Related Files

html\_converter.py (from github)

styles.css (from github, gives styling for html tags created with html\_converter.py.)

readme.md(from github, refer to this for general guide to run)

.indd files(found on box, need to be converted to HTML)

.html files(converted from .indd with InDesign, need to be ran through html\_converter.py)

Converted .html files (often uploaded to Box again for Savannah (or whoever the social media person is) to upload to wordpress)

How To Run/Test

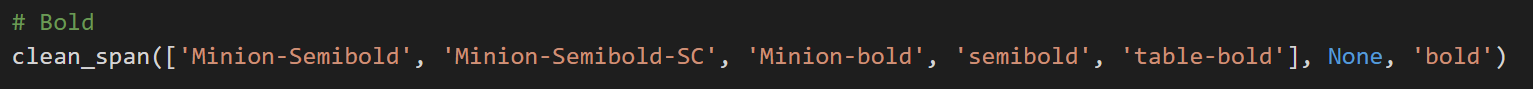
Refer to the readme on github for general steps. I’ll give some more specific steps here.

NOTE: I recommend using PyCharm as the editor for this, make your life a lot easier - Camy

1. Convert .indd files to html with InDesign. These .indd files are found in box, go to BYUS-Web/BYUS-Journal-Archive-2018-11-12/your chosen volume/your chosen issue. Your also going to need to download InDesign on your machine. I remember needing to login to Adobe Creative Cloud with my student email (@byu.edu). Derek also needed to do something on his end to give me permission to download InDesign for free.

After you get InDesign on your machine your going to need to download a script to do mass conversion of .indd to .html. I always refer to this guide here to do this. <https://redokun.com/resources/batch-convert-pdf-file>

I might need to send you the script, just shoot me a message on slack if you need it.

1. Make sure you are on “camy” branch. This is the newest branch with all the filter tag scripts that I wrote. Place html files you converted from .indd in "HTML Converter"/html.
2. Getting html\_converter.py to run is probably going to take some work. I normally run from a command line environment, I use windows so tried to use WSL but I could never get that to work. I eventually got Anaconda and PowerShell to work. Whatever you end up using, I run the file with the command “python3 html\_converter.py”. You’re most likely going to need to install some libraries like beautiful soup. I can’t provide a comprehensive list of everything you’ll need to install since it’ll be different for everyone. For example, you may need the module bs4, and I installed this with the command “pip3 install beautifulsoup4”. This step will probably take a lot of googling on your part unfortunately.
3. Unrecognized tags will show up in the terminal, will need to adjust code to accomodate unrecognized tags. Files with unrecognized tags will move to a "semi-clean" folder. This might seem a little vague, but basically when we run a file through the converter, there are some tags that it may not have seen before. For example, we try to filter all these variations of ‘bold’ tags down to one thing. 

If there are unrecognized tags, but you know that it’s basically just a ‘bold’ tag with a weird name, you would add it to this list. There are also lists for italics, subscripts ect.

Current Work Done

This python script was written by a former co-worker of mine who left awhile ago. The script works well in converting tags, but when we migrated to the wordpress site, we apparently have had some styling issues. The way the script encapsulates the tags (normally with a span tag) isn’t recognized on the wordpress site. I’ve talked to Derek and he’s said that he would like to have native tags (for example italics being <em>) to actually be native tags instead of something like span class = italics, which is what the current converter does. I still haven’t found a solution to this problem.

To do from here

* Adjusting html\_converter.py to filter out new unrecognize tags. Make sure if your working on this to create a separate branch on the github that way we can keep the version control intact.