CS 345B Homework 1: XML and Databases

Donald Kossmann Daniela Florescu Anish Das Sarma

INSTRUCTIONS

Hand out: 01/28/2007
Hand in: 02/04/2007

Please email your homework assignment to **anishds@stanford.edu** by midnight on the hand in date. You may use text, doc, or pdf format.

EXERCISE 1: WELL-FORMEDNESS AND VALIDATION

Consider the following snippet of an XML document.

```
<BOOK ISBN="DB123">
    <TITLE>
     Database Systems - The Complete Book.
    </TITLE>

    <AUTHOR ID="3"><LASTNAME>Widom</AUTHOR>

    <AUTHOR ID="1">
    <LASTNAME>Garcia-Molina<FIRSTNAME>HECTOR</FIRSTNAME></LASTNAME>
    </AUTHOR>

    <AUTHOR><LASTNAME>Ullman</LASTNAME></AUTHOR>

    <INTRODUCTION><TITLE>Introduction to Databases</TITLE></INTRODUCTION>

    <CHAPTER>
    <NUM>1<NUM>
    <TITLE>The Worlds of Database Systems </TITLE>
    <TEXT>We overview the evolution of database systems.</TEXT>
    </CHAPTER>

    <CHAPTER>
    <NUM>1</NUM>
    <TITLE>The Entity-Relationship Data Model</TITLE>
    <TEXT>We explain the elements of the ER model.</TEXT>
    <TEXT>Then we move to design principles</TEXT>
    </CHAPTER>
</BOOK>
```

(a) Is this portion of the XML document well-formed? If not, identify the mistakes and minimally change them to make it well-formed.

(b) Consider the following DTD for the above XML snippet.

```
<!DOCTYPE BOOK [
    <!ELEMENT BOOK (TITLE, AUTHOR?, INTRODUCTION?,CHAPTER+)>
    <!ELEMENT AUTHOR (LASTNAME?, FIRSTNAME?)>
    <!ATTLIST BOOK ISBN CDATA #REQUIRED>
    <!ATTLIST AUTHOR ID #REQUIRED>
    <!ELEMENT TITLE   (#PCDATA)>
    <!ELEMENT INTRODUCTION (TITLE, TEXT)>
    <!ELEMENT CHAPTER (NUM, TITLE, TEXT)>
    <!ELEMENT NUM (#PCDATA)>
    <!ELEMENT TEXT (#PCDATA)>
    <!ELEMENT LASTNAME (#PCDATA)>
    <!ELEMENT FIRSTNAME (#PCDATA)>
]>
```

Is your modified XML from part (a) valid for the DTD above? If not:

i. Minimally change the DTD to fit your XML

ii. Minimally change the XML to fit the DTD

EXERCISE 2: XML FUNDAMENTALS

(a) Give three differences between XML and HTML.

(b) Give two advantages of using elements over attributes.

(c) Illustrates with an example why XML namespaces are useful.

(d) Give one advantage of DTDs over XMLSchema and one advantage of XMLSchema over DTDs.

EXERCISE 3: USE SCENARIO

In this exercise we shall explore a simplified use case scenario for XML: storing medical information. Given the following specifications:

> *We are interested in storing basic information about hospitals, doctors, patients and their illness. Each hospital houses several patients. Patients may be registered with multiple hospitals (possibly for different illnesses). Each doctor is employed by a hospital, and may attend to multiple patients in the hospital.*
>
> *We want to store some basic information about each of the entities involved. For instance, a hospital may be characterized by an id, name, and location, a patient may have an id, date of birth, information about illnesses, hospitals attending, with joining dates, etc.*

(a) Write a simple DTD for this application.

(b) Create a sample XML document respecting your DTD, and containing information about a few hospitals, patients, and doctors.

(c) Rewrite the DTD from part (a) above in XMLSchema. Now make at least two improvements to the schema, which could not be enforced using DTDs.

Note: This is an open-ended question with no correct answer. So please don't worry about answering it "perfectly", to capture every single real-world scenario.

EXERCISE 4: XML TOOLS

Download and install (the free trial version of) one of oXygen, XML Spy, and Stylus Studio. Use it to model the data and write your schema from Exercise 3 above. Play around with the tool.