XML Schema Design Exercise

Foundations of Data Curation - Jared Colbert

The document I selected for this project is a surgical technique or instruction manual for a Johnson and Johnson (J&J) knee implant system designed for use in revision total knee procedures. Occasionally, after a knee replacement has been performed on a patient any number of failure modalities may occur requiring the implant to be removed and replaced with a new implant. The document I selected represents the definitive instruction manual for such a procedure.

In the surgical setting a "technique" is synonymous with "instructions" or "workflow". The use of the word "technique" herein is used in the surgical sense.

Schema Entities

There are two schema entities defined in the Document Type Definition (DTD). Both entities will be static throughout the entire corpus of surgical techniques offered by J&J.

1. **Manufacturer** is the static definition of the equipment manufacturer.
2. **Address** is the location of the manufacturer's headquarters.

Schema Elements

The elements I defined are as follows:

1. **Technique** - consists of one or more **pages**
2. **Page -** contains the elements: **number, title, heading, text, figure, table, note, compatibility, footer** and **instrument_layout**
3. **Number** represents the page number.
4. **Title** represents the title of the page.
5. **Heading** represents a bold label at the beginning of a **note.**
6. **Text** represents descriptive verbiage which serves as instruction.
7. **Figure** represents an image depicting an instruction in the **text.**
8. **Table** represents an grid of values specific to the instruction in the **text.**
9. **Note** represents a succinct message highlighting an important tip or cautionary warning.
10. **Compatibility** represents a relationship between equipment outlined in the **text** that may or may not work together.
11. **Footer** represents a text element found at the bottom of each page next to the page number.

12. **Instrument_layout** represents an image displaying the equipment provided in each tray and how the tools are arranged inside.

Schema Attributes

1. The **figure** element has the attributes: **img_path** and **subscript. Img_path** represents the name of the file while **subscript** represents the label located just beneath each **figure.**
2. The **table** element has the attributes: **title, img_path** and **subscript. Title** represents the descriptive text element found at the top of the **table** while **img_path** and **subscript** are identical to the definitions in the **figure** attribute.
3. The **note** element has the attributes: **informative** or **warning** which are the two values the attribute may hold.
4. The **compatibility** element has the attributes: **compatible** and **incompatible** which indicate if certain implants are appropriate to utilize in the same construct. These values are also the two controlled options the attribute may represent.

Data Representation

While I chose a single page of the 84 page surgical technique to represent in the associated XML file, the DTD can successfully represent all the pages in the technique and likely all the techniques in the J&J corpus.

The elements selected for the structure provide a very natural nesting in the overall document. A surgical technique contains one or more pages. Each page contains a title, page number and footer. A page may also contain any combination of text, table, figure, note, heading, compatibility and instrument layout elements.

Design Challenges

The structure of the selected document comes from a larger formula utilized in creating similar documents that naturally lends itself to a DTD type definition. Many documents of this nature maintain a similar layout which has become standardized across the industry making it somewhat easy to generate an independent composition that may be extended to similar works. The structure of the definition was identified quite naturally though there were a couple challenging decisions. Selection of the attributes was challenging from the perspective of identifying options that may be future-proof. That is, the attribute options will hopefully be relevant to elements that may be added to techniques in the future.

Data Independence

The structure of the DTD supports data independence since it can be extrapolated to any page in the corpus of surgical techniques. The structure is ubiquitous and though there may be elements added to future versions of the document, simply adding such elements to the definition will allow the definition to continue to successfully serve the logical schema.

Data Curation

The structure and options selected in the DTD help ensure that the objectives of data curation are adhered to. Because the attributes of each element are narrowly defined and contain pre-defined options, there is little room for erroneous entries in later XML implementations of the DTD.