

Client: Paul Mather
Virginia Tech CS4624, Blacksburg
May 1, 2014

By Nathanael Bice, Scott Brink & Adam Piorkowski

BTD IMPORTER

What is it the BTD Importer?

- The Bound Thesis Dissertation Importer is a PHP script that fetches metadata for a thesis and imports the thesis and XML files to the Electronic Thesis Dissertation Database

The Current Process

- Hard copy bound theses are scanned into an electronic PDF form
- Script looks for new or updated PDF files to add to database
- PDF file name includes Call Number
- Script fetches metadata for thesis based on Call Number
- PDFs with metadata are uploaded to database of theses

Our Job

- Rewrite the way PDFs have their metadata looked up
- Replace the use of Airpac Classic look up system with the new Sierra API for metadata
- Create an output file structure and place PDF and XML files in correct folder
- A new importer will then take our file structure output and upload the files to a database

Project Modules

- Fetching and Extracting PDF name
- Fetching metadata from Addison
- Building XML files
- Moving PDF and XML files to output directory
- Error Handling

Issues Faced

- Sierra API was not ready in time.
- Originally planned to recode in Java but decided to stick with PHP
- Lack of documentation on current system
- Not being able to test against actual data

Future Plans

- Sierra
- Validating XML against Schema
- Using code to import 13,000+ BTD's

DEMO