LBSC 690 Final Exam **Fall 2007**

Name: _____

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.

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

- 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 spend 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	10	
2	15	
3	25	
4	10	
5	20	
6	20	
TOTAL	100	

1. [10 points total] Multimedia. Let's say you are shopping for digital cameras and are comparing 8 megapixel cameras to 10 megapixel cameras. You know what a pixel is (from Lecture 9) and you know that the megapixel rating describes the number of pixels in the sensor inside the camera that captures the image.

So, the more pixels, the better, right? You ask a friend for advice, and she says, "Nahh, if you're just going to print standard 4x6 photos, 8 versus 10 megapixel cameras aren't going to make much of a difference."

A. [5 points] Is she right? Explain in terms of properties of the human perception system discussed in Lecture 9.

Your friend is correct. Recall that pixels are simply "dots" that comprise an image; if you have small enough dots that are spaced close enough together, the viewer sees a coherent image. Making the dots smaller doesn't improve image quality.

However, your friend is quick to add "But if you're thinking of printing 8x10 photos or larger, then go with as many megapixels as you can afford."

B. [5 points] Once again, is this good advice?

Yes. When you increase the size of raster images, the size of each pixel increases. Thus, with larger images you would benefit from more pixels.

2. [15 points total] Information Retrieval. This question is about inverted indexes, which we discussed in Session 11. For any inverted index built from a collection of documents:

A. [3 points] How many rows are there going to be? (describe in terms of characteristics of the collection)

The inverted index will contain as many rows as there are unique words in the collection.

B. [3 points] How many columns are there going to be? (describe in terms of characteristics of the collection)

The inverted index will contain as many columns as there are documents in the collection.

Let's say I add one more document to the inverted index:

C. [4 points] How is the number of columns going to change?

The number of columns is going to increase by one.

D. [5 points] How is the number of rows going to change?

The number of rows is going to increase by the number of words in the new document that is new to the entire collection (i.e., that don't exist in any other documents).

3. [25 points total] Database design.

A. [10 points] Design a relational database for storing bibliographic entries such as the following:

> Dina Demner-Fushman and Jimmy Lin. Answering Clinical Questions with Knowledge-Based and Statistical Techniques. Computational Linguistics, 33(1):63-103, 2007.

To make the task more tractable, focus only on standard journals (with volume, number, pages, year). Sketch out your table structure with a few sample citations.

Citation Table

CitationID	Title	Journal	Vol	Num	Pages	Year
1	Answering Clinical Questions with Knowledge-Based and Statistical Techniques	Computational Linguistics	33	1	63-103	2007
2	An Exploration of the Principles Underlying Redundancy-Based Factoid Question Answering.	ACM Transactions on Information Systems	27	2	1-55	2007
…						

Author Table

AuthorID	First	Middle	Last
1	Jimmy		Lin
2	Dina		Demner-Fushman
…			

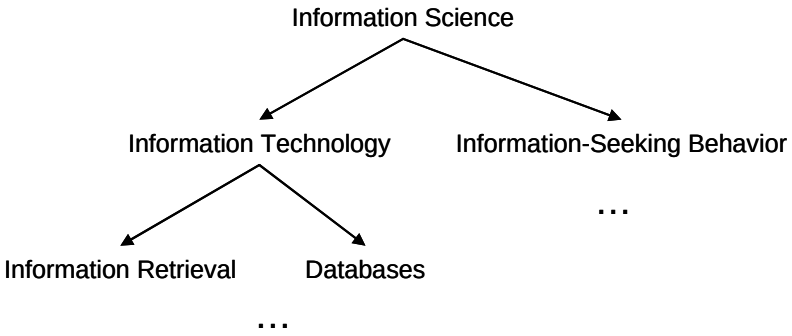
Authorship Table

AuthorID	CitationID
1	1
2	1
1	2
…	

B. [5 points] Describe the SQL query you would execute to find out the titles of all of my publications.

```
select Title from Author, Authorship, Citation
where  Author.AuthorID = Authorship.AuthorID &&
       Citation.CitationID = Authorship.Citation ID &&
       First = "Jimmy" && Last = "Lin"
```

C. [6 points] It would also be a good idea to assign topic descriptors to research articles. For example, consider the following fragment from a hypothetical topic hierarchy:



Design a database that would be able to store hierarchically-arranged topic descriptors such as the example above. Focus only on the hierarchy itself; don't worry about the actual part of assigning topics to articles. HINT: you can do this with one table.

Hierarchy Table

Descriptor	Parent
Information Science	
Information Technology	Information Science
Information-Seeking Behavior	Information Science
Information Retrieval	Information Technology
Databases	Information Technology

D. [4 points]
Describe the SQL query you would execute to find the parent of a particular descriptor.

```
select Parent from Hierarchy where Descriptor = "Information Technology"
```

Describe the SQL query you would execute to find the children of a particular descriptor.

```
select Descriptor from Hierarchy where Parent = "Information Technology"
```

4. [10 points total] Moving bytes around, again… Let's say I'm able to move 1 MB of data from point A to point B in 10.2 seconds.

A. [4 points] If I know my latency is 0.2 seconds, what's my transfer rate?
You can assume that 1MB = 1000KB.

100 KB/second.

B. [3 points] How long will it take to transfer a 5 MB file?

50.2 seconds

C. [3 points] How long will it take to transfer a 2 byte file? (it says, "Hi")

0.2 seconds

6. [20 points total] XML, again. The Poughdunk County library system prides itself on being a technologically-sophisticated operation. For example, it uses an XML standard to store catalog records. Recently, Ben Bitdiddle, head of IT services, had the idea of augmenting the catalog records with the size of the books, to feed into some new fancy shelving algorithm. He came up with something like this:

```
<size>
  <height>11 inches</height>
  <width>8.5 inches</width>
  <thickness>1.75 inches</thickness>
</size>
```

It seemed like a good idea until entries like the following started showing up in catalog records (excerpts):

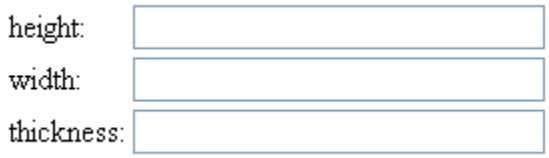
- … <thickness>1 inches</thickness>…
- … <height>8.5 in</height>…
- … <height>14 ½ in.</height>…
- … <width>5'</width>…
- … <width>25 cm</width>…

You've been hired as a high-powered consultant (haha) to fix the problem.

A. [5 points] What is one problem with inconsistently-formatted values?

Inconsistently-formatted values make automatic processing of XML records difficult. For example, it would be difficult to find all oversized books.

B. [5 points] You look at the Web-based interface for data entry designed by Ben Bitdiddle. It looks like this:



Why is this a poorly-designed interface?

The interface doesn't provide any visible constraints for entry of the data.

C. [5 points] Describe a purely interface-based solution to the problem.

Provide a drop-down list that lets the user select inches or centimeters. Provide an example entry.

D. [5 points] Revise the above XML so that size can be captured in a more consistent manner. That is, show what a better XML standard might look like.

```
<size>
   <height unit="inches" format="decimal">11.0</height>
   <width unit="inches" format="decimal">8.5</width>
   <thickness unit="inches" format="decimal">1.75</thickness>
</size>
```

6. [20 points total] JavaScript.

The funny operator % (called "modulo") gives you the remainder of a division operation. So, for example:

- 4 % 3 = 1, since you get one left over when you divide 3 into 4
- 5 % 3 = 2, since you get two left over when you divide 3 into 5
- 6 % 3 = 0, since 3 divides evenly into 6

A. [4 points] Consider the following function:

```
function bar(i) {
  if ( i % 2 == 0 ) {
    return true;
  } else {
    return false;
  }
}
```

What is the result of bar(1)? **false**

What is the result of bar(2)? **true**
What is the result of bar(3)? **false**
What is the result of bar(4)? **true**

B. [2 points] The function bar tells you if the number is *what*? (That is, what property of the number does the function test for?)

The function tells you if the number is even.

C. [7 points] What is the output of the following code?

```
var i = 0;

for ( i=1; i<=10; i++ ) {
  if ( bar(i) ) {
     document.writeln("tock\n");
  } else {
     document.writeln("tick\n");
  }
}
```

tick
tock
tick
tock
tick
tock
tick
tock
tick
tock

D. [7 point] Write a Javascript function to test if a number is evenly divisible by both 3 and 5.

```
function baz(i) {
  if ( i % 3 == 0 && i % 5 == 0) {
    return true;
  } else {
    return false;
  }
}
```