

Настройка для парсера на MacOS

1. Скачиваем и устанавливаем Python с официального сайта по ссылке
<https://www.python.org/ftp/python/3.11.6/python-3.11.6-macos11.pkg>
2. После чего откройте терминал, и выполните команду для перехода в папку с парсером (Далее все команды выполняются находясь в папке с парсером)

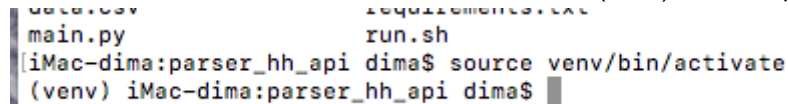
`cd путь/до/папки/с/парсером`

Вводим следующие команды, параллельно устанавливая утилиты которые предлагает МакОс

`python3 -m venv venv`

`source venv/bin/activate`

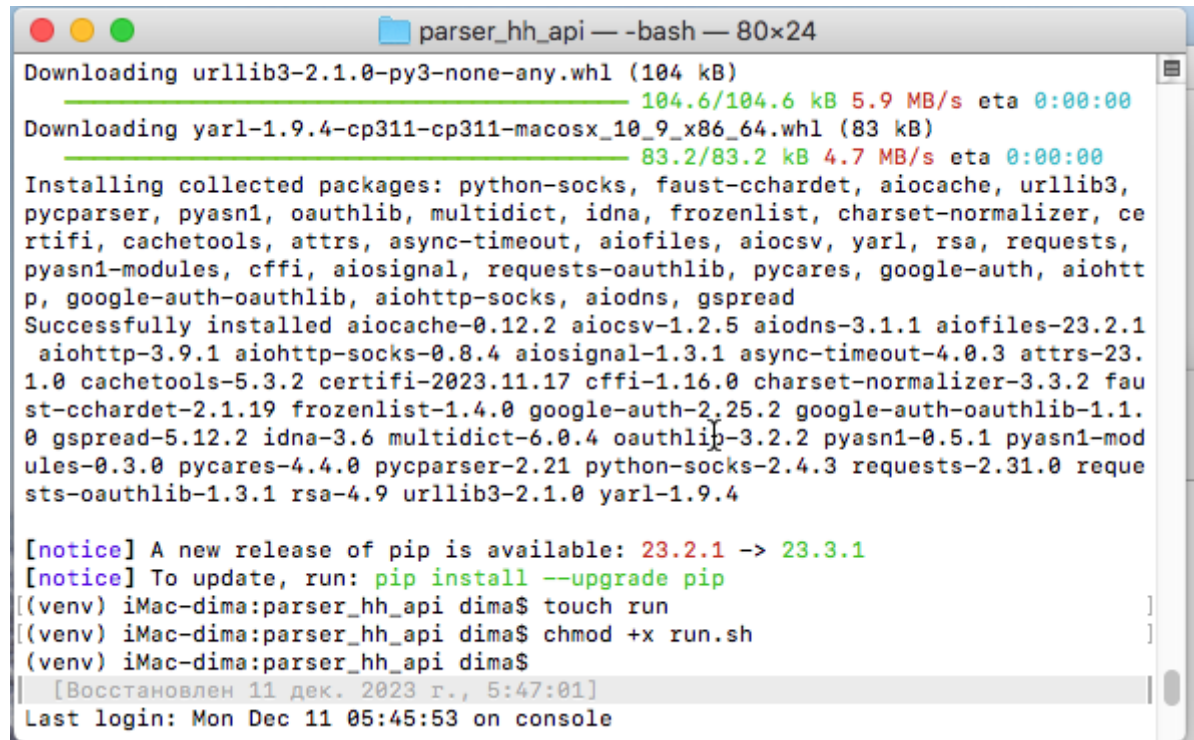
Возле пользователя должна появиться надпись (venv) как на скриншоте



```
main.py      requirements.txt
[1] iMac-dima:parser_hh_api dima$ source venv/bin/activate
(venv) iMac-dima:parser_hh_api dima$
```

Далее введите команду `pip install -r requirements.txt`

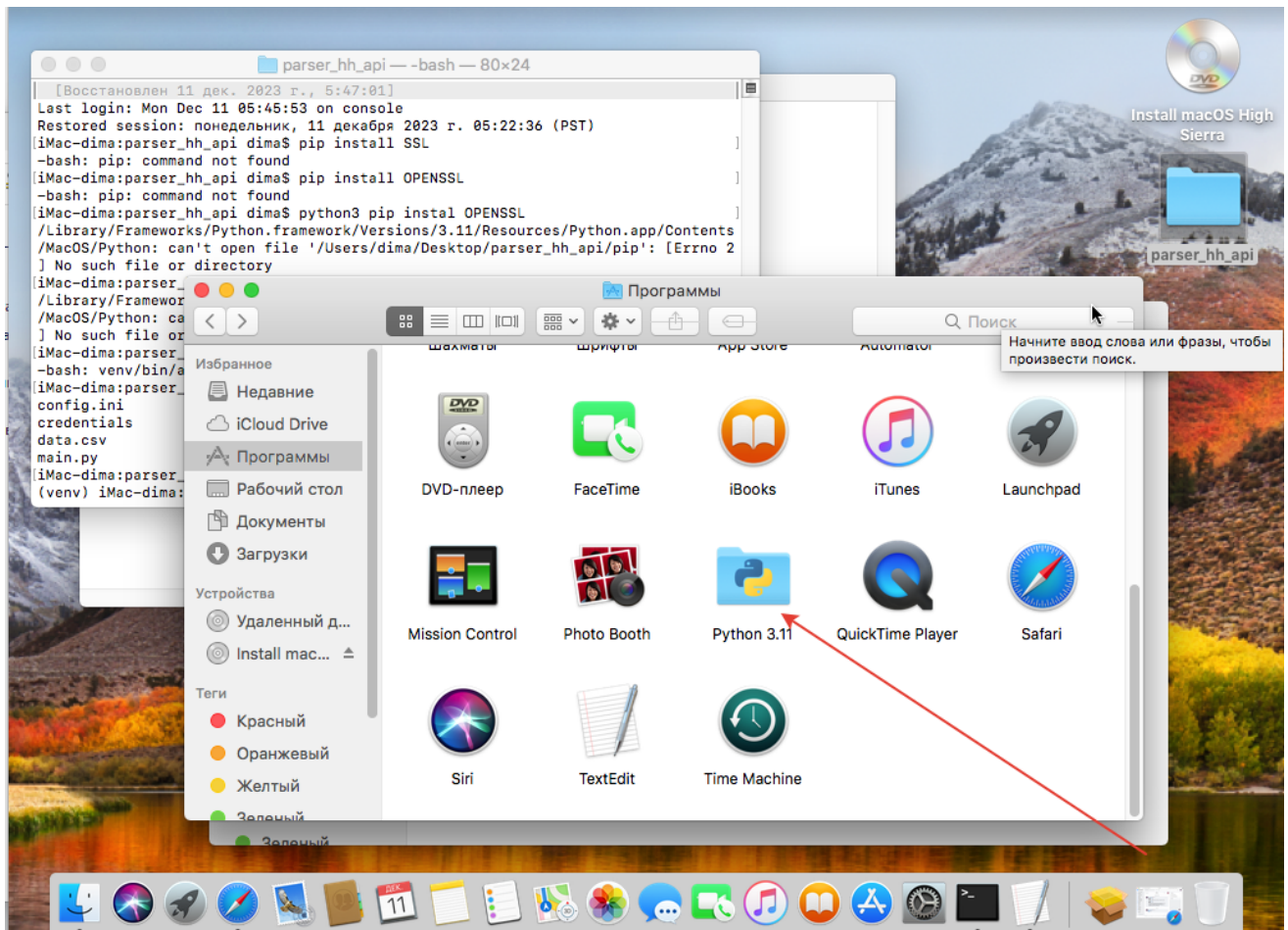
Начнут скачиваться и устанавливаться зависимости, все зависимости будут установлены в изолированную среду, так что сама система засоряться не будет, выглядеть будет примерно как на скриншоте



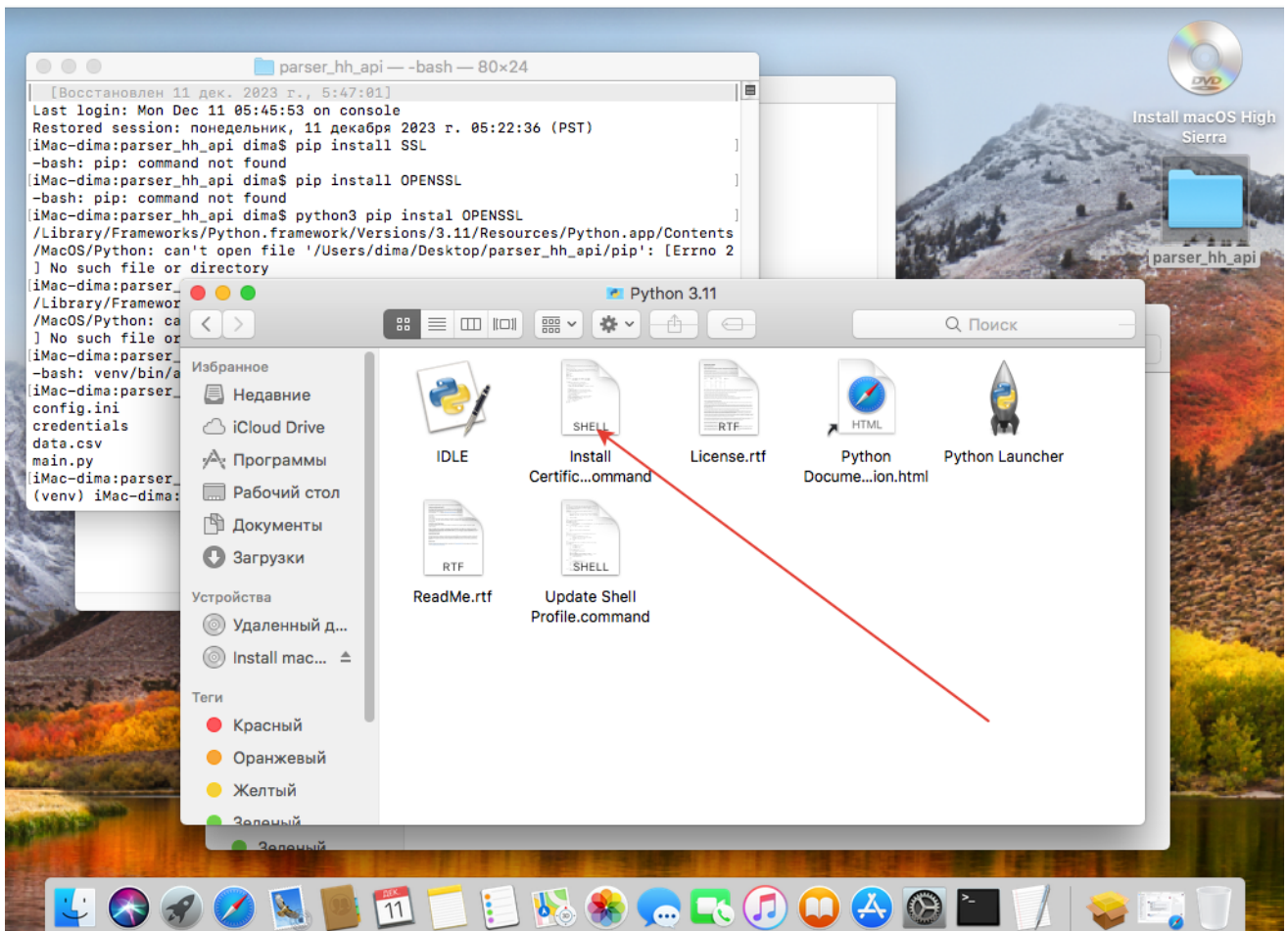
```
parser_hh_api — -bash — 80x24
Downloading urllib3-2.1.0-py3-none-any.whl (104 kB)
104.6/104.6 kB 5.9 MB/s eta 0:00:00
Downloading yarl-1.9.4-cp311-cp311-macosx_10_9_x86_64.whl (83 kB)
83.2/83.2 kB 4.7 MB/s eta 0:00:00
Installing collected packages: python-socks, faust-cchardet, aiocache, urllib3,
pycparser, pyasn1, oauthlib, multidict, idna, frozenlist, charset-normalizer, ce
rtifi, cachetools, attrs, async-timeout, aiofiles, aiocsv, yarl, rsa, requests,
pyasn1-modules, cffi, aiosignal, requests-oauthlib, pycares, google-auth, aiohtt
p, google-auth-oauthlib, aiohttp-socks, aiодns, gspread
Successfully installed aiocache-0.12.2 aiocsv-1.2.5 aiодns-3.1.1 aiofiles-23.2.1
aiohttp-3.9.1 aiohttp-socks-0.8.4 aiosignal-1.3.1 async-timeout-4.0.3 attrs-23.
1.0 cachetools-5.3.2 certifi-2023.11.17 cffi-1.16.0 charset-normalizer-3.3.2 fau
st-cchardet-2.1.19 frozenlist-1.4.0 google-auth-2.25.2 google-auth-oauthlib-1.1.
0 gspread-5.12.2 idna-3.6 multidict-6.0.4 oauthlib-3.2.2 pyasn1-0.5.1 pyasn1-mod
ules-0.3.0 pycares-4.4.0 pycparser-2.21 python-socks-2.4.3 requests-2.31.0 reque
sts-oauthlib-1.3.1 rsa-4.9 urllib3-2.1.0 yarl-1.9.4

[notice] A new release of pip is available: 23.2.1 -> 23.3.1
[notice] To update, run: pip install --upgrade pip
(venv) iMac-dima:parser_hh_api dima$ touch run
(venv) iMac-dima:parser_hh_api dima$ chmod +x run.sh
(venv) iMac-dima:parser_hh_api dima$
[Восстановлен 11 дек. 2023 г., 5:47:01]
Last login: Mon Dec 11 05:45:53 on console
```

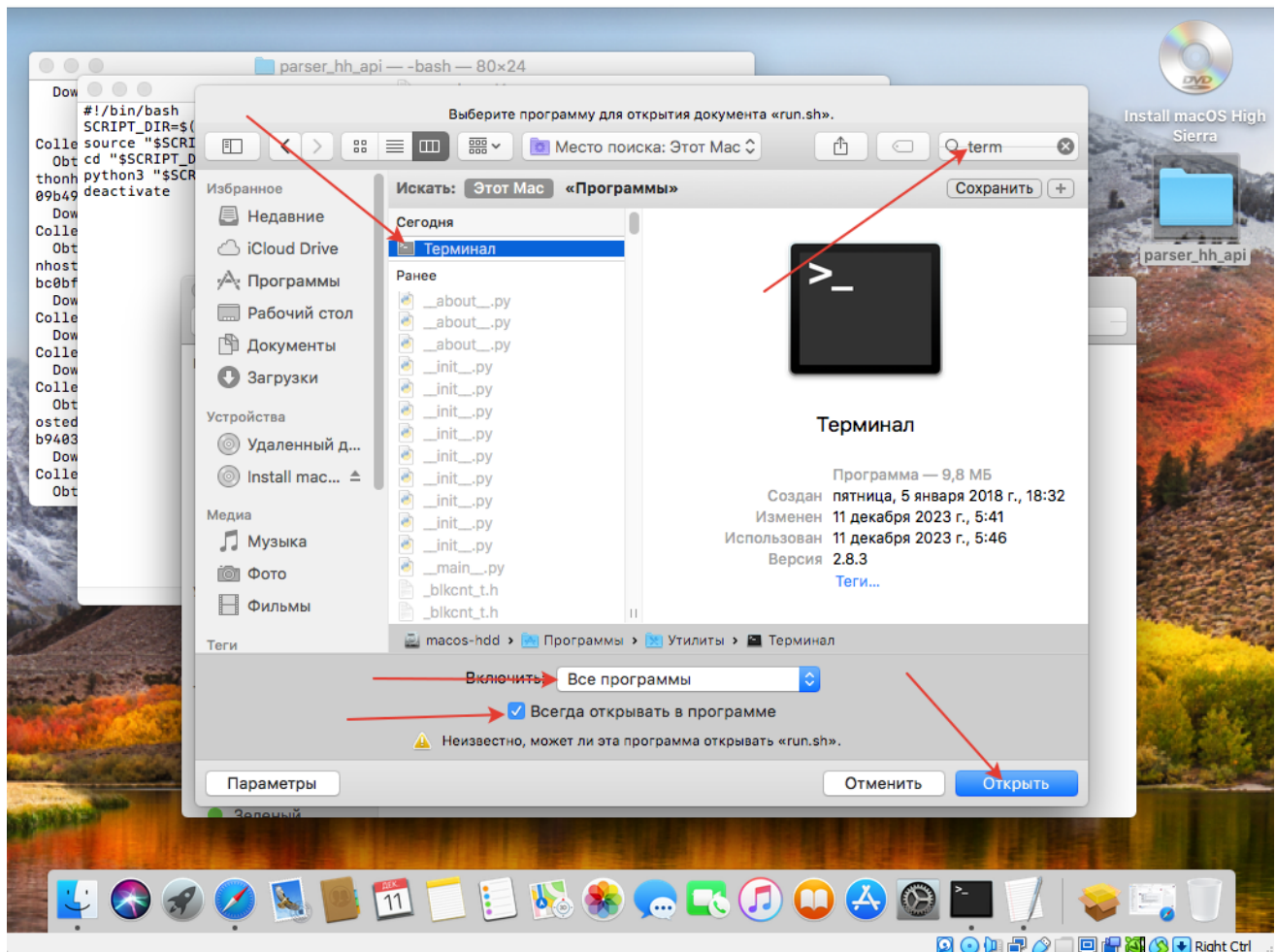
Далее нужно сделать файл запуска "run.sh" исполняемым, для этого введите в терминале `chmod +x run.sh`
После проделанных действий заходим в папку с установленным Python



и запускаем вот этот файл



3. Запускается парсер файлом run.sh, открывать файл нужно при помощи терминала, можно выбрать его для этого файла по умолчанию



При первом запуске парсера он откроет браузер, где нужно будет авторизоваться через гугл в том аккаунте, в котором вы создавали API. Все разрешения, которые он будет просить нужно дать и везде поставить галочки.

В файле proxies.json вы вводите свои прокси согласно шаблону уже там находящихся, внутри уже лежат 3 прокси, можете ими пользоваться, они будут действовать еще около 20 дней, после чего нужно будет старые удалить и добавить свои. Покупать стоит российские прокси IPv4 и использовать для них порт socks5(указан на скриншоте)

#	Адрес прокси	Порт HTTP(S)	Порт SOCKS5	Логин (user, если есть, входит в логины)	Пароль	login:password@ip:port (HTTP)
1	194.32.126.179	7484	17484	user141851	194hsw	user141851:194hsw@194.32.126.179:7484
2	195.69.150.83	7484	17484	user141851	194hsw	user141851:194hsw@195.69.150.83:7484
3	45.11.22.40	7484	17484	user141851	194hsw	user141851:194hsw@45.11.22.40:7484

Я покупал прокси на данном сайте: <https://proxys.io/ru/> , но где покупать не особо важно
Для экономии денег можно брать IPv4 Shared прокси.

В файле конфига вы указываете в tags - тэги в формате:

tags = ПервыйЗапросТэг1 ПервыйЗапросТэг2 ПервыйЗапросТэг3, ВторойЗапросТэг1 ВторойЗапросТэг2
ВторойЗапросТэг3, ТретийЗапросТэг1 ТретийЗапросТэг2

т.е. тэги для одного запроса должны быть разделены пробелами, а сами запросы - запятыми

В table указывается название таблицы, если такая таблица в вашем аккаунте уже существует, то он ее перезапишет, если такой таблицы нет - то создаст. Также в папке с парсером создается файл data.csv, это excel файл с данными которые записывают в гугл таблицу, он является промежуточным, но его можно использовать если нужно.

Покупать прокси больше чем запросов не имеет смысла, распределяются запросы автоматически, т.е. если у вас 3 прокси и 6 запросов, то на каждый прокси будет отдаваться по 2 запроса, но как минимум 1 прокси должен быть. Задержка между отправкой запроса на каждом прокси 30 сек.

Не забудьте файл с данными API переименовать в 'credentials' и положить в папку 'credentials'