

news-analyzer-TianweiCao

****I did not successfully upload it to aws, so this website can only use locally(but all function has been tested on my pc****

steps to test:

0 make sure you have implemented the following modules:

pip install beautifulsoup4

pip install lxml

pip install html5lib

also need these modules:

PyPDF2

pandas

1 please download all files in the dictory and place them in correct folders, how to do this is showed in 'code structure' part

2 open and run app.py

3 copy the link <http://127.0.0.1:8083/> to start

function implemented:

1 User register

2 Store user info in database

3 User login

4 User upload txt files

5 search files by keyword

6 search files by title

7 show the files on website

8 catch news from 'http://www.chinadaily.com.cn/world' saved them

9 search saved news by keyword

10 search saved news by filename, display them on screen

11 show emotion of content

user story

I, the investigator, need to login to a secure service to upload my content

I, the investigator, have a number of PDF files that I want to upload to my account.

I, the investigator, want to search in all files I have uploaded.

I, the investigator, want to discover content from the WEB to enhance story

I, the investigator, want to know find the common keywords in subset of my documents.

I, the investigator, want to know all names, locations, institutions and address in my documents.

I, the investigator, want to know any relation between names, locations and institutions.

I, the investigator, want to find relationship between keywords in the ingested and discovered content.

I, the investigator, want to find sentiment (positive, negative or neutral) of different parts of

the content

I, the investigator, want to modify the sentiment calculated by the system in case I disagree.

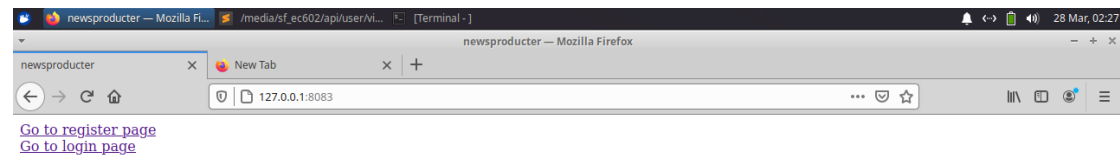
I, the investigator, want to search based on keywords and sentiments

I, the investigator, want the system to calculate similarities or correlation between key wor

Provide snapshots and summary of documents

I, the investigator, want to upload different types of files (CSV, DOC, etc.)

DISPLAY THE WEBSITE:

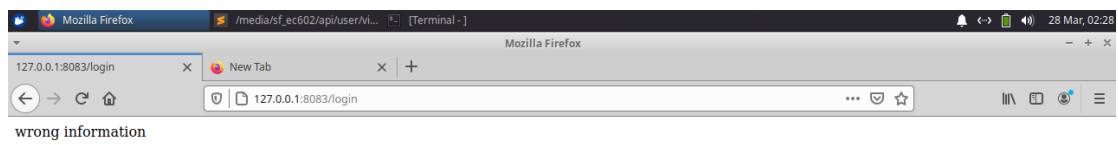


regist and login

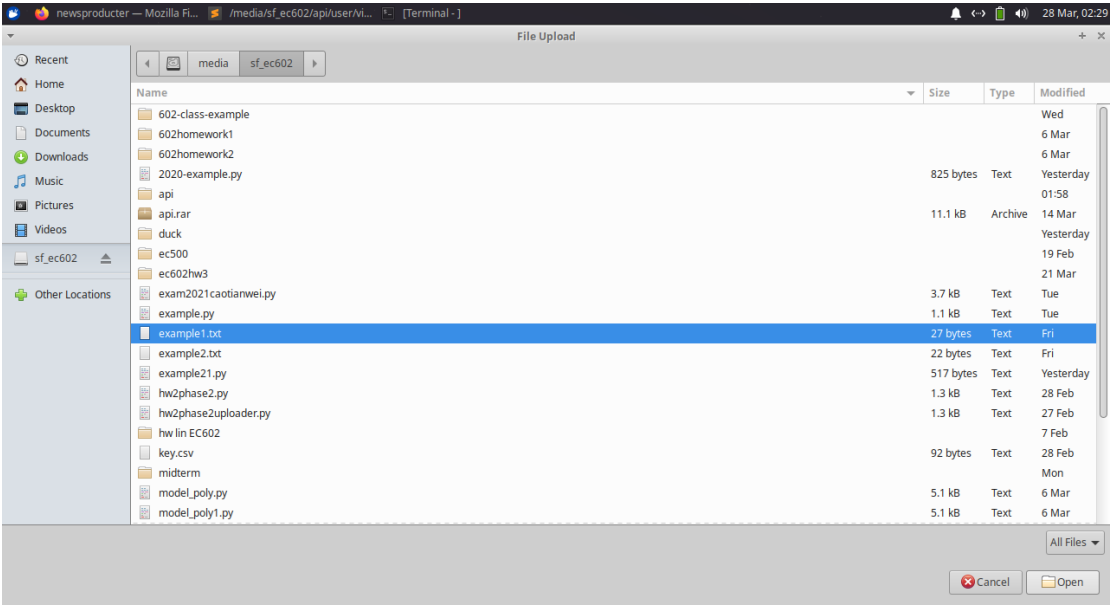
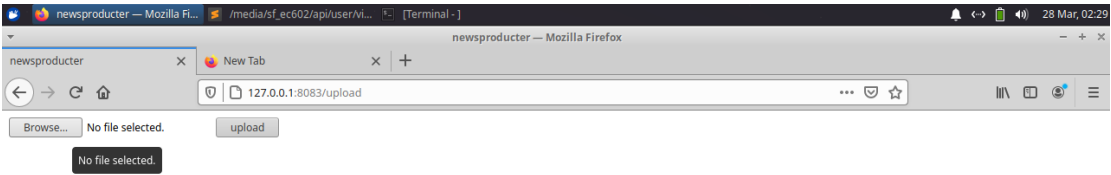
A screenshot of the registration form on the newsproducer website. The form is displayed in a web browser window with the address bar showing '127.0.0.1:8083/regist'. The form contains the following fields and elements:

- User name:** A text input field with the value '1'.
- Password:** A text input field with the value '1'.
- Repassword:** A text input field with the value '1'.
- Submit:** A button labeled 'Submit'.

Below the form, there is a text label: 'Enter your information to regist'.

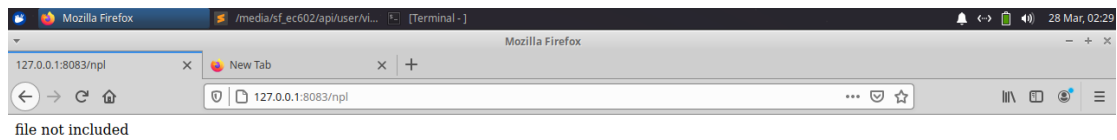


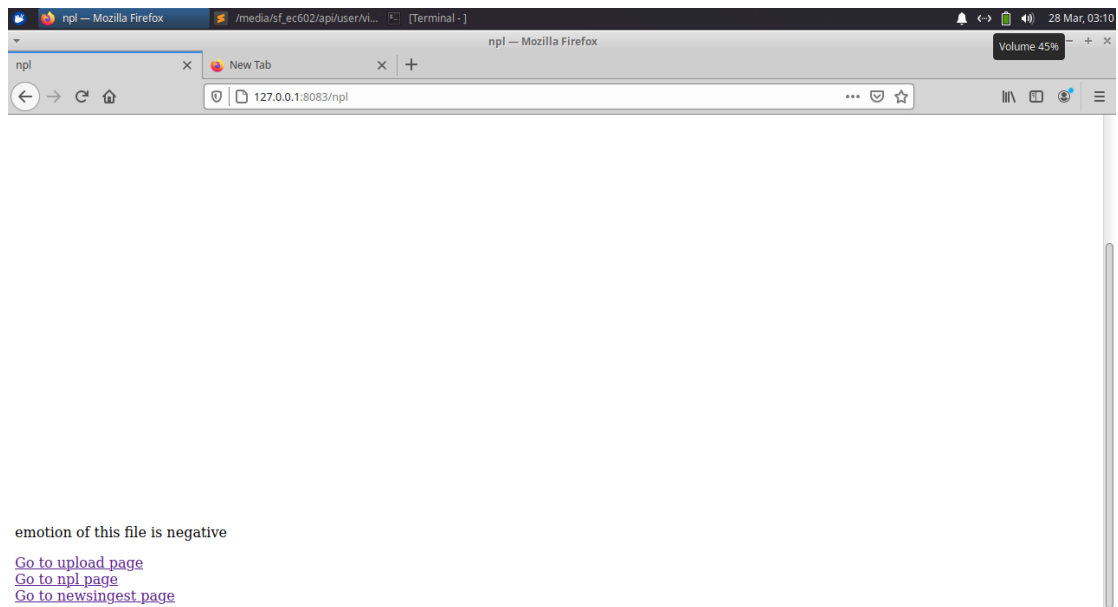
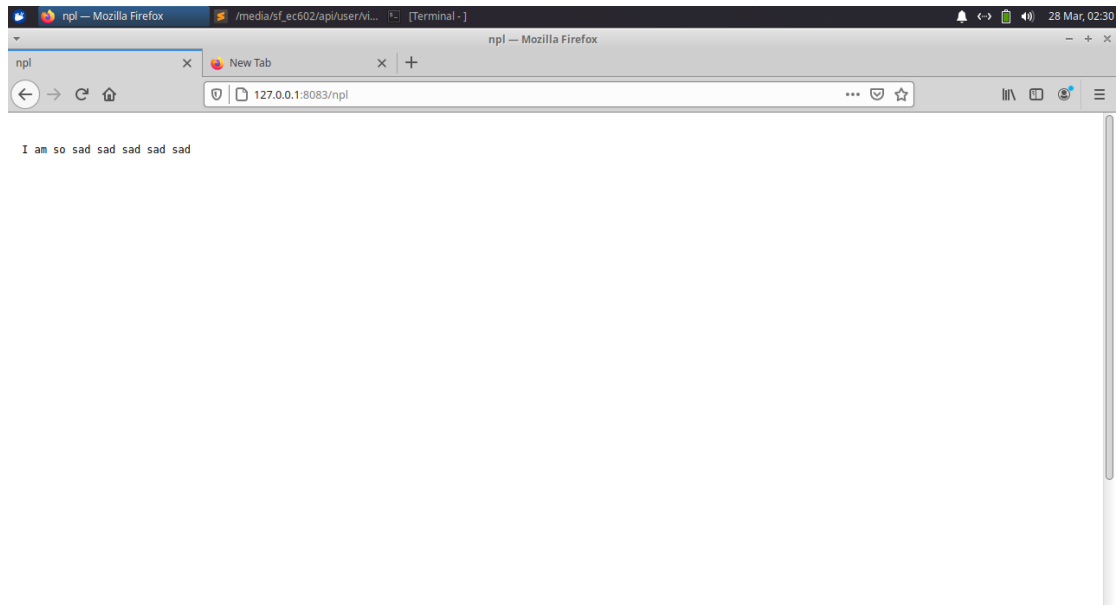
Upload



Npl

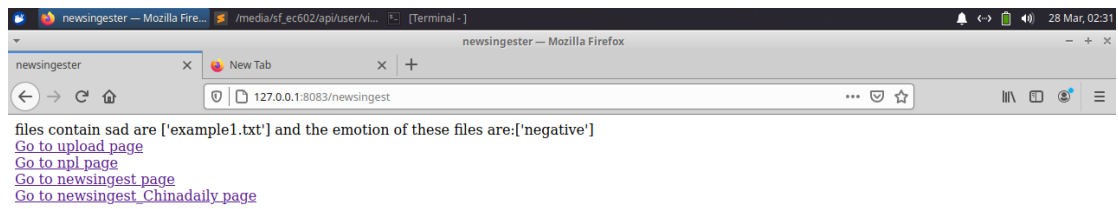
:





Newsingest

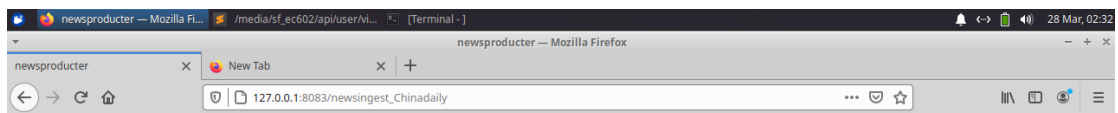
:



127.0.0.1:8083/newsingest_Chinadaily

Chinadailynewsingest

:

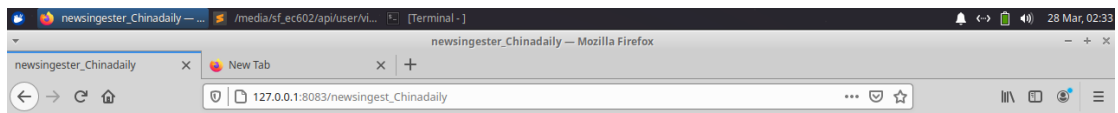


File keyword:

China

Submit

Enter File keyword



files contain China are ["Nation's vaccine aid is turning tide against pandemic.txt", 'US urged to enhance ties with China.txt'] and the emotion of these files are: ['positive', 'positive']

[Go to upload page](#)

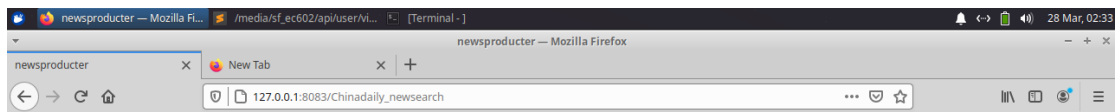
[Go to npl page](#)

[Go to newsingest page](#)

[Go to newsingest_Chinadaily page](#)

[Go to newsearch_Chinadaily page](#)

127.0.0.1:8083/Chinadaily_newsearch

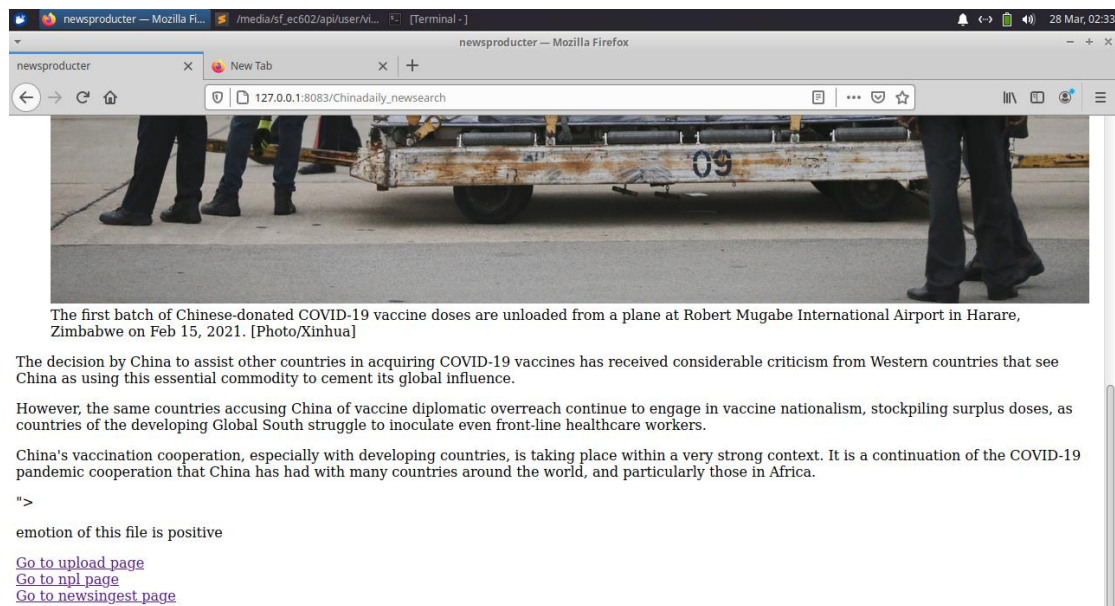


File name:

ring tide against pandemic.txt

Submit

Enter File name



code sturcture:

the right structure should be like this however I have not yet figured out how to build folders on github, so I just put them all together sry for that.

| - templates

| - homepage.html

| - regist.html

| - login.html

| - upload.html

| - newsingest.html

| - npl.html

| - newsongest_Chinadaily.html

| - Chinadaily_newssearch.html

| - static

| - example1.txt

| - example2.txt

| - __init__.py

| - setting.py

| - app.py

| - user

| - __init__.py (leave empty)

| - __view__.py

| - news_ingest

| -

templates

this is a folder for all html files.

- 1.homepage.html construct the homepage of my website which works as a map that guide users to all other pages(login, regist...)
- 2.regist.html construct the register page, where you can create user and userinformation will be saved(now is in a list and saved as dictionaries)
- 3.login.html construct the loginpage where you input you information and I will check if your information is correct.
- 4.upload.html after login you can go to other pages like upload, in upload page you can upload your files which would be saved in static folder
- 5.npl.html enable user to search the file they upload according to filename and display the file and emotion on website
- 6.newsingest.html enable users to search news they upload according to key word, return all files contain the keyword and its emotion
- 7.newsingest_Chinadaily.html catch all news on 'http://www.chinadaily.com.cn/world', user can search news according to keyword and the website return name of files contain the keyword, also shows its emotion
- 8.Chinadaily_newssearch.html enable user to search news caught on 'http://www.chinadaily.com.cn/world', display the first 10 lines of news(include images) and the emotion of the news

static

this is a folder to save files that I upload.

to test its function, please place example1.txt and example2.txt in it

app.py, __init.py__,setting.py

this three programs are bond with one another,

app.py is the core project, all other modules are connected to it

__init__.py is the init project for app.py(all python modules have a init.py) it cannot be empty

setting.py is the setting information for app.py, we can turn up DEBUG module or do other setting stuffs here

user

this is a python module contains __init.py__ and view.py

__init__.py is allowed to be empty

view.py is connected to app.py using blueprint, it deal with upload, regist login and all work related to user module

I should have devided view.py into several modules, but dont have enough time to do so, sorry for that