

[Back to Week 5](#) [Lessons](#)

This Course: CS 598: Foundations of Data Curation

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.

3. 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,