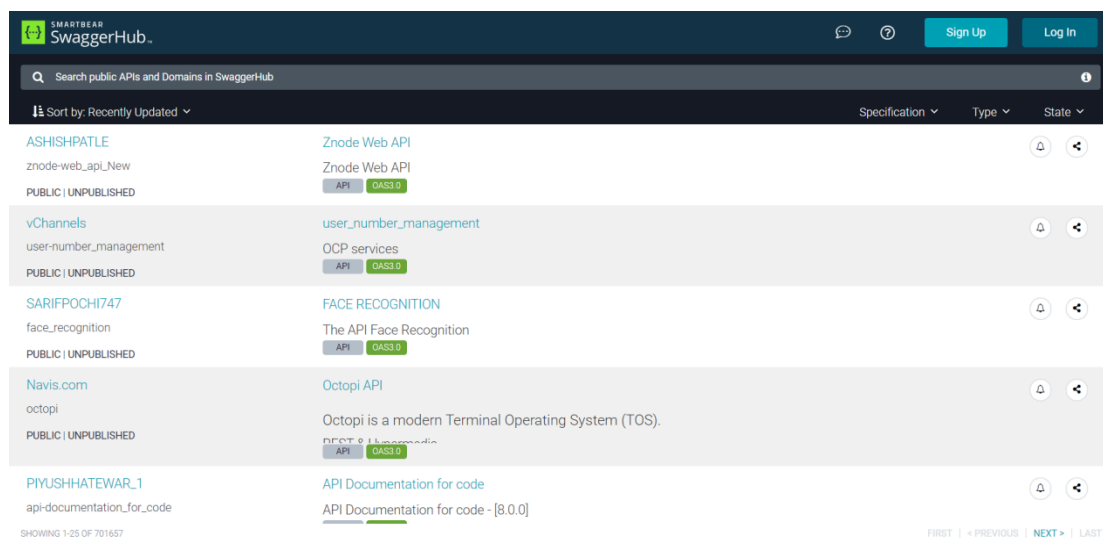


Documentation: Working with SwaggerHub API

SwaggerHub is a platform designed to help teams collaboratively design, build, document, and manage APIs. With a repository of over 700,000 APIs, it offers a comprehensive suite of tools to support the entire API lifecycle. The platform enables synchronized design and documentation workflows, which helps to ensure consistency and quality across all APIs. It supports a range of API standards, including OpenAPI (formerly Swagger), and provides features for hosting API documentation, automating the coding process, and generating client libraries. By facilitating collaboration and automation, SwaggerHub aims to streamline API development and help organizations efficiently scale their API strategies.



Overview

This comprehensive documentation outlines the procedures and scripts utilized for extracting and processing data from the SwaggerHub API. The objective is to compile a detailed and structured CSV database of APIs. The document is designed to guide users through various stages, from initial data retrieval to the final transformation and enhancement of the dataset, ensuring clarity and ease of use.

1. Downloading JSON Files from SwaggerHub

The process commenced with downloading resolved JSON files from SwaggerHub, focusing on obtaining complete and accurate data.

The files were thoroughly inspected to decode the structure and ascertain the types of data provided, such as API names, versions, and related metadata.

This step was crucial to lay the groundwork for subsequent data processing and extraction.

Identifying Base URLs

- This step involved meticulously searching within the downloaded JSON files to identify necessary base URLs.
- Base URLs are essential as they represent the foundational endpoints for API functionalities, aligning with the team's specific data requirements.
- The accurate identification of these URLs was pivotal for the success of the overall data extraction process.

2. API Request for Data Retrieval

To gather a comprehensive dataset, an API request was executed, designed to paginate through the SwaggerHub data. In the process of API data retrieval from SwaggerHub, a specialized API request is made to systematically navigate through the vast repository of over 700,000 APIs. This retrieval operation is meticulously designed to handle pagination, which is essential given the substantial volume of data. The resulting output provides a rich array of details for each API, including but not limited to, the API name, description, version information, creation and modification timestamps, access specifics like privacy settings, and OpenAPI Specification versions. Additional information such as server details and contact points are also gleaned, ensuring a holistic dataset that includes base URLs and other critical metadata to support a variety of applications and analyses.

```
GET /apiproxy/specs?sort=UPDATED&order=DESC&page=2&limit=25 HTTP/2
Host: app.swaggerhub.com
Cookie: authenticated=false; publicShared=null; liveChatShared=null; session=psM7Hf76ZdVfzGpo04mk-Q.Jkdj_xvDHCobK9aicrYjg8out6-wh9Ihmb5IiPCH20knQ3L9rzrk4z8ozK9Wf6KtCgGVtIpmMvNtB3fD
Sec-Ch-Ua: "Not A Brand";v="8", "Chromium";v="120"
Accept: application/json
X-Csrf-Token: 1Q29DuKx-klwId2SY3acQTcbF6BLjqAOKWI
Sec-Ch-Ua-Mobile: ?0
User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/120.0.6099.71 Safari/537.36
Sec-Ch-Ua-Platform: "Windows"
Sec-Patch-Site: same-origin
Sec-Patch-Mode: cors
Sec-Patch-Dest: empty
Referer: https://app.swaggerhub.com/search?page=1
Accept-Encoding: gzip, deflate, br
Accept-Language: en-GB,en-US;q=0.9,en;q=0.8
Priority: u=1, i
```

Conversion to Text File :

- The initially retrieved data was stored in a text file, setting the stage for further data transformation. This step was crucial as it provided a manageable format for handling large volumes of data, serving as an intermediate storage solution before the final conversion to CSV.

```
{
  "name": "Default listing",
  "description": "Default registry listing",
  "url": "specs/",
  "offset": 25,
  "totalCount": 698079,
  "apis": [
    {
      "name": "Wasapi",
      "description": "Wasapi API documentation.",
      "summary": "",
      "tags": [],
      "properties": [
        {
          "type": "Swagger",
          "url": "https://api.swaggerhub.com/apis/JUAN_1/Wasapi/1.0.0"
        },
        {
          "type": "X-Version",
          "value": "1.0.0"
        },
        {
          "type": "X-Created",
          "value": "2022-07-31T15:19:39Z"
        },
        {
          "type": "X-Modified",
          "value": "2024-01-03T04:19:51Z"
        },
        {
          "type": "X-Published",
          "value": "false"
        },
        {
          "type": "X-Versions",
          "value": "1.0.0"
        },
        {
          "type": "X-Private",
          "value": "false"
        }
      ]
    }
  ]
}
```

Script for JSON to CSV Conversion

A dedicated Python script was crafted for the conversion of the JSON data from the text file into a structured CSV format.

This script is crucial for transforming the raw JSON data into a more accessible and analyzable CSV format, making it suitable for various data analysis tasks.

Enhancing the CSV with Additional Data

To enrich the dataset, a second Python script was developed to append additional details to the existing CSV file.

- This script significantly enhances the dataset by adding server and contact information for each API, providing a more comprehensive view of the data.

Documentation and Scripts

The documentation was meticulously compiled to encompass each script and its purpose.

	Name	Description	URL	Version	Created	Modified	Published	Versions	Private	OASVersion	Specification	Notification	Standardization	CreatedBy	Servers	Contact	Servers	Contact
1	Wasapi	Wasapi API	https://api-1.1.0.0	2022-07-31	2024-01-03	FALSE	1.0.0	FALSE	3.0.0	openapi-3.0	FALSE	{}		juan21	{}	(url: 'https://api-1.1.0.0')	{}	(email: 'info@vinixcode.com')
2	Netmonk Hi Kettik prop	https://api-2.2.0	2023-11-16	2024-01-03	FALSE	1.0.0.1.1.0.1	FALSE	3.0.0	openapi-3.0	FALSE	{}			dennetmonk	{}	(description: 'email: 'yo	{}	(description: 'email: 'you@your-company.com')
3	Simple Invei	This is a sim	https://api-1.1.0.0	2024-01-03	2024-01-03	FALSE	1.0.0	FALSE	3.0.0	openapi-3.0	FALSE	{}		thannam27	{}	(description: 'email: 'yo	{}	(description: 'email: 'you@your-company.com')
4	DataService	DataService	https://api-1.1.0.0	2023-10-04	2024-01-03	FALSE	1.0.0	FALSE	3.0.0	openapi-3.0	FALSE	{}		brydenolive	{}	(description: 'email: 'yo	{}	(description: 'email: 'you@your-company.com')
5	Sample App	This is an ex	https://api-1.1.0.0	2024-01-03	2024-01-03	FALSE	1.0.0	FALSE	3.0.0	openapi-3.0	FALSE	{}		minsukoo	{}	(description: 'email: 'yo	{}	(description: 'email: 'you@your-company.com')
6	Simple Invei	This is a sim	https://api-1.1.0.0	2024-01-03	2024-01-03	TRUE	*1.0.0	FALSE	3.0.0	openapi-2.0	FALSE	{}		CamiloGuer	{}	(email: 'yo	{}	(email: 'you@your-company.com')
7	RedWeek Pi	RedWeek Pi	https://api-2.1.1.19	2023-12-19	2024-01-03	FALSE	2.1.0-beta.2	FALSE	3.0.0	openapi-3.0	FALSE	{}		ann2	{}	(url: 'http: 'we	{}	(url: 'http: 'webmaster@redweek.com')
8	Sample App	This is an ex	https://api-1.1.0.0	2024-01-03	2024-01-03	FALSE	1.0.0	FALSE	3.0.0	openapi-3.0	FALSE	{}		minsukoo	{}	(description: 'email: 'yo	{}	(description: 'email: 'you@your-company.com')
9	uasService	uasService	https://api-0.0.1	2024-01-03	2024-01-03	FALSE	0.0.1	FALSE	3.0.0	openapi-3.0	FALSE	{}		trongem	{}	(url: '//loc	{}	(url: '//loc
10	Registro de API	que per	https://api-1.1.0.0	2023-12-20	2024-01-03	FALSE	1.0.0	FALSE	3.0.0	openapi-3.0	FALSE	{}		ArreolaSant	{}	(description: 'name: 'La	{}	(description: 'name: 'Laboratorio de Multimedia e Intern
11	Pijar Sekolah	This is an ex	https://api-1.1.0.0	2023-11-21	2024-01-03	FALSE	1.0.0, sim-pl	FALSE	3.0.0	openapi-3.0	FALSE	{}		doc.pijarsek	{}	(description: 'email: 'yo	{}	(description: 'email: 'you@your-company.com')
12	Taxonomy	This is a sim	https://api-1.1.0.0	2023-12-21	2024-01-03	FALSE	1.0.0	FALSE	3.0.0	openapi-3.0	FALSE	{}		geeta.k	{}	(description: 'email: 'pa	{}	(description: 'email: 'payal.s@vseerv.com')
13	home-iot-aj	The API for	https://api-1.1.0.0	2024-01-03	2024-01-03	FALSE	1.0.0	FALSE	3.0.0	openapi-3.0	FALSE	{}		AnthonyTor	{}	(description: 'email: 'yo	{}	(description: 'email: 'you@your-company.com')
14	ecommerce	ecommerce	https://api-1.1.0.0	2024-01-03	2024-01-03	FALSE	1	FALSE	3.0.0	openapi-3.0	FALSE	{}		Neik01	{}	(description: 'email: 'yo	{}	(description: 'email: 'you@your-company.com')
15	SupportBox	API pro inte	https://api-1.1.0.0	2023-11-20	2024-01-03	FALSE	1.0.1.0.0	FALSE	3.0.0	openapi-3.0	FALSE	{}		infobred454	{}	(description: 'email: 'va	{}	(description: 'email: 'vaclav@makes.cz')
16	API SaaS	apilace	[SupportBo															

3. Conclusion :

This documentation offers a complete and detailed guide to the methodology employed in extracting and processing data from the SwaggerHub API.

It includes in-depth descriptions of each stage, supplemented with Python scripts developed for converting JSON data to CSV format and enhancing the dataset with additional information.

The thorough approach ensures that the dataset is robust, comprehensive, and useful for further analysis.