# Web crawler workshop

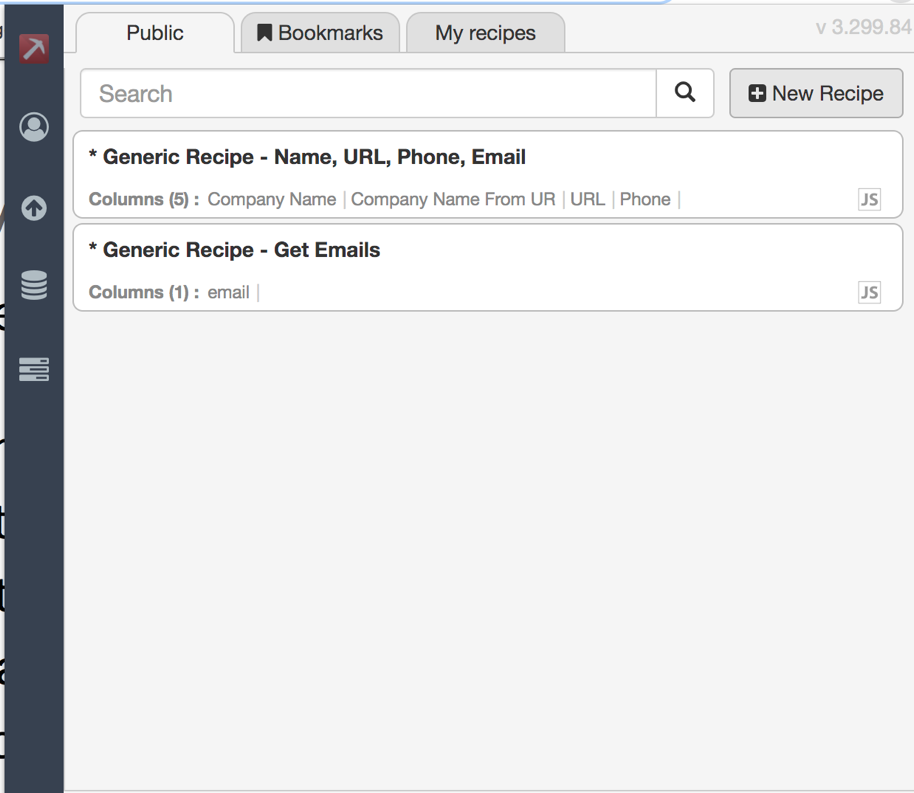
1. Plug-in tool data miner
2. Python installation and crawler
3. Crawler tool use

# Plug-in tool data miner

1. Install chrome browser
2. Prepair a google account
3. Find data miner plug-in and install it



1. Start a new recipe



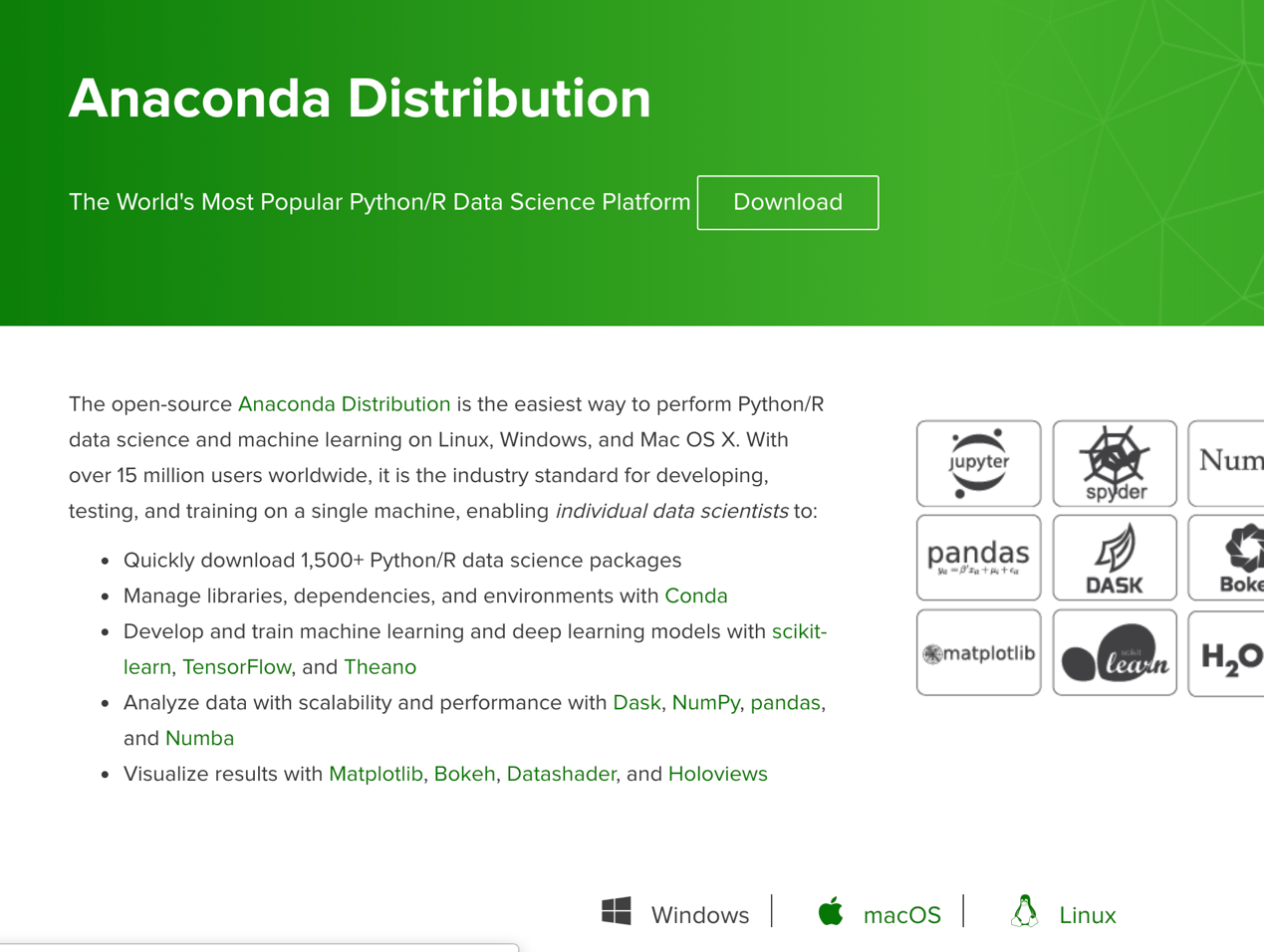
1. Tutorials available

<https://www.youtube.com/watch?v=bpyX8G7wopQ>

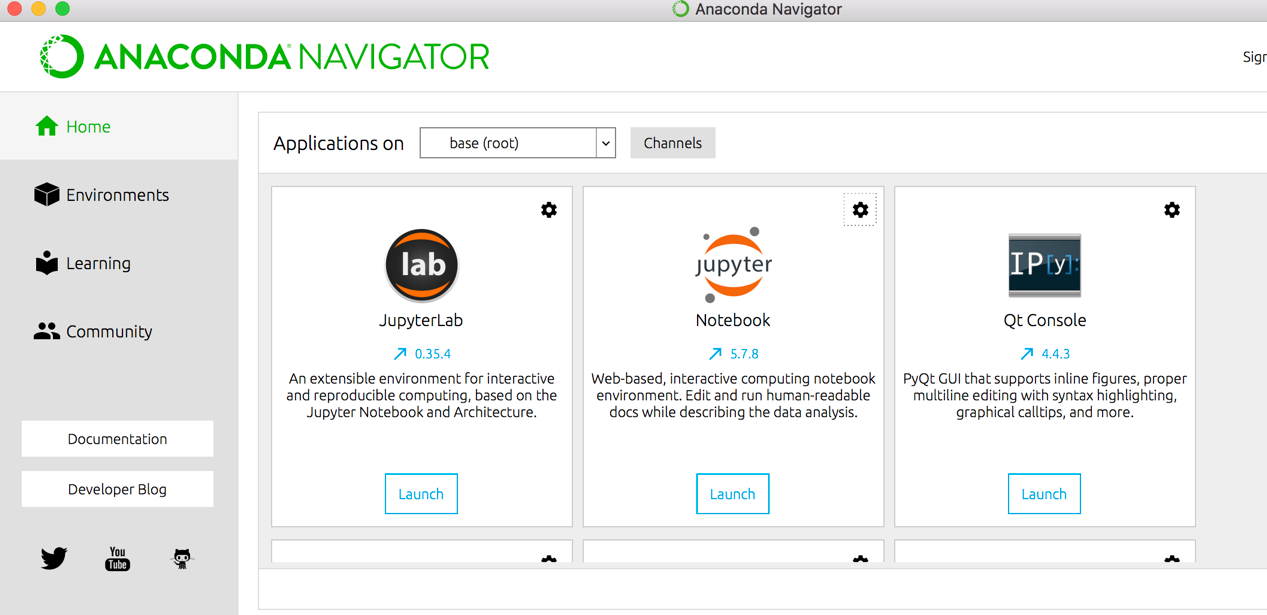
# Python installation and crawler

1. Download and install anaconda

<https://docs.anaconda.com/anaconda/install/>



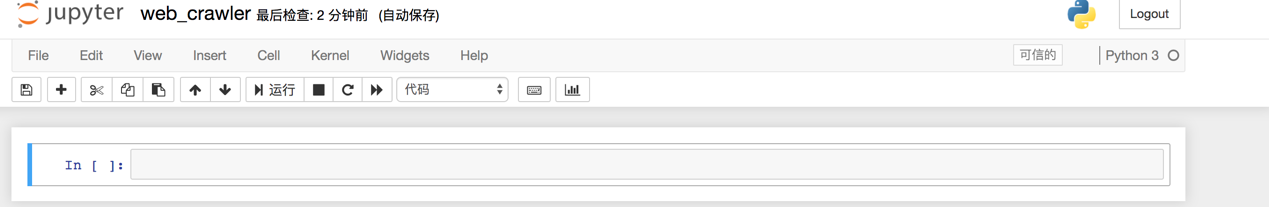
1. Open jupyter notebook



1. Start first coding experience with python
2. Create a new folder



1. Start a new python3 file



# Crawler tool use: BAZHUAYU

1. Setup software environment

For windows:

Down load from <https://www.octoparse.com/> (English version)

<https://www.bazhuayu.com/download> ) (chinese version)

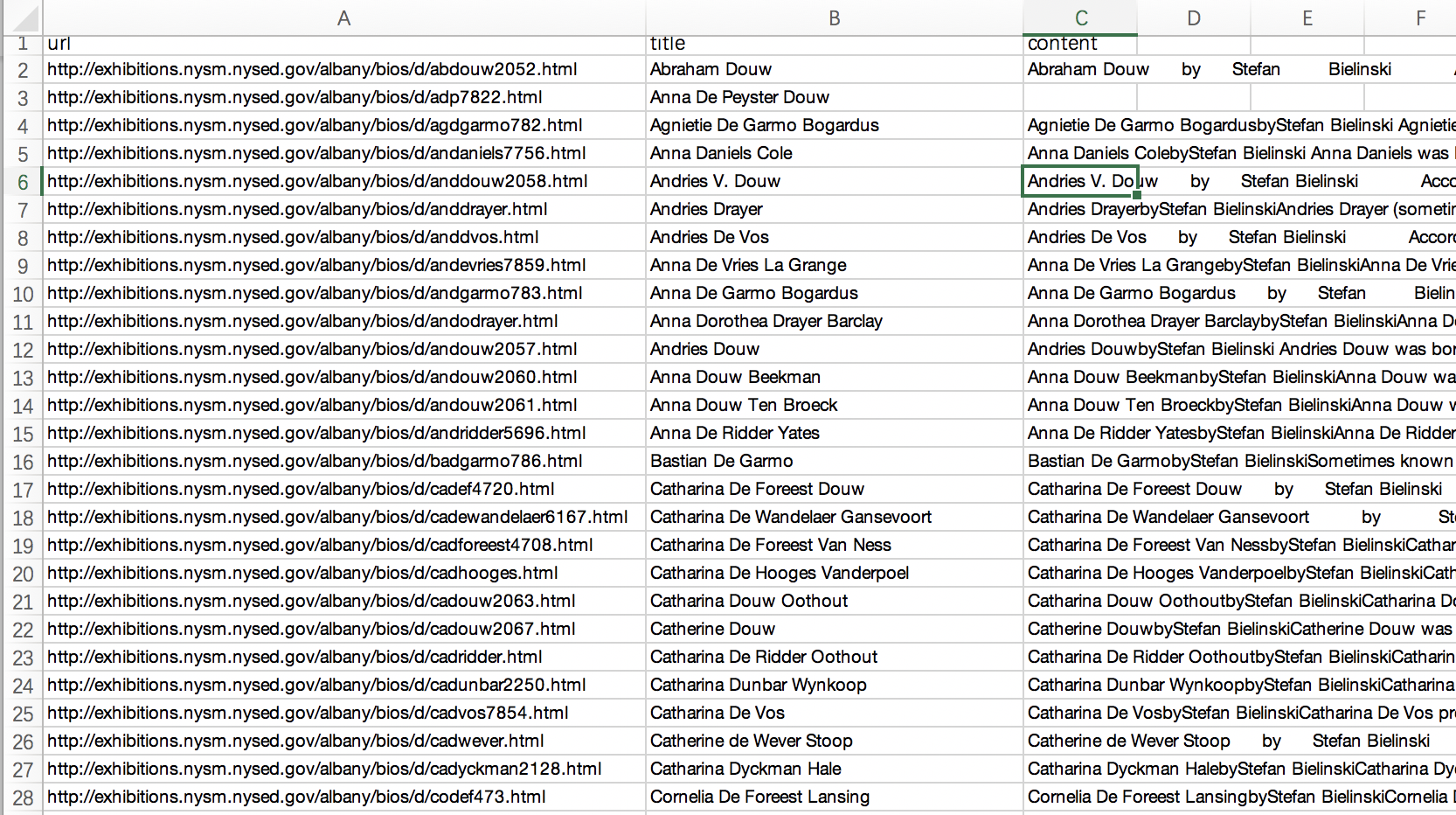
Install .net3.5 sp1

<https://www.microsoft.com/en-us/download/details.aspx?id=22>

For mac:

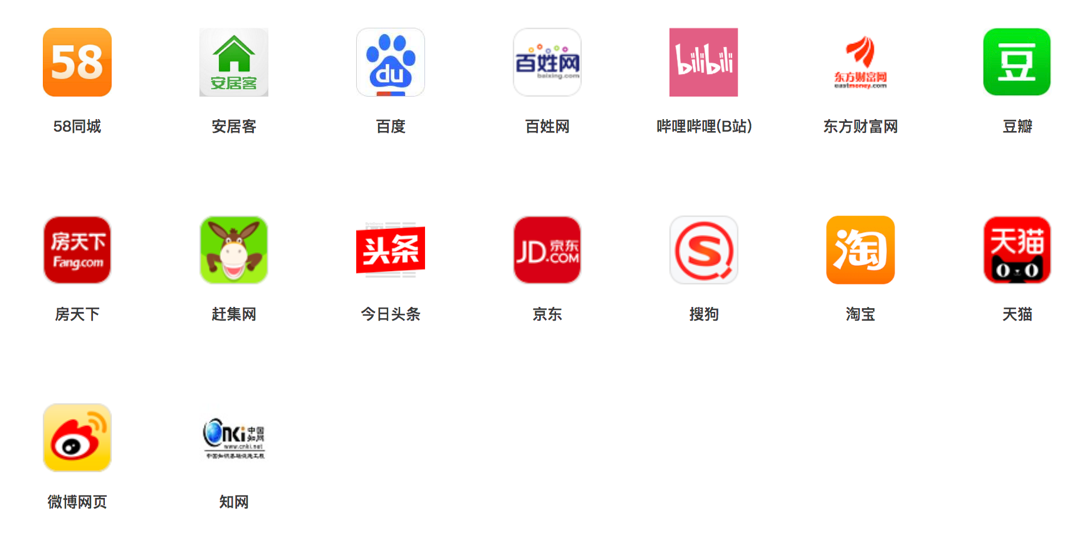
Need to first install vmware. <https://www.vmware.com/> to install win10 system

2.a demo of crawlering title and content from a web page



3. other useful information

Templets: <https://www.bazhuayu.com/app/mission/models>



Tutorial:

(english)

<https://www.youtube.com/channel/UCweDWm1QY2G67SDAKX7nreg/playlists>

(chinese)

<https://www.bazhuayu.com/tutorialIndex>