PA 3 design document
CSCI.724.01 - Web Serv & Serv Oriented Comp

Dhruv Sharma, ds7042@g.rit.edu

1. A document that describes the design of the data structure and the query system. It should contain the screenshots of the data in the database and testing scenarios of the query system.

The data model is structured with the help of Pydantic models , (the application uses fastapi) We have the following models as specified in models.py:

APIModel: Represents a Web API with a comprehensive set of attributes such as title, summary, rating, name, and many more that detail its capabilities, usage, and metadata. This model is designed to store a wealth of information about each API, including its type, download URLs, modification dates, related tags and categories, protocols it uses, as well as various other properties related to its deployment, documentation, and usage constraints.

APIUsed: A simple model that represents an API used within a mashup, including the API's name and url. This model facilitates tracking which APIs are utilized by different mashups.

MashupModel: Represents a mashup, a composite application that uses data from one or more APIs. It shares several fields with APIModel, such as title, summary, description, and tags, which describe the mashup itself. Unique to mashups, the apis field is a list of APIUsed instances, indicating the APIs this mashup incorporates. Additional fields such as useCount and numComments might be used to store usage statistics and engagement metrics.

The query system gives the user the option to select the query type and then presents various forms based on the parameters that can be used for querying. We have the following 6 endpoints for handling each query, based on the user selection and params we use the fetch to access these endpoints and retrieve data.

The following is more information about the query endpoints :

1st Query: Search APIs by Criteria
This endpoint enables users to search for APIs based on various criteria including updated_year, protocols, category, min_rating, max_rating, and tags. We used regex, partial and case insensitive matching to get matches here.

Default size : 10000

2nd Query: Search Mashups by Criteria

This endpoint allows users to discover mashups based on updated_year, the names of used_apis, and tags. Implementation is similar to the first query.

Default size : 10000

3rd Query: Search APIs by Keywords
This query enables users to perform keyword searches across multiple fields (title, summary, description) of APIs. It's designed to ensure that all provided keywords must be present in at least one of the searchable fields for an API to be included in the results.

Default size : 1000

4th Query: Search Mashups by Keywords
Similar to the 3rd query but focused on mashups, this endpoint allows users to search for mashups using keywords that may appear in their title, summary, or description. It adopts a comprehensive search strategy to match mashups containing all the provided keywords in any of the specified fields

Default size : 1000

5th Query: Top Used APIs
This innovative endpoint identifies the top K most frequently used APIs in mashups, providing insights into popular or trending APIs within the mashup ecosystem. It achieves this by aggregating usage counts directly from the mashup documents, offering a dynamic view of the API landscape based on actual usage data. We use unwind in MongoDB to get the different apis used and then order them by frequency to get this result.

Default size : 10

6th Query: Top API-rich Mashups
Focusing on mashups that incorporate a large number of APIs, this endpoint ranks mashups based on the number of distinct APIs they utilize.

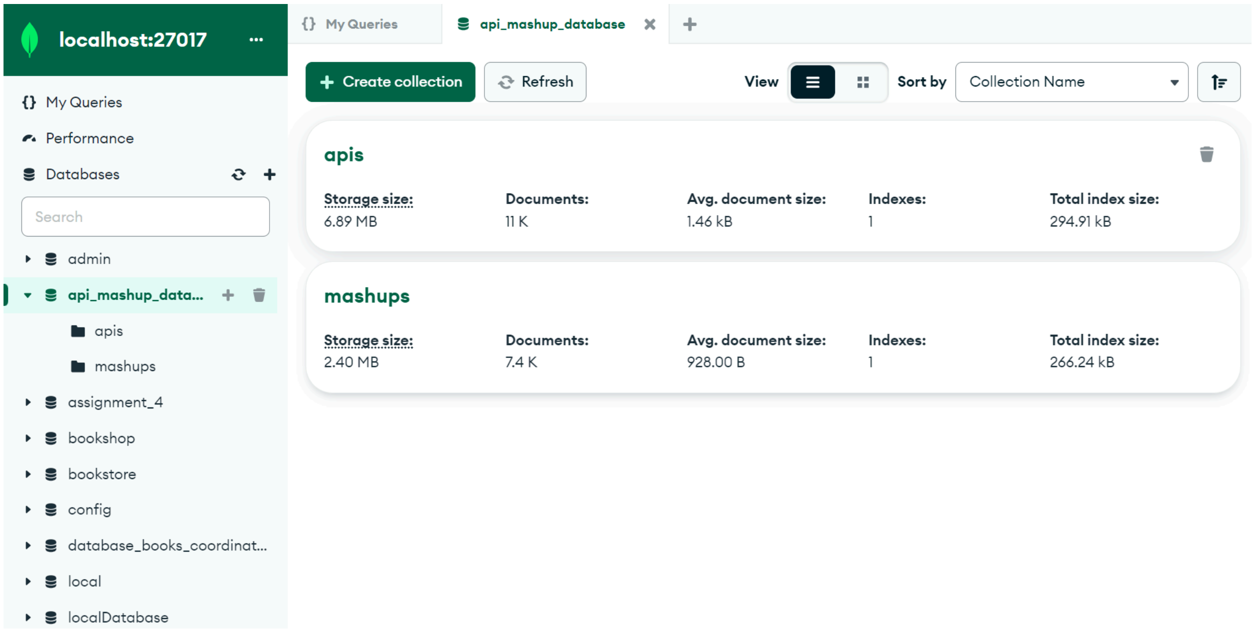
Default size : 10

The default size is the default number of records inspected so that there is a limit to the display. The last 2 queries have the option of K though (default if not specified is 10).

Screenshots :
For the parsing and the population we run the parsing.py that utilizes the other endpoints in our fastapi app for population.

Results:

Database



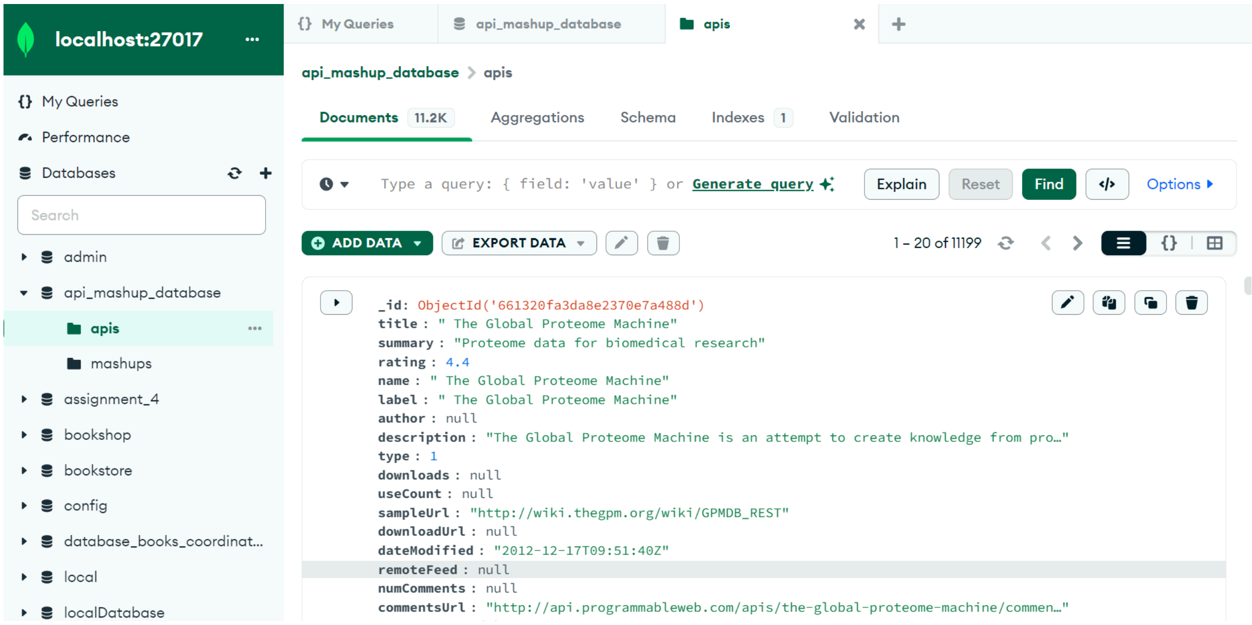
apis

Storage size:	Documents:	Avg. document size:	Indexes:	Total index size:
6.89 MB	11 K	1.46 kB	1	294.91 kB

mashups

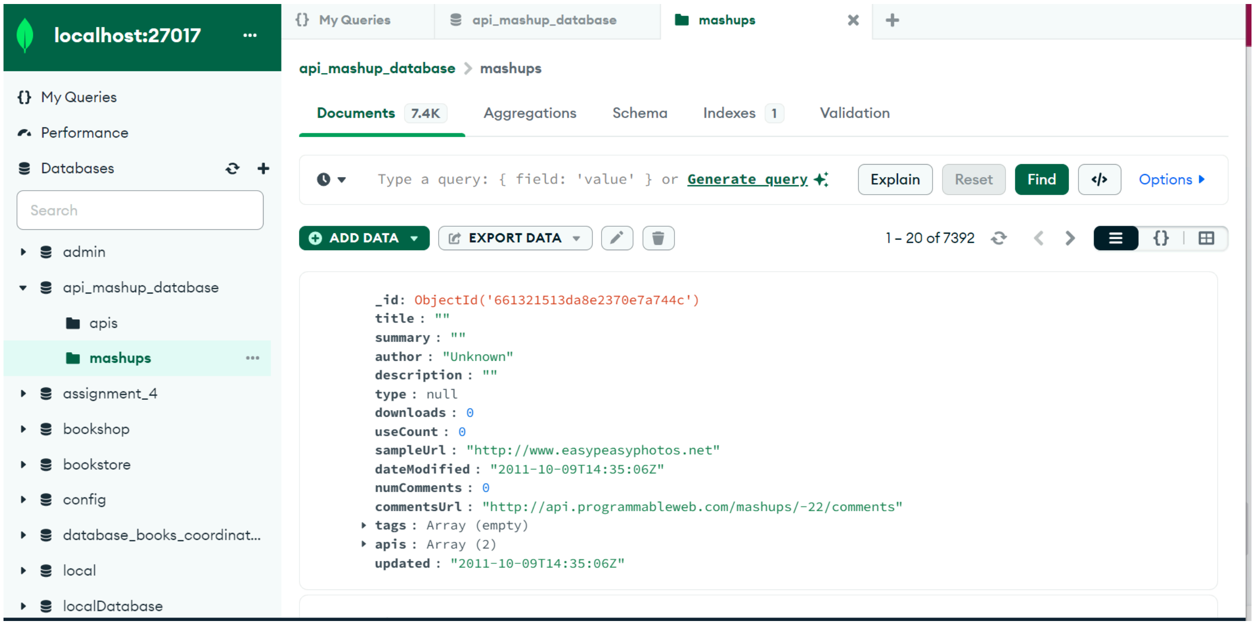
Storage size:	Documents:	Avg. document size:	Indexes:	Total index size:
2.40 MB	7.4 K	928.00 B	1	266.24 kB

Apis collection



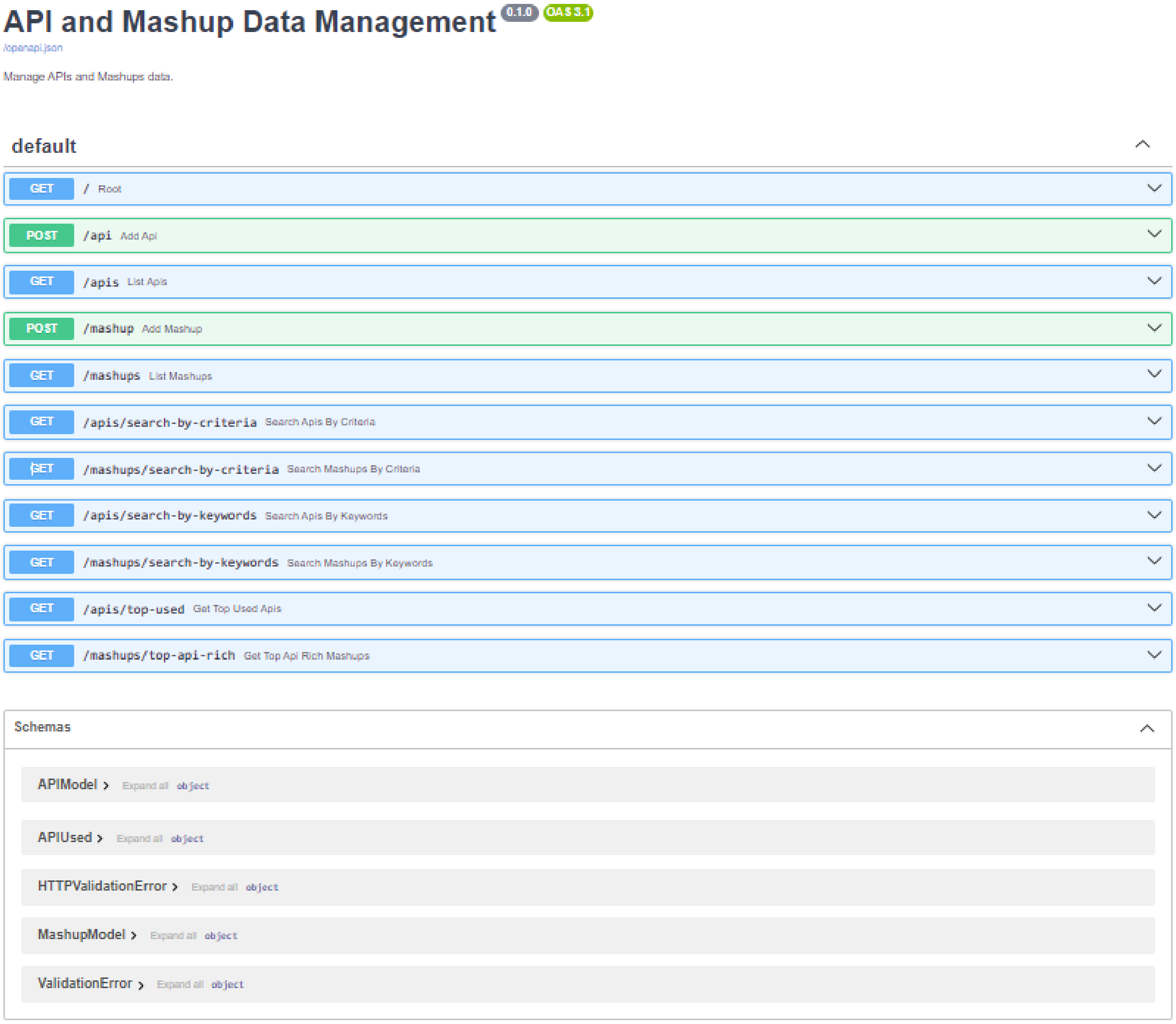
```
_id: ObjectId('661320fa3da8e2370e7a488d')
title : " The Global Proteome Machine"
summary : "Proteome data for biomedical research"
rating : 4.4
name : " The Global Proteome Machine"
label : " The Global Proteome Machine"
author : null
description : "The Global Proteome Machine is an attempt to create knowledge from pro…"
type : 1
downloads : null
useCount : null
sampleUrl : "http://wiki.thegpm.org/wiki/GPMDB_REST"
downloadUrl : null
dateModified : "2012-12-17T09:51:40Z"
remoteFeed : null
numComments : null
commentsUrl : "http://api.programmableweb.com/apis/the-global-proteome-machine/commen…"
```

Mashups collection



```
_id: ObjectId('661321513da8e2370e7a744c')
title : ""
summary : ""
author : "Unknown"
description : ""
type : null
downloads : 0
useCount : 0
sampleUrl : "http://www.easypeasyphotos.net"
dateModified : "2011-10-09T14:35:06Z"
numComments : 0
commentsUrl : "http://api.programmableweb.com/mashups/-22/comments"
tags : Array (empty)
apis : Array (2)
updated : "2011-10-09T14:35:06Z"
```

Api docs used for populating the database and handling queries



API and Mashup Data Management 0.1.0 OAS 3.1

/openapi.json

Manage APIs and Mashups data.

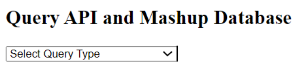
default

- GET / Root
- POST /api Add Api
- GET /apis List Apis
- POST /mashup Add Mashup
- GET /mashups List Mashups
- GET /apis/search-by-criteria Search Apis By Criteria
- GET /mashups/search-by-criteria Search Mashups By Criteria
- GET /apis/search-by-keywords Search Apis By Keywords
- GET /mashups/search-by-keywords Search Mashups By Keywords
- GET /apis/top-used Get Top Used Apis
- GET /mashups/top-api-rich Get Top Api Rich Mashups

Schemas

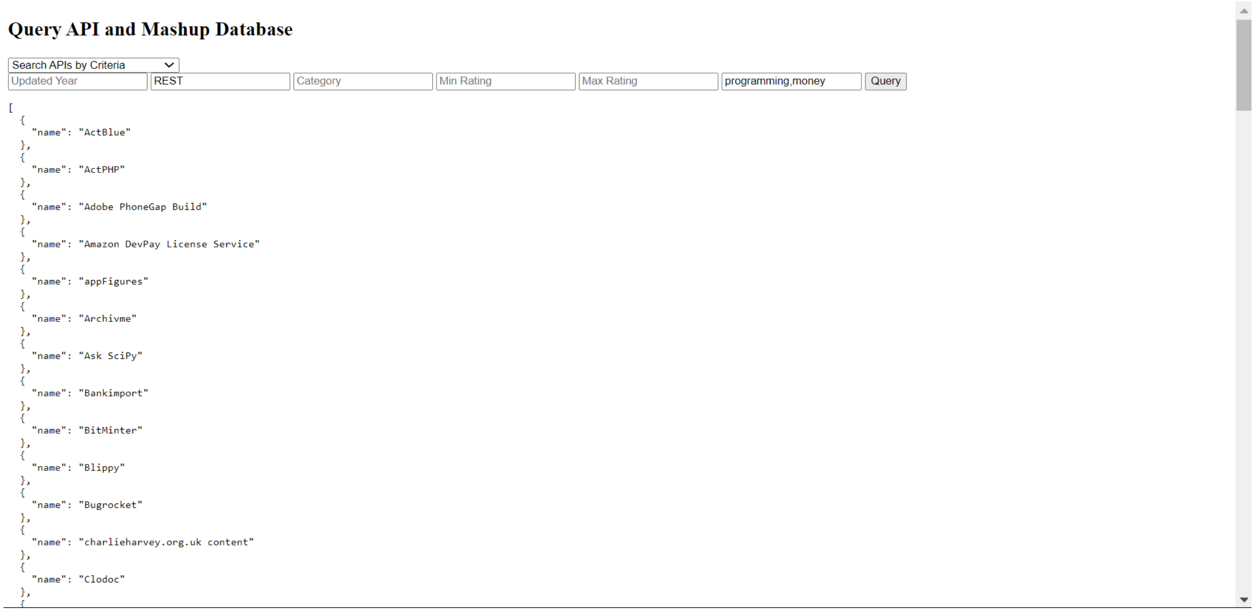
- APIModel › Expand all object
- APIUsed › Expand all object
- HTTPValidationError › Expand all object
- MashupModel › Expand all object
- ValidationError › Expand all object

Query system home



Query API and Mashup Database

Query testing, type specified by user using dropdown for all 6 types are below (not in order)

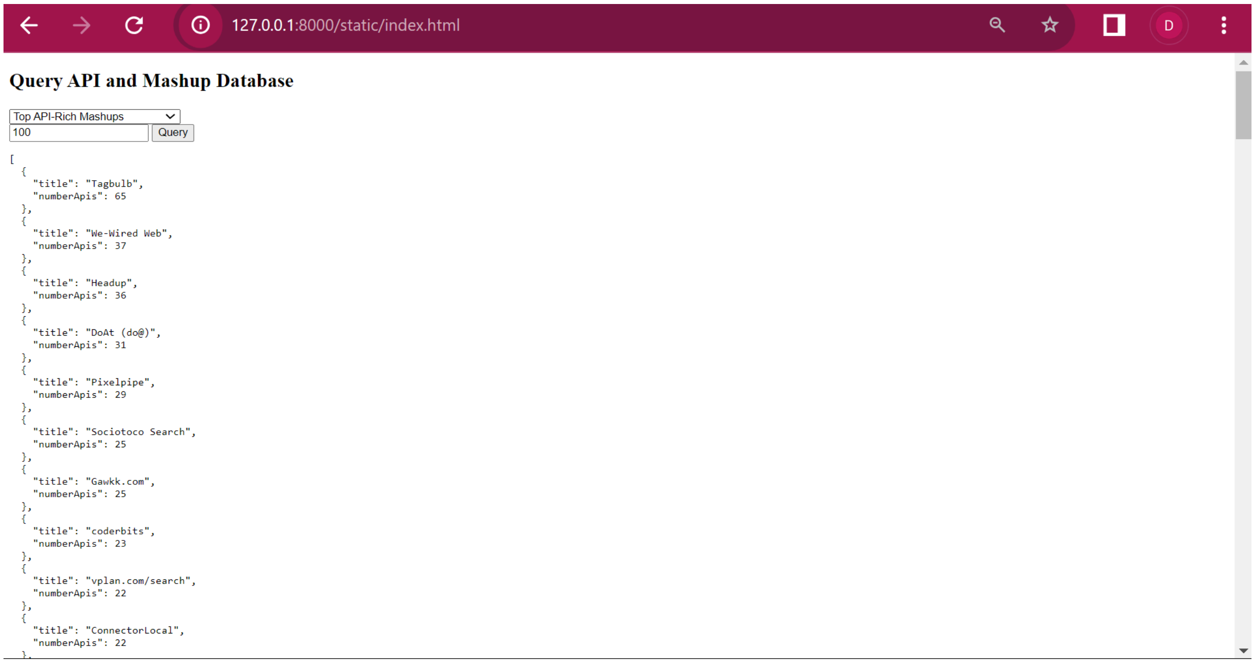


Query API and Mashup Database

Search APIs by Criteria

Updated Year | REST | Category | Min Rating | Max Rating | programming,money | Query

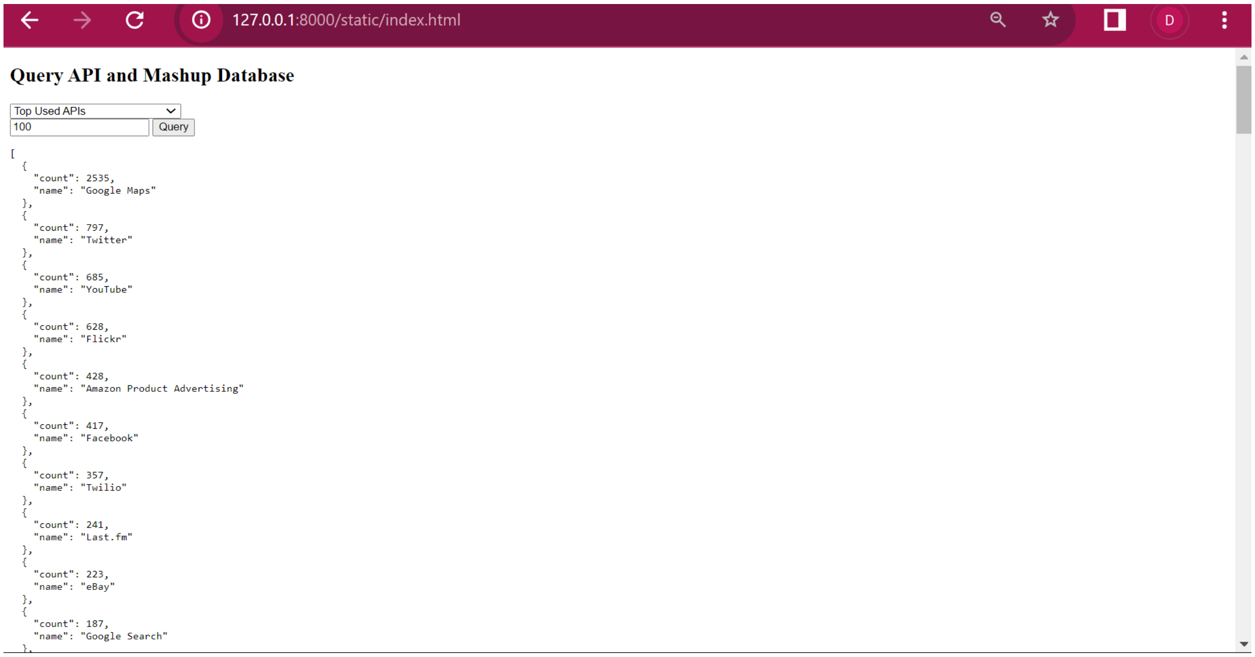
```
[
  {
    "name": "ActBlue"
  },
  {
    "name": "ActPHP"
  },
  {
    "name": "Adobe PhoneGap Build"
  },
  {
    "name": "Amazon DevPay License Service"
  },
  {
    "name": "appFigures"
  },
  {
    "name": "Archivme"
  },
  {
    "name": "Ask SciPy"
  },
  {
    "name": "Bankimport"
  },
  {
    "name": "BitMinter"
  },
  {
    "name": "Blippy"
  },
  {
    "name": "Bugrocket"
  },
  {
    "name": "charlieharvey.org.uk content"
  },
  {
    "name": "Clodoc"
  },
  {
```



Query API and Mashup Database

Top API-Rich Mashups
100 Query

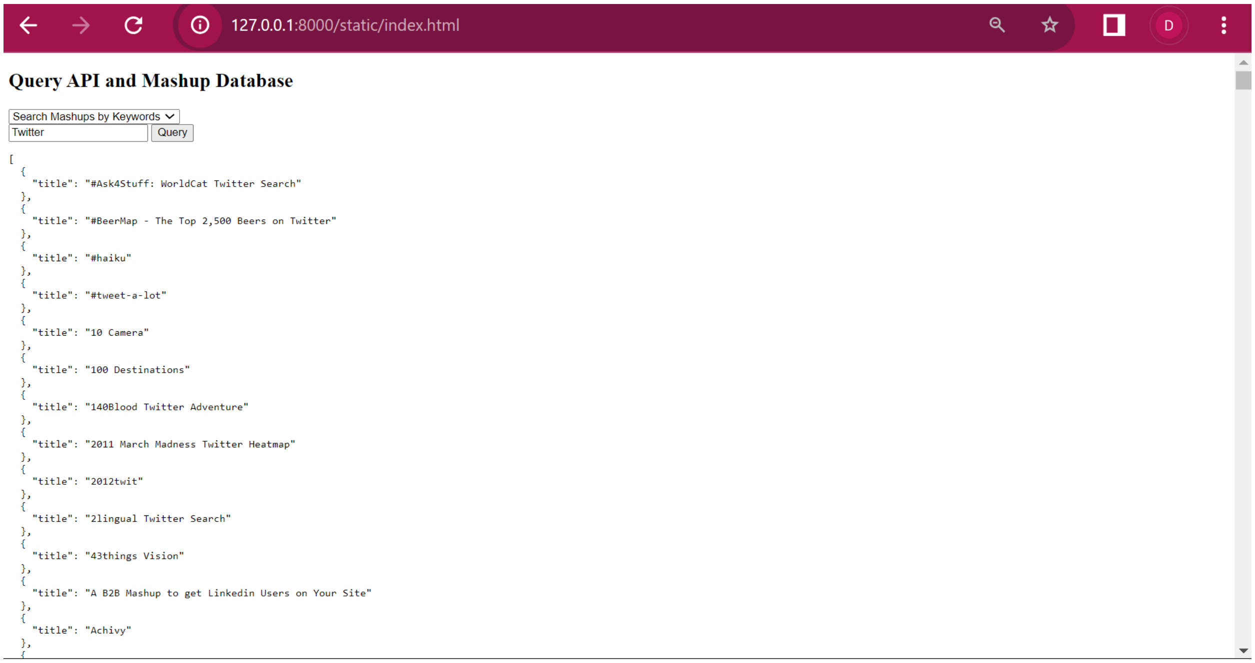
```
[
  {
    "title": "Tagbulb",
    "numberApis": 65
  },
  {
    "title": "We-Wired Web",
    "numberApis": 37
  },
  {
    "title": "Headup",
    "numberApis": 36
  },
  {
    "title": "DoAt (do@)",
    "numberApis": 31
  },
  {
    "title": "Pixelpipe",
    "numberApis": 29
  },
  {
    "title": "Sociotoco Search",
    "numberApis": 25
  },
  {
    "title": "Gawkk.com",
    "numberApis": 25
  },
  {
    "title": "coderbits",
    "numberApis": 23
  },
  {
    "title": "vplan.com/search",
    "numberApis": 22
  },
  {
    "title": "ConnectorLocal",
    "numberApis": 22
  },
```



Query API and Mashup Database

Top Used APIs
100 Query

```
[
  {
    "count": 2535,
    "name": "Google Maps"
  },
  {
    "count": 797,
    "name": "Twitter"
  },
  {
    "count": 685,
    "name": "YouTube"
  },
  {
    "count": 628,
    "name": "Flickr"
  },
  {
    "count": 428,
    "name": "Amazon Product Advertising"
  },
  {
    "count": 417,
    "name": "Facebook"
  },
  {
    "count": 357,
    "name": "Twilio"
  },
  {
    "count": 241,
    "name": "Last.fm"
  },
  {
    "count": 223,
    "name": "eBay"
  },
  {
    "count": 187,
    "name": "Google Search"
  },
```



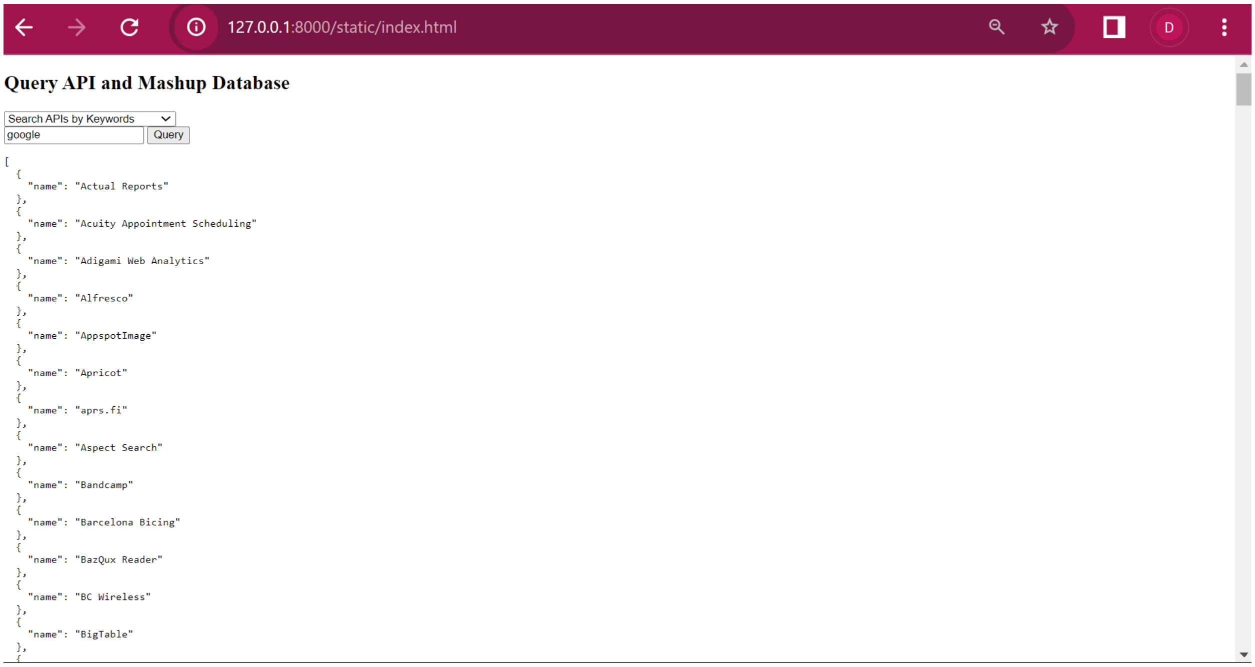
127.0.0.1:8000/static/index.html

Query API and Mashup Database

Search Mashups by Keywords

Twitter

```
[
  {
    "title": "#Ask4Stuff: WorldCat Twitter Search"
  },
  {
    "title": "#BeerMap - The Top 2,500 Beers on Twitter"
  },
  {
    "title": "#haiku"
  },
  {
    "title": "#tweet-a-lot"
  },
  {
    "title": "10 Camera"
  },
  {
    "title": "100 Destinations"
  },
  {
    "title": "140Blood Twitter Adventure"
  },
  {
    "title": "2011 March Madness Twitter Heatmap"
  },
  {
    "title": "2012twit"
  },
  {
    "title": "2lingual Twitter Search"
  },
  {
    "title": "43things Vision"
  },
  {
    "title": "A B2B Mashup to get Linkedin Users on Your Site"
  },
  {
    "title": "Achivy"
  },
  {
```



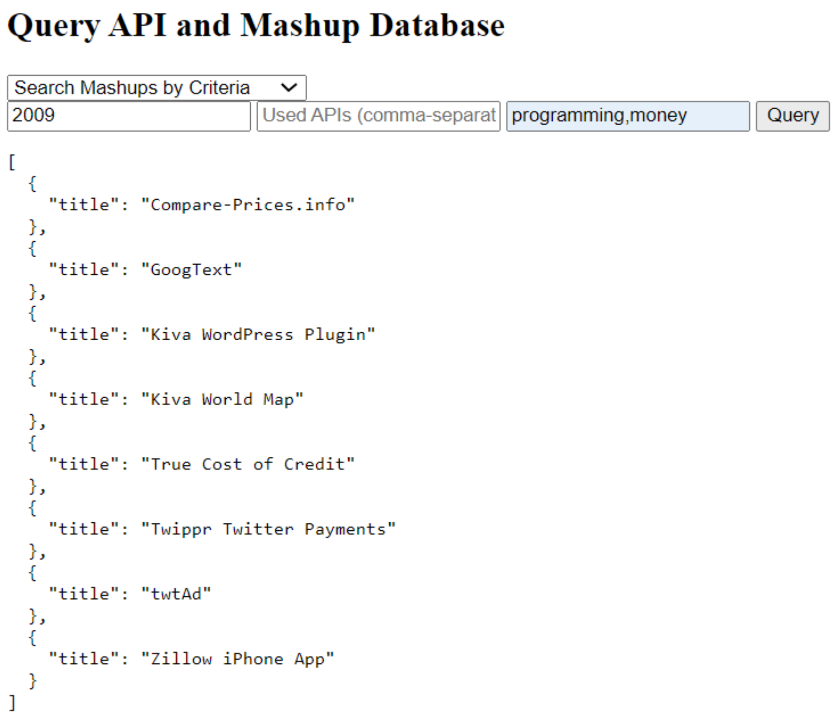
127.0.0.1:8000/static/index.html

Query API and Mashup Database

Search APIs by Keywords

google

```
[
  {
    "name": "Actual Reports"
  },
  {
    "name": "Acuity Appointment Scheduling"
  },
  {
    "name": "Adigami Web Analytics"
  },
  {
    "name": "Alfresco"
  },
  {
    "name": "AppspotImage"
  },
  {
    "name": "Apricot"
  },
  {
    "name": "aprs.fi"
  },
  {
    "name": "Aspect Search"
  },
  {
    "name": "Bandcamp"
  },
  {
    "name": "Barcelona Bicing"
  },
  {
    "name": "BazQux Reader"
  },
  {
    "name": "BC Wireless"
  },
  {
    "name": "BigTable"
  },
  {
```

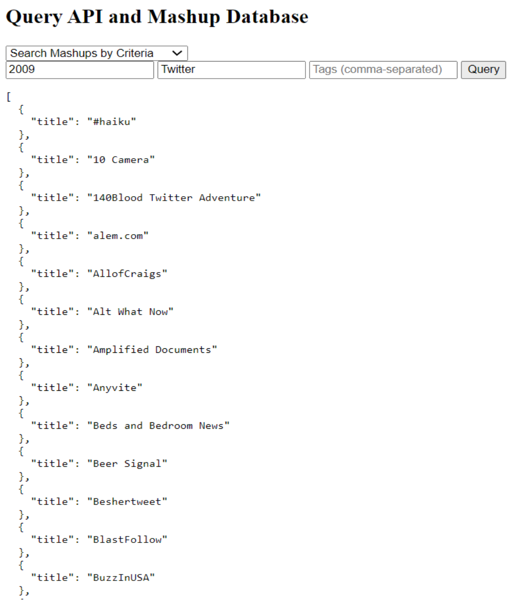


Query API and Mashup Database

Search Mashups by Criteria

2009 | Used APIs (comma-separat | programming,money

```
[
  {
    "title": "Compare-Prices.info"
  },
  {
    "title": "GoogText"
  },
  {
    "title": "Kiva WordPress Plugin"
  },
  {
    "title": "Kiva World Map"
  },
  {
    "title": "True Cost of Credit"
  },
  {
    "title": "Twippr Twitter Payments"
  },
  {
    "title": "twtAd"
  },
  {
    "title": "Zillow iPhone App"
  }
]
```

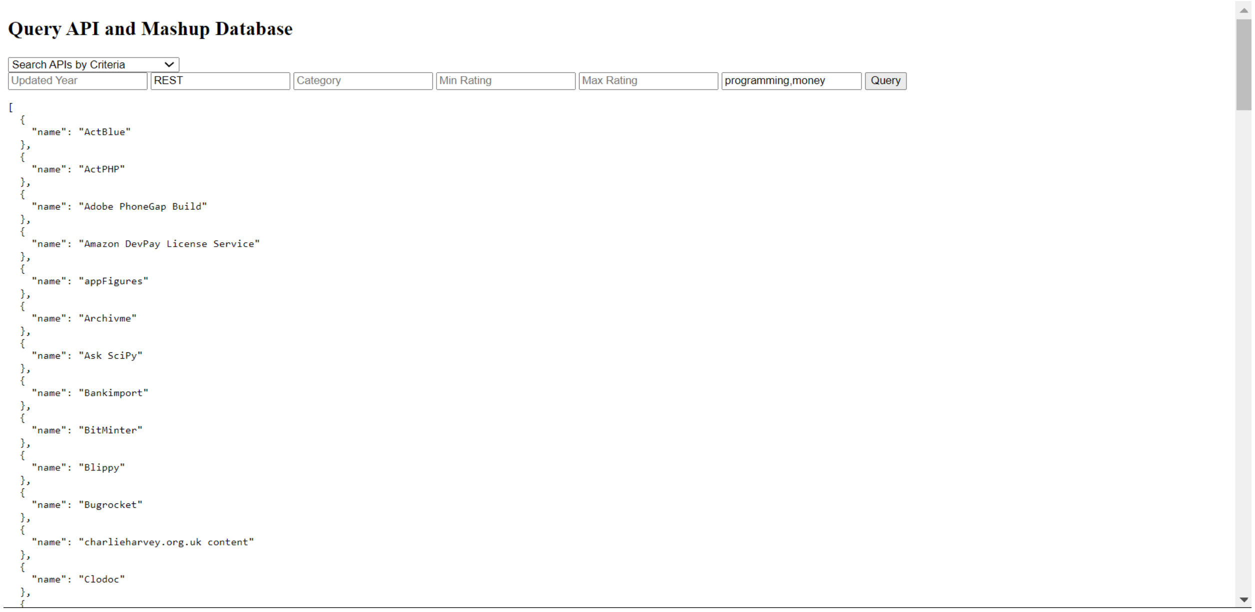


Query API and Mashup Database

Search Mashups by Criteria

2009 | Twitter | Tags (comma-separated) | Query

```
[
  {
    "title": "#haiku"
  },
  {
    "title": "10 Camera"
  },
  {
    "title": "140Blood Twitter Adventure"
  },
  {
    "title": "alem.com"
  },
  {
    "title": "AllofCraigs"
  },
  {
    "title": "Alt What Now"
  },
  {
    "title": "Amplified Documents"
  },
  {
    "title": "Anyvite"
  },
  {
    "title": "Beds and Bedroom News"
  },
  {
    "title": "Beer Signal"
  },
  {
    "title": "Beshertweet"
  },
  {
    "title": "BlastFollow"
  },
  {
    "title": "BuzzInUSA"
  },
```



Query API and Mashup Database

Search APIs by Criteria

Updated Year | REST | Category | Min Rating | Max Rating | programming,money | Query

```
[
  {
    "name": "ActBlue"
  },
  {
    "name": "ActPHP"
  },
  {
    "name": "Adobe PhoneGap Build"
  },
  {
    "name": "Amazon DevPay License Service"
  },
  {
    "name": "appFigures"
  },
  {
    "name": "Archivme"
  },
  {
    "name": "Ask SciPy"
  },
  {
    "name": "Bankimport"
  },
  {
    "name": "BitMinter"
  },
  {
    "name": "Blippy"
  },
  {
    "name": "Bugrocket"
  },
  {
    "name": "charlieharvey.org.uk content"
  },
  {
    "name": "Clodoc"
  },
  {
```