

使用说明：

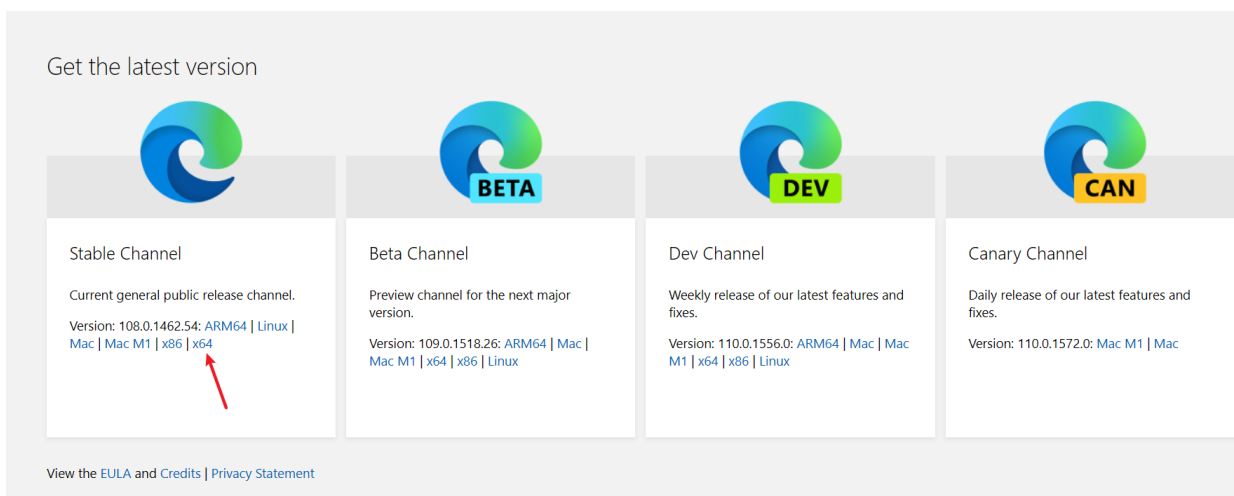
1、安装所需要的工具包，缺什么安装什么，pip install 包名

2、下载msedgedriver.exe这个很重要，我用的edge浏览器，所以下载的这个，其他浏览器同样要找适配的。

- 2.1、首先查看浏览器的版本



- 2.2、然后访问[Microsoft Edge WebDriver - Microsoft Edge Developer](#) 即可看到对应浏览器版本，选择下载。



- 然后要把压缩包解压出来得到msedgedriver.exe，文件名更改为MicrosoftWebDriver.exe。把它放入和你python解释器同级的文件夹中

LICENSE_PYTHON.txt	2021/8/31 3:02	文本文档	14 KB
MicrosoftWebDriver.exe	2022/12/16 2:29	应用程序	16,211 KB
msvcp140.dll	2020/9/8 18:10	应用程序扩展	577 KB
msvcp140_1.dll	2020/9/8 18:10	应用程序扩展	31 KB
msvcp140_2.dll	2020/9/8 18:10	应用程序扩展	190 KB
msvcp140_codecvt_ids.dll	2020/9/8 18:10	应用程序扩展	28 KB
opencv_videoio_ffmpeg455_64.dll	2022/1/10 20:39	应用程序扩展	20,398 KB
pyodbc.pyi	2021/7/4 5:41	Python 源文件	12 KB
python.exe	2021/9/17 6:12	应用程序	93 KB
python.pdb	2021/9/17 6:12	VisualStudio.pd...	460 KB

3、都安装完毕此时就可以正常使用了，注意在分词完毕后，是使用到了Hadoop平台，利用MapReduce框架编写java程序，打成jar包，运行可执行jar包得到统计词频的。考虑到适用性，这里我新加入了一个python脚本词频统计.py。可以利用它模拟执行jhadoop平台WordCount.jar程序功能。

词频统计结果输出为out.txt,可在程序中自行修改输出名字，在词频排序程序中要注意用的文件是哪个，改为out.txt（你修改的）即可

```
# @File:词频排序.py
# @Software:PyCharm

with open('../词频统计结果/outcome.txt', 'r', encoding='utf-8') as f:
```

4、可视化部分要将自己的数据放入即可

注：（合集.txt为爬取新闻的综合，手动综合）