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 PosgreSQL
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

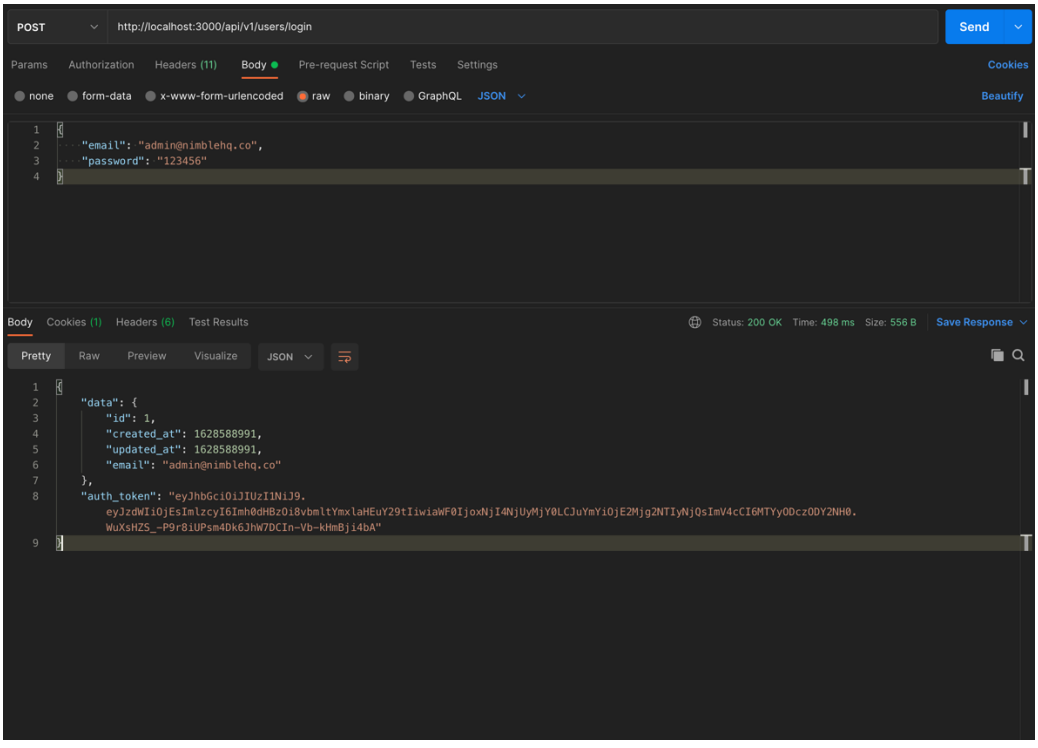
1. Clone project
2. Duplicate .env from .env.example
3. Add Google Cloud Credential JSON into *config/secrets.yml*
4. Build image (docker build . -t tamle:lasted)
5. Compose container (docker-compose up -d)
6. Access *localhost:3000*

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.

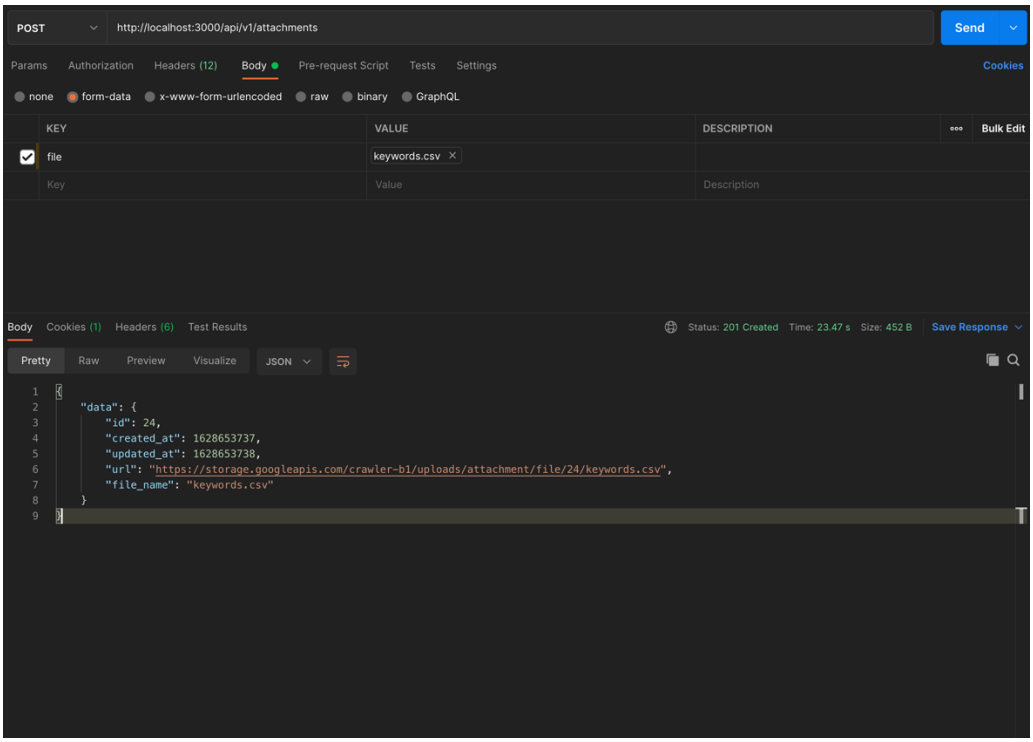
1. Login to get JWT token

`POST api/v1/users/login`



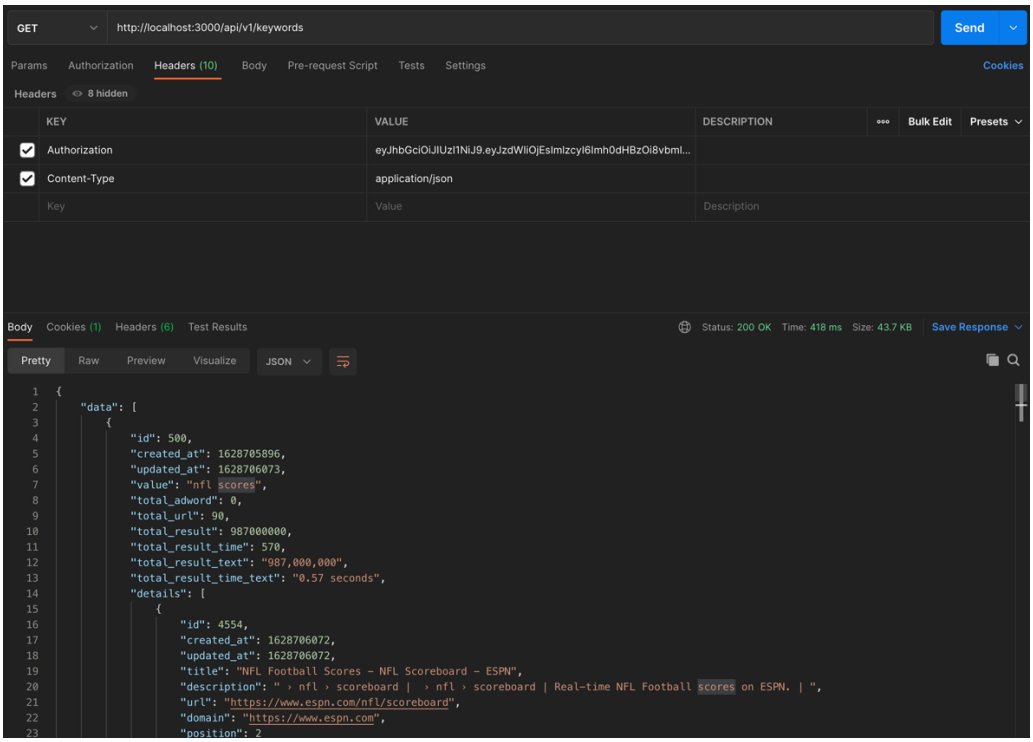
2. Upload a keywords csv file

`POST api/v1/attachments`



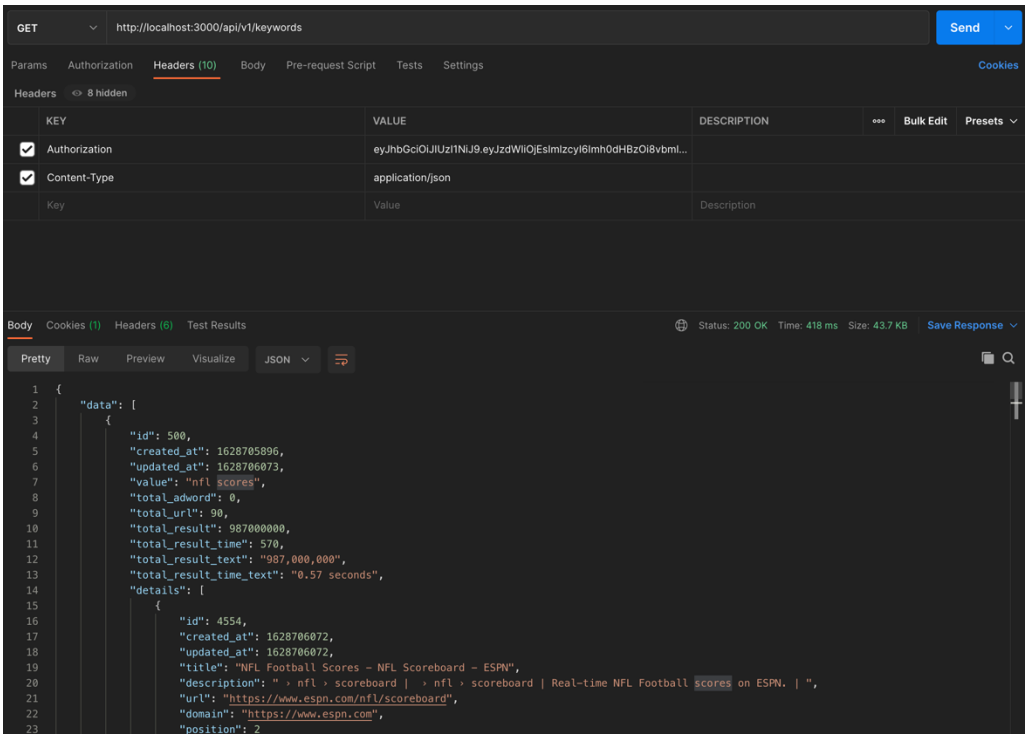
3. Get list of keywords

`GET api/v1/keywords`



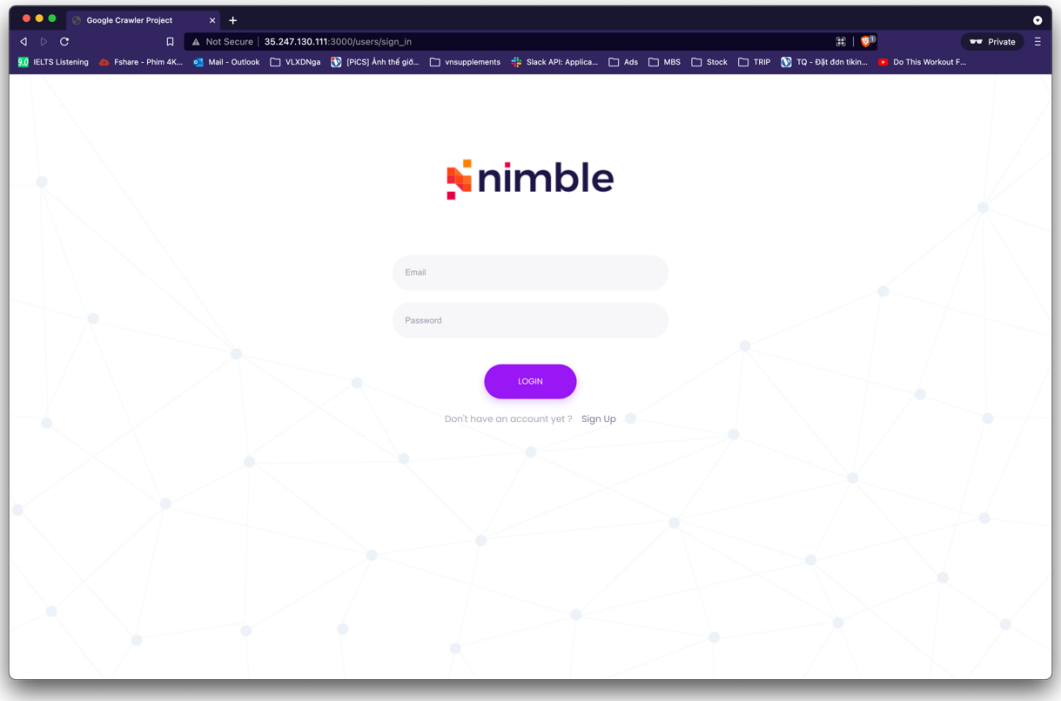
4. Get search results for each keyword

`GET api/v1/keywords/:id`

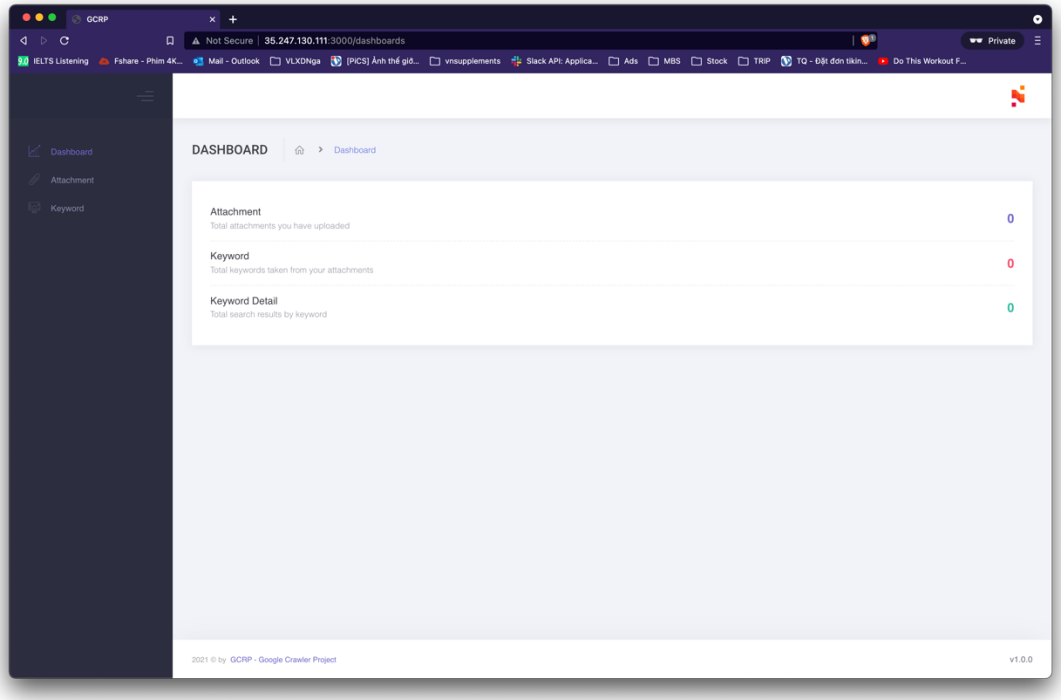


V. UI

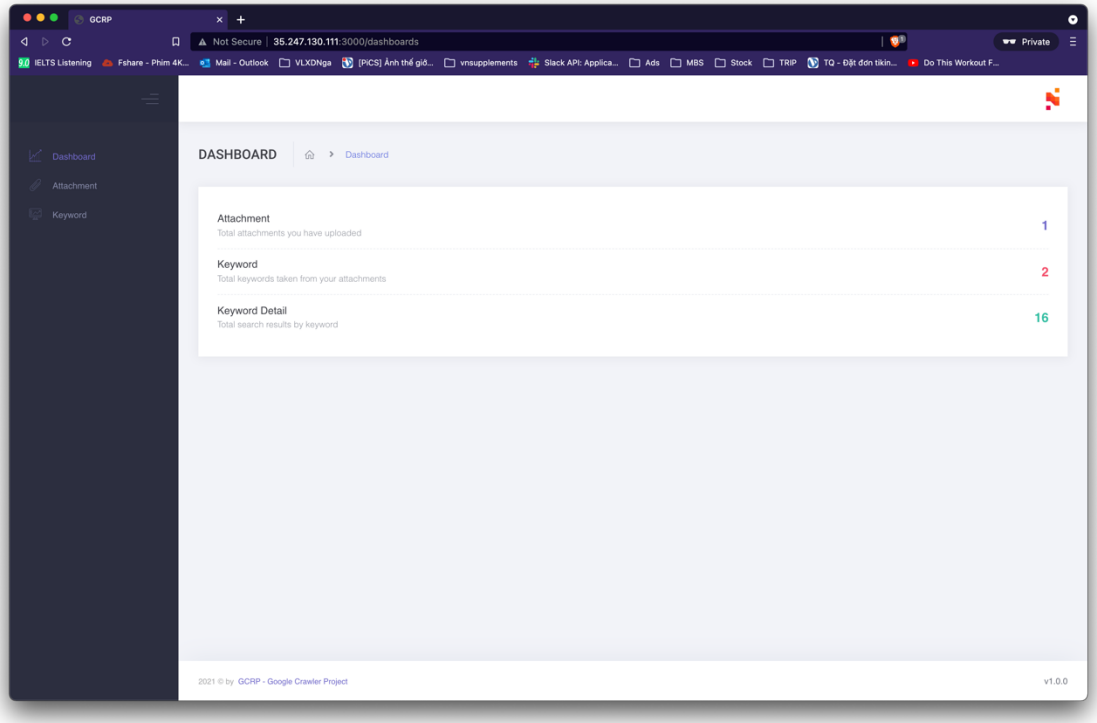
1. Sign in



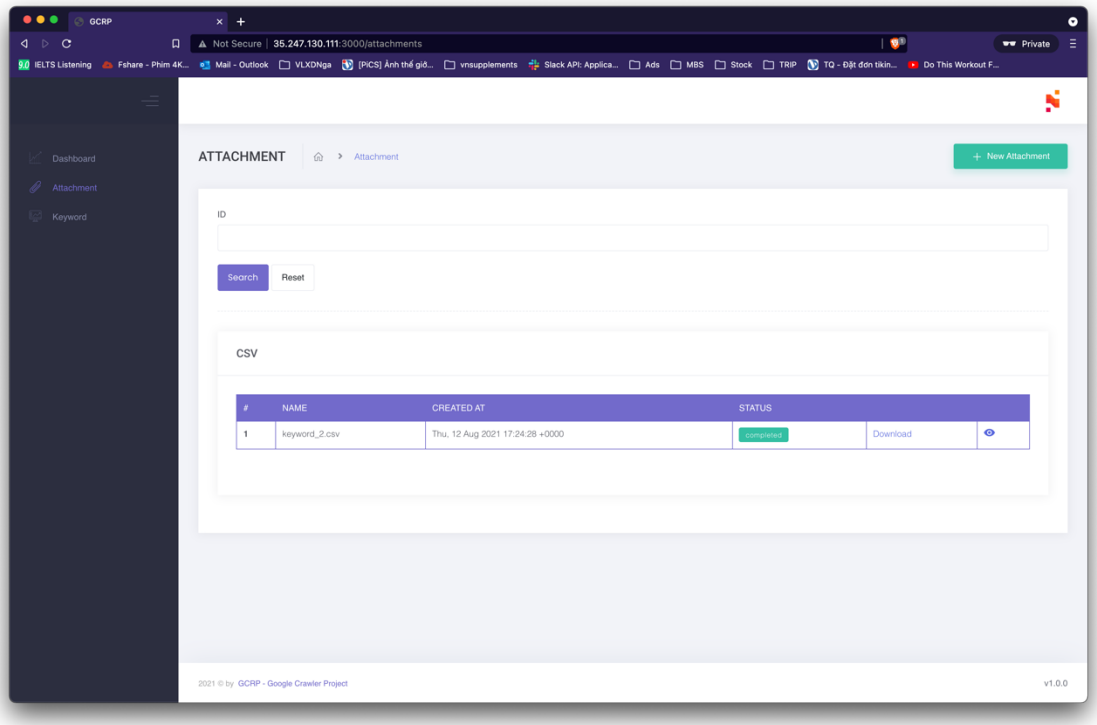
2. Sign up



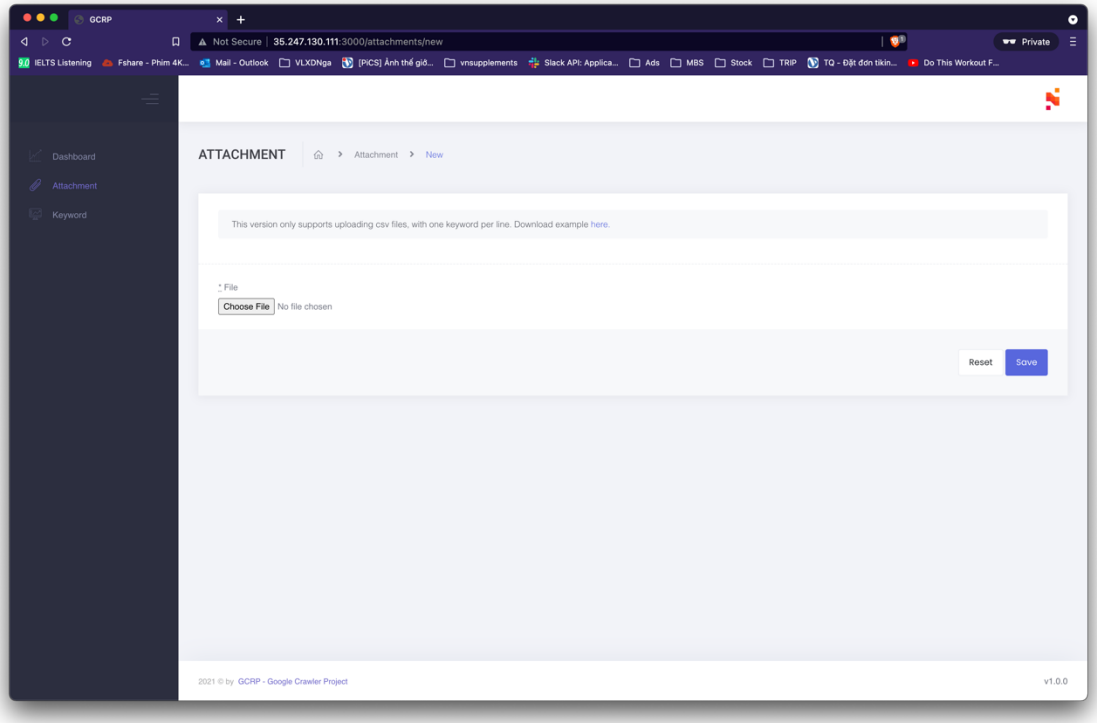
3. Dashboard



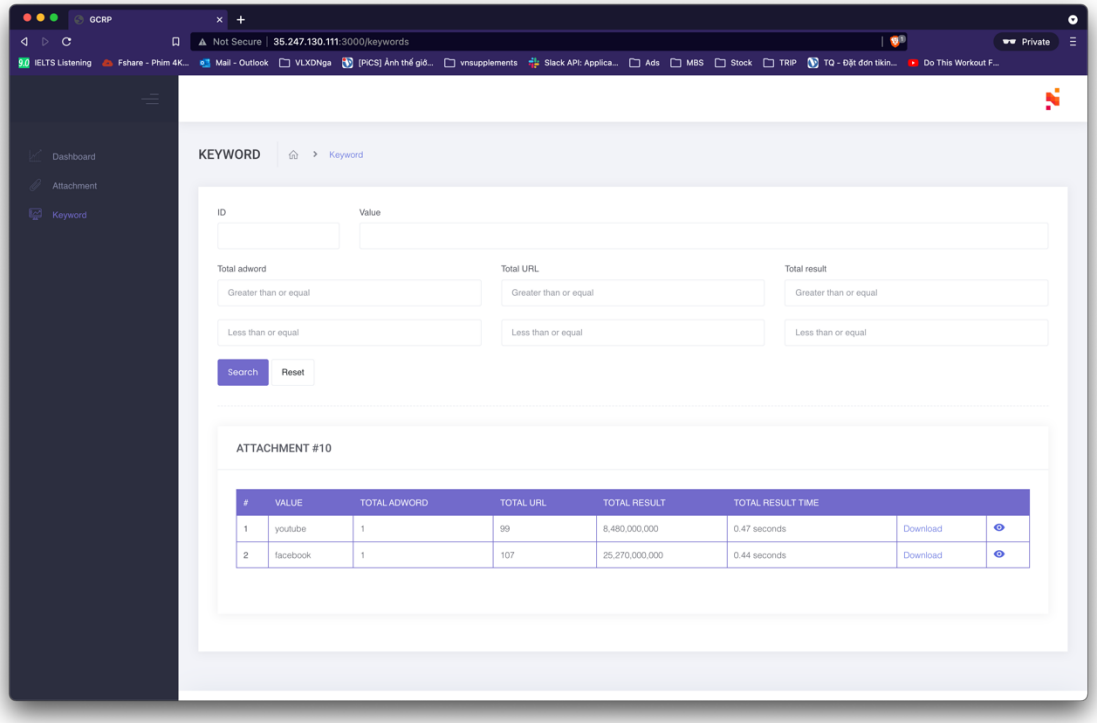
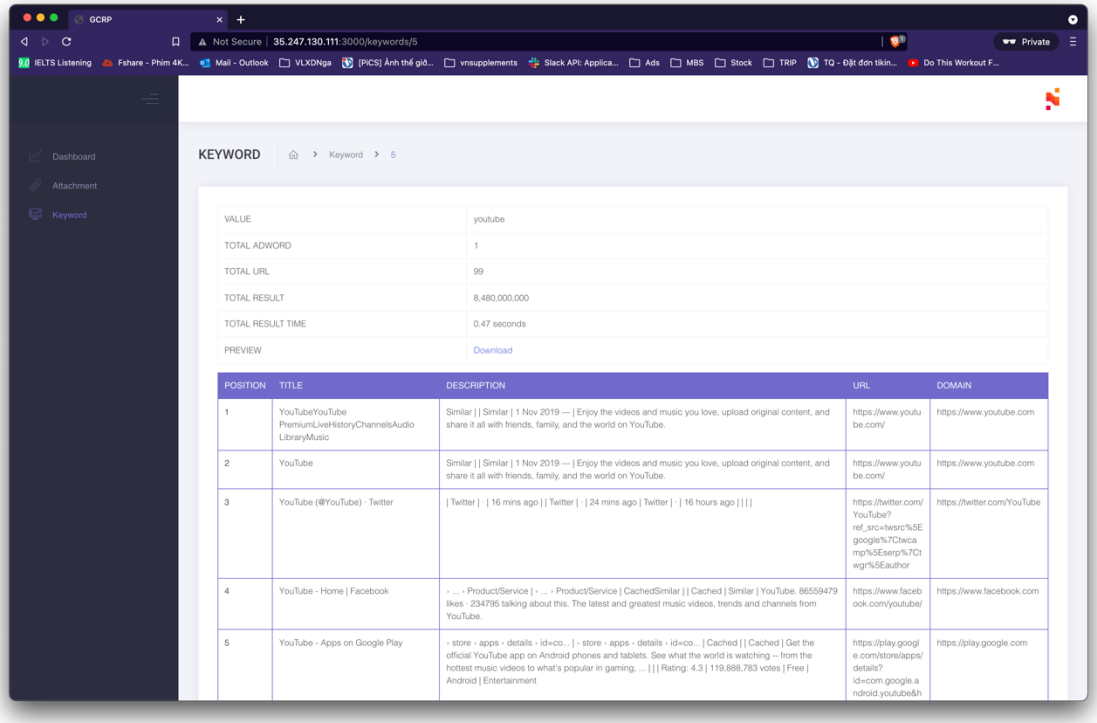
4. List attachments



5. Upload an attachment



6. List keywords



7. Keyword details