

How to use?

Version 1.2

Structure

— Only four files are important, please keep this structure.

```
AutoNoticeSpider.py -- main function of this app
sender.py -- email related function
start_spider.sh -- for cron mechanism to schedule this app
quotes_spider.py - grab data from webpage
```

```
.
├─ AutoNoticeSpider.py
├─ README.md
├─ how\ to\ use.pdf
├─ quotes.json
├─ scrapy.cfg
├─ sender.py
├─ start_spider.sh
├─ tutorial
│   ├─ __init__.py
│   ├─ __pycache__
│   │   ├─ __init__.cpython-36.pyc
│   │   └─ settings.cpython-36.pyc
│   ├─ items.py
│   ├─ middlewares.py
│   ├─ pipelines.py
│   ├─ settings.py
│   └─ spiders
│       ├─ __init__.py
│       ├─ __pycache__
│       │   ├─ __init__.cpython-36.pyc
│       │   └─ quotes_spider.cpython-36.pyc
│       └─ quotes_spider.py
```

Set up

A. AutoNoticeSpider.py — no need to modify.

B. sender.py — three places need to be modified.

```
sender_email = "testonly@gmail.com" # Enter sender address
password = 'asdfasdfasdf' # pwd for testonly@gmail.com
receiver_email = "yourPersonalEmail@gmail.com" # Enter receiver address, usually your personal address.
```

```
-- sender_email
```

The most easiest way is to apply a new gmail.

- * Create a new Google account.

- * Turn **Allow less secure apps to ON**. Be aware that this makes it easier for others to gain access to your account.

(<https://myaccount.google.com/lesssecureapps>)

-- password

It's the sender's password.

-- receiver_email

It's your personal email, you want to use this one to receive the result.

C. quotes_spider.py — two places need to be modified.

```
caseId = 1234567
```

```
num = 2
```

-- caseId

the case id you want to check.

case id format: z1234567, **only** put the number part here.

-- num

Batch check is supported.

It is a search range which bases on caseId.

e.g. :

if num is 2, caseId is z1234567, two cases will be checked, they are z123456**7** and z123456**8**.

Install related python package on your PC before run this code

A. Python 3: <https://realpython.com/installing-python/>

B. Scrapy : <https://docs.scrapy.org/en/latest/intro/install.html>

Run Run Run!

Go to root of this project where AutoNoticeSpider.py is.

Then run this,

```
$ python AutoNoticeSpider.py
```

Check your email!

Note: this way only run the app **once**.

If we want to run the app repeatedly, please check below part.

Schedule to run spider

way 1, set a dead loop in AutoNoticeSpider.py:

```
def main():
    # it is a dead loop, press ctrl c to exit!
    while True:
        p = Process(target=mySpider)
        p.start()
        p.join()
        time.sleep(43200)
        # time.sleep(10)
```

This is a naive way. We can still use **\$python**

AutoNoticeSpider.py to run this app.

But, if the OS was broken, after OS restarting, our spider service is gone.

way 2, use cron (not corn.. 🐛)

The software utility cron is a time-based job scheduler in Unix-like computer operating systems.

crontab

```
1 SHELL=/bin/sh
2 PATH=/Users/xiaoyangzhang/anaconda3/bin:/usr/bin:/bin
3 */1 * * * * sh /Users/xiaoyangzhang/git_base/Spider/tutorial/start_spider.sh
4 #*/1 * * * * sh /Users/xiaoyangzhang/Downloads/tutorial/start_app.sh
```

start_spider.sh

```
1 #!/bin/bash
2 cd /Users/xiaoyangzhang/git_base/Spider/tutorial
3 #PATH=$PATH:/Users/xiaoyangzhang/anaconda3/bin
4 #export PATH
5 python AutoNoticeSpider.py
```

More details

email related info: <https://realpython.com/python-send-email/#sending-your-plain-text-email>

CSS filter: <http://scrapingauthority.com/2016/09/07/xpath-and-css-selectors/>

crontab: https://blog.csdn.net/qq_36743482/article/details/83616436

<https://blog.csdn.net/ethanzhao/article/details/4406017>

<http://man7.org/linux/man-pages/man5/crontab.5.html>