

UCLA
 Computer Science Department
 Winter 2011

Instructor: J. Cho

Student Name and ID: _____

CS144 Midterm: Closed Book, 90 minutes

(IMPORTANT PLEASE READ **):**

- *Simplicity and clarity of your solutions will count.* You may get as few as 0 point for a problem if your solution is far more complicated than necessary, or if we cannot understand your solution.
- If you need to make any assumption to solve a question, please write down your assumptions.
- To get partial credits, you may want to write down how you arrived at your answer step by step.
- You may use one sheet of double-sided notes during exam. You are also allowed to use a calculator.
- Please, write your answers neatly. Attach extra pages as needed. Write your name and ID on the extra pages.

Problem	Score	
1	10	
2	10	
3	30	
4	15	
Total	65	

Problem 1: 10 points

Yesterday when you were surfing the Web through the “W3C-standard-compliant” browser, named “IWillBeBack”, you happen to visit the “comeagain.com” site. This is the first time you visit the Web site and the followings are the HTTP request and response that were exchanged during the first visit:

Request:

```
GET / HTTP/1.1
Host: comeagain.com
User-Agent: IWillBeBack
Referrer: http://too-easy-for-me.com/problem1
```

Response:

```
HTTP/1.1 200 OK
Set-Cookie: uid=5; expires=Wed, 31-Dec-2050 23:59:59 GMT
Content-Length: 40
Connection: close
Content-Type: text/html
```

```
<html>
<body>Welcome!</body>
</html>
```

Just before coming to this midterm, you visited the `comeagain.com` site again. In particular, you visited the registration page at `http://comeagain.com/users/register.html` that has the following content:

```
<html>
<head><title>Register your name and phone number</title></head>
<body>
Please provide your name and phone number to enter our weekly lottery!
<form action="register.php" method="GET">
  Your name:
  <input type="text" name="name"><br/>
  Your phone number:
  <input type="text" name="phone"><br/>
  <input type="submit">
</form>
</body>
```

From this page, you submitted your name Jason and your phone number 310-555-1234 by filling in your name and phone number at the appropriate fields and clicking the submit button.

Please write down the HTTP request sent to the server when you clicked the submit button.

In writing your answer, assume that you used the same “IWillBeBack” browser and made no changes to its settings. Thus, any optional HTTP header field was present in the earlier request should stay as long as the field is still relevant. Before writing down your answer, please carefully read the first HTTP exchange and determine which header field should stay.

NOTE: unless necessary, please do NOT make any other assumptions about optional HTTP header fields.

ANSWER:

```
GET /users/register.php?name=Jason&phone=310-555-1234 HTTP/1.1
Host: comeagain.com
Cookie: uid=5
User-Agent: IWillBeBack
Referrer: http://comeagain.com/users/register.html
```

Problem 2: 10 points

Consider the following web page:

```
<html>
  <head><style type="text/css">
    p {
      font-family: "Tahoma";
      color: red;
      font-size: 20pt;
      font-weight: normal;
      background: blue;
    }
    .note {
      font-family: "Times New Roman";
      background: yellow;
      font-weight: bold;
      font-size: 12pt;
    }
    p.note { color: black; }
    #p1 {
      font-family: "Lucida Console";
      background: white;
    }
  </style></head>
  <body>
    <p class="note" id="p1">How do I look?</p>
  </body>
</html>
```

Write down the properties of the text “How do I look?” in the following space:

ANSWER:

Font-family: Lucida Console

Font-size: 12pt

Font-weight: bold

Color: black

Background: white

Problem 3: 30 points

Consider the following XML schema definition for an XML data for bookstores:

```
<?xml version="1.0"?>
<xs:schema xmlns="http://www.w3.org/2001/XMLSchema"
            targetNamespace="http://cs144.ucla.edu/">
  <xs:element name="Products">
    <xs:complexType>
      <xs:element name="Book" minOccurs="0" maxOccurs="unbounded"/>
      <xs:complexType>
        <xs:element name="Title" type="xs:string"/>
        <xs:element name="Author" type="xs:string"/>
        <xs:element name="Review" type="xs:string"
                    minOccurs="0" maxOccurs="unbounded"/>
        <xs:element name="Edition" minOccurs="1" maxOccurs="unbounded">
          <xs:complexType>
            <xs:element name="Ed" type="xs:integer"/>
            <xs:element name="Year" type="xs:integer"/>
          </xs:complexType>
        <xs:attribute name="ISBN" type="xs:string" use="required"/>
      </xs:complexType>
    </xs:element>
    <xs:element name="CD" minOccurs="0" maxOccurs="unbounded"/>
    <xs:complexType>
      <xs:element name="Title" type="xs:string"/>
      <xs:element name="Artist" type="xs:string"/>
      <xs:element name="Review" type="xs:string"
                  minOccurs="0" maxOccurs="unbounded"/>
      <xs:attribute name="AISN" type="xs:string" use="required"/>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

Assume that a book may have multiple editions published in one year. All editions of a book share one ISBN, which is unique to each book. Note that a review is linked to a book, not to a specific edition. That is, all editions of a book are linked to the same reviews. In addition, every CD has a unique AISN number. Note that book title, author, CD title, and artist are not unique.

Remember that the default value for minOccurs or maxOccurs is 1 if these attributes are not present in an element definition.

1. An XML file contains two products, a book and a CD. We summarize relevant information as follows:

- The first edition of the book “Web Applications” written by John Cho in year 2008. Its ISBN is 1423902017 and no review has been written for this book.
- A CD with the title “Circus” performed by Britney Spears. Its AISN is B001GO09MI and there are two reviews with the following content: “She is BACK!!!”, “No, it is not a comeback, she never left us!”

Please follow the XML schema we provided and write down the XML data for the above two products. Note: Pay special attention to the namespace as well as names of each element.

ANSWER:

```
<?xml version="1.0"?>
<Products xmlns="http://cs144.ucla.edu">
  <Book ISBN="1423902017">
    <Title>Web Application</Title>
    <Author>John Cho</Author>
    <Edition><Ed>1</Ed><Year>2008</Year></Edition>
  </Book>
  <CD AISN="B001GO09MI">
    <Title>Circus</Title>
    <Artist>Britney Spears</Artist>
    <Review>She is BACK!!!</Review>
    <Review>No, it is not a comeback, she never left us!</Review>
  </CD>
</Products>
```

2. Some of our programmers know only DTD, not XML schema, so we would like to provide a DTD version of the schema as well. Please write down the DTD definition that is equivalent to the XML schema provided in the first page of this problem. If an equivalent DTD definition is not possible due to DTD's limited expressive power, please provide a definition as close to the original XML schema definition as possible.

ANSWER:

```
<!DOCTYPE Products[
  <!ELEMENT Products (Book*, CD*)>
  <!ELEMENT Book (Title, Author, Review*, Edition+)>
  <!ELEMENT Title (#PCDATA)>
  <!ELEMENT Author (#PCDATA)>
  <!ELEMENT Review (#PCDATA)>
  <!ELEMENT Edition (Ed, Year)>
  <!ELEMENT Ed (#PCDATA)>
  <!ELEMENT Year (#PCDATA)>
  <!ELEMENT CD (Title, Artist, Review*)>
  <!ELEMENT Artist (#PCDATA)>
  <!ATTLIST Book ISBN CDATA #Required>
  <!ATTLIST CD AINS CDATA #Required>
]>
```

3. After reading the XML schema definition, our lead programmer proposed the following table design for storing the book information.

`Book(ISBN, title, author, edition, year, review)`

- (a) Please identify the set of nontrivial functional dependencies and multivalued dependencies that hold on the above table. You do *NOT* need to list all FDs and MVDs. It is OK to list the set of FDs and MVDs that imply all FDs and MVDs that hold on the table.

ANSWER:

`ISBN → title, author`

`ISBN,edition → year`

`ISBN,title,author → review`

- (b) Is the above table definition in BCNF? YES / NO

Is the above table definition in 4NF? YES / NO

If you say no to any of the above two questions, please write down a table design that is in both BCNF and 4NF.

ANSWER:

No. No. `Book(ISBN, title, author), Edition(ISBN, edition, year), Review(ISBN, review)`

Problem 4: 15 points

Consider a corpus consisting of the following five documents:

- 1: "i am best"
- 2: "no i am best"
- 3: "am i best"
- 4: "please index me"
- 5: "me me me me me please"

We decide to use TF-IDF cosine similarity as our query-document similarity metric. For computing the IDF of a term, we use the simple inverse of the document frequency of the term, not the logarithm or any other advanced formula. We do not filter out any stopword, meaning that *all* terms are included when we compute the TF-IDF value.

1. Given a query "*please me*", what is the similarity of document 5 to this query?

ANSWER:

Doc5: TF(please)=1 IDF(please)=1/2, TF(me)=5 IDF(me)=1/2

Query: TF(please)=1 TF(me)=1

Similarity: $\frac{(1/2) \cdot 1 + (5/2) \cdot 1}{\sqrt{(1/2)^2 + (5/2)^2} \sqrt{1^2 + 1^2}}$

Note: It is OK if IDF's are multiplied by 5. It is also OK to multiply IDFs to query terms.

2. Given a query "*i am best*", what is/are the document(s) with the highest TF-IDF similarity?

ANSWER:

1 and 3

3. In order to answer the users' queries quickly, we decided to build an inverted index on the five documents.

(a) Write down the number of entries in the lexicon of the index.

ANSWER:

7

(b) Write down the total number of document IDs in the postings list.

ANSWER:

15