

DATA AGGREGATION FUNNEL

We aggregate, archive and process all the data, content and discussions from relevant platforms (twitter, discord, forums, yt etc) and index them so that they can be referred by the audience for our specific use cases. (see use cases)

PS : I've already applied the ACT Stack v2.0 for the BBAC accelerator program, hoping to hear some good news. Will keep you guys posted

[

image|690x

1764x1088 90.5 KB

](https://global.discourse-cdn.com/apecoin/original/2X/5/59628b88d7b43ac57401c6d37a09bfe2a0d10ac2.png)

TWITTER :

(Tweet, Threads and Spaces)

For archiving Posts and Threads which contain valuable information, Mention @apecoin

tracker with the keyword Capture or Note. This way, the relevant post will be added to our search engine which can later be referenced.

Ex : Capture this thread @apecointracker

or Note this thread @apecointracker

These keywords are custom and can be modified, capture this post

, cover this space

etc are the specific keywords which the bot uses to listen to the incoming stream of data and differentiate be.

We can change it to whatever that makes sense accordingly, we can tie in additional functionality to this workflow. Example, we want some specific action to trigger, we mention the relevant keywords for that task and mention the bot which triggers a certain automation.

For archiving Spaces within the fold, there's nothing you need to do. They are covered automagically

For archiving Spaces outside the fold but containing relevant information, mention @apecointracker

with a combination of keywords cover

and space

so that it can be taken through the pipeline.

Ex : cover this space @apecointracker

Data from this mention goes into a separate file, and then gets processed EoD

Functionality : Browser automation using Selenium

{Go to search the mentioned instances, take the tweet text, s.split

and re

for a combination of individual words and compare with test strings like capture

and tweet

, thread

or post

to differentiate this from other mentioned instances for additional use cases}

Status : Deployed

(Imn when you want to test this out, i'll deploy it accordingly)

DISCORD :

(Message, Thread, Stages and Calls)

For individual messages and Threads, we can use a bot to monitor the chats for a certain action. As soon as that action is detected, it'll perform a predetermined set of actions. For example, when I react to a message or a thread with a ' ' emoji, it takes that message and puts it through the data pipeline.

For Voice calls, the bot can be activated by a certain command and told which vc needs to be recorded, at the end of which it'll return the audio recording.

There are many off the shelf tools we can use for this specific case like [craig.chat](#)

For Stages, we can use the tool [craig.net](#)

to record and then refer to it at a later date.

Functionality : Python + Discord API along with other tools mentioned

Status : Ideation Stage

YOUTUBE :

(Videos / Podcasts)

Since videos and podcasts share the same url structure and the shorts url structure is also compatible with the videos url structure (but in a video format, instead of a scrolling carousel format)

There is a case to be made about filtering out short form content cause of their relevance or the lack thereof on an individual (channel) basis

Functionality : Demonstrated

Status : Deployed

This includes the channels who signed up for coverage [here](#)

REDDIT :

My analysis of apecoin related activity on reddit is pretty much non existent, there are some stagnant communities (r/apecoin) and only one or two are active enough to get one or two posts a day. (r/apecoin_dao, r/apeointoken and r/apecoins)

But we've included this avenue of data aggregation for when activity picks up there.

will explore mention style post archival for twitter posts outside of these subreddits.

Status : Not viable atm, shelved until further notice

FORUMS :

The AIPs can be broadly categorised into three categories :

- Upcoming (All the tags associated with upcoming aips, and those going through the process up until its being voted upon)
- Approved (AIPs with the approved tag)
- Withdrawn (Rejected and Withdrawn AIPs)

The AIPs, along with the user feedback (comments, vote mandate etc) and associated tags can decide how that data is being perceived.

We can use off the shelf Industrial Scale data scrapers for this instead of a custom script, but I'd prefer the latter as it'll offer us considerable savings over the long term.

Status : In the Works

INTERNET :

Users submit urls for consideration and after filtering through them (criteria = relevance) and then we scrape the url for

information and add it to the Index along with the relevant identifiers and metadata.

Same goes for this, we can either use off the shelf scrapers for a fee or use our custom script which'll save us a considerable amount over the long term.

Status : In the Works

USE CASES :

RaG Assistant :

A RaG (Retrieval Augmented Generation) powered assistant which helps you answer any query with context from the UCI. It fetches the relevant data it needs and responds accordingly.

The backend of it is the same and it can be integrated with any platform of one's choice, for example as a discord bot, telegram or twitter bot.

(See Twitter data aggregation portion, you could mention it with the keyword help, ex : help @apecointracker

) followed by your query and it'll respond accordingly

It can even be integrated into a website or act as a standalone one.

Status : Complete

(gave [@bigbull](#) among others a demo a while back)

Unified Content Index :

The data aggregation pipeline ultimately feeds into the Unified content Index where information is processed, sorted and quantified. Here's where we associate the data with relevant identifiers and metadata so that we can further improve the quality of our data.

Think of it as a central Library which houses all discussions which are held. The content inside which would be sorted accordingly and would be searchable

By keeping all this in a single place, one would be able to search through the content library for relevant discussions and then be able to refer to that particular discussion with the backlinks for that particular snippet.

Data Aggregation from multiple sources to ensure we take in as much relevant data as humanly possible alongside its metadata identifiers to classify data accordingly. It also keeps track of how the ideas being discussed evolve over time and provides a way for users to reference that.

UCI powers both the RaG assistant and the Search Engine

Search Engine Powered by Unified Content Index :

Now imagine this, You want to search more about a specific topic, let's say Voting Reforms

. I'll search for it and get all the instances where that was mentioned across platforms and I wouldn't have to go on and individually try to gather this information from all the different sources. This would not only save me the time, but it'll work even in cases where I can't search through (like twitter spaces) individual instances of a certain topic being mentioned.

Not only will I get a response to answer my original question, but also get the sources (links to said discussions, timestamped audio archives) so I can continue my research (using the unified content Index)

Apart from getting a response to my original question, I'll also get the sources (links to said discussions, timestamped audio archives) so I can continue my research (using the unified content Index) and reach to a conclusion faster.