Matthew Owens

IS531 Assignment 2

9/28/2017

**Write prose documentation for each element, attribute, and attribute value.**

Elements:

* Bill: This is the root element of the document. A single bill element represents a piece of legislation moving through Congress. It must contain an Identifier, summary, header, and at least one section.
* Identifier: This element, although it does not contain any data within it, is important because it provides the minimum information necessary to identify a specific bill. This information is held in the congress, chamber, session, type, and number attributes, all of which are required.
* Header: The header contains procedural information amount the bill that serves as an introduction to the bill itself. It must contain a chamber, date, and sponsor. It may contain any number of cosponsors.
* Chamber: This element includes the declaration of the chamber the bill was introduced in.
* Date: The date the bill was introduced.
* Sponsor: The sponsor is the representative/senator who introduced the bill.
* Cosponsor: Bills can have many cosponsors or none, these elements contain names of other representatives/senators.
* Summary: The summary is intended to be a short, overarching view of the bill.
* Section: This is the first major element including the actual content of the bill. At the top level, the bill is divided into sections with other elements inside them.
* Paragraph: A paragraph element is exactly like a paragraph in normal text. It can contain plain text or lists.
* Subsection: A subsection is another organizational element within a section.
* List: Bills frequently contain lists, and this element supports that structure.
* ListItem: This element allows data users to easily access the items in a list.

Attributes:

* Identifier
  + Congress: This attribute identifies which two-year period of the legislature
  + Session: Each congress is broken down into several sessions (meetings) of the legislative body
  + Type: This attribute identifies what kind of bill the document represents. There are several different types of bills that can be introduced in each body of Congress. This serves as a prefix to the bill number. It must be one of the following list:

S,S.J.,S.Con.Res,S.Res.,H.R.,H.J.Res.,H.Con.Res.,H.Res.

* + Number: The number of a bill is simply a numeric identifier of the bill within a particular session of congress.

**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?**

I decided to include the identity element to make searching for a bill easier. This way, every document contains a known element that will uniquely identify any bill in the history of the United States Congress. For the other elements which actually contain the text of the bill, I spent time studying the sample bill I picked and tried to find a list of elements that would easily fit the structure of the document. I wanted the information to be easy to reference, as bills frequently reference parts of previous legislation. This is what led me to include the section and subsection elements with corresponding id and title attributes and the paragraph element. This makes it easy to reference specific parts of the bill. The rest of the elements contain what amounts to metadata: the bill title, sponsor, cosponsor, and date of introduction.

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

Easily the most difficult decision I had to make designing this schema was determining the constraints on the elements contained within a bill element. Each bill in Congress follows a similar general structure, but there are many differences between them, so I wanted to have a formal structure without being too restrictive. Because of this, I decided force the element to contain basic information, but I left the actual text of the bill open to variation. The section element can contain almost any text structure, which is what I intended.

**How does your DTD design support data independence?**

This design completely abstracts away from the storage mechanism of the bills. Any program that needs to access the contents or descriptive information of the bill can do so through the DTD. If there is always a mechanism to translate the stored format of the bill into a corresponding XML document, client applications will always be able to access the bill.

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

This design makes it easy to reference specific portions of a bill, which useful for preservation and provenance. Provenance is especially important for legislation because new legislation frequently amends existing laws, so it is important to be able to track the changes to the laws over time, much like file history in a version control system. The identifier element makes discoverability easier because every document will contain this basic information. Compliance could also be an important factor because of federal record-keeping laws.