**Technical Documentation-Whats Happenin**

In this document, the team updates all technical processes that might have to be repeated by team members on different systems.

1. Process to connect with google cloud through terminal
   1. Initial google cloud in terminal
   2. Kubectl config view
   3. gcloud container clusters get-credentials what --region us-west2 --project rapid-burner-415905
   4. kubectl port-forward cb-example-0000 8091:8091
   5. Add data to database using add document.
2. Setting up environment in vsc:

Tip: If the activate command generates the message "Activate.ps1 is not digitally signed. You cannot run this script on the current system.", then you need to temporarily change the PowerShell execution policy to allow scripts to run (see About Execution Policies in the PowerShell documentation): Set-ExecutionPolicy -ExecutionPolicy RemoteSigned -Scope Process

A screenshot of a computer

Description automatically generated

1. Getting direnv, then wsl then linux distribution to use direnv:
   1. Direnv: https://gist.github.com/rmtuckerphx/4ace28c1605300462340ffa7b7001c6d
   2. WSL: <https://learn.microsoft.com/en-us/windows/wsl/install>
   3. Linux distribution, Ubuntu: <https://learn.microsoft.com/en-us/windows/wsl/install?source=recommendations>  
      username: shanya

Password: shanya

In the ubuntu shell, everytime you start, type: sudo apt update && sudo apt upgrade

1. How to clear the socket when an application is using 8000 port
   1. Error : ERROR: [WinError 10013] An attempt was made to access a socket in a way forbidden by its access permissions  
      on Powershell find what process is occupying 8000 port.
   2. Netstat -ano | findstr :8000
   3. LISTENING xxxxx
   4. Press Ctrl+Shft+Esc
   5. Go to details
   6. Find the port and end the task
2. How to update branch and make a pull request using gitbash

Git checkout main

Git pull origin main

Git checkout MLmanager

Git merge main

Git status

Git add .

Git commit -m ‘commit comment’

Git push origin MLmanager

Git status

1. Commands to run API:
2. Open gitbash terminal
3. cd Topic\_Modeling
4. source .venv/Scripts/Activate
5. pip install -r requirements.txt
6. python api\_call\_script.py oid
7. Explanation of MongoDB data before processing:

A screenshot of a computer

Description automatically generated

* 1. \_id (ObjectId): Object Id used to access the MongoDB document
  2. location (str): query parameter used to make API call to newscatcherAPI, need this field to remove unnecessary articles from the data fetched from newscatcher api
  3. articles (List[Dict]): Raw data from newscatcherAPI with all the fields that are fetched, Prateek feel free to call for less fields.

1. Explanation of MongoDB data after processing:

A screen shot of a computer

Description automatically generated A screenshot of a computer code

Description automatically generated

* 1. \_id (ObjectId): Object Id used to access the MongoDB document
  2. location (str): query parameter used to make API call to newscatcherAPI, need this field to remove unnecessary articles from the data fetched from newscatcher api.
  3. articles (List[Dict]): List of dictionary with processes articles and topics from topic modelling. Each dictionary contains the following keys: rank, location, title, excerpt, summary, link, author, published\_date, image\_link, topic.
  4. data\_for\_bubble (Dict): Contains the dictionary with the topics and its respective percentage out of total. Each value in the dictionary would correspond to the size of the bubble in the bubble chart
  5. top\_10\_by\_topics (Dict): Dictionary of the topic name and respective list of indices of top x article in ascending order. One can use these indices to grab articles from “articles” field in MongoDB. This will need to be accessed when the user clicks the topic bubble. Bill will need the articles at these indices and the following keys from each of the article dictionaries: title, link, author, published\_date, image\_link
  6. top\_x\_all\_cat (List)-Default: List of indices of the top x ranked articles. This is the list that would be used to display the top articles when no topic is chosen. Bill will need the articles at these indices and the following keys from each of the article dictionaries: title, link, author, published\_date, image\_link