

Name (please print):_solution

You have 2 hours to complete this exam. Time will begin after we have read through it together.

- You may use your own personal copy of text book along with any notes or other files that you have brought with you, and anything that you find on the Internet that existed before the exam started.
- You may use any program on your own computer or the computer you have borrowed from the college, and you may use a calculator if you brought one.
- You may NOT communicate with any other person during this exam, (except the professor), either in person or using an electronic means.
- Do not discuss this exam with anyone until specific permission is given by the professor or all the students have taken the exam. Discussing them in any forum where the discussion might be overheard by a person who has not yet taken the exam would constitute academic dishonesty.

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

- If you find a question to be ambiguous, you may ask about it privately by coming to the front of the room. If the confusion is not resolved to your satisfaction, please explain your confusion along with your answer so that we can consider it during grading.
- You are more likely to get partial credit for an answer that is not completely correct 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	50	
2	25	
3	25	
TOTAL	100	

1. [50 points] Short answers. Answer any FIVE of the following questions. If you write ANYTHING for more than five questions, only the first five you answered will be graded. Your answer should fit in the space provided (although there is no penalty for longer answers)

a. [10 points] You have used one of the three software design models for your project. Briefly explain why the “waterfall” model does not work well for your Website design, and suggest an alternative that would result in the creation of better Websites in about the same amount of time.

(a) To use the waterfall model effectively, the user’s requirements must be well-specified. Unfortunately, when working with a client, they are rarely aware of all of their requirements, or how to articulate their requirements, at the beginning of a project. Due to the time constraints of the project, we cannot wait until the requirements are clearly specified.

(b) An alternative is the rapid prototyping model. The rapid creation of an prototype website based on the client’s preliminary requirements allows the client to further specify his/her requirements and leaves us more time to adjust and debug our design.

b. [10 points] When you search in your LSU email account (through PAWS), which type of system are you searching: a relational database or an information retrieval system? What are the major differences between these two types of systems? How do you tell whether it is a relational database or an information retrieval system?

(a) Accessing my PAWS account is through a relational database (matching an exact user ID and a password). Searching in the email system is performed by an information retrieval system.

(b) Major differences: a database manages structured information, executes unambiguous structured queries (in SQL syntax), and retrieves either exact answers or none. An IR system manages unstructured information, takes ambiguous queries (often in natural language), and retrieves a list of relevant or non-relevant documents.

(c) I typed in a natural language query in the Subject search box, and got a list of relevant and non-relevant emails.

c. [10 points] Louisiana State University must retain student records for 100 years with an assurance of authenticity. Briefly how asymmetric encryption can be used to verify the authenticity of those records when they are accessed in the future. One important risk of using encryption in archival applications

is that the key might be lost, so briefly explain also how using an asymmetric encryption technique can mitigate that risk.

(a) Asymmetric encryption would allow verification of the authenticity of the records because the only way the records could have been encrypted in the first place was by LSU using its private key. LSU would use its private key to encode the student records, and announce the public key to authorized decoders, or get it registered with a certificate authority.

(b) The public key is made public and registered with a certificate authority, so it cannot be lost. If LSU lost its private key, authorized decoders (e.g., students, employers) could still decrypt the already encrypted records using the public key. (LSU would need to create a new private key since the old one was lost.)

d. [10 points] Briefly explain why database-driven Web pages are hard for traditional Web search engines to correctly crawl and index; then suggest an approach to allow Web search engines to crawl a database-driven Web page that you really like.

(a) Traditional Web search engines crawl and index static Web pages on the surface Web. Database-driven Web pages are generated by user's queries that are not linked by any Web page, which makes crawlers difficult to crawl.

(b) To allow Web search engines to crawl a database-driven Web page, I would make a link from my home page to that page (so that it is on the surface Web).

e. [10 points] Identify one threat (other than cookies) to the privacy of Internet users, identify the capabilities and limitations of one technical means of addressing that threat, then identify one policy option that might be used to address the threat through federal legislation or regulation.

(a) One threat to privacy of Internet users is the recording and analysis of query logs, which tracks and documents the search habits and histories of users (Recall the AOL user who loved dogs).

(b) One technical means of addressing that threat is to remove the IP addresses from the query log. A limitation of the means is that intelligence analysts might still be able to associate queries with a specific person by analyzing the location of computer, searching pattern and time, etc.

(c) One policy option to this privacy threat is to in some way limit, or completely prohibit, the keeping of query logs, at least remove IP addresses from the query log.

- f. [10 points] What does SMIL do? What is the following SMIL program meant to do? Show enough to convey your understanding of the program.

```
<>
< src="http://www.hq.nasa.gov/alsj/a17/a17a1412742.ra"
  clipBegin="195s" dur="15s" />
<>
< src="http://www.hq.nasa.gov/alsj/a17/10100929.jpg" dur="5s" />
< src="http://www.hq.nasa.gov/alsj/a17/10100930.jpg" dur="5s" />
< src="http://www.hq.nasa.gov/alsj/a17/10100931.jpg" dur="5s" />
</>
</>
```

Synchronized Multimedia Integration Language (SMIL) is a markup language that allows for the integration of multiple forms of media, such as text, audio, and images. The SMIL program above will run the audio found at <http://www.hq.nasa.gov/alsj/a17/a17a1412742.ra>, beginning at the second of the sound file. The audio will last for 15 seconds. At the same time, three images (found at the specific web addresses above) will be shown for 5 seconds each, corresponding with the audio. The total result is a multimedia production lasting 15 seconds which includes both audio and images.

- g. [10 points] Use human computer interaction (HCI) guidelines and principles to evaluate the Website of the East Baton Rouge Parish Library ([://www.ebr.lib.la.us/](http://www.ebr.lib.la.us/)) and identify 3-5 problems.

A good answer will use the CRAP guidelines to evaluate the Website and find 1-2 problems from C. R. A. P. You need to discuss what CRAP is.

2. [25 points] Mirror (i.e., rehost – don't just link to) the Home page you find at <http://slis.lsu.edu> on your SLIS Web account (or your project Web account) and print the URL here:

You need to open that home page, view its source, copy the source and the most important images, check the path of every image, fix any problems. Once everything is fine, upload them to the server, then do a final check.

For full credit, you should mirror the photographs as well, but you don't need to mirror the content of all the linked pages. Make sure your URL is legible and

accessible, otherwise you will receive 0 points. (A printed copy of the home page is attached for your reference.)

3. [25 points] Every state organizes its book festival in a somewhat different way, but it seems that the events in the book festivals are often organized around featured authors. Scan the following Featured Author and Schedule Web pages of Louisiana Book Festival and Texas Book Festival to see their similarity and difference:

[://www.louisianabookfestival.org/authors](http://www.louisianabookfestival.org/authors). (LA Authors)

[://www.texasbookfestival.org/Authors](http://www.texasbookfestival.org/Authors). (TX Authors)

[://www.louisianabookfestival.org/At A Glance](http://www.louisianabookfestival.org/At_A_Glance). (LA Schedule)

[://www.texasbookfestival.org/Calendar](http://www.texasbookfestival.org/Calendar). (TX Schedule)

Imagine there were an XML language for sharing book festival event schedules between the states. Let's call this fictional event schedule XML language EventXML. (Printed schedules will be distributed if we lose Internet access on the exam day.)

- a. [6 points] Design the DTD of the EventML language. You do not need to design a full-grown DTD, but show enough to convey your understanding of what DTD is. Minor syntax errors are allowed.

```
<!ELEMENT BOOKTALK ( (BOOK, AUTHOR, LOCATION, DATE, TIME)* )>
<!ELEMENT BOOK ( (TITLE, AUTHOR, PUBLISHER)* )>
<!ELEMENT TITLE (#PCDATA) >
<!ELEMENT AUTHOR (FIRSTNAME, LASTNAME) >
<!ELEMENT FIRSTNAME (#PCDATA) >
<!ELEMENT LASTNAME (#PCDATA) >
<!ELEMENT PUBLISHER (#PCDATA) >
<!ELEMENT LOCATION ( (ROOM, BUILDING)* )>
<!ELEMENT ROOM (#PCDATA) >
<!ELEMENT BUILDING (#PCDATA) >
<!ELEMENT DATE (#PCDATA) >
<!ELEMENT TIME (#PCDATA) >
```

- b. [8 points] What might a book festival event schedule encoded in the EventML language look like? Show enough to convey your understanding of what XML is. (Points will be deducted for syntax errors.)

```
<BOOKTALK>
  <BOOK>
    <TITLE>The World's End</TITLE>
    <AUTHOR> <FIRSTNAME>J.K.</FIRSTNAME>
      <LASTNAME>Luddite</LASTNAME>
    </AUTHOR>
    <PUBLISHER>Raccort Press</PUBLISHER>
  </BOOK>
```

```
<LOCATION>
  <ROOM>Conference Room A </ROOM>
  <BUILDING> Auditorium </BUILDING>
</LOCATION>
<DATE>12 Aug 2008</DATE>
<TIME> 12:00 PM</TIME>
</BOOKTALK>
```

- c. [5 points] If you have book festival events encoded in EventML, what might you use XSL (or XSLT) for?

XSLT (Extensible Style Language Transformations) is the style sheet for XML documents. How an XML document is displayed must be defined and “transformed” through XSLT. As with CSS in XHTML documents, XSLT defines how the XML elements are rendered.

- d. [6 points] If every state encoded their book festival events in EventML, name two capabilities that the book festival organizers (and/or authors and visitors) would gain.

(a) If every state used the EventML language described above, this would establish a standardized means of sharing and exchanging information among book festival administrators across all the states.

(b) Since all data would be described in the same way using the established XML language, all data from all the states can be extracted automatically and stored in a database, which allow administrators, authors, and visitors to search across all the states.

(c) In addition, this would also facilitate the reuse of data (author, title, publisher) that remained the same at different festivals, or at the same festival in a future year. The coded data could be maintained and reapplied where appropriate, saving the time and energy of re-entering the data from scratch. Extracting data from XML documents (and even storing in a database) and Posting on the Web can be automated.