Department of Computer Science and Engineering
University of California, San Diego

Midterm Examination (REVISED)

Wednesday October 31, 4:40pm to 6pm

Your full name:

Instructions: Look through the whole exam and answer the questions that you find easiest first. Answer each question in the space below the question, using the backs of the pages for extra space as necessary. There are three questions, three pages, and 50 points in total. Whenever you make an assumption, state it clearly.

You may bring and use the following materials: one PHP book and one MySQL book, your own personal hand-written notes, documents handed out in class, and a printed copy of the published lecture notes. You may not use any other materials.

(Question 1) [15 points] Definition: An identifier is a series of mixed letters and digits, of length one or more, where the first character is a letter. Assume that identifiers are as long as possible, so two identifiers cannot overlap. Any characters that are not digits and not letters may appear between identifiers, not just white space.

Your answer to each of the following parts should be a PHP function that is efficient and understandable.

(a) [5 points] Carefully write a PHP function using regular expressions that returns the first identifier present anywhere in a long string.

(b) [5 points] Carefully write a PHP function that returns in an array all the identifiers in a string.

(c) [5 points] Carefully write a PHP function that returns the last identifier in a string. (For efficiency, this function must not be based on extracting all the identifiers in the string.)

(Question 2) [20 points] This question refers to a table named `president` with schema as follows:

```
+------------+-------------+------+-----+------------+-------+
| Field      | Type        | Null | Key | Default    | Extra |
+------------+-------------+------+-----+------------+-------+
| last_name  | varchar(15) |      |     |            |       |
| first_name | varchar(15) |      |     |            |       |
| state      | char(2)     |      |     |            |       |
| city       | varchar(20) |      |     |            |       |
| birth      | date        |      |     | 0000-00-00 |       |
| death      | date        | YES  |     | NULL       |       |
+------------+-------------+------+-----+------------+-------+
```

If desired, you may use subselects, even though our version of MySQL does not provide these.

(a) [2 points] Write an SQL query that returns the full names of presidents who are still alive.

(b) [3 points] Write an SQL query that finds the state in which the greatest number of presidents have been born.

(c) [4 points] Can you write a query that finds the states in which no presidents have been born? Explain your answer.

(d) [4 points] Write an SQL query that finds the full name of every president with the same last name as some other president.

(e) [4 points] Why is the following query likely to be slow, regardless of what indexes are available?

```
select * from president where 2001 - year(birth) < 75
```

Explain how to write a different query that computes the same result but is much faster.

(f) [3 points] When you are testing the speed of an SQL query with and without an index, it is important to run each version of the query several times. Explain why.

(Question 3) [15 points]

The description of the current project says "The content browser should use the information already in the database at the time of each request. You should not implement any real-time retrieval of news" (Design A). This design uses a central database.

Alternatively, to answer a user's query, your software could gather information directly from other web sites, in real-time (Design B). This design does not need a database for storing information from the other web sites.

Neither design is always better. Each design has advantages and disadvantages.

Along each of the following dimensions, explain at least one important advantage or disadvantage for Design A. Also explain at least one important advantage or disadvantage for Design B.

If an advantage for one corresponds to a disadvantage for the other, you should still give a brief explanation separately for each design.

(a) [3 points] scalability;

(b) [3 points] convenience of implementing alternative presentation media (for example web-based and telephone-based);

(c) [3 points] privacy for users;

(d) [3 points] one important aspect of response quality (any aspect of the service that makes users happy or unhappy);

(e) [3 points] a second different important aspect of response quality.