

Name: _Jimmy Lin_____

You have 150 minutes (2.5 hours) to complete this exam. There are seven questions. 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 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	10	10
2	20	20
3	10	10
4	15	15
5	8	8
6	7	7
7	10	10
TOTAL	80	80

Question 1. By land, air, or sea? (10 points)

Let's say I have access to two separate ways for sending data across the pond to London, with the following characteristics:

- Satellite link: higher latency, higher bandwidth
- Undersea cable: lower latency, lower bandwidth

Part A. (3 points) Of the two, which would you choose for having a telephone conversation with someone in London? In a sentence, explain why.

In a telephone conversation, latency (delay between speaking and hearing the response) is more important; the amount of bandwidth needed to transfer human voice is comparatively small. As such, the undersea cable would be the better option.

Part B. (3 points) Of the two, which would you choose for sending a DVD's worth of information to someone in London? In a sentence, explain why.

Since we're sending lots of data, bandwidth is more important. As such, the satellite link will be the better option.

Part C. (4 points) Let's say I put data on a hard drive and ship it to London via overnight mail. Characterize this method of transmitting information: (less than ten words for each)

Latency: high

Bandwidth: high

Question 2. Databases (20 points)

Develop a database for managing a gift list. You will need to keep track of people (make up some properties like “age”) and information about gifts (e.g., make up some properties like “price”). Importantly, each person may want more than one gift, and the same gift may be desired by more than one person.

Part A. (12 points) Design this database. Sketch out your table structure; identify primary/foreign keys. Make up sample data to fill in a few rows. To receive full credit, the examples should demonstrate a person wanting more than one gift and a gift being desired by more than one person.

Table 1-Wisher Table: Wisher ID [primary key], Name, Address, Email

Table 2-Gift Table: Gift ID [primary key], Gift, Price

Table 3-Gift Match: Wisher ID [foreign key], Gift ID [foreign key]

Table 1-Wisher Table:

Wisher ID	Name	Address	Email
1	Jane Smith	123 K St WDC	smithj@gifts
2	Joe Allen	345 C St WDC	allen@gifts
3	Peggy Jones	456 E St WDC	Jones@gifts

Table 2-Gift Table:

Gift ID	Gift	Price
1	Top	\$5.00
2	Ball	\$3.00
3	Doll	\$4.00

Table 3-Gift Match:

Wisher ID	Gift ID
1	1
2	2
3	1
1	3

Part B. (2 points) The relationship between people and gifts is: (pick one)

one-to-one

one-to-many

(many-to-many)

Part C. (6 points) Pick a person. Write a database query to list the gift or gifts that he or she wants. Show the results of the query. Either an SQL statement or a query in terms of conceptual operators (join, project, restrict) is acceptable.

Join tables 1, 2, and 3 (Wisher ID and Gift ID as join criteria)

Restrict Name to “Jane Smith”

Project Name and Gift

Jane Smith wants a top and a doll.

Question 3. Buffering and Rebuffering (10 points)

Believe it or not, I actually like to watch Congressional testimonies online at C-SPAN.org (just kidding!). Anyway, I go to the site, click on a link that pulls up RealPlayer. However, the video does not start immediately. It says “buffering” in the status window of the player. In at most a couple of sentences, tell me what the software is doing.

RealPlayer is caching the video in a local buffer. The cache is a mechanism to recover from hiccups in the network connection. If the network suddenly drops, data will continue to stream from the buffer.

I’m happily watching the exciting testimony, and all of a sudden it stops. The status window of the player says “rebuffering”. In at most a couple of sentences, tell me what has happened and what the software is doing.

The buffer is empty due to a prolonged loss of network connection. The player is refilling the buffer.

Question 4. Programming (15 points)

Consider the following JavaScript program:

```
var n = 4;
var i = 0;
var total = 0;
for (i=0; i<n; i++) {
    total += i * i;
}
```

Part A. (5 points) What does the program do?

Sum up the square of numbers.

Part B. (5 points) What value is stored in the variable **total** after the program completes execution?

$(0 \times 0) + (1 \times 1) + (2 \times 2) + (3 \times 3) = 14$

Part C. (5 points) What is the importance of “**i++**” in the **for** statement? What would happen if I took it out?

It increments the variable i after each iteration of the loop. Taking the statement out will result in an infinite loop.

Question 5. Search (8 points)

Name **two** natural language issues that make keyword-based querying difficult. In one sentence each, briefly explain what they are. For each, describe **one specific** strategy, device, technique, or method that attempts to address the difficulty (something overly broad like “conceptual search” or “natural language understanding” does not count).

Synonymy – multiple words, same meaning

Use **Boolean OR** to include term variants, e.g., (cat OR feline)

Polysemy – one word, multiple meanings

Use **Boolean AND** to define a context: e.g., (bank AND river) for “side of river”, (bank AND money) for “financial institution”.

Question 6: Huh? (7 points)

The perfect query paradox: In order to formulate a good query to get what you’re looking for, you have to already know what it is you’re looking for. However, if you already know what you’re looking for, why are you searching in the first place?

In at most a few sentences, critique this “paradox”.

The paradox is resolved through interaction and iteration. The user starts with an imperfect query that reflects an incomplete understanding of the problem. Both evolve through user-system interactions (by, for example, browsing retrieved results).

Question 7. Policy and Privacy (10 points)

Congress is interested in protecting the privacy of Internet users and your congressional representative has asked for your opinion on that subject.

- Identify **one** threat to the privacy of Internet users.
- Identify the capabilities and limitations of **one** technical means for addressing that threat.
- Identify **one** policy option that might be used to address that threat through legislation or regulation, and then state and justify your opinion on whether the legislation or regulation is a good idea.

Please answer this question carefully. You will only receive full credit if you address **all** bullet points. You will not receive extra credit for providing more than one response when only one was asked for. I am looking for a well-crafted, succinct argument, not everything you know about privacy.

Threat: Phishing email attacks that attempt to steal your identity

Technical solution: Spam filters, but they are not perfect. Some Phishing emails appear very convincing

Policy option: Make spam illegal (already done).