
My_Study_Spider Documentation

Release v0.1

zhaojiedi1992@outlook.com

Jul 29, 2018

1	urllib	1
1.1	1
1.2	1
1.3	2
1.4	4
1.5	4
1.6	4
1.7	5
2	firefox	7
3	chrome	9
4	bs4	11
4.1	bs4 test1	11
4.2	bs4 test2	11
4.3	bs4 test3	11
5	selenium	13
5.1	selenium test1	13
5.2	selenium test2	13
5.3	selenium test3	13
6	pyspider	15
7	scrapy	17
8	scrapy-redis	19

CHAPTER 1

urllib

1.1

```
# pythonurllib
import urllib.request
url="https://docs.scrapy.org/en/latest/"
response = urllib.request.urlopen(url)
# decodeutf-8
html = response.read().decode('utf-8')
print(html)
```

1.2

```
# pythonurllib
import urllib.request
url="http://www.linuxpanda.tech/images/avatar.jpg"
response = urllib.request.urlopen(url)
if response.getcode == 200:
    with open ("sijia.png", 'bw') as f :
        f.write(response.read())
```

sijia.png



1.3

1.112 打开浏览器调试功能，切换到network选项卡

3 点击翻译

2. 输入你要翻译的内容

4. 查看这个请求

有道翻译 免费、即时的多语种在线翻译

检测到: 英语 > 中文

翻译 人工翻译

good boy

好男孩

good boy
外星神犬 (电影名)
查看完整结果 >>

有道人工翻译 / 母语润色
全球最优质的译员实时接单
专业、精准、地道

有道翻译
27种语言的智能机器翻译
最小更强大的翻译神器

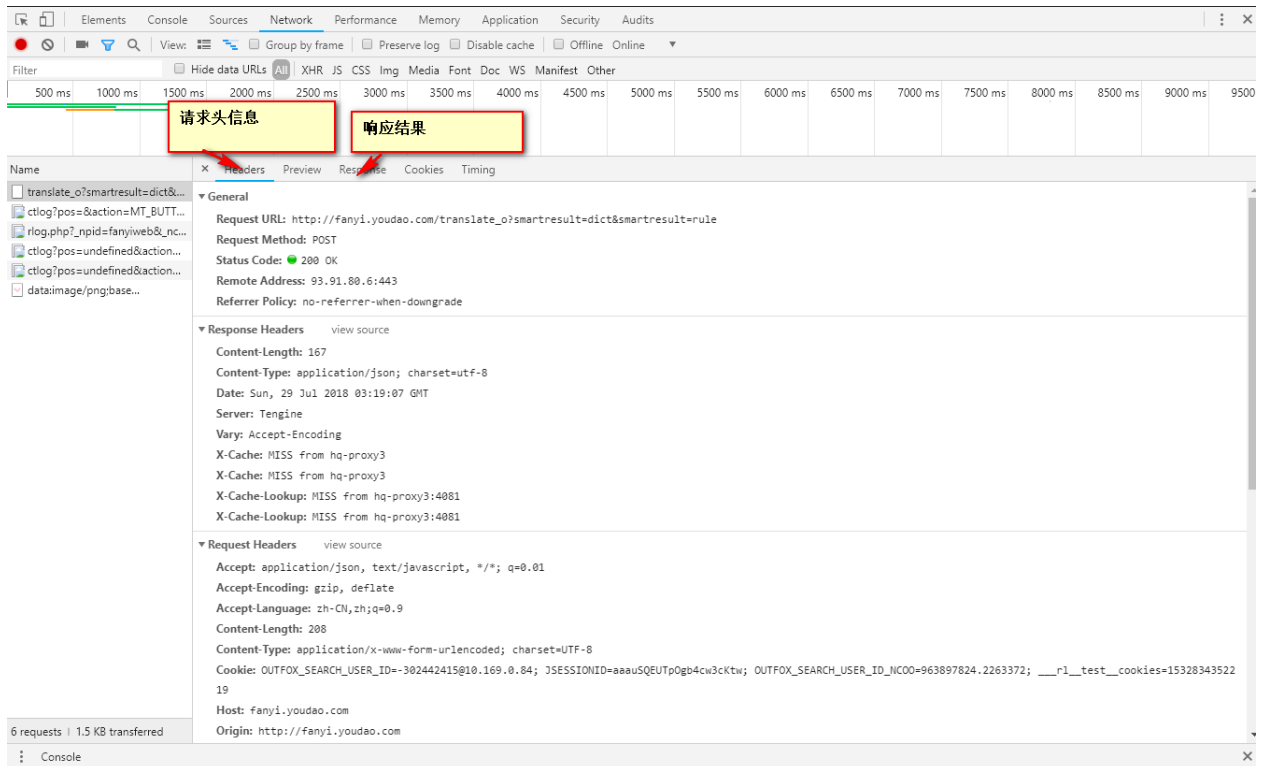
有道翻译官 APP
支持海量翻译和拍照翻译
107种语言的随身翻译

有道首页 | 有道智选 | 有道精品课 | 关于有道 | 官方微博

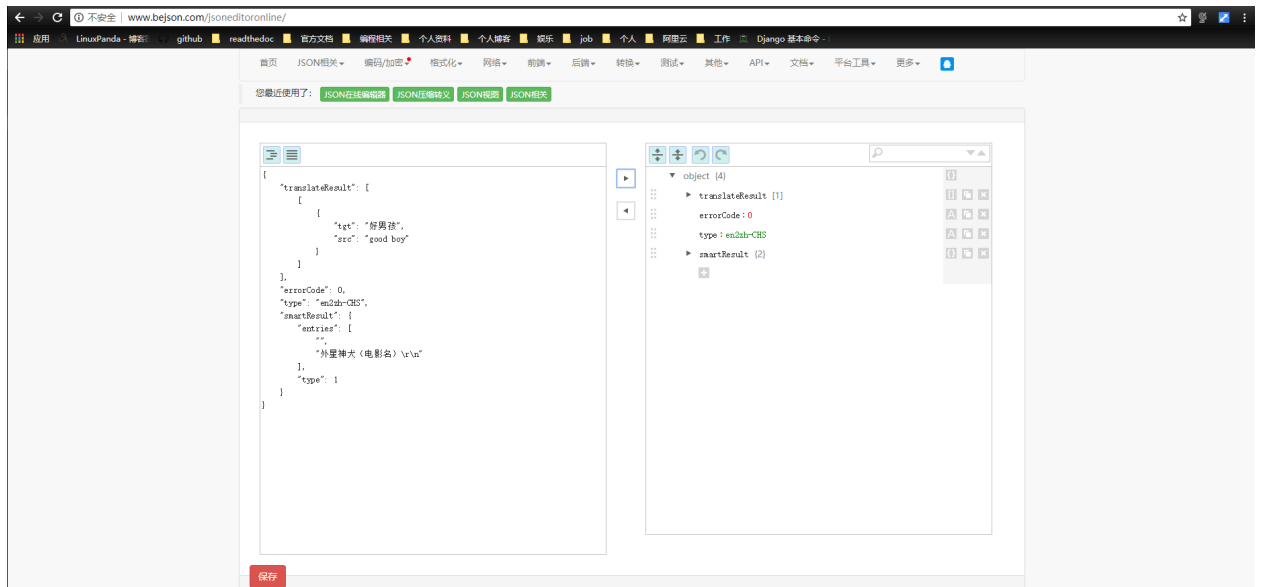
© 2018 网易公司 京ICP证080268号

6 requests | 1.5 KB transferred

Name	Status	Type	Initiator	Size	Time	Waterfall
translate_o/smart...	200	xhr	fany...	475 B	436...	
ctlog/post=8&actio...	200	text...	fany...	263 B	864...	
ctlog.php?_sigle=...	200	text...	fany...	297 B	814...	
ctlog/post=sundef...	200	text...	fany...	263 B	790...	
ctlog/post=sundef...	200	text...	fany...	263 B	1.0...	
dataimage/png/b...	200	png	Other	(fro...	0 ms	



jsonjson



```
# pythonurllib
import urllib.request
import urllib.parse
import json
# url,translate_o
url="http://fanyi.youdao.com/translate?smartresult=dict&smartresult=rule"

# dataform data
```

(continues on next page)

(continued from previous page)

```
data={
    'i': 'good boy',
    'from': 'AUTO',
    'to': 'AUTO',
    'smartresult': 'dict',
    'client': 'fanyideskweb',
    'salt': '1532835301042',
    'sign': '861dbcf1c6a959f3d0933003d0b8e81d',
    'doctype': 'json',
    'version': '2.1',
    'keyfrom': 'fanyi.web',
    'action': 'FY_BY_CLICKBUTTON',
    'typoResult': 'false'
}
data= urllib.parse.urlencode(data).encode('utf-8')
response = urllib.request.urlopen(url,data)
# decodeutf-8
html = response.read().decode('utf-8')
json_result = json.loads(html)
print(json_result['translateResult'][0][0]["tgt"])
```

Note: saltsign

1.4

urllibuser-agentPython-urllib/2.6

```
# head
headers={
    "User-Agent": "Mozilla/5.0 (X11; U; Linux i686) Gecko/20071127 Firefox/2.0.0.11"
}
req= urllib.request.Request(url,headers=headers)
# heads,Request.add_header(key, val)head
response = urllib.request.urlopen(req)
```

1.5

```
req= urllib.request.Request(url,headers=headers)
req.add_header('Referer', 'http://www.python.org/')
req.add_header('User-Agent', 'Mozilla/5.0 (X11; U; Linux i686) Gecko/20071127 Firefox/
↪2.0.0.11')
response = urllib.request.urlopen(req)
```

1.6


```
import urllib.request
# Create an OpenerDirector with support for Basic HTTP Authentication...
auth_handler = urllib.request.HTTPBasicAuthHandler()
auth_handler.add_password(realm='PDQ Application',
                          uri='https://mahler:8092/site-updates.py',
                          user='klem',
                          passwd='kadidd!ehopper')
opener = urllib.request.build_opener(auth_handler)
# ...and install it globally so it can be used with urlopen.
urllib.request.install_opener(opener)
urllib.request.urlopen('http://www.example.com/login.html')
```

install_openeropener,urlopenopenerProxyHandler

1.7

ip

```
import urllib.request
url="http://www.ip.cn"
proxy_handler = urllib.request.ProxyHandler({"http":"91.147.221.160:41766"})
opener = urllib.request.build_opener(proxy_handler)
opener.addheaders = [('User-agent', 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/67.0.3396.99 Safari/537.36')]
response = opener.open(url)
html = response.read().decode('utf-8')
print(html)
```

ip

CHAPTER 2

firefox

CHAPTER 3

chrome

4.1 bs4 test1

4.1.1 bs4 testtest

4.2 bs4 test2

4.3 bs4 test3

5.1 selinium test1

5.1.1 selinium testtest

5.2 selinium test2

5.3 selinium test3

CHAPTER 6

pyspider

CHAPTER 7

scrapy

CHAPTER 8

scrapy-redis
