Elaboration II

Georgia Rutherford

August 2019

1 Problem

I recently started using hypothes.is for my note taking process. I like that I am able to make notes on the actual document I am using, see exactly how these notes fit into the document as a whole, and tag my notes. I want to be able to back up my notes locally in case anything happens to hypothes.is and for version control in case the websites that I am annotating change. However, the exporting process on hypothes.is is a hassle.

The easiest way to export notes manually is through a third party website called Facet. This process involves:

1. Open document that needs annotating

2. Turn on hypothes.is browser extension

3. Highlight section of interesting text

4. Make a note as to why it is interesting in provided space

5. Tag annotations

6. Make more general notes in "Page Notes" section

7. Finish annotating document

8. Copy URL

9. Open https://jonudell.info/h/facet/?max=50

10. Paste URL in the space provided

11. Click Save document as HTML

12. Rename file from "hypothesis" to something more appropriate

1.1 Potential Solution

Originally I thought I could solve the difficulty of exporting my notes by creating a tool that would interact with Facet's API to download my notes in HTML format, rename the file automatically, and then perhaps interact with Github's API to back up my notes there as well. The first issue I ran into when testing this solution was that while the HTML exporting option was the clearest for me to read, Github doesn't easily allow for viewing HTML files. This issue could be alleviated through extremely good commit messages and using http://htmlpreview.github.io/ to preview the file when necessary. However, this situation was certainly not ideal. The issue that really put the nail in the coffin of this potential solution was that I could not find an API for the Facet website. Without an API it would become extremely difficult for my program to interact with the Facet website. At this stage I turned my attention towards searching for another solution.

1.2 New Problems and Solutions

While using Facet to export my notes definitely had its limitations, one benefit was that the tool would format my notes for me. Without Facet I would need to format my notes myself. Having realised this new requirement I changed the scope of my project. My new project is to make a tool that:

1. Allows for user input of URL of annotations

2. Allows for user input of file name

3. Interacts with hypothes.is API to get annotations

4. Would reformat annotations to be human readable and allow for analysis

I suspected that for this project to be successful I would need to use a coding language such as R for steps 1, 2 and 4, and and API for step 3. However, I decided to begin testing using bash commands as I am more familiar with them. My specific plan for testing the validity of this project was to:

1. Discover if hypothes.is has an API

2. Try to get some annotation information using the hypothes.is API

3. Try to get the annotation information using git bash or SWAN terminal

4. Rename the file using the mv bash command

5. Reformat the document into a .CSV using the mv bash command

6. Reformat the document using bash commands such as grep or sed

7. Make a rough script that puts together any successful commands I have been testing

2 Testing Solution

To test the validity of this solution I decided to focus specifically on one annotation I made on the site: https://plato.stanford.edu/entries/self-consciousness/. I will now give a brief overview of my testing and results. For a more detailed account please see my learning journal.

My first step was to check if hypothes.is had an API which I could access, and to get some information off it. I was able to get the annotation I wanted by physically pasting https://api.hypothes.is/api/annotations/id into a browser with the annotation id instead of id. I was able to get this same information using the command wget https://api.hypothes.is/api/annotations/id in SWAN terminal on cloudstor. The command wget https://api.hypothes.is/api/annotations/id was unsuccessful in git bash. Therefore the rest of my testing will be done using SWAN terminal.

Next I wanted to test if I could use SWAN terminal to change the formating of the file. I renamed the original document hypothesistestnotes.txt using the mv command and then made a copy of it called hypothesiscopy.txt using the cp command. I changed the extension of the copy to .csv using the mv command. This formatting was slightly clearer. However as a .csv file separates the information based on commas, my quotation and notes were also separated where they used a comma.

After changing the file extension was unsuccessful, I turned my focus to editing the text within the file. I used the grep and sed commands to copy a section of text and remove unwanted words and output the results into a new file. I also wrote these in a rough script to test that it would all work together as expected. This is the code I ended up with:

```
# Will download notes from hypothesis, rename original, move info into a new
file separated by category and without unwanted sections
# Usage: bash ReformatTest.sh annotationID Filename
wget https://api.hypothes.is/api/annotations/"$1"
grep -o '"id".*"created"' "$1" > output.txt
sed 's/"created"//g' output.txt > "$2".txt
grep -o '"created".*"updated"' "$1" > output.txt
sed 's/"updated"//g' output.txt > > "$2".txt
grep -o '"updated".*"user"' "$1" > output.txt
sed 's/"user"//g' output.txt > > "$2".txt
grep -o '"user".*"uri"' "$1" > output.txt
sed 's/"uri"//g' output.txt > > "$2".txt
grep -o '"uri".*"text"' "$1" > output.txt
sed 's/"text"//g' output.txt > > "$2".txt
grep -o '"text".*"tags"' "$1" > output.txt
sed 's/"tags"//g' output.txt > > "$2".txt
```

```
grep -o '"tags".*"group"' "$1" > output.txt
sed 's/"group"//g' output.txt >> "$2".txt
grep -o '"exact".*"prefix"' "$1" > output.txt
sed 's/"prefix"//g' output.txt >> "$2".txt
rm output.txt
mv "$1" "$2-original.txt"
```

When the command "bash ReformatTest.sh K_JrMsbiEem5hutSGynvDw reformattest" is given the output is two separate files. The original file is kept and renamed reformattest-original.txt. Keeping the original text untouched is important in case the program ever produces unexpected outputs. The new file made is called reformattest.txt, and it contains the new formatting. As you can see below, each section is placed on a new line and unnecessary text is removed.

Script output:

"id": "K_JrMsbiEem5hutSGynvDw",
"created": "2019-08-25T02:43:51.327033+00:00",
"updated": "2019-08-30T08:46:31.519316+00:00",
"user": "acct:georgiarutherford@hypothes.is",
"uri": "https://plato.stanford.edu/entries/self-consciousness/",
"text": "People to look at for my idea of the self and other: Fichte (1794u20131795; Wood 2006), Hegel (1807; Pippin 2010), and, from a somewhat different perspective, Mead (1934; Aboulafia 1986)",
"tags": ["Hegel", "self-consciousness", "self"],
"exact": "Another, related tradition has argued that an awareness of subjectsnother than oneself is a necessary condition of self-consciousness (see n u00a74.4).n Historical variations on such a view can be found in Fichten(1794u20131795; Wood 2006), Hegel (1807; Pippin 2010), and, from ansomewhat different perspective, Mead (1934; Aboulafia 1986).",

3   Conclusions

The testing I have done demonstrates that I can create a script using SWAN terminal that gets annotations from hypothes.is and changes their format to be clearer. I am interested in clearing up the formatting further, and improving this process so that it works for multiple annotations.