CYBV- 473 Final

Your final script will scrape the https://casl.website/ within CyberApolis.

IMPORTANT: you will not follow any links that veer outside of CyberApolis

IMPORTANT: you are allowed to use standard Python libraries and any 3rd party library we have used during the class. (keep these basic)

Your script will generate a report that contains the following information.

1) Unique URLs of all the pages found on the website
2) Unique URL links to images found on the website
3) Extract phone numbers found on the website
4) Extract all text content from each of the pages and store them in a string variable
5) Extract any zip codes

NOTE: for items 6-8, use NLTK to process all the text found on the website, using the text content you extracted during item 4 above.

6) A list of all unique vocabulary found on the website
7) A list of all possible verbs
8) A list of all possible nouns

NOTE: REGEX PATTERN HELP

ZipCode Regex

```
zipPatt = re.compile(b'\d{5}(?:-\d{4})?')   # Zip code regex
```

Phone Number Regex

```
phonePatt = re.compile(b'\(?\d{3}\)?-? *\d{3}-? *-?\d{4}')  # Phone number regex
```