Taha Rashid’s 2024 4M Comp. Eng CCA Notes!

Monday May 27, 2024:

* Proof-of-concept, made my own web scraper in Javascript to see if this idea could work
* Researched and saw Anki used python for extensions
  + Will make everything in Python now

Tuesday May 28, 2024:

* Made a python version of the web scraper
* Started work on the Anki plugin code, learning how to use it with the documentation
* With my after school time, I figured out how to properly make the Anki plugin work

Wednesday May 29, 2024:

* Created a GUI (copying some other code, I have no idea what's going on)
  + I will make my own version based off of this person’s code in the future if I have time
* Cleaning up web scraper code
* Created a “cleaning function” (with the help of ChatGPT)
  + I will make my own/learn how to in the future (this was just proof-of-concept) if/when I have time to learn more about regex

Thursday May 30, 2024:

* Tried to understand how to optimize my code
* Trying to make an async web scraper (not completed yet)
* Downloading/trying to manually install dependencies for async web scraping
  + Note: We cannot use pip if we want this extension to work on different machines out of the box, which is why we are manually installing dependencies

Friday May 31, 2024:

* Finished making my async web scraper code
* Roughly has decreased the time from ~30 min to ~1 min for 6k cards
  + It has become REALLY fast compared to last time
  + It becomes slower everytime we run it though…
* Still need to improve the scraper’s performance… trying to find different ways to achieve this
  + For some reason, it does not like to work with my big set of data…

Monday June 3, 2024

* For now, my plan is to find/make more performance optimizations, clean up the code
* I also need to make my own “cleaner” function (hiragana/katakana), and be able to take in different kinds of readings (Smartly find what I need)
* I think I will move away from a web scraper to checking a local dictionary (for reliability/speed)
  + Started work on this… Made a separate python file to test with its provided sample code
  + Learned how to manually installed dependencies via pip
* I plan to make (tomorrow) the equivalent of my web scraper’s function with the dictionary

Tuesday Jun 4, 2024:

* Finished making the equivalent scraper code with the dictionary
* Need to test performance and implement into my main addon
  + See if it is faster than scraping

Wednesday June 5 2024:

* Made the dictionary thing it a function
  + Very fast (0.03 sec instead of 0.3 sec with web scraper)
* Made it asynchronous (yet to test async speed)
* Trying to use regex to get the first definition (in a list of related definitions)

Thursday June 6 2024:

* Dictionary async speed was tested
  + Between 0.020 and 0.90sec for individual word
  + Between 0.060 and 0.075sec for 4 words (async, at the same time)
  + Memory import does not work.. Figure this out later?
* Look at making a progress bar (tqdm library, if possible)
* Look at implementing it with the actual Anki plugin, and test/compare speed (tomorrow)

Friday June 7 2024

* Implemented within anki module
* Debugging, some weird error with the answer list
* ERROR IS PROBABLY THE DELIMITER (Not the right index, It's literally giving us the printed error when we do list[300] (Note: THE EXCEPTION IS THE STRING!!!)
  + Fix the error!

Monday June 10 2024:

* Fixed delimiter error
* Implemented tons of if statements and possible return strings (so I know exactly what is happening)
* The code runs so far (no error), I am checking IF I get my desired results now
  + It looks like it works, but for some reason in my code to grab all results, I will sometimes get a duplicate +- a position from itself (? Is there a physical, duplicate card?)
  + Check tomorrow for how many dupes there are
* Make the program work for individual cards as well

Tuesday June 11 2024:

* Duplicate problem wasn't even relevant
* Confirmed that my code works
* Trying to optimize for speed (automatically skip cards if not a verb or if scrapped already)
* Trying to “follow” coding conventions (1 return statement/function), trying to clean up/make sense of my code
* Throwing error, something to do with new “write” variable check
  + Might have fixed it?

Wednesday June 12 2024:

* Everything works
* Now, I am implementing individual card selection (in the editor menu)

Thursday June 13, 2024:

* Trying to figure out how to implement a way to select specific cards in the card browser and get their IDS
  + I ultimately could not do this 🙁

Friday June 14, 2024:

* Cleaned up my code for actual use (ex. Getting rid of debug statements, print commands)
* Implemented the actual “modification” part of my code
  + Up until now, all I was doing was getting the data for each card, scraping and finding the relevant information BUT NOT appending it/updating each card
* I mostly implemented the final functions (minus “pushing” the actual changes to save, because realistically I would do more testing)

Main Resources:

* <https://jamdict.readthedocs.io/en/latest/tutorials.html> (jamdict library)
* <https://addon-docs.ankiweb.net/the-anki-module.html> (anki module)
* <https://note.nkmk.me/en/python-split-rsplit-splitlines-re/#split-a-string-by-regex-resplit> (regex)
* chat.openai.com (for debugging purposes)
* <https://docs.python.org/3/library/asyncio-task.html#coroutines> (asyncio)