# WEBSITE\_ANALYSIS API

## 1.Obtain data from DB :-

### a) If Data is already present in DB:-

This API is used to fetch data from DB, based on the url specified. We should also pass the Token in the header to obtain the correct data. Otherwise there will be authentication error.

|  |  |
| --- | --- |
| TITLE | ACCESS THE DATA |
| URL | http://127.0.0.1:8000//website\_analysis/ |
| RETURN TYPE | JSON |
| METHOD | GET |
| HEADER | Authorization:  Token axxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxe |
| INPUT PARAMETER | url:(For which we need data)  Eg: url=http://www.us-consulting.de |
| SAMPLE OUTPUT | {  "blog\_urls": [],  "blog\_posts": [],  "rss\_feeds": [],  "under\_construction": false,  "new\_technologies": [  "index\_p.css",  "usc\_g.css"  ],  "social\_media\_links": [],  "page\_count": 2,  "technologies": [],  "emails": [],  "direct\_emails": [],  "latest\_update": "",  "postal\_addresses": [],  "service\_offered": [  "Beratung",  "E",  "ANA",  "Dienstleistung",  "ST",  "Eiche",  "SC",  "er",  "Dienstleistungen"  ],  "jobs\_exist": false,  "time": [],  "responsive": false,  "direct\_phone\_numbers": [],  "about\_us": "",  "image\_count": 0,  "phone\_numbers": [],  "website\_url": "http://www.us-consulting.de"  } |

### b)Adding url to Scraping list :

If the url specified is new, ie, if it is not present in the url list (which we will scrap using scrapyhub), then the new url will be added to the url list .This url will be scraped automatically by the scrapyhub.

|  |  |
| --- | --- |
| TITLE | Adding that to url list and scrap |
| URL | http://127.0.0.1:8000//website\_analysis/ |
| RETURN TYPE | JSON |
| METHOD | GET |
| HEADER | Authorization:  Token axxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxe |
| INPUT | url=https://www.sayonetech.com/ |
| SAMPLE OUTPUT | {  "spider\_status": "Updated"  } |

c) Checking the url before getting added:

This will enable the checking of url before getting added to the url list, ie, if the url is not valid then, that url will not be added to the list and that new url will be got rejected.

|  |  |
| --- | --- |
| TITLE | INVALID URL |
| URL | http://127.0.0.1:8000//website\_analysis/ |
| RETURN TYPE | JSON |
| METHOD | GET |
| HEADER | Authorization:  Token axxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxe |
| INPUT PARAMETERS | url=www.sayonetech.com |
| SAMPLE OUTPUT | {  "url\_status": false  } |

## 2) Inserting data into DB:-

This is used to insert the scraped data into the DB from scrapyhub. Data will be a combination of url and scraped\_data.

|  |  |
| --- | --- |
| TITLE | INSERT TO DB |
| URL | http://127.0.0.1:8000//website\_analysis/ |
| RETURN TYPE | JSON |
| METHOD | POST |
| HEADER | Authorization:  Token axxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxe  Content-type : application/json |
| INPUT | data={'blog\_urls': [], 'blog\_posts': [], 'rss\_feeds': [], 'under\_construction': False, 'new\_technologies': [u'assets/css/cd075ba4b74c.css', u'http://fast.fonts.net/cssapi/a4f0b4dc-08e1-44d8-bee3-284052276f63.css', u'system/modules/dlh\_googlemaps/html/dlh\_googlemaps.css'], 'social\_media\_links': [{u'url': u'https://www.facebook.com/InzoneHairgroup', u'type': u'facebook'}], 'page\_count': 4, 'technologies': [], 'emails': [{u'email': u'info@bajac.de.'}], 'direct\_emails': [], 'latest\_update': '', 'postal\_addresses': [], 'service\_offered': [], 'jobs\_exist': False, 'time': [u'zeitenPreiseInterieur\xa0BAJACTreuebonusKamp62,O'], 'responsive': True, 'direct\_phone\_numbers': [], 'about\_us': u'Junge und kreative Stylisten f\xc3\xbcr Deinen abgefahrenen Look. Wir sind der trendige und junge Friseur f\xc3\xbcr Dich in Osnabr\xc3\xbcck und Bielefeld. Cut & Go, modisch, schnell, g\xc3\xbcnstig, gut. Die jungen Friseure mit handwerklichem K\xc3\xb6nnen f\xc3\xa4rben und schneiden Deine Haare ganz nach Wunsch. Wir geben Dir den modernen, pflegeleichten Look.', 'image\_count': 121, 'phone\_numbers': [{u'phone\_number': u'0541-9986666', u'type': u'undefined'}], 'website\_url': u'[http://www.inzone.de](http://www.inzone.de/)'}  (data scraped from website)  url=[http://www.inzone.de](http://www.inzone.de/)  (url of website) |
| SAMPLE OUTPUT | {"data\_status":true} |