API Documentation

Introduction

A self hosted API for web metadata retrieval, offers the web-scraping functionality when any valid URL is provided. The identified stakeholders are:

- Developers (Who contributes to this open-source project)
- Companies/Developers (Who integrates our service with their product(Web/Mobile App))
- End-users of the system (Everyone who uses our service integrated products.)

This will be a very useful tool for those who are interested in extracting the data and use it for any analysis.

Description
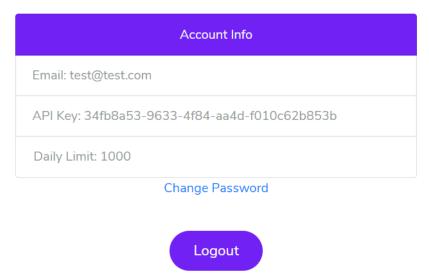
1. **Registration** :
   - Enter Email address
   - Password
   - Confirm password

(Password should be min 8 characters. Shouldn't match to email address, and cannot be too common. If so, then it will show an error message in red letters.)

Register

Email address

Password

Confirm Password

☐ Show Password

This password is too short. It must contain at least 8 characters.

This password is too common.

This password is entirely numeric.
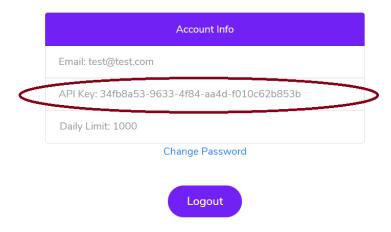
Proceed

Already have an Account? Login

◀ Go back to Home

Following screen will appear once registration is done successfully.

Account Info

Email: test@test.com

API Key: 34fb8a53-9633-4f84-aa4d-f010c62b853b

Daily Limit: 1000

Change Password

Logout

2. **Consume API**

Following are the ways to consume the self-hosted API:

○ **API_Key**: captured from the registration screen

Account Info

Email: test@test.com

API Key: 34fb8a53-9633-4f84-aa4d-f010c62b853b

Daily Limit: 1000

Change Password

Logout

○ Valid Web **URL** that is required to scraped

○ **Cached**= true/false (optional): It is useful for optimising the access time.

3. **Accessing with URL**

❖ Url to access API Endpoint:
https://web-metadata.herokuapp.com/api/

❖ Input Parameters for GET Request.:

❖ api_key (Enter your valid API key)
❖ web_page (Enter a valid website)
❖ cached (Enter true/false) (default=true) (optional)

❖ Metadata API Url:

https://web-metadata.herokuapp.com/api/?api_key=34fb8a53-9633-4f84-aa4d-f010c62b853b&format=api&web_page=https%3A%2F%2Fwww.youtube.com%2Fwatch%3Fv%3DXQgXKtPSzUI

❖ Success Response:

```
{
"web_page": "https://www.youtube.com/watch?v=XQgXKtPSzUI",

"title": "Intro to Web Scraping with Python and Beautiful Soup",

"description": "Web scraping is a very powerful tool to learn for any data professional. With web scraping the entire internet becomes your database. In this tutorial we sho...",

"thumbnail": "https://i.ytimg.com/vi/XQgXKtPSzUI/maxresdefault.jpg"

}
```

★ *Following is the sample output→ when we have only entered the registered API key and valid URL:*

★ *Following is the expected output when we add a parameter: cached = true. (when you cached=true, it will bring data from the cache.)*

★ *Following is the expected output when we add a parameter: cached = false. (when cached=false, it will scrape the page irrespective of already existing in the cache)*

Metadata Api

Metadata Api   OPTIONS   GET ▾

GET /api/?api_key=34fb8a53-9633-4f84-aa4d-f010c62b853b&web_page=https%3A%2F%2Fwww.youtube.com%2Fwatch%3Fv%3DXQgXKtPSzUI&cached=false

```
HTTP 200 OK
Allow: GET, OPTIONS
Content-Type: application/json
Vary: Accept

{
    "web_page": "https://www.youtube.com/watch?v=XQgXKtPSzUI",
    "title": "Intro to Web Scraping with Python and Beautiful Soup",
    "description": "Web scraping is a very powerful tool to learn for any data professional. With web scraping the entire internet becomes your database. In this tutorial
    "thumbnail": "https://i.ytimg.com/vi/XQgXKtPSzUI/maxresdefault.jpg"
}
```
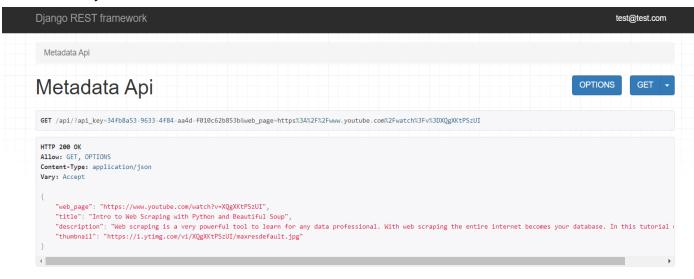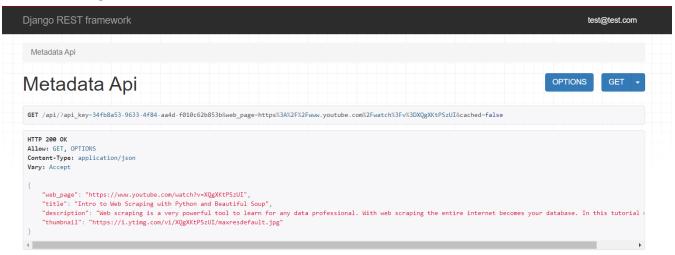
★ *Output for another valid URL:*

Metadata Api

Metadata Api   OPTIONS   GET ▾

GET /api/?api_key=34fb8a53-9633-4f84-aa4d-f010c62b853b&web_page=https://www.youtube.com/watch?v=kyJCzFNRK4s

```
HTTP 200 OK
Allow: GET, OPTIONS
Content-Type: application/json
Vary: Accept

{
    "web_page": "https://www.youtube.com/watch?v=kyJCzFNRK4s",
    "title": "🌏 BEST LOCKDOWN BIRTHDAY SURPRISE! (Gujarati) | P.S. This will MELT YOUR HEART! | Dr. Aayushi & Isha",
    "description": "3rd August, 1969 was when a fairy, an angel, a princess and the Queen of Hearts was born! 👩🍒🎉Presenting to you the wonder-woman that we're celebrat
    "thumbnail": "https://i.ytimg.com/vi/kyJCzFNRK4s/maxresdefault.jpg"
}
```

★ *Output: For an Invalid URL*

web-metadata.herokuapp.com/a × +

← → C ⚠ Not secure | web-metadata.herokuapp.com/api/?api_key=34fb8a53-9633-4f84-aa4d-f010c62b853b&web_page=https://www.youtubeeee.com/watch?v=kyJCzFNR... ☆

Server Error (500)

❖ **Daily limit**: Every user is provided 1000 as a daily limit of accessing the API.

★ *See below, after calling the API Key 4 times, on checking the daily limit: we can see it has been reduced to 996 out of 1000.*