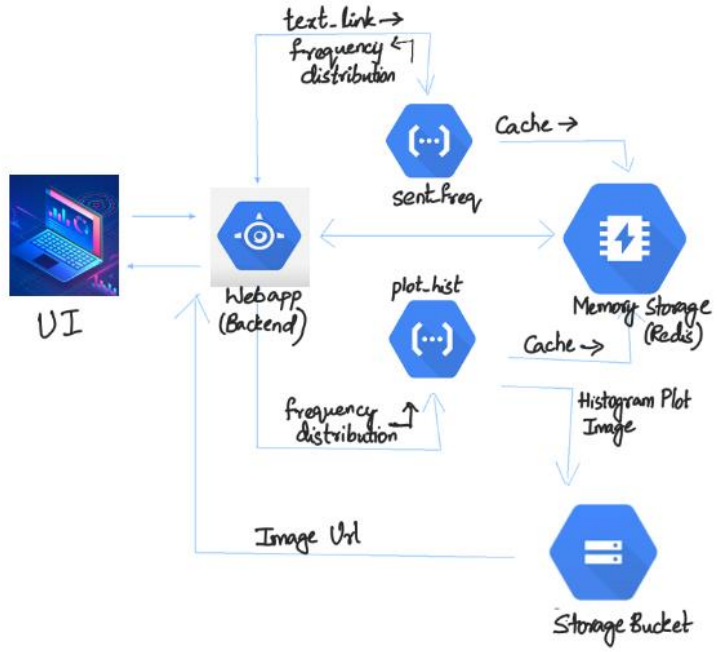
Assignment 5: Function as a Service

Report – 12/2/2020

In this assignment, implemented a serverless application using Google Cloud functions with code in python. Designed and deployed an end-to-end web application on Google App Engine which uses these cloud functions and maintains cache of results in a Memory Store of GCP. The functionality of this application is to take an input as url of an ebook and outputs the distribution of the lengths of sentences in that book. This application gives both the frequency distribution as text and histogram plot on the Web UI. URL of this application: https://vijaysai.uc.r.appspot.com/

Design details:

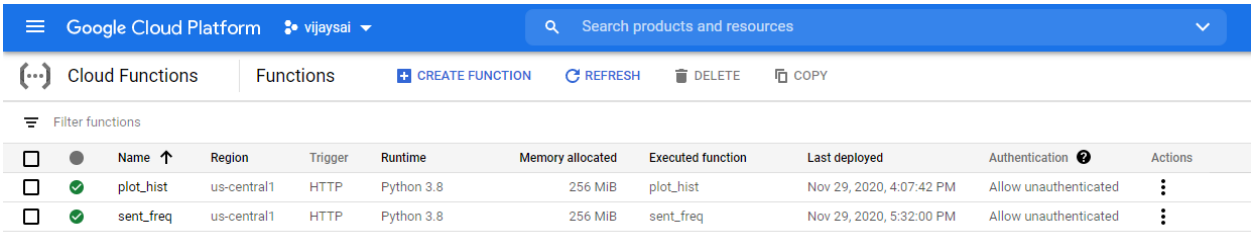
- Web Application takes url from user and runs 2 cloud functions sent_freq and plot_hist sequentially
- sent_freq – cloud function reads text from given url and outputs the frequency distribution of sentence lengths from the text
- plot_hist – cloud function takes frequencies as input (which are output from sent_freq function) and generates a histogram plot which is saved on cloud bucket and returns the public url
- Web app inturn gets these two outputs, that is frequencies from sent_freq function and url of histogram image from plot_hist and displays then on the ui
- Actually, web app first checks if the redis memory store cache already has url provided by the user, if yes, then it directly displays the results saved, else it runs both the cloud functions as mentioned in above steps



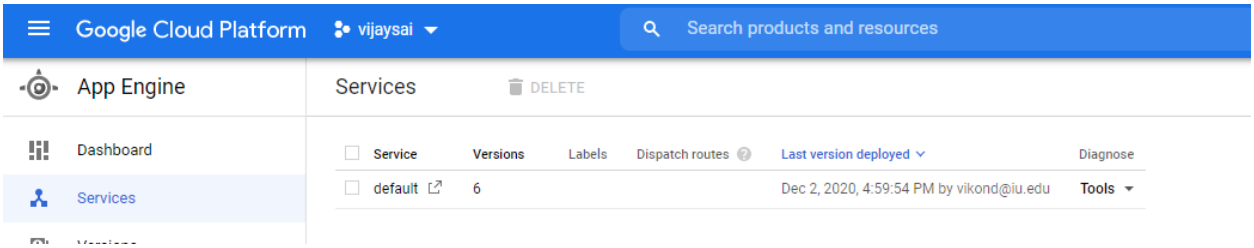
Cache – Used memory store of GCP to have a key-value store in redis. This cache stores key as book url and value as a list of two elements. First is frequency distribution in the format of a dictionary but saved as string and second item in value list is histogram image url. This implementation is very efficient and robust. Both SET and GET operations give results in milliseconds. More details in the performance section.

Cloud APIs used:

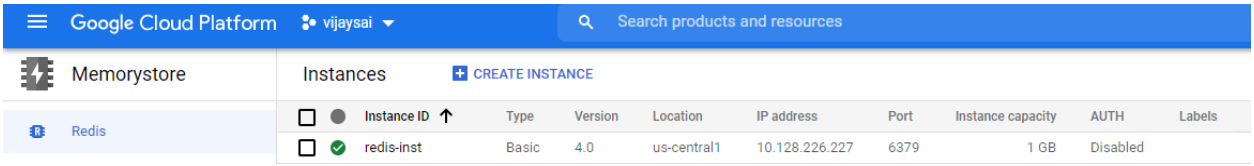
- Cloud Functions



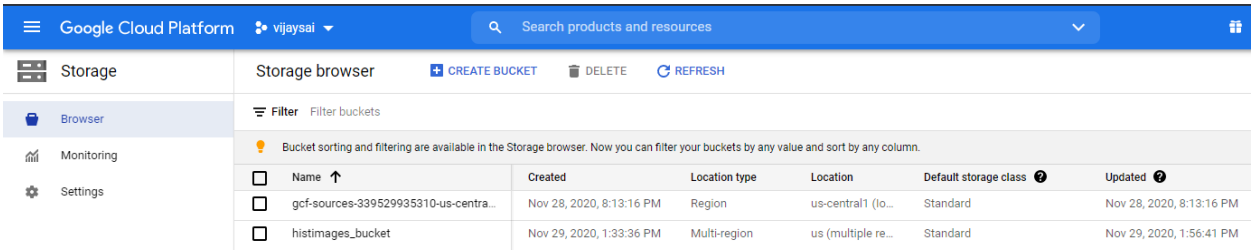
- App Engine(to host web app)



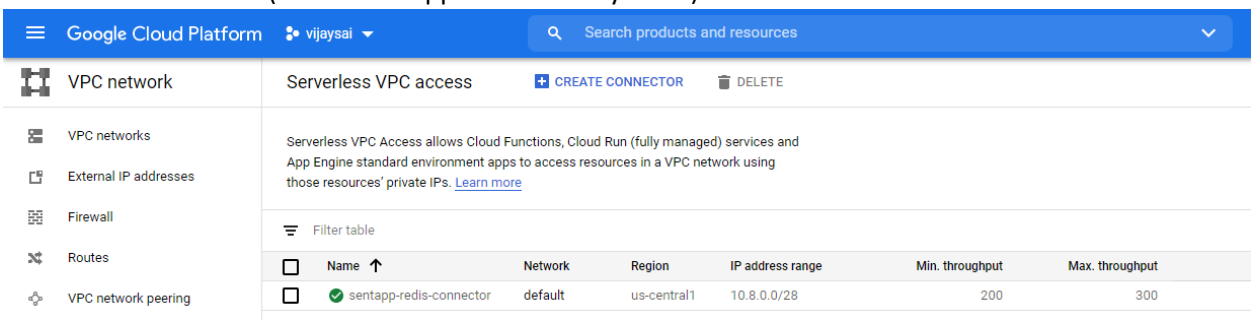
- Memorystore(to cache results)



- Storage (to store histogram images)



- Serverless VPC access (to connect app with memory store)



Sample Gcloud Logs:

sent_freq function logs

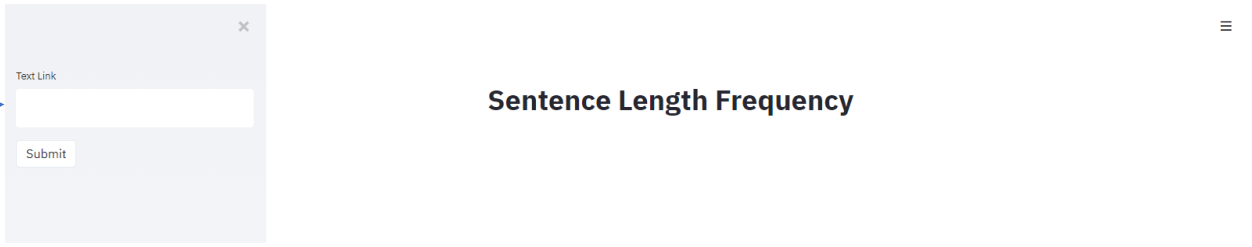
insertId	labels.exe	logName	receiveTimestamp	resource.labe	resource.lab	resource.label	resource.type	severity	textPayload	timestamp	trace
000000-6536d79f-5(	2zz0yf361l	projects/v	2020-12-02T22:31:22	sent_freq	vijaysai	us-central1	cloud_function	DEBUG	Function execution took 3683 ms, finished with status code: 200	2020-12-02T22:31:15.206782415Z	projects/vij
000000-f65f0aaa-33	2zz0yf361l	projects/v	2020-12-02T22:31:22	sent_freq	vijaysai	us-central1	cloud_function		[nltk_data] Unzipping tokenizers/punkt.zip.	2020-12-02T22:31:14.515Z	projects/vij
000000-f50789f2-3a	2zz0yf361l	projects/v	2020-12-02T22:31:22	sent_freq	vijaysai	us-central1	cloud_function		[nltk_data] Downloading package punkt to /root/nltk_data...	2020-12-02T22:31:13.959Z	projects/vij
000000-743e0025-7	2zz0yf361l	projects/v	2020-12-02T22:31:22	sent_freq	vijaysai	us-central1	cloud_function		[2020-12-02 22:31:13 +0000] [6] [INFO] Booting worker with pid: 6	2020-12-02T22:31:13.799Z	projects/vij
000001-c9ddf52a-0!	2zz0yf361l	projects/v	2020-12-02T22:31:22	sent_freq	vijaysai	us-central1	cloud_function		[2020-12-02 22:31:13 +0000] [1] [INFO] Using worker: threads	2020-12-02T22:31:13.749Z	projects/vij
000000-5f9b1433-a:	2zz0yf361l	projects/v	2020-12-02T22:31:22	sent_freq	vijaysai	us-central1	cloud_function		[2020-12-02 22:31:13 +0000] [1] [INFO] Listening at: http://0.0.0.0:8080 (1)	2020-12-02T22:31:13.749Z	projects/vij
000000-c6dc199c-c8	2zz0yf361l	projects/v	2020-12-02T22:31:22	sent_freq	vijaysai	us-central1	cloud_function		[2020-12-02 22:31:13 +0000] [1] [INFO] Starting gunicorn 20.0.4	2020-12-02T22:31:13.748Z	projects/vij
000000-ca828528-34	2zz0yf361l	projects/v	2020-12-02T22:31:22	sent_freq	vijaysai	us-central1	cloud_function	DEBUG	Function execution started	2020-12-02T22:31:11.524478747Z	projects/vij
000000-79b87fe9-4	rcxm6emn	projects/v	2020-12-02T22:29:41	sent_freq	vijaysai	us-central1	cloud_function	DEBUG	Function execution took 9720 ms, finished with status code: 200	2020-12-02T22:29:40.546104802Z	projects/vij
000000-2b04d3e9-9	rcxm6emn	projects/v	2020-12-02T22:29:41	sent_freq	vijaysai	us-central1	cloud_function		[nltk_data] Package punkt is already up-to-date!	2020-12-02T22:29:31.131Z	projects/vij
000000-20196eb3-b	rcxm6emn	projects/v	2020-12-02T22:29:41	sent_freq	vijaysai	us-central1	cloud_function		[nltk_data] Downloading package punkt to /root/nltk_data...	2020-12-02T22:29:30.927Z	projects/vij
000000-2365299f-b(	rcxm6emn	projects/v	2020-12-02T22:29:41	sent_freq	vijaysai	us-central1	cloud_function	DEBUG	Function execution started	2020-12-02T22:29:30.826811912Z	projects/vij
000000-74047137-2(	rcxminpot	projects/v	2020-12-02T22:17:38	sent_freq	vijaysai	us-central1	cloud_function	DEBUG	Function execution took 5311 ms, finished with status code: 200	2020-12-02T22:17:33.249454113Z	projects/vij
000000-069726d5-a	rcxminpot	projects/v	2020-12-02T22:17:38	sent_freq	vijaysai	us-central1	cloud_function		[nltk_data] Unzipping tokenizers/punkt.zip.	2020-12-02T22:17:31.409Z	projects/vij
000000-9341dcd0-3	rcxminpot	projects/v	2020-12-02T22:17:38	sent_freq	vijaysai	us-central1	cloud_function		[nltk_data] Downloading package punkt to /root/nltk_data...	2020-12-02T22:17:30.615Z	projects/vij
000000-1f446979-0(	rcxminpot	projects/v	2020-12-02T22:17:38	sent_freq	vijaysai	us-central1	cloud_function		[2020-12-02 22:17:30 +0000] [6] [INFO] Booting worker with pid: 6	2020-12-02T22:17:30.423Z	projects/vij
000001-3bce5404-d	rcxminpot	projects/v	2020-12-02T22:17:38	sent_freq	vijaysai	us-central1	cloud_function		[2020-12-02 22:17:30 +0000] [1] [INFO] Using worker: threads	2020-12-02T22:17:30.362Z	projects/vij
000000-19d5168f-b4	rcxminpot	projects/v	2020-12-02T22:17:38	sent_freq	vijaysai	us-central1	cloud_function		[2020-12-02 22:17:30 +0000] [1] [INFO] Listening at: http://0.0.0.0:8080 (1)	2020-12-02T22:17:30.362Z	projects/vij
000000-45f1c339-52	rcxminpot	projects/v	2020-12-02T22:17:38	sent_freq	vijaysai	us-central1	cloud_function		[2020-12-02 22:17:30 +0000] [1] [INFO] Starting gunicorn 20.0.4	2020-12-02T22:17:30.357Z	projects/vij
000000-56324be5-f(	rcxminpot	projects/v	2020-12-02T22:17:38	sent_freq	vijaysai	us-central1	cloud_function	DEBUG	Function execution started	2020-12-02T22:17:27.939475225Z	projects/vij
000000-7eba7e86-a	pxjl0gvtl1s	projects/v	2020-12-02T22:02:26	sent_freq	vijaysai	us-central1	cloud_function	DEBUG	Function execution took 6561 ms, finished with status code: 200	2020-12-02T22:02:22.813128858Z	projects/vij
000000-64c07c69-b(	pxjl0gvtl1s	projects/v	2020-12-02T22:02:26	sent_freq	vijaysai	us-central1	cloud_function		[nltk_data] Unzipping tokenizers/punkt.zip.	2020-12-02T22:02:19.801Z	projects/vij
000000-3a93d609-9	pxjl0gvtl1s	projects/v	2020-12-02T22:02:26	sent_freq	vijaysai	us-central1	cloud_function		[nltk_data] Downloading package punkt to /root/nltk_data...	2020-12-02T22:02:19.151Z	projects/vij
000000-f907c512-95	pxjl0gvtl1s	projects/v	2020-12-02T22:02:26	sent_freq	vijaysai	us-central1	cloud_function		[2020-12-02 22:02:18 +0000] [6] [INFO] Booting worker with pid: 6	2020-12-02T22:02:18.841Z	projects/vij
000001-ec5cbbe5-d	pxjl0gvtl1s	projects/v	2020-12-02T22:02:26	sent_freq	vijaysai	us-central1	cloud_function		[2020-12-02 22:02:18 +0000] [1] [INFO] Using worker: threads	2020-12-02T22:02:18.724Z	projects/vij
000000-cb0b6ee7-2	pxjl0gvtl1s	projects/v	2020-12-02T22:02:26	sent_freq	vijaysai	us-central1	cloud_function		[2020-12-02 22:02:18 +0000] [1] [INFO] Listening at: http://0.0.0.0:8080 (1)	2020-12-02T22:02:18.724Z	projects/vij

plot_hist function logs

insertId	labels.execution_id	logName	receiveTimestamp	resource.lab	resource.l	resource.labels	resource.typ	severity	textPayload	timestan
000000-3efc6	q3iyptkj3nhk	projects/vija	2020-12-02T22:31:25.577035108Z	plot_hist	vijaysai	us-central1	cloud_functi	DEBUG	Function execution took 1861 ms, finished with status code: 200	2020-12-(
000000-e481(	q3iyptkj3nhk	projects/vija	2020-12-02T22:31:25.577035108Z	plot_hist	vijaysai	us-central1	cloud_functi	DEBUG	Function execution started	2020-12-(
000000-db7ff	q3iymq93o8wd	projects/vija	2020-12-02T22:29:50.775297513Z	plot_hist	vijaysai	us-central1	cloud_functi	DEBUG	Function execution took 3528 ms, finished with status code: 200	2020-12-(
000000-83fe0	q3iymq93o8wd	projects/vija	2020-12-02T22:29:50.775297513Z	plot_hist	vijaysai	us-central1	cloud_functi	DEBUG	Function execution started	2020-12-(
000000-ecef8	q3iyk90pt5xr	projects/vija	2020-12-02T22:17:45.582687286Z	plot_hist	vijaysai	us-central1	cloud_functi	DEBUG	Function execution took 4929 ms, finished with status code: 200	2020-12-(
000000-2def6	q3iyk90pt5xr	projects/vija	2020-12-02T22:17:45.582687286Z	plot_hist	vijaysai	us-central1	cloud_function		[2020-12-02 22:17:38 +0000] [10] [INFO] Booting worker with pid: 10	2020-12-(
000001-03455	q3iyk90pt5xr	projects/vija	2020-12-02T22:17:45.582687286Z	plot_hist	vijaysai	us-central1	cloud_function		[2020-12-02 22:17:38 +0000] [1] [INFO] Using worker: threads	2020-12-(
000000-7697c	q3iyk90pt5xr	projects/vija	2020-12-02T22:17:45.582687286Z	plot_hist	vijaysai	us-central1	cloud_function		[2020-12-02 22:17:38 +0000] [1] [INFO] Listening at: http://0.0.0.0:8080 (1)	2020-12-(
000000-43343	q3iyk90pt5xr	projects/vija	2020-12-02T22:17:45.582687286Z	plot_hist	vijaysai	us-central1	cloud_function		[2020-12-02 22:17:38 +0000] [1] [INFO] Starting gunicorn 20.0.4	2020-12-(
000000-dca7d	q3iyk90pt5xr	projects/vija	2020-12-02T22:17:45.582687286Z	plot_hist	vijaysai	us-central1	cloud_function		OpenBLAS WARNING - could not determine the L2 cache size on this system, assuming 256k	2020-12-(
000000-e126(	q3iyk90pt5xr	projects/vija	2020-12-02T22:17:45.582687286Z	plot_hist	vijaysai	us-central1	cloud_functi	DEBUG	Function execution started	2020-12-(
000000-65e68	81gy1x2gq6xv	projects/vija	2020-12-02T22:02:34.675342846Z	plot_hist	vijaysai	us-central1	cloud_functi	DEBUG	Function execution took 7972 ms, finished with status code: 200	2020-12-(
000000-e464(	81gy1x2gq6xv	projects/vija	2020-12-02T22:02:34.675342846Z	plot_hist	vijaysai	us-central1	cloud_function		[2020-12-02 22:02:30 +0000] [11] [INFO] Booting worker with pid: 11	2020-12-(
000000-d6cb8	81gy1x2gq6xv	projects/vija	2020-12-02T22:02:34.675342846Z	plot_hist	vijaysai	us-central1	cloud_function		[2020-12-02 22:02:30 +0000] [1] [INFO] Using worker: threads	2020-12-(
000000-67aed	81gy1x2gq6xv	projects/vija	2020-12-02T22:02:34.675342846Z	plot_hist	vijaysai	us-central1	cloud_function		[2020-12-02 22:02:30 +0000] [1] [INFO] Listening at: http://0.0.0.0:8080 (1)	2020-12-(
000000-7db01	81gy1x2gq6xv	projects/vija	2020-12-02T22:02:34.675342846Z	plot_hist	vijaysai	us-central1	cloud_function		[2020-12-02 22:02:30 +0000] [1] [INFO] Starting gunicorn 20.0.4	2020-12-(
000000-def2(	81gy1x2gq6xv	projects/vija	2020-12-02T22:02:34.675342846Z	plot_hist	vijaysai	us-central1	cloud_function		OpenBLAS WARNING - could not determine the L2 cache size on this system, assuming 256k	2020-12-(
000000-df378	81gy1x2gq6xv	projects/vija	2020-12-02T22:02:34.675342846Z	plot_hist	vijaysai	us-central1	cloud_functi	DEBUG	Function execution started	2020-12-(
000000-81221	bm4ck84eizir	projects/vija	2020-12-02T19:07:05.628710310Z	plot_hist	vijaysai	us-central1	cloud_functi	DEBUG	Function execution took 8207 ms, finished with status code: 200	2020-12-(
000000-bb55	bm4ck84eizir	projects/vija	2020-12-02T19:07:05.628710310Z	plot_hist	vijaysai	us-central1	cloud_function		[2020-12-02 19:07:02 +0000] [10] [INFO] Booting worker with pid: 10	2020-12-(
000000-e17e(	bm4ck84eizir	projects/vija	2020-12-02T19:07:05.628710310Z	plot_hist	vijaysai	us-central1	cloud_function		[2020-12-02 19:07:02 +0000] [1] [INFO] Using worker: threads	2020-12-(
000000-cc0d8	bm4ck84eizir	projects/vija	2020-12-02T19:07:05.628710310Z	plot_hist	vijaysai	us-central1	cloud_function		[2020-12-02 19:07:02 +0000] [1] [INFO] Listening at: http://0.0.0.0:8080 (1)	2020-12-(
000000-53074	bm4ck84eizir	projects/vija	2020-12-02T19:07:05.628710310Z	plot_hist	vijaysai	us-central1	cloud_function		[2020-12-02 19:07:02 +0000] [1] [INFO] Starting gunicorn 20.0.4	2020-12-(
000000-9db19	bm4ck84eizir	projects/vija	2020-12-02T19:07:05.628710310Z	plot_hist	vijaysai	us-central1	cloud_function		OpenBLAS WARNING - could not determine the L2 cache size on this system, assuming 256k	2020-12-(
000000-d6dd(	bm4ck84eizir	projects/vija	2020-12-02T19:07:05.628710310Z	plot_hist	vijaysai	us-central1	cloud_functi	DEBUG	Function execution started	2020-12-(

Screenshots of Application:

Before entering any input:



User Input

Sample Outputs:

Text Link

http://www.gutenberg.org/cache/epub

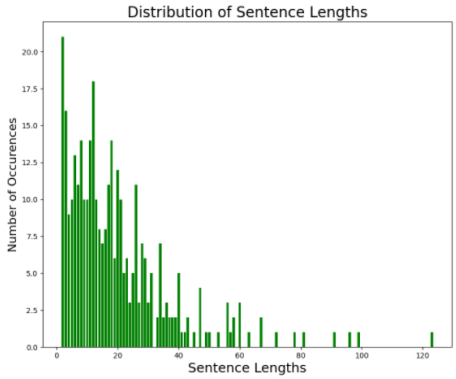
Submit

Sentence Length Frequency

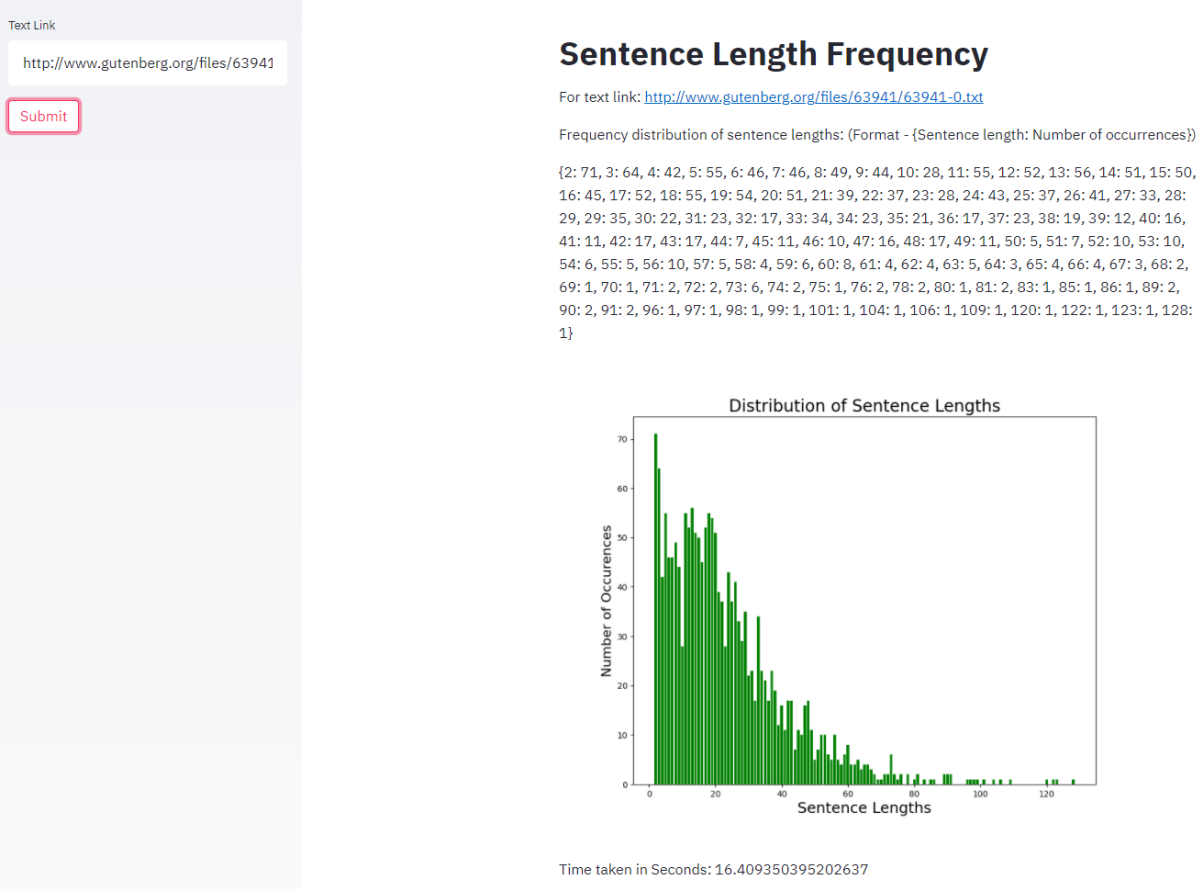
For text link: http://www.gutenberg.org/cache/epub/63934/pg63934.txt

Frequency distribution of sentence lengths: (Format - {Sentence length: Number of occurrences})

{2: 21, 3: 16, 4: 9, 5: 10, 6: 13, 7: 11, 8: 14, 9: 10, 10: 10, 11: 14, 12: 18, 13: 10, 14: 8, 15: 7, 16: 8, 17: 11, 18: 14, 19: 6, 20: 12, 21: 10, 22: 5, 23: 6, 24: 3, 25: 5, 26: 11, 27: 3, 28: 7, 29: 6, 30: 3, 31: 5, 33: 2, 34: 7, 35: 2, 36: 3, 37: 2, 38: 2, 39: 2, 40: 5, 41: 1, 42: 1, 43: 2, 45: 1, 47: 4, 49: 1, 50: 1, 53: 1, 56: 3, 57: 1, 58: 2, 60: 3, 63: 1, 67: 2, 72: 1, 78: 1, 81: 1, 91: 1, 96: 1, 99: 1, 123: 1}



Time taken in Seconds: 0.0038633346557617188



Performance:

1. Time taken by application for books of different sizes run for the first time (results not in cache)

Book Url	Size	Time Taken (in sec)
http://www.gutenberg.org/files/63941/63941-0.txt	269 KB	16.4
http://www.gutenberg.org/files/63940/63940-0.txt	690 KB	11.2
http://www.gutenberg.org/cache/epub/63937/pg63937.txt	431 KB	6.7
http://www.gutenberg.org/files/1250/1250-0.txt	128 KB	4.2
http://www.gutenberg.org/cache/epub/63930/pg63930.txt	36 KB	2.4
http://www.gutenberg.org/files/76/76-0.txt	602 KB	8.67

Observations

- Irrespective of size of book, initial run of function takes more time
- As the size of book increases, time taken is increasing

2. Comparing the time taken by cloud functions to output sentence length frequency vs getting it from cache

Book Url	Size	Time(in sec) by functions	Time(in sec) from cache
http://www.gutenberg.org/files/63941/63941-0.txt	269 KB	16.4	0.0047
http://www.gutenberg.org/files/63940/63940-0.txt	690 KB	11.27	0.0059

We can clearly see the best performance of cache using redis in memory store

3. Also tested on a sample Wikipedia page:
    - https://en.wikipedia.org/wiki/Indiana_University_Bloomington - time_taken - 15.4s

Cost of running experiments:

- Cost of running these experiments is low although I used many different Cloud APIs
- Initial cost credit was $45. Credits remaining after thoroughly testing the application is $41

Implementation Improvements:

- In both the cloud functions and web app exceptions need to be handled
- sent_freq can be made robust to take any url but just scrape text data and give sentence lengths. At present, this function fails if the link provided has images etc.
- NLP part of splitting the text into sentences and sentences into words can be improved for better performance

List of Folder/Files:

1. 'sentence_length_function' folder has python function that was used for sent_freq cloud function
2. 'plot_histogram' folder has python function that was used for plot_hist function
3. 'webapp' folder has all files to host the web app
4. 'logs' folder has all relevant sample log files
5. sample_cache.txt – has cache of few user inputs. This is exported from GCP Memory Store