

GOOGLE CRAWLER PROJECT

TAM LE | DEVELOPER | TAMLE.DEV@OUTLOOK.COM | 12/08/2021

I. WHAT IS IT?

This web application that will extract large amounts of data from the Google search results page.

This version only supports uploading csv file, with one keyword per line. Download example file below.

II. SYSTEM DEPENDENCIES

- Ruby 2.6.2
- Rails 5.2.6
- PostgreSQL
- Google Cloud Storage
- Sidekiq
- Redis

III. SETUP & CONFIGURATION

i. MANUAL

1. Install PostgreSQL
2. Install Ruby 2.6.2 (recommend use rvm)
3. Install Geckodriver
4. Clone project
5. Duplicate *.env* from *.env.example*
6. Add Google Cloud Credential JSON into *config/secrets.yml*
7. Install dependencies
8. Create database
9. Migrations and Seed data
10. Start rails application
11. Access *localhost:3000*

ii. DOCKER

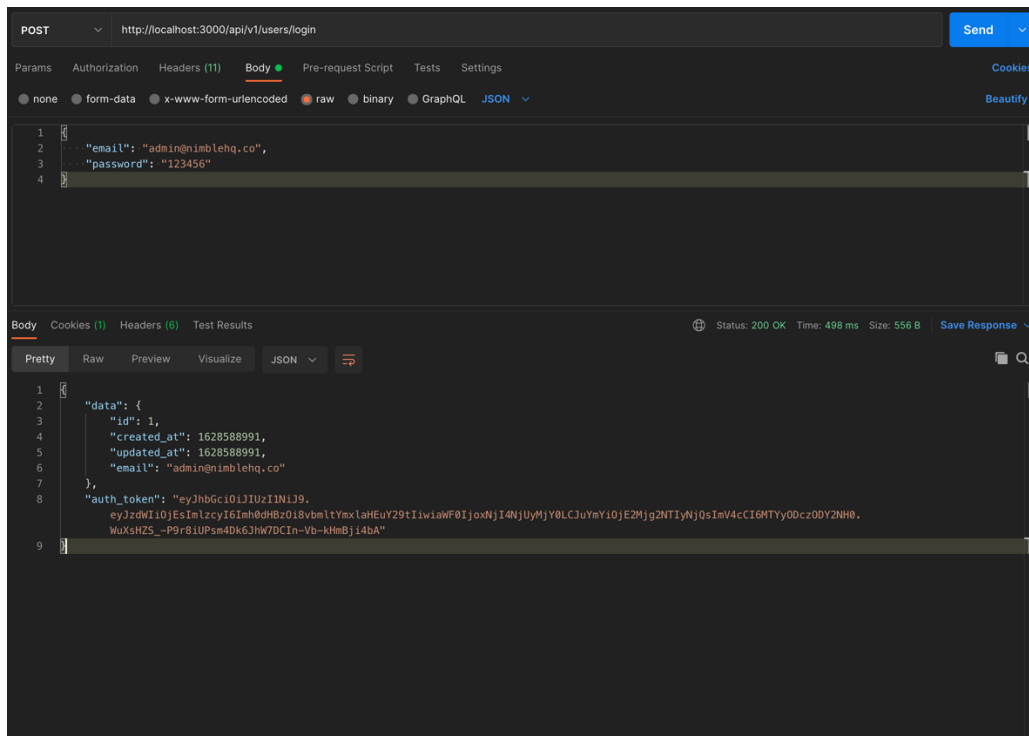
1. Clone project
2. Duplicate `.env` from `.env.example`
3. Add Google Cloud Credential JSON into `config/secrets.yml`
4. Build image (`docker build . -t tamle:lasted`)
5. Compose container (`docker-compose up -d`)
6. Access `localhost:3000`

IV. API

GCRP API is organized around REST which has predictable resource-oriented URLs, accepts form-encoded request bodies, returns JSON-encoded responses, and uses standard HTTP response codes, authentication, and verbs.

1. Login to get JWT token

POST api/v1/users/login



2. Upload a keywords csv file

POST api/v1/attachments

The screenshot shows a REST client interface with a POST request to `http://localhost:3000/api/v1/attachments`. The request body is a file named `keywords.csv`. The response is a JSON object with the following structure:

```
1 {
2   "data": {
3     "id": 24,
4     "created_at": 1628653737,
5     "updated_at": 1628653738,
6     "url": "https://storage.googleapis.com/crawler-b1/uploads/attachment/file/24/keywords.csv",
7     "file_name": "keywords.csv"
8   }
9 }
```

The status bar indicates a 201 Created response with a time of 23.47 s and a size of 452 B.

3. Get list of keywords

GET api/v1/keywords

The screenshot shows a REST client interface with a GET request to `http://localhost:3000/api/v1/keywords`. The request headers include `Authorization` and `Content-Type: application/json`. The response is a JSON object with the following structure:

```
1 {
2   "data": [
3     {
4       "id": 500,
5       "created_at": 1628705896,
6       "updated_at": 1628706073,
7       "value": "nfl scores",
8       "total_adword": 0,
9       "total_url": 90,
10      "total_result": 987000000,
11      "total_result_time": 570,
12      "total_result_text": "987,000,000",
13      "total_result_time_text": "0.57 seconds",
14      "details": [
15        {
16          "id": 4554,
17          "created_at": 1628706072,
18          "updated_at": 1628706072,
19          "title": "NFL Football Scores - NFL Scoreboard - ESPN",
20          "description": "> nfl > scoreboard | > nfl > scoreboard | Real-time NFL Football scores on ESPN. | ",
21          "url": "https://www.espn.com/nfl/scoreboard",
22          "domain": "https://www.espn.com",
23          "position": 2
24        }
25      ]
26    }
27  ]
28 }
```

The status bar indicates a 200 OK response with a time of 418 ms and a size of 43.7 KB.

4. Get search results for each keyword

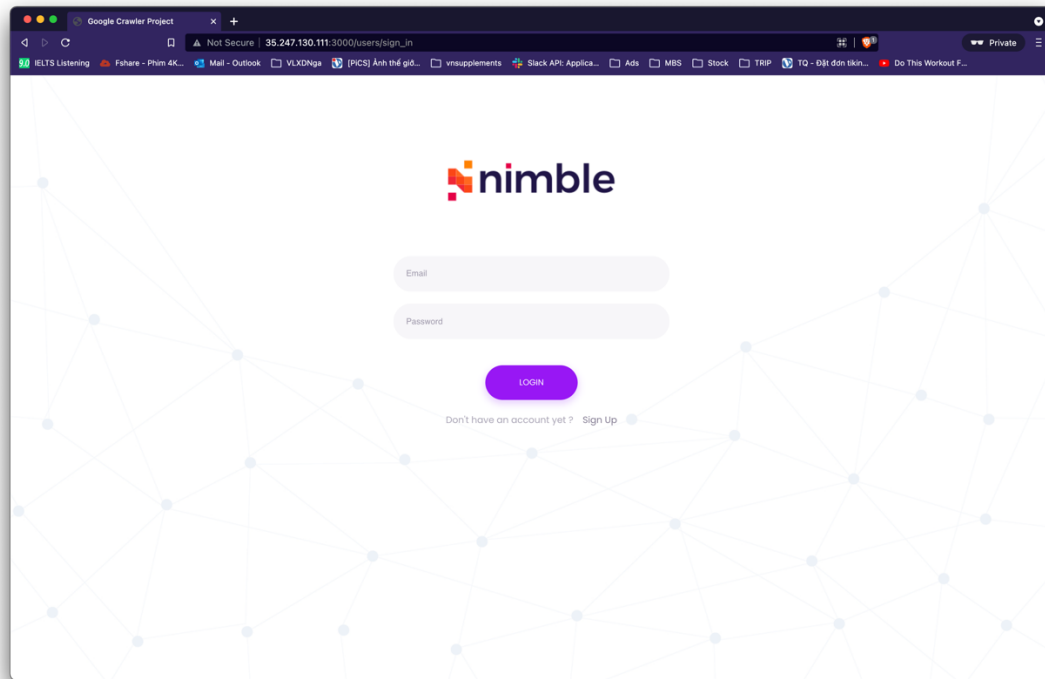
GET `api/v1/keywords/:id`

The screenshot shows a REST client interface with a GET request to `http://localhost:3000/api/v1/keywords`. The request headers are set to `Authorization: eyJhbGciOiJIUzI1Ni9yZzdWIl0iEslmizcyf6lmh0dHBzOi8vbml...` and `Content-Type: application/json`. The response status is 200 OK, with a time of 418 ms and a size of 43.7 KB. The response body is a JSON object containing an array of keyword data.

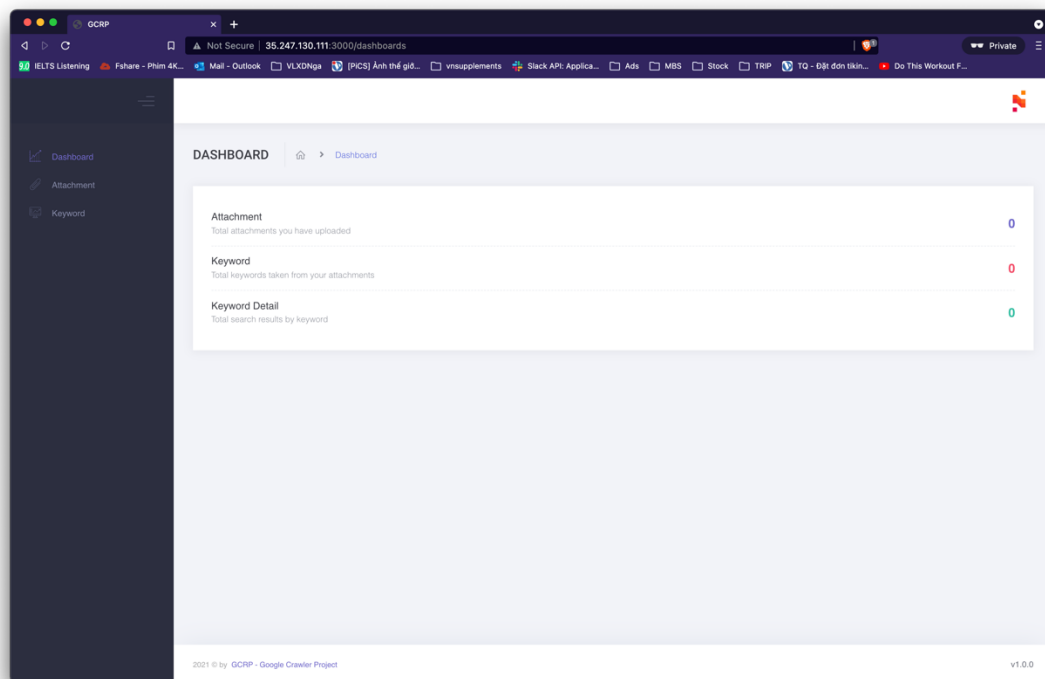
```
{
  "data": [
    {
      "id": 500,
      "created_at": 1628705896,
      "updated_at": 1628706073,
      "value": "nfl scores",
      "total_adword": 0,
      "total_url": 90,
      "total_result": 987000000,
      "total_result_time": 570,
      "total_result_text": "987,000,000",
      "total_result_time_text": "0.57 seconds",
      "details": {
        "id": 4554,
        "created_at": 1628706072,
        "updated_at": 1628706072,
        "title": "NFL Football Scores - NFL Scoreboard - ESPN",
        "description": "nfl > scoreboard | > nfl > scoreboard | Real-time NFL Football scores on ESPN. | ",
        "url": "https://www.espn.com/nfl/scoreboard",
        "domain": "https://www.espn.com",
        "position": 2
      }
    }
  ]
}
```

V. UI

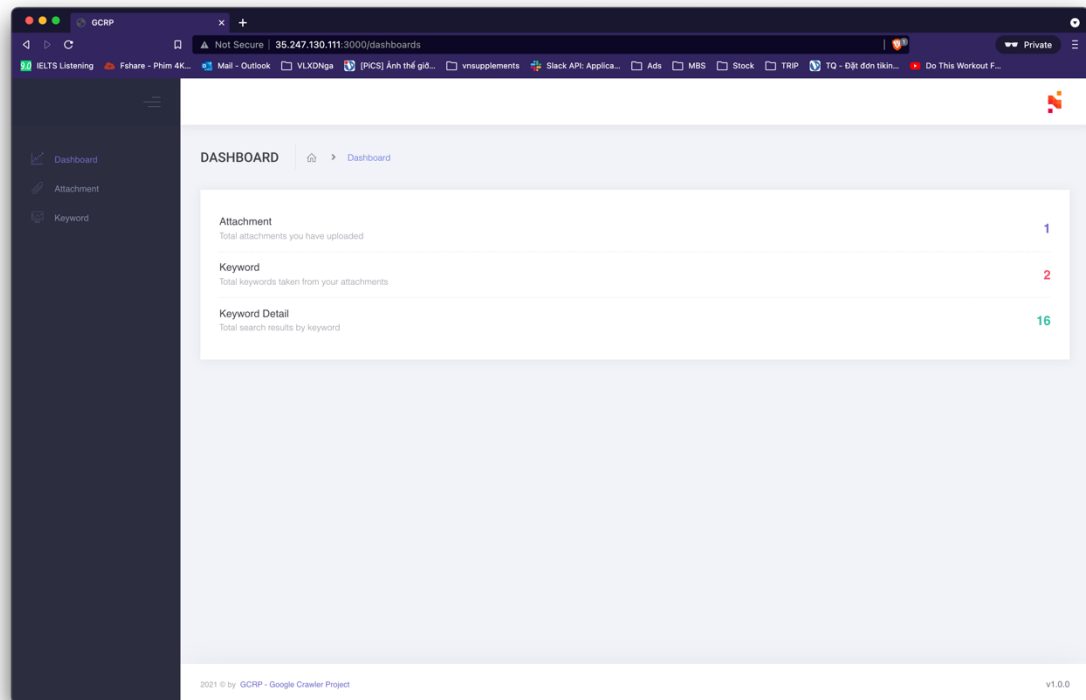
1. Sign in



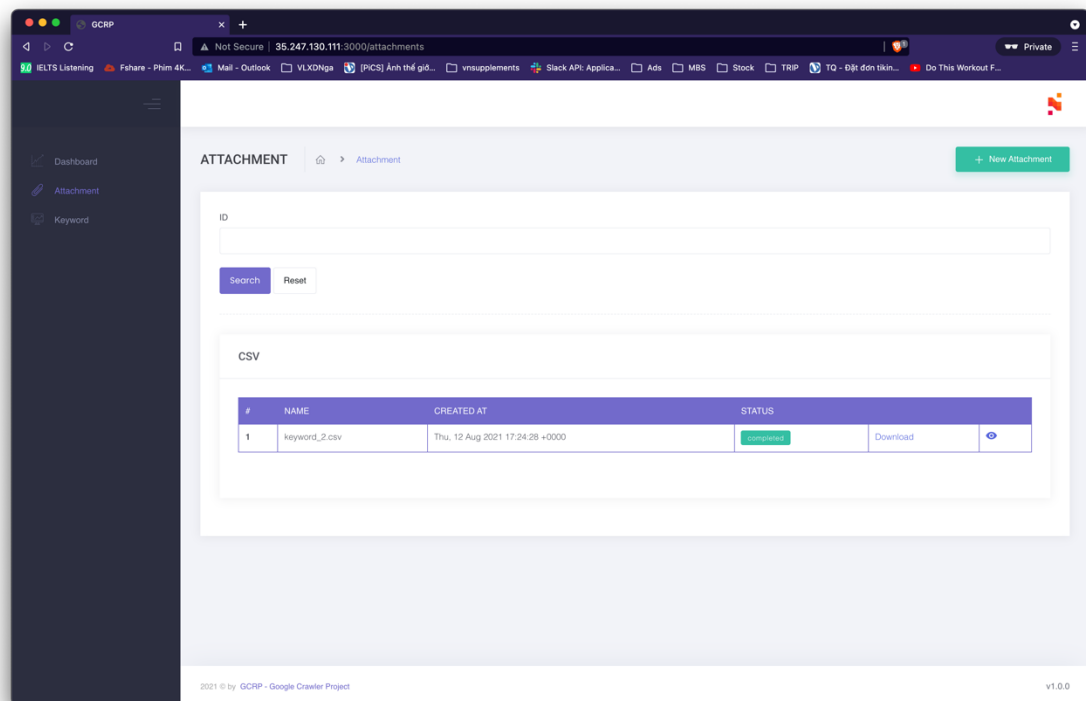
2. Sign up



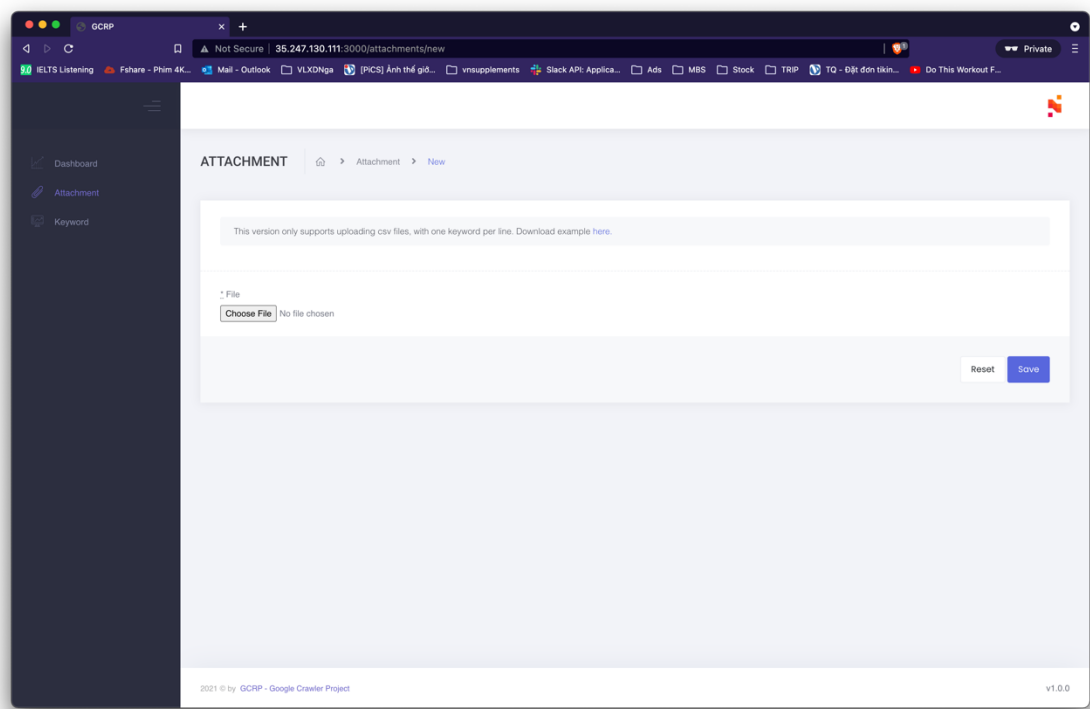
3. Dashboard



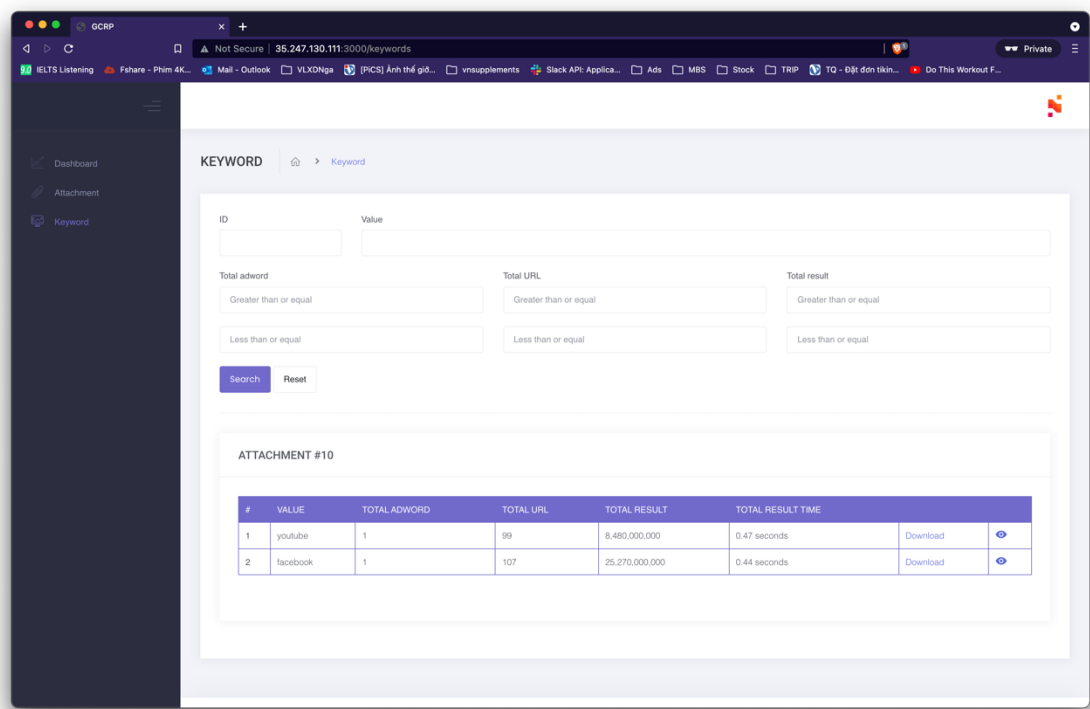
4. List attachments



5. Upload an attachment



6. List keywords



7. Keyword details

Dashboard

Attachment

Keyword

KEYWORD

Home

Keyword

5

VALUE	youtube
TOTAL ADWORD	1
TOTAL URL	99
TOTAL RESULT	8,480,000,000
TOTAL RESULT TIME	0.47 seconds
PREVIEW	Download

POSITION	TITLE	DESCRIPTION	URL	DOMAIN
1	YouTubeYouTube PremiumLiveHistoryChannelsAudio LibraryMusic	Similar Similar 1 Nov 2019 — Enjoy the videos and music you love, upload original content, and share it all with friends, family, and the world on YouTube.	https://www.youtube.com/	https://www.youtube.com
2	YouTube	Similar Similar 1 Nov 2019 — Enjoy the videos and music you love, upload original content, and share it all with friends, family, and the world on YouTube.	https://www.youtube.com/	https://www.youtube.com
3	YouTube (@YouTube) · Twitter	Twitter · 16 mins ago Twitter · 24 mins ago Twitter · 16 hours ago	https://twitter.com/YouTube?ref_src=twsrc%5Egoogle%7Ctwamp%5Eserp%7Ctwg%5Eauthor	https://twitter.com/YouTube
4	YouTube - Home Facebook	· ... · Product/Service · ... · Product/Service CachedSimilar Cached Similar YouTube. 86559479 likes · 234795 talking about this. The latest and greatest music videos, trends and channels from YouTube.	https://www.facebook.com/youtube/	https://www.facebook.com
5	YouTube - Apps on Google Play	· store · apps · details · id-co... · store · apps · details · id-co... Cached Cached Get the official YouTube app on Android phones and tablets. See what the world is watching -- from the hottest music videos to what's popular in gaming... Rating: 4.3 119,888,783 votes Free Android Entertainment	https://play.google.com/store/apps/details?id=com.google.android.youtube&hl	https://play.google.com