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

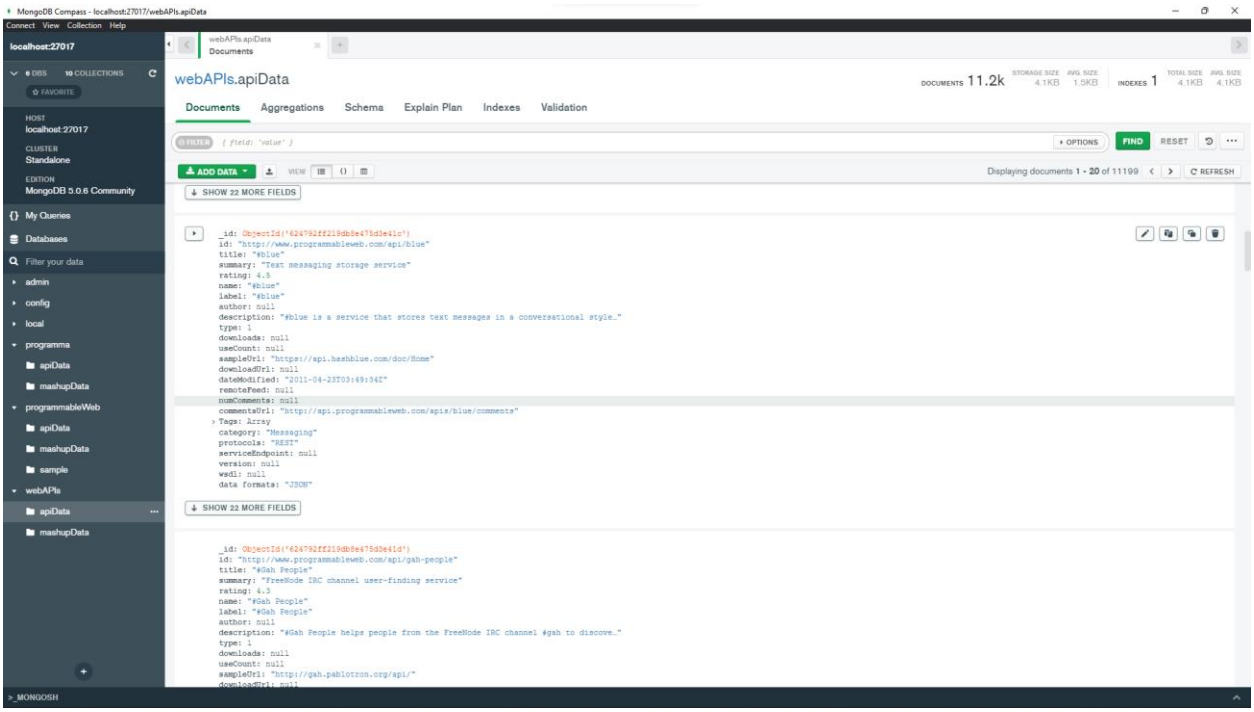
For queries 3 and 4:

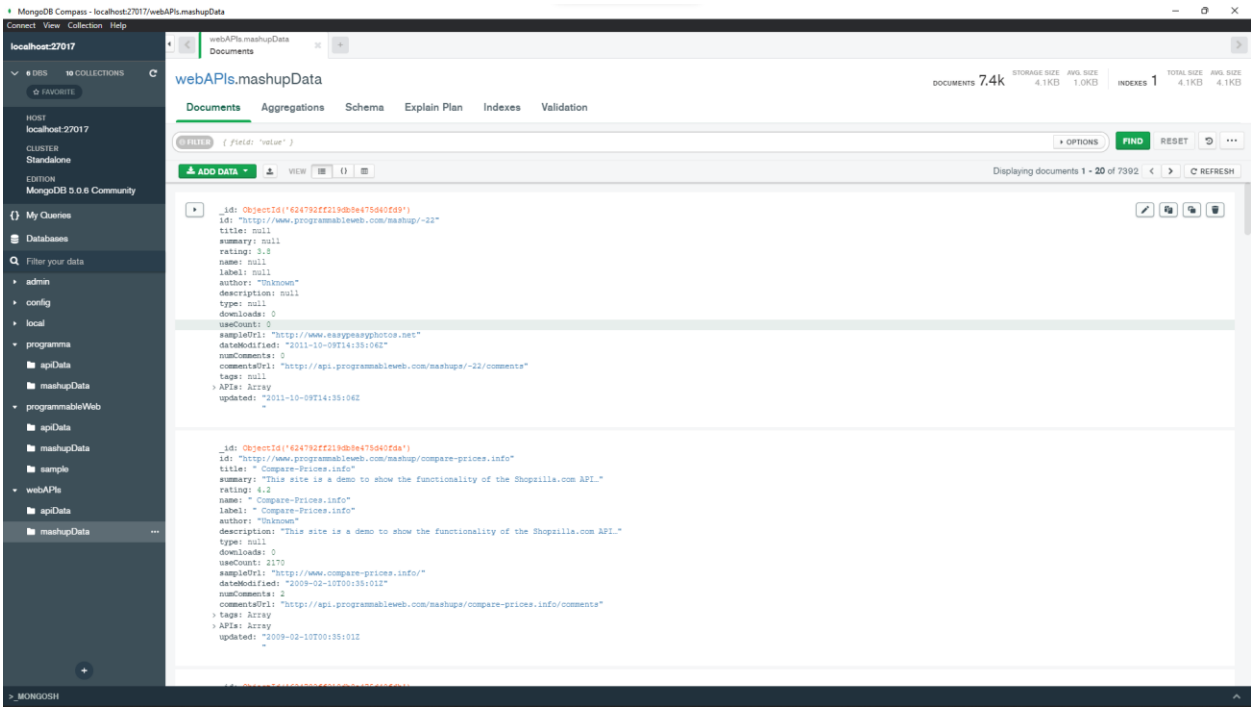
Using regex, all the keywords (keywords are *case insensitive*) are checked one after another in a for loop if they are present in one of the 3 fields at least (using '*$or*'), and then included in the pipeline.

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:

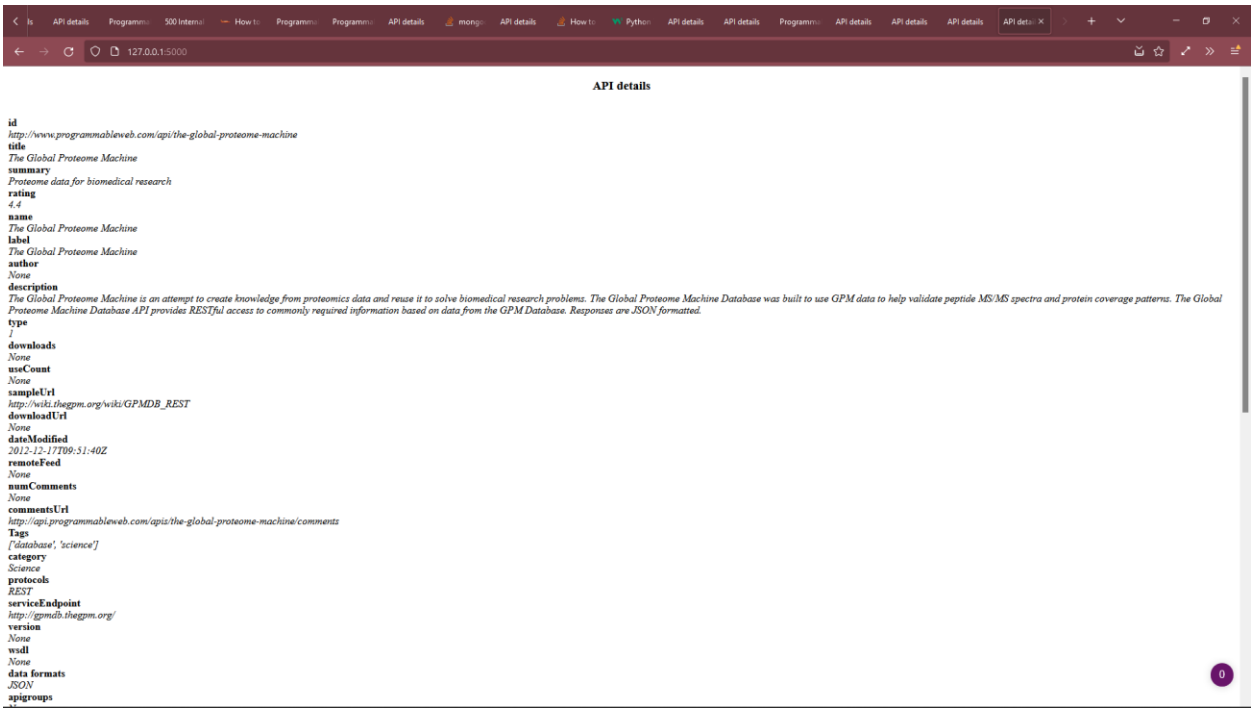
Database





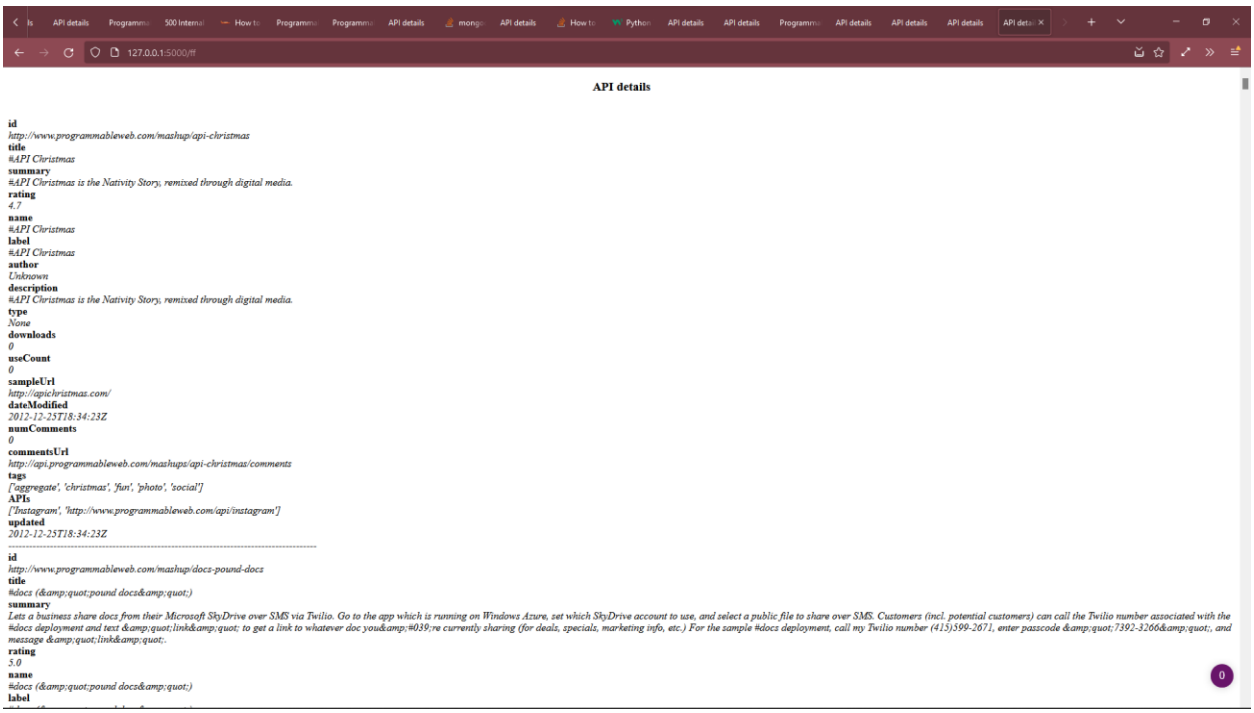
Query 1:


127.0.0.1:5000
FORM FOR API DATA
Enter updated year 2012
Enter protocols REST
Enter category Science
Enter rating greater than 2
Enter rating lesser than 3
OR
Enter rating equal to 4.4
Enter tags database, science
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

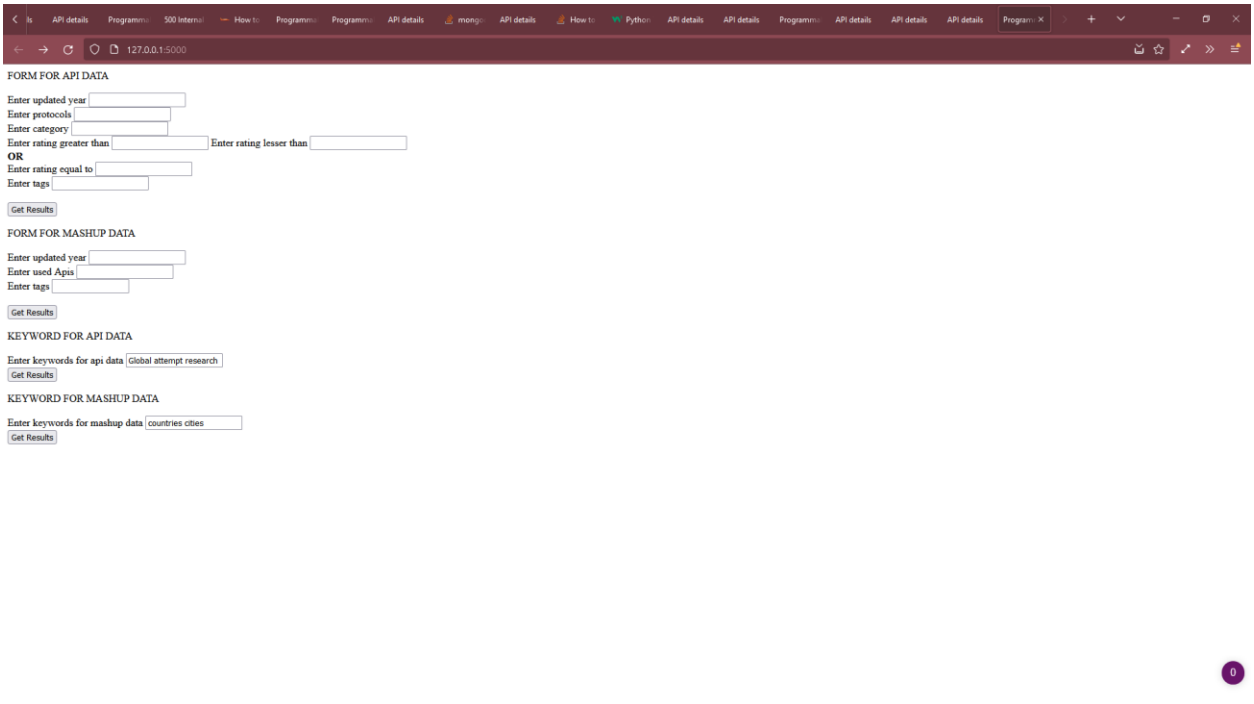

127.0.0.1:5000
API details
id
http://www.programmableweb.com/api/the-global-proteome-machine
title
The Global Proteome Machine
summary
Proteome data for biomedical research
rating
4.4
name
The Global Proteome Machine
label
The Global Proteome Machine
author
None
description
The Global Proteome Machine 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

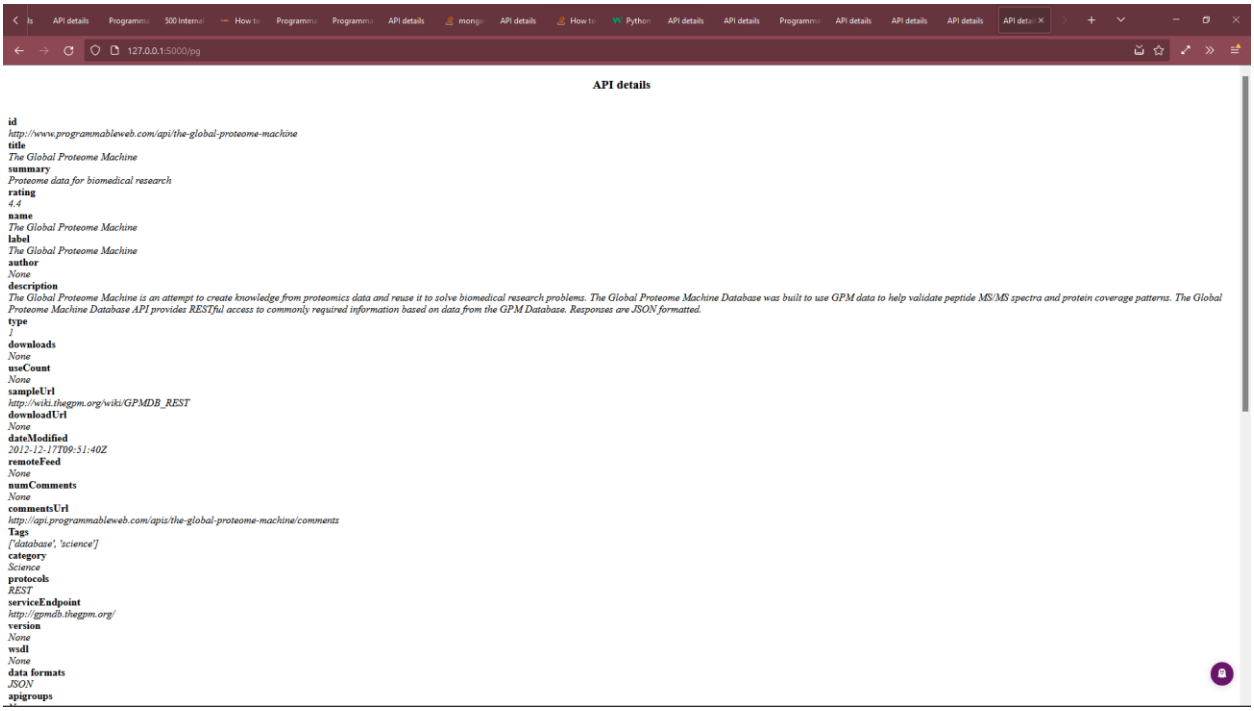
Query 2:


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 2012
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

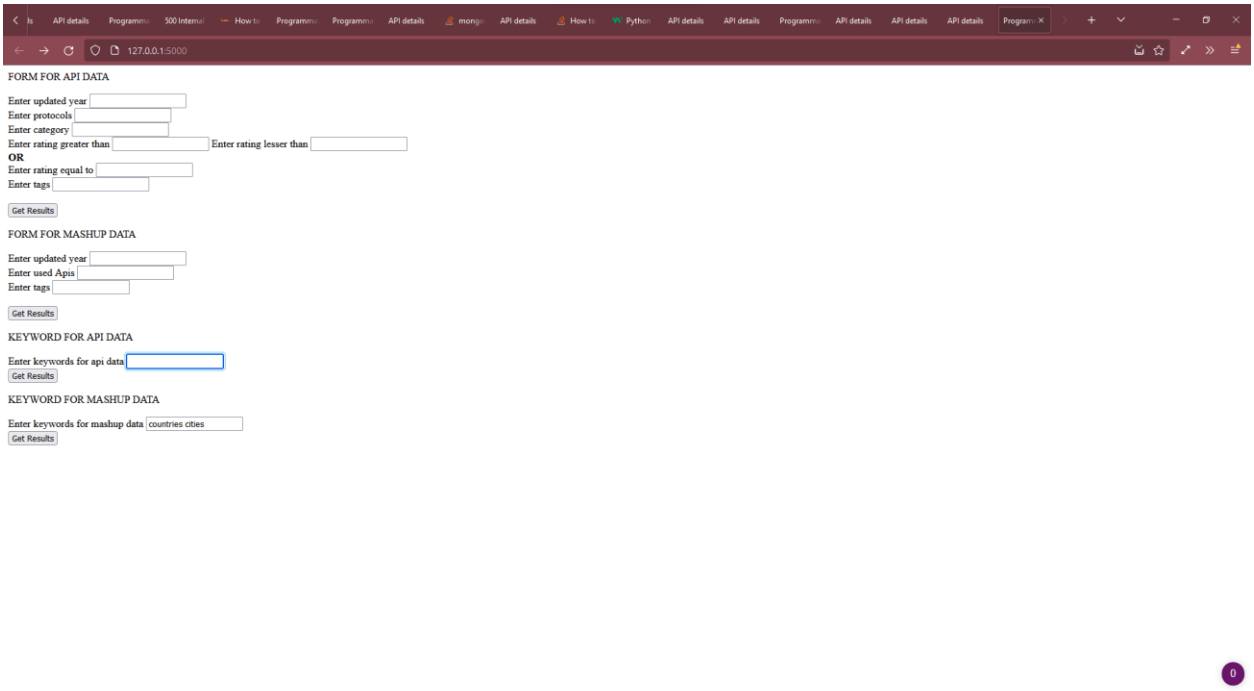

127.0.0.1:5000/ff
API details
id
http://www.programmableweb.com/mashup/api-christmas
title
#API Christmas
summary
#API Christmas is the Nativity Story, remixed through digital media.
rating
4.7
name
#API Christmas
label
#API Christmas
author
Unknown
description
#API Christmas is the Nativity Story, remixed through digital media.
type
None
downloads
0
useCount
0
sampleUrl
http://apichristmas.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/mashup/docs-pound-docs
title
#docs ("pound docs")
summary
Lets 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
5.0
name
#docs ("pound docs")
label

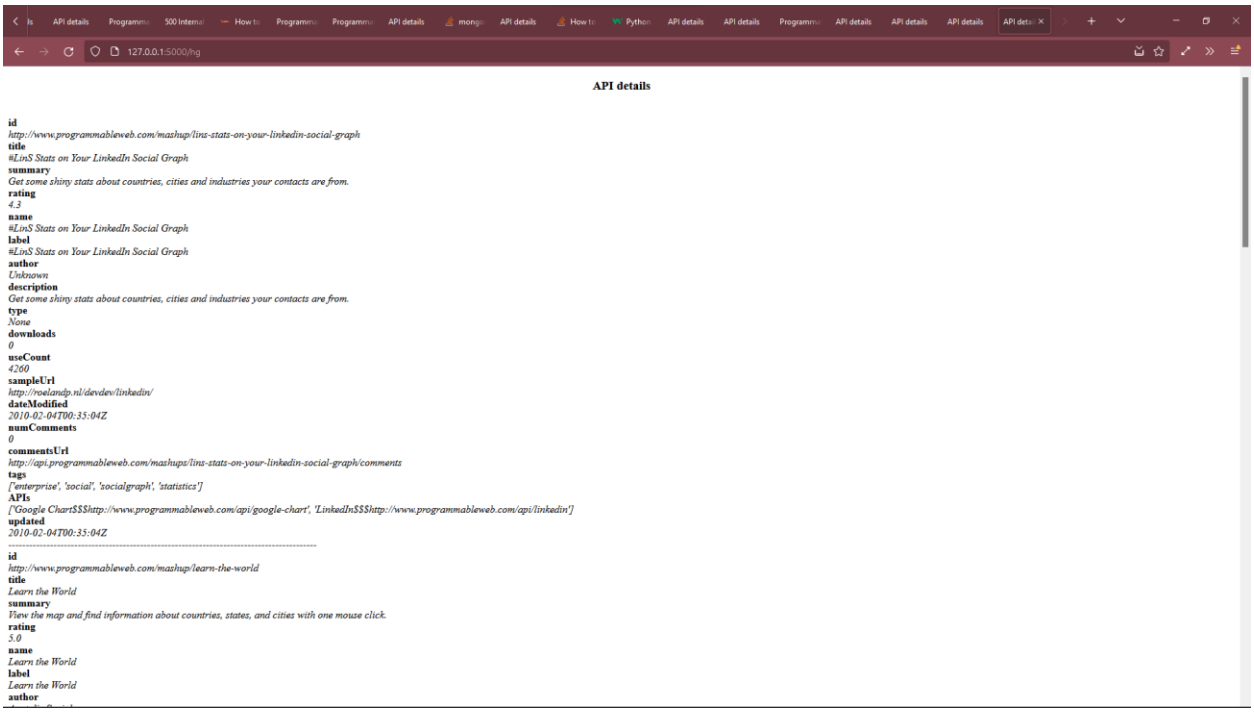
Query 3:


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
Global attempt research
Get Results
KEYWORD FOR MASHUP DATA
Enter keywords for mashup data
countries cities
Get Results


127.0.0.1:5000/pg
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

Query 4:


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
countries cities
Get Results


127.0.0.1:5000/hg
API details
id
http://www.programmableweb.com/mashup/lins-stats-on-your-linkedin-social-graph
title
#LinS Stats on Your LinkedIn Social Graph
summary
Get some shiny stats about countries, cities and industries your contacts are from.
rating
4.3
name
#LinS Stats on Your LinkedIn Social Graph
label
#LinS Stats on Your LinkedIn Social Graph
author
Unknown
description
Get some shiny stats about countries, cities and industries your contacts are from.
type
None
downloads
0
useCount
4260
sampleUrl
http://roelandp.nl/devdev/linkedin/
dateModified
2010-02-04T00:35:04Z
numComments
0
commentsUrl
http://api.programmableweb.com/mashups/lins-stats-on-your-linkedin-social-graph/comments
tags
['enterprise', 'social', 'socialgraph', 'statistics']
APIs
['Google Chart$$$http://www.programmableweb.com/api/google-chart', 'LinkedIn$$$http://www.programmableweb.com/api/linkedin']
updated
2010-02-04T00:35:04Z
id
http://www.programmableweb.com/mashup/learn-the-world
title
Learn the World
summary
View the map and find information about countries, states, and cities with one mouse click.
rating
5.0
name
Learn the World
label
Learn the World
author