

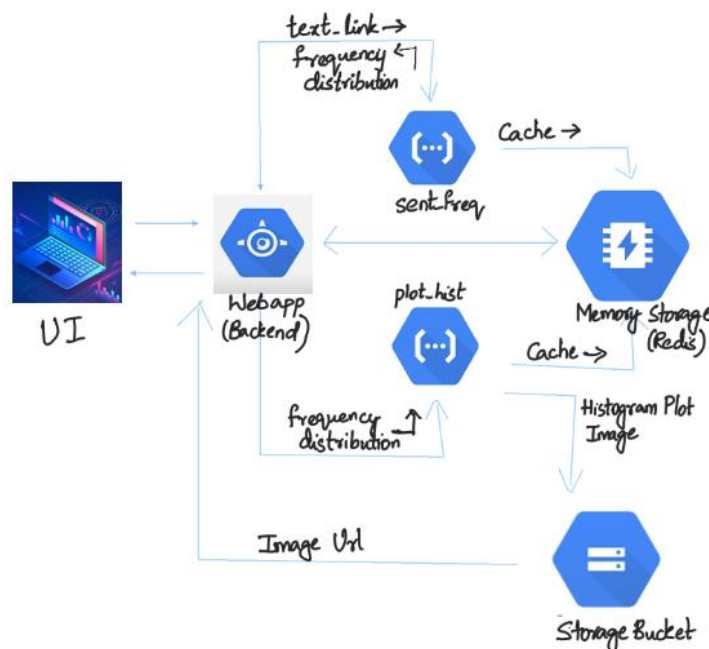
Assignment 5: Function as a Service

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:

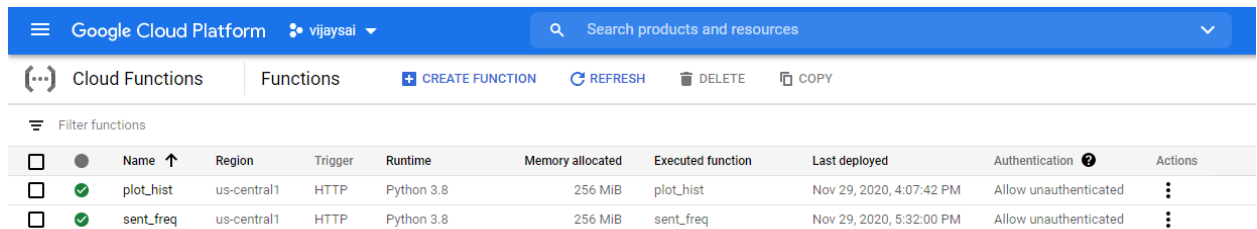
- Web Application takes url from user and runs 2 cloud functions `sent_freq` and `plot_hist` sequentially
- `sent_freq` – cloud function reads text from given url and outputs the frequency distribution of sentence lengths from the text
- `plot_hist` – cloud function takes frequencies as input (which 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:

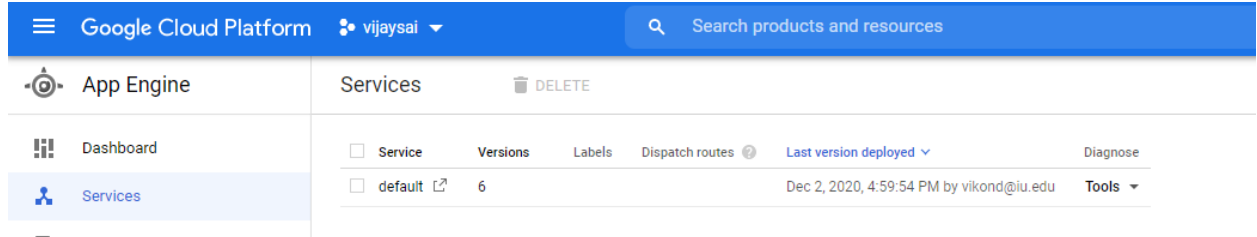
- Cloud Functions



The screenshot shows the Google Cloud Platform interface for Cloud Functions. The top navigation bar includes the Google Cloud Platform logo, the user name 'vijaysai', and a search bar. The left sidebar shows the 'Cloud Functions' menu. The main content area displays a table of functions.

Name	Region	Trigger	Runtime	Memory allocated	Executed function	Last deployed	Authentication	Actions
plot_hist	us-central1	HTTP	Python 3.8	256 MiB	plot_hist	Nov 29, 2020, 4:07:42 PM	Allow unauthenticated	⋮
sent_freq	us-central1	HTTP	Python 3.8	256 MiB	sent_freq	Nov 29, 2020, 5:32:00 PM	Allow unauthenticated	⋮

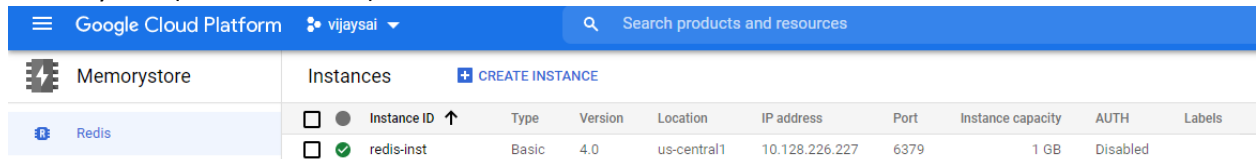
- App Engine(to host web app)



The screenshot shows the Google Cloud Platform interface for App Engine Services. The top navigation bar includes the Google Cloud Platform logo, the user name 'vijaysai', and a search bar. The left sidebar shows the 'App Engine' menu. The main content area displays a table of services.

Service	Versions	Labels	Dispatch routes	Last version deployed	Diagnose
default	6			Dec 2, 2020, 4:59:54 PM by vikond@iu.edu	Tools

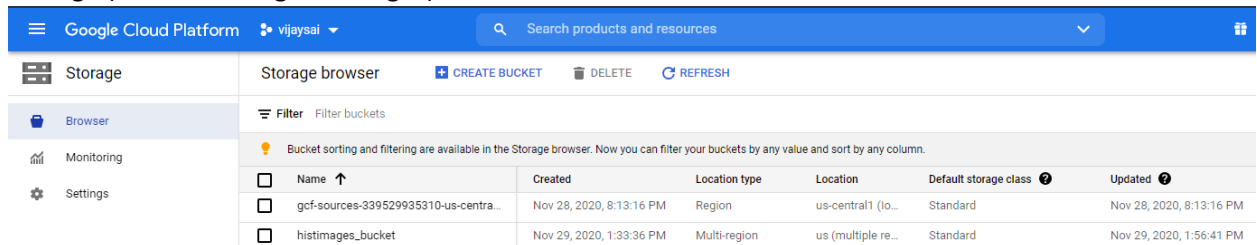
- Memorystore(to cache results)



The screenshot shows the Google Cloud Platform interface for Memorystore Redis Instances. The top navigation bar includes the Google Cloud Platform logo, the user name 'vijaysai', and a search bar. The left sidebar shows the 'Memorystore' menu. The main content area displays a table of instances.

Instance ID	Type	Version	Location	IP address	Port	Instance capacity	AUTH	Labels
redis-inst	Basic	4.0	us-central1	10.128.226.227	6379	1 GB	Disabled	

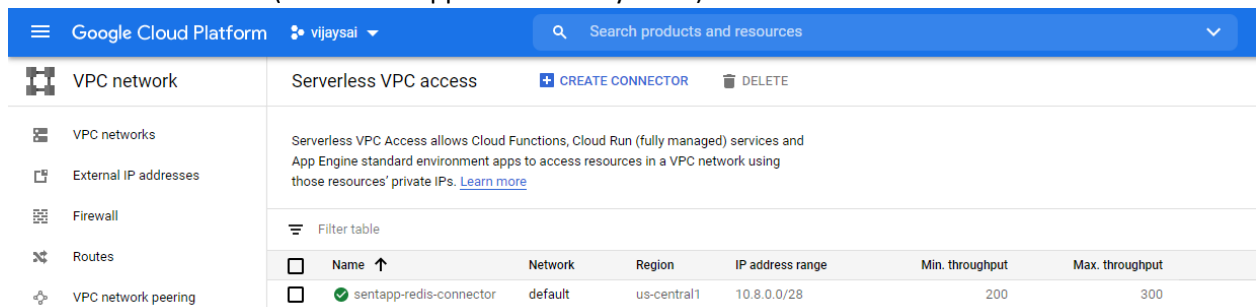
- Storage (to store histogram images)



The screenshot shows the Google Cloud Platform interface for Storage Browser. The top navigation bar includes the Google Cloud Platform logo, the user name 'vijaysai', and a search bar. The left sidebar shows the 'Storage' menu. The main content area displays a table of buckets.

Name	Created	Location type	Location	Default storage class	Updated
gcf-sources-339529935310-us-centra...	Nov 28, 2020, 8:13:16 PM	Region	us-central1 (lo...	Standard	Nov 28, 2020, 8:13:16 PM
histimages_bucket	Nov 29, 2020, 1:33:36 PM	Multi-region	us (multiple re...	Standard	Nov 29, 2020, 1:56:41 PM

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



The screenshot shows the Google Cloud Platform interface for VPC network Serverless VPC access. The top navigation bar includes the Google Cloud Platform logo, the user name 'vijaysai', and a search bar. The left sidebar shows the 'VPC network' menu. The main content area displays a table of connectors.

Name	Network	Region	IP address range	Min. throughput	Max. throughput
sentapp-redis-connector	default	us-central1	10.8.0.0/28	200	300

Sample Gcloud Logs:
sent_freq function logs

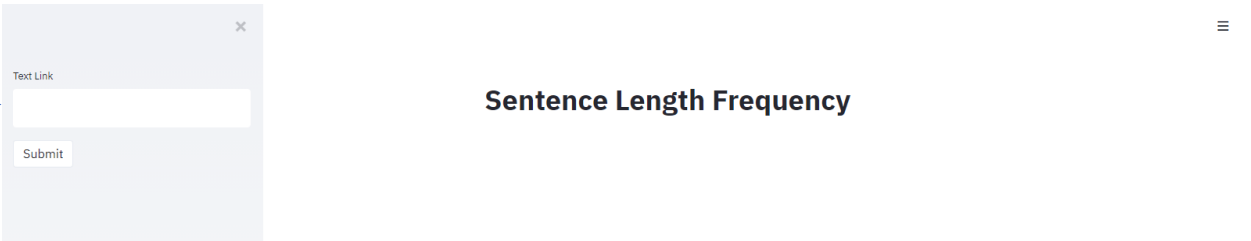
insertid	labels.exe	logName	receiveTimestamp	resource.labels	resource.labels	resource.type	severity	textPayload	timestamp	trace	
000000-6536d79f-54	z2z0yf361i	projects/vij	2020-12-02T22:31:22	sent_freq	vijaysai	us-central1	cloud_function	DEBUG	Function execution took 3683 ms, finished with status code: 200	2020-12-02T22:31:15.206782415Z	projects/vij
000000-f65f0aaa-33	z2z0yf361i	projects/vij	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-f65789f2-3a	z2z0yf361i	projects/vij	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	z2z0yf361i	projects/vij	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-c9dd52a-0	z2z0yf361i	projects/vij	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	z2z0yf361i	projects/vij	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
000000-6dc1198c-c	z2z0yf361i	projects/vij	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-ca828538-3	z2z0yf361i	projects/vij	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	rcxm6emr	projects/vij	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	rcxm6emr	projects/vij	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	rcxm6emr	projects/vij	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	rcxm6emr	projects/vij	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	rcxm6emr	projects/vij	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
000000-06972605-a	rcxm6emr	projects/vij	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	rcxm6emr	projects/vij	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	rcxm6emr	projects/vij	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	rcxm6emr	projects/vij	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-b	rcxm6emr	projects/vij	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
000000-45f1c339-5	rcxm6emr	projects/vij	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
000000-56324b2e-5	rcxm6emr	projects/vij	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-ax	jll0gtv1s	projects/vij	2020-12-02T22:02:28	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-bx	jll0gtv1s	projects/vij	2020-12-02T22:02:28	sent_freq	vijaysai	us-central1	cloud_function		[nltk_data] Unzipping tokenizers/punkt.zip.	2020-12-02T22:02:19.801Z	projects/vij
000000-3a930609-bx	jll0gtv1s	projects/vij	2020-12-02T22:02:28	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-f907c512-9x	jll0gtv1s	projects/vij	2020-12-02T22:02:28	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-ec5cb8e5-dx	jll0gtv1s	projects/vij	2020-12-02T22:02:28	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-cb06ee7-2x	jll0gtv1s	projects/vij	2020-12-02T22:02:28	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.labels	resource.labels	resource.type	severity	textPayload	timestamp	
000000-3a6cf0d	q3iyp7k3jnhk	projects/vij	2020-12-02T22:31:25.7770351082	plot_hist	vijaysai	us-central1	cloud_function	DEBUG	Function execution took 1861 ms, finished with status code: 200	2020-12-02
000000-e481f	q3iyp7k3jnhk	projects/vij	2020-12-02T22:31:25.7770351082	plot_hist	vijaysai	us-central1	cloud_function	DEBUG	Function execution started	2020-12-02
000000-dbf7f	q3iymq93o8wd	projects/vij	2020-12-02T22:29:50.7752975132	plot_hist	vijaysai	us-central1	cloud_function	DEBUG	Function execution took 3528 ms, finished with status code: 200	2020-12-02
000000-83fef	q3iymq93o8wd	projects/vij	2020-12-02T22:29:50.7752975132	plot_hist	vijaysai	us-central1	cloud_function	DEBUG	Function execution started	2020-12-02
000000-ec2ef	q3iyl90pt5xr	projects/vij	2020-12-02T22:17:45.5826872862	plot_hist	vijaysai	us-central1	cloud_function	DEBUG	Function execution took 4929 ms, finished with status code: 200	2020-12-02
000000-2def0	q3iyl90pt5xr	projects/vij	2020-12-02T22:17:45.5826872862	plot_hist	vijaysai	us-central1	cloud_function		[2020-12-02 22:17:38 +0000] [10] [INFO] Booting worker with pid: 10	2020-12-02
000001-0345f	q3iyl90pt5xr	projects/vij	2020-12-02T22:17:45.5826872862	plot_hist	vijaysai	us-central1	cloud_function		[2020-12-02 22:17:38 +0000] [1] [INFO] Using worker: threads	2020-12-02
000000-7697c	q3iyl90pt5xr	projects/vij	2020-12-02T22:17:45.5826872862	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)	2020-12-02
000000-4334d	q3iyl90pt5xr	projects/vij	2020-12-02T22:17:45.5826872862	plot_hist	vijaysai	us-central1	cloud_function		[2020-12-02 22:17:38 +0000] [1] [INFO] Starting gunicorn 20.0.4	2020-12-02
000000-dca7c	q3iyl90pt5xr	projects/vij	2020-12-02T22:17:45.5826872862	plot_hist	vijaysai	us-central1	cloud_function		OpenBLAS WARNING - could not determine the L2 cache size on this system, assuming 256k	2020-12-02
000000-e126c	q3iyl90pt5xr	projects/vij	2020-12-02T22:17:45.5826872862	plot_hist	vijaysai	us-central1	cloud_function	DEBUG	Function execution started	2020-12-02
000000-65e68	81gy1x2gq6v	projects/vij	2020-12-02T22:02:34.6753428462	plot_hist	vijaysai	us-central1	cloud_function	DEBUG	Function execution took 7972 ms, finished with status code: 200	2020-12-02
000000-e464c	81gy1x2gq6v	projects/vij	2020-12-02T22:02:34.6753428462	plot_hist	vijaysai	us-central1	cloud_function		[2020-12-02 22:02:30 +0000] [11] [INFO] Booting worker with pid: 11	2020-12-02
000000-d6cb8	81gy1x2gq6v	projects/vij	2020-12-02T22:02:34.6753428462	plot_hist	vijaysai	us-central1	cloud_function		[2020-12-02 22:02:30 +0000] [1] [INFO] Using worker: threads	2020-12-02
000000-67ae8	81gy1x2gq6v	projects/vij	2020-12-02T22:02:34.6753428462	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)	2020-12-02
000000-7db0d	81gy1x2gq6v	projects/vij	2020-12-02T22:02:34.6753428462	plot_hist	vijaysai	us-central1	cloud_function		[2020-12-02 22:02:30 +0000] [1] [INFO] Starting gunicorn 20.0.4	2020-12-02
000000-def2f	81gy1x2gq6v	projects/vij	2020-12-02T22:02:34.6753428462	plot_hist	vijaysai	us-central1	cloud_function		OpenBLAS WARNING - could not determine the L2 cache size on this system, assuming 256k	2020-12-02
000000-df378	81gy1x2gq6v	projects/vij	2020-12-02T22:02:34.6753428462	plot_hist	vijaysai	us-central1	cloud_function	DEBUG	Function execution started	2020-12-02
000000-8122d	bm4ck84eizir	projects/vij	2020-12-02T19:07:05.6287103102	plot_hist	vijaysai	us-central1	cloud_function	DEBUG	Function execution took 8207 ms, finished with status code: 200	2020-12-02
000000-bb55f	bm4ck84eizir	projects/vij	2020-12-02T19:07:05.6287103102	plot_hist	vijaysai	us-central1	cloud_function		[2020-12-02 19:07:02 +0000] [10] [INFO] Booting worker with pid: 10	2020-12-02
000000-e17ef	bm4ck84eizir	projects/vij	2020-12-02T19:07:05.6287103102	plot_hist	vijaysai	us-central1	cloud_function		[2020-12-02 19:07:02 +0000] [1] [INFO] Using worker: threads	2020-12-02
000000-c0d8f	bm4ck84eizir	projects/vij	2020-12-02T19:07:05.6287103102	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)	2020-12-02
000000-5307d	bm4ck84eizir	projects/vij	2020-12-02T19:07:05.6287103102	plot_hist	vijaysai	us-central1	cloud_function		[2020-12-02 19:07:02 +0000] [1] [INFO] Starting gunicorn 20.0.4	2020-12-02
000000-9db1d	bm4ck84eizir	projects/vij	2020-12-02T19:07:05.6287103102	plot_hist	vijaysai	us-central1	cloud_function		OpenBLAS WARNING - could not determine the L2 cache size on this system, assuming 256k	2020-12-02
000000-d6dd4	bm4ck84eizir	projects/vij	2020-12-02T19:07:05.6287103102	plot_hist	vijaysai	us-central1	cloud_function	DEBUG	Function execution started	2020-12-02

Screenshots of Application:

Before entering any input:



Sample Outputs:

Text Link

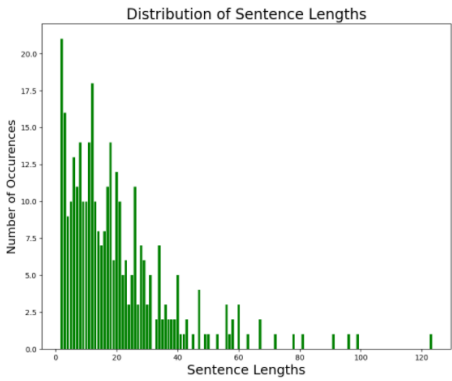
Submit

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

Text Link

<http://www.gutenberg.org/files/63941>

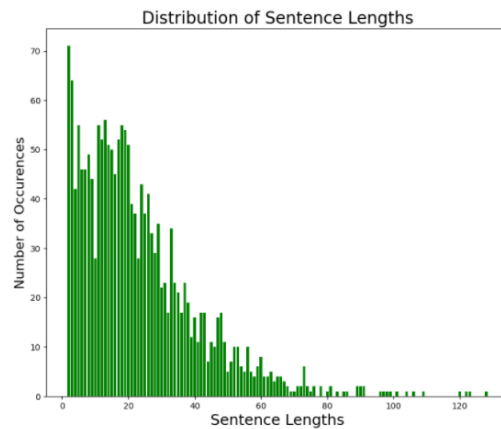
Submit

Sentence Length Frequency

For text link: <http://www.gutenberg.org/files/63941/63941-0.txt>

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

```
{2: 71, 3: 64, 4: 42, 5: 55, 6: 46, 7: 46, 8: 49, 9: 44, 10: 28, 11: 55, 12: 52, 13: 56, 14: 51, 15: 50,
16: 45, 17: 52, 18: 55, 19: 54, 20: 51, 21: 39, 22: 37, 23: 28, 24: 43, 25: 37, 26: 41, 27: 33, 28:
29, 29: 35, 30: 22, 31: 23, 32: 17, 33: 34, 34: 23, 35: 21, 36: 17, 37: 23, 38: 19, 39: 12, 40: 16,
41: 11, 42: 17, 43: 17, 44: 7, 45: 11, 46: 10, 47: 16, 48: 17, 49: 11, 50: 5, 51: 7, 52: 10, 53: 10,
54: 6, 55: 5, 56: 10, 57: 5, 58: 4, 59: 6, 60: 8, 61: 4, 62: 4, 63: 5, 64: 3, 65: 4, 66: 4, 67: 3, 68: 2,
69: 1, 70: 1, 71: 2, 72: 2, 73: 6, 74: 2, 75: 1, 76: 2, 78: 2, 80: 1, 81: 2, 83: 1, 85: 1, 86: 1, 89: 2,
90: 2, 91: 2, 96: 1, 97: 1, 98: 1, 99: 1, 101: 1, 104: 1, 106: 1, 109: 1, 120: 1, 122: 1, 123: 1, 128:
1}
```



Time taken in Seconds: 16.409350395202637

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