Assignment 2 Information

Assignment 2: XML Schema Design Exercise

Learning objectives

- Become familiar with data representation and DTD.
- Gain experience analyzing a document and implementing an XML schema.
- Construct convincing arguments on data representation and the importance of DTD in data curation.

Instructions

1. Start by reviewing all 7 sections of this DTD tutorial: https://www.w3schools.com/xml/xml_dtd_intro.asp

2. Choose a document or piece of text that you're interested in from your workplace. This document can be structured or unstructured. You may choose a text of any sort and any length, as long as it is long enough to meet the following encoding criteria. Be sure to include the text as a separate file in your upload (or if it is online, you may provide a link to it) [10 points].

3. [20 points] Make an XML DTD for that text. Your DTD should specify the following:

- At least 10 different elements
- At least 5 different attributes (you may have multiple attributes per element; not every element requires attributes). For at least 2 attributes, a controlled list of values the attribute may take.
- Indicate, of course, whether each element and attribute is required, optional, an either/or, repeated, etc.
- Indicate how elements may nest.

Your goal is to write a schema that best represents the document, and in particular its important or essential components and aspects, for your purposes.

Do not use a DTD generator to generate drafts of your DTDs. The spirit of this exercise is to challenge you to consider your options of attributes, constraints, and default values and justify your decisions. DTD generators tend to generate attributes, constraints, and default values that do not accurately reflect the possibilities and requirements for your XML documents. You will lose points if we see useless attributes, inaccurate constraints, or unjustified default values in your DTD.

4. [10 points] Write prose documentation for each element, attribute, and attribute value.

5. [10 points] Mark up the text you have chosen according to the DTD you designed. Make sure you include either a <!DOCTYPE> reference to an external DTD file, or an internal DTD within the <!DOCTYPE> element in the XML document. Be sure to validate your document (see below) as validation is one of the areas you will be assessed on. Be sure to validate your document (see below) as validation is one of the areas you will be assessed on.

6. [25 points] Write a narrative about this process, answering the following reflection questions:

- 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?
- What were the hardest decisions you had to make in this design process?
- How does your DTD design support data independence?
- How may your DTD design support the overarching goals of data curation (revisit objectives and activities of Week 1)?
- What are the pros and cons of your DTD design?

You can test whether your document will validate against your DTD here: http://xmlvalidator.new-studio.org/

Please note that you must select "DTD" as your validation type. This validation tool requires an internal DTD, which means you must paste your DTD into your XML document, below the XML declaration and above the root element of your XML document. For example:

7. Submit your documents to Assignment 2 Peer Review for peer grading. Your assignment should include

- 3 documents if you used an internal DTD: a copy of the original text, without markup; your marked-up text (an XML file) with internal DTD; and your prose documentation along with your narrative in a separate document; OR
- 4 documents if you used an external DTD: a copy of the original text, without markup; your marked-up text; an external DTD document; and your prose documentation along with your narrative in a separate document.

8. Each student will be required to grade the submissions of 5 of their peers. Submissions will be graded based on the following criteria. Write a constructive and professional review and post to the course forum replying to the individual's submission.

a. Is everything represented?

b. Is it clearly written? Is the scheme and data clearly presented?

c. What are the pros and cons of this representation?

d. How could it be done differently? How could it be improved?

9. Revise and submit all documents to Assignment 2 Submission for instructor grading.

If you have questions about the assignment, use the course Forum. This is a great place to ask questions and also help your fellow classmates.

Instructor Grading Rubric

Criteria	**Max points**
Evidence of in-depth examination of text document (steps 2 & 4)	**20 points**
Evidence of understanding data representation, schemas and independence (steps 3 & 5)	**30 points**
Discussion of curation activities, needs, and decisions (step 6)	**25 points**
Overall quality analysis and completeness (missing parts?)	**25 points**
Total	