with lower rates of error.

You may not know this, but Apple iTunes stores metadata about your music collection in XML. Briefly describe 4 new capabilities that this enables. No more than a couple of sentences per capability. Organize as bullet points.

Note that this question focuses on potential capabilities; i.e., what could be done, which is not necessarily what *is* currently being done.

- Export of metadata to other applications, e.g., databases (to organize your personal music library).
- Alternative methods to search and browse your music collection. For example, through XSLT transforms.
- Ability to share metadata with others over the Web. For example, to fill in missing track information.
- Ability to aggregate listening habits. For example, to make music recommendations.

5. [15 points total] Potpourri. True or false?
(If you find a question ambiguous, explain why you think so.)

T The point of caching is to keep data needed by the processor in faster memory.

F In a packet-switched network, packets may arrive out of order because the DNS server tells the router to send the packet to different places.
The actual packet routes have nothing to do with DNS.

F From an HTML page on your local computer, it is possible to create a link to an HTML on your live homepage using a relative path.

F In a network with a star topology, failures of different computers affect the network in the same way.
If the hub goes down, so goes the entire network.

F A routing table must have a separate entry for every IP address.
Routing entries can have wildcards.

T One can transform an XHTML page into another XHTML page using XSLT.
XHTML is XML, so it can be transformed using XSLT. Why would you do this? For example, to reformat the appearance of your website.

F Since in a typical computer RAM is smaller than the hard drive, processors are optimized to directly access data on the hard drive.