Data Curation

Assignment 2 Narrative

Deven Jarvis

October, 1st, 2017

For this assignment I choose to build an XML DTD to structure example documentation for a machine data platform called Splunk. This platform is one used at my workplace, and provides a good example for well-structured documentation. The hope is by putting some definition to this structure it can be replicated for internal product documentation. I have provided a text version of the document, though some attributes such as links were lost in this translation. To see the original document, please navigate here: <http://docs.splunk.com/Documentation/Forwarder/6.6.3/Forwarder/Configuretheuniversalforwarder>

I choose the elements and attributes that I did, based on the logical grouping of the provided documentation. All Splunk documentation appears to follow a generally similar format, so this specific document was used as a digestible sample. The goal was to capture a reusable, and effective structure when documenting technical platforms. Specifically, I began with the stand-out text elements such as title, headers, and sub headers. These were what shaped the sections and flow of the documentation. Next, I dug into the underlying text, where the actual process and information about the platform was documented and began dissecting the important elements that gave this information clarity and structure. Combining this process, with the end goal of creating a reproducible documentation format, I came up with the elements and attributes found in my DTD.

The hardest decisions made during this process were around the attributes. Given we only have a sample size of one, I had to determine which elements required attributes that allowed for customization from document-to-document. This led to the decision to lock in font size, but offer some options around font type. It was also difficult to determine how fine-grained the elements should be, as I wanted to provide a structural guideline for documentation without requiring every document to follow this exact same format (some documentation may require additional steps, or may be more information-based rather than process-based).

The DTD provided should support data independence as the structure of the document can be reproduced and utilized without having any direct ties to the text data underneath. This same DTD can now be pulled into internal documentation processes, without relying on any of the specifics of the Splunk document that we structured for this assignment. This DTD should also support the goals of data curation, by not only providing greater structure to our documentation efforts, but also by providing some meta-data around the documentation itself enabling reuse, organization, storage, discoverability, access, integration, reformatting, and modification activities. By providing a clear structure for our documentation data, documentation becomes easier to find and use, in turn hopefully adding to the data curatorial activities for the systems that they are here to document.