

# MODULES USED IN THE AUTOMATION PROJECT

## 1. Re [ Usage -> import re ]

Re is the module used for executing regular expression in python , in our project we have used this module to filter the links out of all the links pulled out from the source code.

```
links = re.findall('https?://\S+'),linkgroup)
```

Above piece of code has been used in our project, and it pulls out all http and https links out of all links.

## 2. Urllib and Urllib2 [Usage -> import urllib ,urllib2 ]

Urllib2, and urllib are python modules which are extensible library for opening URLs, in our project we have used these module to open up the WebPages using the links stored in a text file

```
opener=urllib2.build_opener()
```

Above piece of code has been used in the project ,which perform the similar operation described above.

## 3.Mechanize [Usage-> Import mechanize]

Mechanize is a module which can be consider as an substitute for urllib , which is more efficient than the same , in our project we have use this module to traverse through a web site and fetch out all the links present in the source code of corresponding web site .

```
link= mechanize.Browser()
```

link.open(url)Above piece of code has been used in the project which perform the similar operation described above

## 4.String [Usage->import string]

string module is a module which consist of many method to process the python strings , in our project we have used this string module to perform some basic string operations.

## **5. Cookieslib and CookieJar [Usage-> from cookielib import CookieJar ]**

Cookies lib is a module used to manage http requests and replay , in our project we have used cookieslib module to find the particular webpage using cookies sent from the server to the client machine, and open up the webpage to fetch out the source code .

```
cj = CookieJar()
```

```
opener = urllib2.build_opener(urllib2.HTTPCookieProcessor(cj))
```

Above piece of code has been used in the project ,which perform the similar operation described above.

## **6. itertools , ifilter and imap [Usage-> from itertools import ifilter, imap]**

ifilter is a module used to filter the null values in the text or a piece of data and it used itertools to iterate through the data to remove the null values , in our project we have used the module to remove the null values found in the data pulled out from the webpage .

```
Data= filter(None,Predata)
```

Above piece of code has been used in the project ,which perform the similar operation described above.

## **7.BeautifulSoup or bs4 [ import BeautifulSoup from BeautifulSoup ]**

Beautifulsoup is a module used to do web scripting , in our project we have used this module to fetch out the desired content from the source code .

```
soup.findAll("div",{ 'class':'col-xs-12'}):
```

```
content= (add.get_text().encode('utf-8'))
```

Above piece of code has been used in the project ,which perform the similar operation described above.

-----THANK YOU-----