Harsiddh Patel
CS 598 Data Curation

Assignment 2

Step 4

Elements:

assignementInformation - The element tag encapsulates the whole assignment along with title, subheading, instructions

title – describes the title of the assignment document

subheader - describes the subheading of the documentation and has bold attribute

instructions – contains all steps elements within the element tag

step1 – contains instructions of step 1

step2 – contains instructions of step 2

step3 – contains instructions of step 3

step4 – contains instructions of step 4

step5 – contains instructions of step 5

step6 – contains instructions of step 6

step7 – contains instructions of step 7

step8 – contains instructions of step 8

step9 – contains instructions of step 9

description – encapsulates the actual content of the assignment instructions/data

bulletPoints – every line within each of these elements will have bullet tags which will contain content of each bullet point along with showBulletPoints attributes

bullet & bullet_alphabetic – the element indicate there is either bullet points or alphabetic bullet points

image – this element will not have any content or element within it but it only have src attribute which will contain path to the physical image

table – encapsulates table caption, heading, rows, and rows' cells element tags

caption – describes table title

table_heading – contains table heading columns along with bold attribute

table_row – contains cells elements in table_row tag

th & td – these elements actually contain content of the individual heading and rows data and also has bold attribute

Attributes:

bold – the attribute makes any text to bold to the assigned element

hasUrl – the true/false attribute indicates the element has url in the text

src – contains path to the physical image

urlVal – the attribute has the url value which is same as what it is in the text

showBulletPoints – the attribute indicates the elements inside the assigned element has bullet/alphabetic bullet points

italic – the attribute *italics* text in the assigned element

bulletPointsOccurrence - the attribute indicates the occurrence of the bullet points set in the XML file

Step 6

1) How did you decide to represent the data in the way that you did? Why did you choose the elements and attributes that you did?

The XML file is made up of the assignment 2 instructions document. The way I decided to represent the data is to separate out the individual steps (bold, hasUrl, urlVal, showBulletPoints, italic) into different elements (steps elements) and encapsulate content inside each element plus attributes are assigned to appropriate elements. This helps to capture all information in a structured manner which will make the data more readable to applications that will use the XML file.

2) What were the hardest decisions you had to make in this design process?

One of the hardest decisions was how to represent bullet points and alphabetic bullet points in the XML so I do not lose any structure information and plus application or the user can easily read the file. To distinguish this, I used two different element names to represent different bullet points. Plus, I added hasUrl & urlval attributes to these elements so the application can retrieve all bullet elements in one go if it demands.

3) How does your DTD design support data independence?

One way the DTD design supports data independence is via separated steps elements. Meaning for each step in the instruction, it has its own elements. For example, step 3 in the document, there is a step3 element in the XML file which has different body than the other step elements. Updating step 3 will not have any impact on the other elements.

4) How may your DTD design support the overarching goals of data curation (revisit objectives and activities of Week 1)?

The design makes sures it supports the goals of data curation since any students/TAs can utilize the schema for their purpose, any application can access all parts of the document even utilizing attributes to better filter lookup since different elements are identified and able accessible. The schema was designed in a way that covered all scenarios students/TAs can bring.

5) What are the pros and cons of your DTD design?

One of the pros of the DTD design is that it provides flexibility to reader since all elements are independent, they can easily lookup content for a specific element or even lookup via attributes for a specific element. Another pro is that users can easily incorporate the DTD design for next XML file. In contrary, one con is that elements like step3bulletpoints element can be difficult to parse if there are multiple elements since it would require multiple steps to retrieve text for a specific text.