

Design of Programming Assignment 3

Tech stack used:

flask, mongodb compass and html are used to create this web application.

Using flask_pymongo I was able to connect python and mongodb. At the time of establishing connection I have created a new database called 'webAPIs'

```
(app.config["MONGO_URI"] = "mongodb://localhost:27017/webAPIs")
```

index.html is design of webpage where the user can insert filters. User can choose to run 1 of 4 queries.

home.html bears the results (all the fields of documents returned except '*_id*') of the search provided

Parsing the data:

Given 2 text files, *api.txt* and *mashup.txt*, I have parsed the data with given delimiters as dictionaries and appended them to a list (implemented in *parse.py*). Created field names in the back end according to the instructions provided in the manual for both the collections, *apiData* and *mashupData* collections

These lists have been imported to *main.py* where they are being loaded into the database into respective collections

Running Application:

I have created 4 different functions for each query

For queries 1 and 2:

I have used pipeline where I am checking for the given inputs

- a. *year*: Since we need to check for the year alone, I left the field 'updated', which is in string format as is, and performed string match using regex for 'year'
- b. *protocol/ category*: Matching the given inputs with the aforementioned fields in the database
- c. *tag*: all the tags are saved as list and the given input is checked if the tag is present or not in every document of that collection (tags are case sensitive)
- d. *rating*: Higher precedence is given to 'rating equal to' over greater than and less than, if the former is inputted along with the later ones, app only checks for rating equal to field. This was enforced on back end.

If the rating equal to is not provided only then, the other 2 fields are considered. All the queries are added to a pipeline which are then run through the database

Query 2 also was executed similarly against *mashupData* collection

While returning all the fields, I have omitted '*_id*' and rest of the fields that are part of original document are returned on *home.html* page

For queries 3 and 4:

Using regex, all the keywords (keywords are *case insensitive*) are checked one after another in a for loop if they are present in one of the 3 fields at least (using '*\$or*'), and then included in the pipeline.

Once all the keywords are checked, the pipeline is run through the database and the results(all the fields of the document except '*_id*') are fetched and returned on *home.html* page.

Screenshots:

Database

This screenshot shows the MongoDB Compass interface for the 'webAPIs.apiData' collection. The left sidebar displays the database structure, including the 'webAPIs' database and the 'apiData' collection. The main panel shows the 'Documents' tab with a list of documents. The first document is expanded, showing its JSON structure. The document contains fields such as '_id', 'id', 'title', 'summary', 'rating', 'name', 'label', 'author', 'description', 'type', 'download', 'useCount', 'sampleUrl', 'dateModified', 'remoteFeed', 'numComments', 'commentsUrl', 'tags', 'category', 'protocol', 'serviceEndpoint', 'version', 'wadi', and 'data format'.

```
{
  "_id": "ObjectID('424792ff219db0e475d9e41c')",
  "id": "http://www.programmableweb.com/api/blue",
  "title": "#blue",
  "summary": "Text messaging storage service",
  "rating": 4.5,
  "name": "#blue",
  "label": "#blue",
  "author": null,
  "description": "#blue is a service that stores text messages in a conversational style.",
  "type": 1,
  "download": null,
  "useCount": null,
  "sampleUrl": "https://api.hashblue.com/doc/home",
  "dateModified": null,
  "remoteFeed": null,
  "numComments": null,
  "commentsUrl": "http://api.programmableweb.com/api/blue/comments",
  "tags": Array,
  "category": "Messaging",
  "protocol": "REST",
  "serviceEndpoint": null,
  "version": null,
  "wadi": null,
  "data format": "JSON"
}
```

This screenshot shows the MongoDB Compass interface for the 'webAPIs.mashupData' collection. The left sidebar displays the database structure, including the 'webAPIs' database and the 'mashupData' collection. The main panel shows the 'Documents' tab with a list of documents. The first document is expanded, showing its JSON structure. The document contains fields such as '_id', 'id', 'title', 'summary', 'rating', 'name', 'label', 'author', 'description', 'type', 'download', 'useCount', 'sampleUrl', 'dateModified', 'numComments', 'commentsUrl', 'tags', 'category', 'protocol', 'serviceEndpoint', 'version', 'wadi', and 'data format'.

```
{
  "_id": "ObjectID('424792ff219db0e475d9e41c')",
  "id": "http://www.programmableweb.com/mashup/-22",
  "title": null,
  "summary": null,
  "rating": 3.5,
  "name": null,
  "label": null,
  "author": "Unknown",
  "description": null,
  "type": null,
  "download": 0,
  "useCount": 0,
  "sampleUrl": "http://www.easypeasyphotos.net",
  "dateModified": "2011-10-09T14:35:04Z",
  "numComments": 0,
  "commentsUrl": "https://api.programmableweb.com/mashups/-22/comments",
  "tags": Array,
  "category": null,
  "protocol": null,
  "serviceEndpoint": null,
  "version": null,
  "wadi": null,
  "data format": null
}
```

Query 1:

← → ↺ 127.0.0.1:5000

FORM FOR API DATA

Enter updated year

Enter protocols

Enter category

Enter rating greater than Enter rating lesser than

OR

Enter rating equal to

Enter tags

Get Results

FORM FOR MASHUP DATA

Enter updated year

Enter used APIs

Enter tags

Get Results

KEYWORD FOR API DATA

Enter keywords for api data

Get Results

KEYWORD FOR MASHUP DATA

Enter keywords for mashup data

Get Results

← API details Programm... 500 Internal... How to... Programm... Programm... API details mong... API details How to... Python API details API details Programm... API details API details API details API details API details X

← → ↺ 127.0.0.1:5000

API details

id

<http://www.programmableweb.com/api/the-global-proteome-machine>

title

The Global Proteome Machine

summary

Proteome data for biomedical research

rating

4.4

name

The Global Proteome Machine

label

The Global Proteome Machine

author

None

description

The Global Proteome Machine is an attempt to create knowledge from proteomics data and reuse it to solve biomedical research problems. The Global Proteome Machine Database was built to use GPM data to help validate peptide MS/MS spectra and protein coverage patterns. The Global Proteome Machine Database API provides RESTful access to commonly required information based on data from the GPM Database. Responses are JSON formatted.

type

1

downloads

None

useCount

None

sampleUrl

http://wiki.thegpm.org/wiki/GPMDB_REST

downloadUrl

None

dateModified

2012-12-17T09:51:40Z

remoteFeed

None

numComments

None

commentsUrl

<http://api.programmableweb.com/apis/the-global-proteome-machine/comments>

Tags

['database', 'science']

category

Science

protocols

REST

serviceEndpoint

<http://gpmdb.thegpm.org/>

version

None

wSDL

None

data formats

JSON

apigroups

1

Query 2:

127.0.0.1:5000

FORM FOR API DATA

Enter updated year

Enter protocols

Enter category

Enter rating greater than Enter rating lesser than

OR

Enter rating equal to

Enter tags

Get Results

FORM FOR MASHUP DATA

Enter updated year

Enter used APIs

Enter tags

Get Results

KEYWORD FOR API DATA

Enter keywords for api data

Get Results

KEYWORD FOR MASHUP DATA

Enter keywords for mashup data

Get Results

API details

id

http://www.programmableweb.com/mashups/api-christmas

title

#API Christmas

summary

#API Christmas is the Nativity Story, remixed through digital media.

rating

4.7

name

#API Christmas

label

#API Christmas

author

Unknown

description

#API Christmas is the Nativity Story, remixed through digital media.

type

None

downloads

0

useCount

0

sampleUrl

http://api.christmas.com/

dateModified

2012-12-25T18:34:23Z

numComments

0

commentsUrl

http://api.programmableweb.com/mashups/api-christmas/comments

tags

['aggregate', 'christmas', 'fun', 'photo', 'social']

APIs

['instagram', 'http://www.programmableweb.com/api/instagram']

updated

2012-12-25T18:34:23Z

id

http://www.programmableweb.com/mashups/docs-pound-docs

title

#docs ("pound docs")

summary

Let's a business share docs from their Microsoft SkyDrive over SMS via Twilio. Go to the app which is running on Windows Azure, set which SkyDrive account to use, and select a public file to share over SMS. Customers (incl. potential customers) can call the Twilio number associated with the #docs deployment and text "link" to get a link to whatever doc you"re currently sharing (for deals, specials, marketing info, etc.) For the sample #docs deployment, call my Twilio number (415)599-2671, enter passcode "7392-3266", and message "link".

rating

3.0

name

#docs ("pound docs")

label

#docs

Query 3:

API detailsProgram: 500 InternalHow to: Program: Program: API detailsmongoAPI detailsHow to: PythonAPI detailsAPI detailsProgram: API detailsAPI detailsAPI detailsAPI detailsProgram: X

127.0.0.1:5000

0

FORM FOR API DATA

Enter updated year

Enter protocols

Enter category

Enter rating greater than

Enter rating lesser than

OR

Enter rating equal to

Enter tags

Get Results

FORM FOR MASHUP DATA

Enter updated year

Enter used APIs

Enter tags

Get Results

KEYWORD FOR API DATA

Enter keywords for api data

Global attempt research

Get Results

KEYWORD FOR MASHUP DATA

Enter keywords for mashup data

countries cities

Get Results

API detailsProgram: 500 InternalHow to: Program: Program: API detailsmongoAPI detailsHow to: PythonAPI detailsAPI detailsProgram: API detailsAPI detailsAPI detailsAPI detailsAPI detailsAPI detailsProgram: X

127.0.0.1:5000/pg

API details

id

<http://www.programmableweb.com/apis/the-global-proteome-machine>

title

The Global Proteome Machine

summary

Proteome data for biomedical research

rating

4.4

name

The Global Proteome Machine

label

The Global Proteome Machine

author

None

description

The Global Proteome Machine is an attempt to create knowledge from proteomics data and reuse it to solve biomedical research problems. The Global Proteome Machine Database was built to use GPM data to help validate peptide MS/MS spectra and protein coverage patterns. The Global Proteome Machine Database API provides RESTful access to commonly required information based on data from the GPM Database. Responses are JSON formatted.

type

1

downloads

None

useCount

None

sampleUrl

http://wiki.thegpm.org/wiki/GPMDB_REST

downloadUrl

None

dataModified

2012-12-17T09:51:40Z

remoteFeed

None

numComments

None

commentsUrl

<http://api.programmableweb.com/apis/the-global-proteome-machine/comments>

Tags

['database', 'science']

category

Science

protocols

REST

serviceEndpoint

<http://gpmdb.thegpm.org/>

version

None

wsdl

None

data formats

JSON

apigroups

Query 4:

FORM FOR API DATA

Enter updated year

Enter protocols

Enter category

Enter rating greater than Enter rating lesser than

OR

Enter rating equal to

Enter tags

FORM FOR MASHUP DATA

Enter updated year

Enter used APIs

Enter tags

KEYWORD FOR API DATA

Enter keywords for api data

KEYWORD FOR MASHUP DATA

Enter keywords for mashup data

0

API details	
id <i>http://www.programmableweb.com/mashup/ins-stats-on-your-linkedin-social-graph</i>	
title <i>#LinkedIn Stats on Your LinkedIn Social Graph</i>	
summary <i>Get some shiny stats about countries, cities and industries your contacts are from.</i>	
rating 4.3	
name <i>#LinkedIn Stats on Your LinkedIn Social Graph</i>	
label <i>#LinkedIn Stats on Your LinkedIn Social Graph</i>	
author <i>Unknown</i>	
description <i>Get some shiny stats about countries, cities and industries your contacts are from.</i>	
type <i>None</i>	
downloads 0	
useCount 4260	
sampleUrl <i>http://roelandp.nl/devdev/linkedin/</i>	
dateModified <i>2010-02-04T00:35:04Z</i>	
numComments 0	
commentsUrl <i>http://api.programmableweb.com/mashup/ins-stats-on-your-linkedin-social-graph/comments</i>	
tags <i>['enterprise', 'social', 'socialgraph', 'statistics']</i>	
APIs <i>['Google Chart\$\$\$http://www.programmableweb.com/api/google-chart', 'LinkedIn\$\$\$http://www.programmableweb.com/api/linkedin']</i>	
updated <i>2010-02-04T00:35:04Z</i>	

id <i>http://www.programmableweb.com/mashup/learn-the-world</i>	
title <i>Learn the World</i>	
summary <i>View the map and find information about countries, states, and cities with one mouse click.</i>	
rating 5.0	
name <i>Learn the World</i>	
label <i>Learn the World</i>	
author <i>Learn the World</i>	