

Configuration for Labs

Bok, Jong Soon
javaexpert@nate.com
<https://github.com/swacademy/NLP>

Overview

- OS
- Tools
- Python
- Anaconda
- Google Colab
- Python Libraries
- OpenJDK



Operating Systems



Ubuntu 18.04 LTS

- 대부분의 내용은 OS에 상관없이 설치 또는 구현 가능.
- 다만, 전처리에서 형태소 분석기 Mecab의 경우에는 Windows version에 따라 설치가 어렵거나 불가할 수도.
- KoNLPy와 mecab-ko 사이의 호환에는 WSL(Windows Subsystem for Linux, Windows 10에 포함) 환경이 필요할 수도.
- 원래 PyTorch는 얼마전까지 Windows 환경을 공식적으로 지원하지 않았음.

Tools



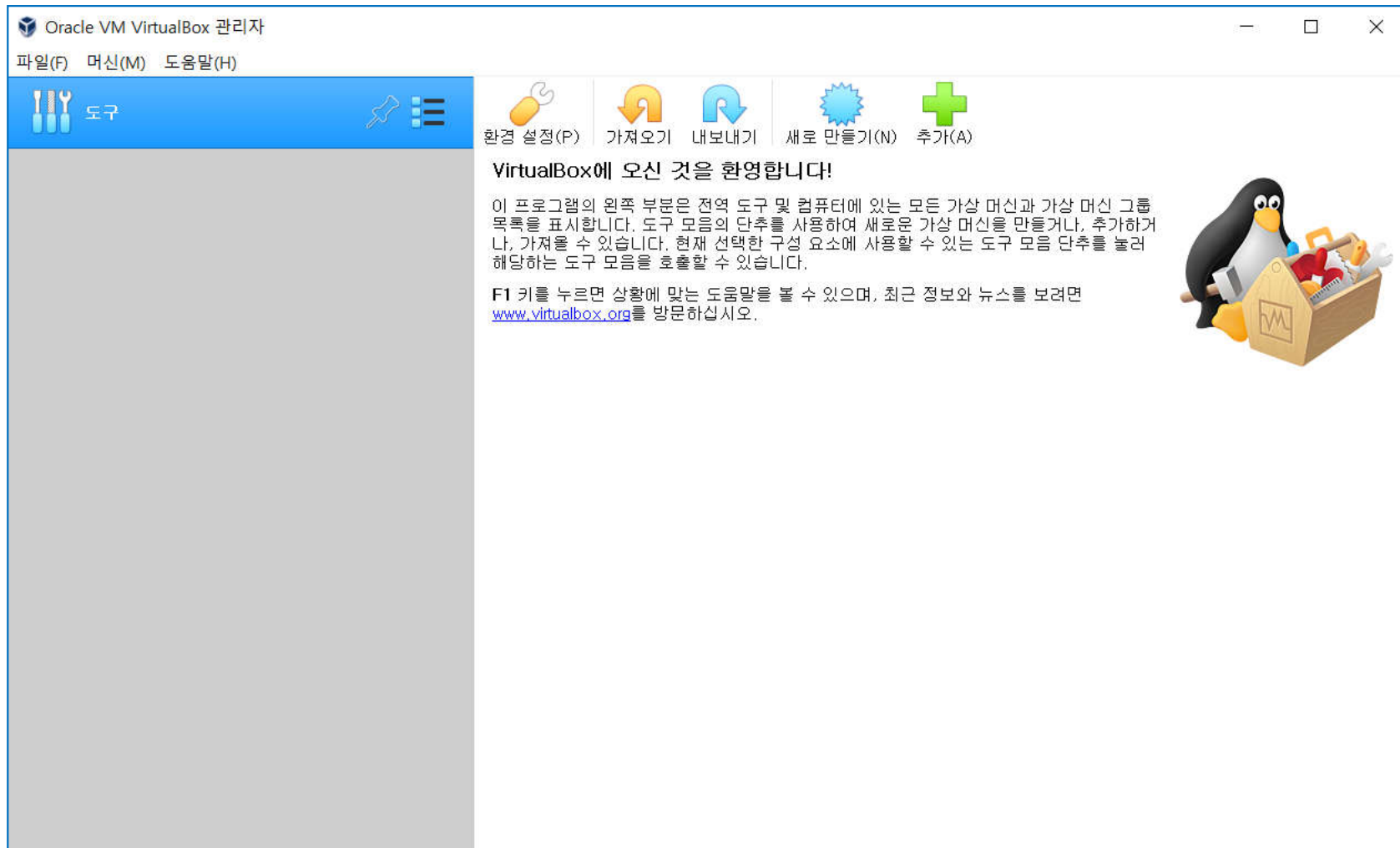
Microsoft Excel 2013 +

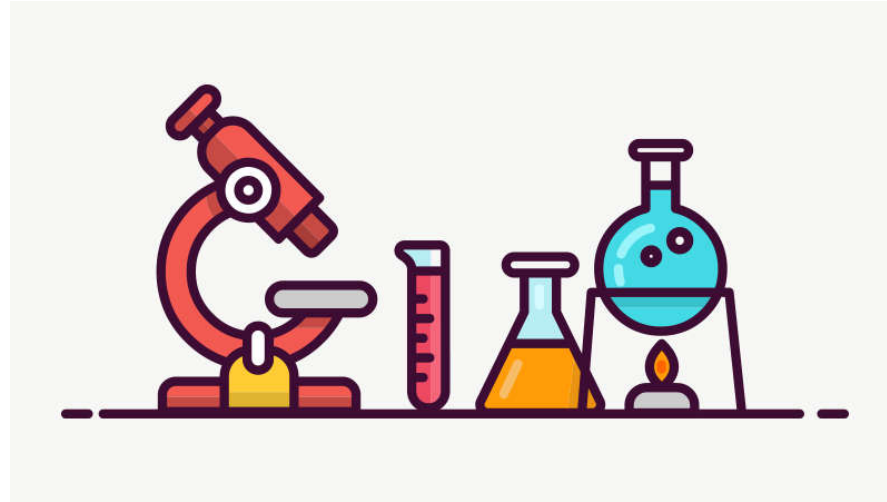
- 데이터 분석 소프트웨어에서 R과 Python 다음으로 많이 사용되는 프로그램.
- 상용.
- Table 중심의 Spreadsheet Program.
- Power Query, Power Pivot Add-in
- 대용량 Data 처리는 가능하나, 속도나 효율성 면에서 R이나 Python보다 취약함.
- 시각화는 Power BI로 진화됨.

Oracle VirtualBox

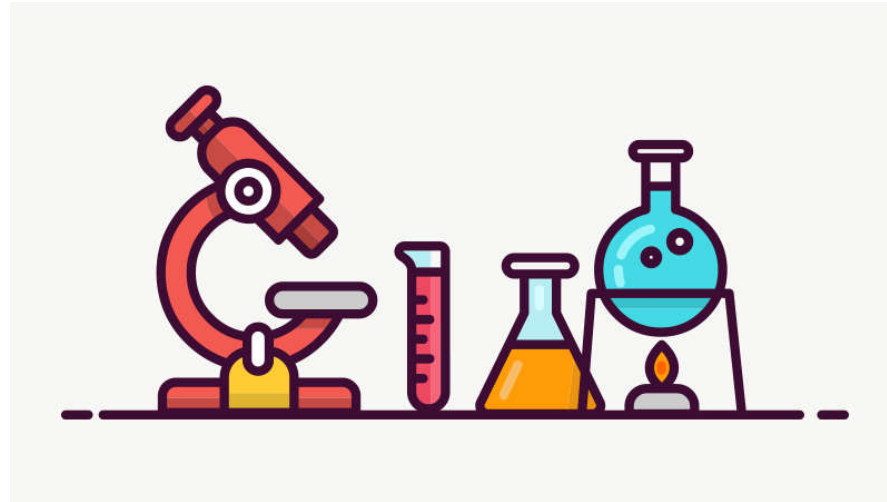
- Is a powerful x86 and AMD64/Intel64 virtualization product.
- Is freely available as Open Source Software under the terms of the GNU General Public License (GPL) version 2.
- Runs on Windows, Linux, Macintosh, and Solaris hosts.
- Supports a large number of guest operating systems.
 - Windows (NT 4.0, 2000, XP, Server 2003, Vista, Windows 7, Windows 8, Windows 10), DOS/Windows 3.x.
 - Linux (2.4, 2.6, 3.x and 4.x).
 - Solaris and OpenSolaris, OS/2, and OpenBSD.

Oracle VirtualBox (Cont.)





Lab. Oracle VirtualBox Installation & Configuration



Lab. Ubuntu 18.04 LTS Installation & Configuration

Git

workflows.



About

The advantages of Git compared to other source control systems.



Documentation

Command reference pages, Pro Git book content, videos and other material.



Downloads

GUI clients and binary releases for all major platforms.



Community

Get involved! Bug reporting, mailing list, chat, development and more.



Pro Git by Scott Chacon and Ben Straub is available to [read online for free](#). Dead tree versions are available on [Amazon.com](#).



Linux GUIs



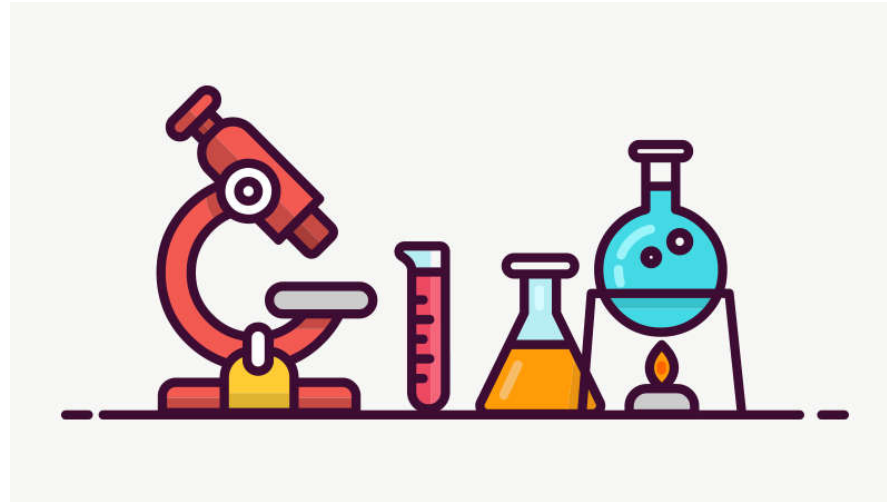
Tarballs



Mac Build



Source Code



Lab. Git and GitHub Handling

Python 3.8.x



■ Installation

- Refer to

- <https://github.com/swacademy/Python/blob/master/2.%20Configure%20for%20Lab.pdf>
- <https://realpython.com/installing-python/>

■ Configuration

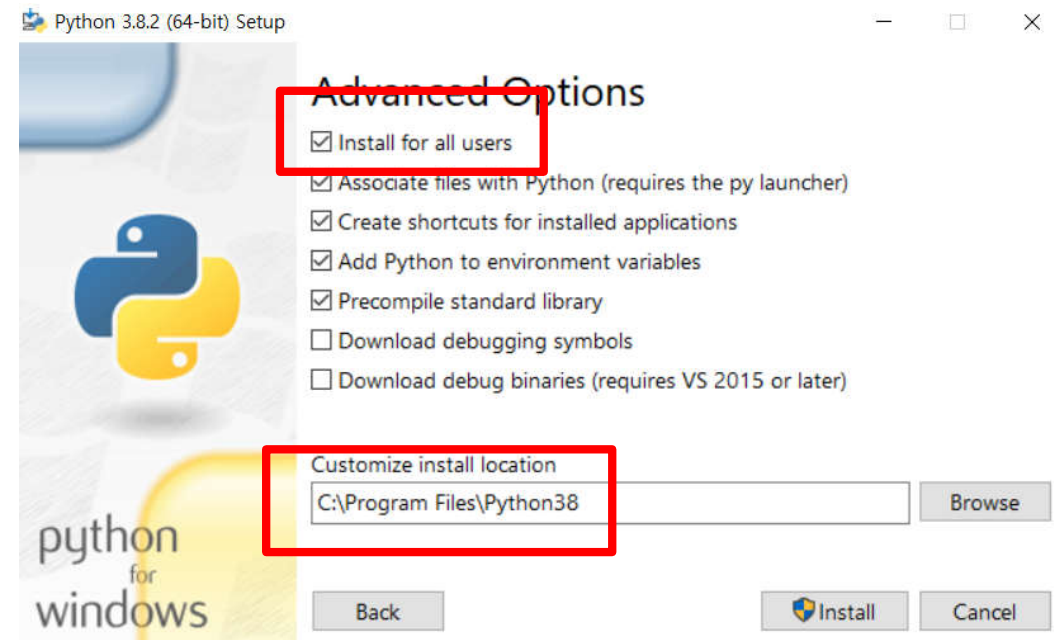
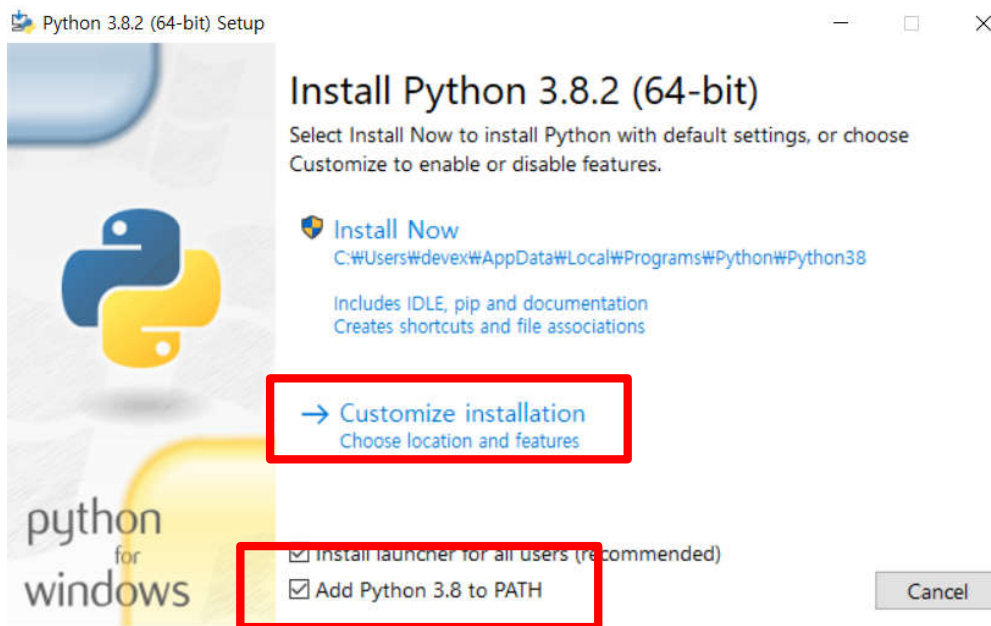
- OS PATH
- Windows : ~\Scripts
- Linux : ~/bin

```
C:\Users\devex>python -V  
Python 3.8.2
```

```
C:\Users\devex>python --version  
Python 3.8.2
```

Python

■ <https://www.python.org/>



Python (Cont.)

환경 변수

devex에 대한 사용자 변수(U)

변수	값
ChocolateyLastPath...	132332422213921823
MOZ_PLUGIN_PATH	C:\Program Files (x86)\Foxit Software\Foxit Reader\plugins\
OneDrive	C:\Users\devex\OneDrive
OneDriveConsumer	C:\Users\devex\OneDrive
Path	C:\Users\devex\AppData\Local\Microsoft\WindowsApps\C:\Program Files\Bandizip\

새로 만들기(N)... 편집(E)... 삭제(D)

시스템 변수(S)

변수	값
NUMBER_OF...	8
OS	Windows_NT
Path	C:\Program Files\Python38\Scripts\;C:\Program Files\Python38\;C:\Python38\Scripts\;C:\Python38\;C...
PATHEXT	.COM;.EXE;.BAT;.CMD;.VBS;.VBE;.JS;.JSE;.WSF;.WSH;.MSC;.PY;.PYW
PROCESSOR ...	AMD64

새로 만들기(W)... 편집(I)... 삭제(L)

확인 취소

```
C:\Users\devex>python -V
Python 3.8.2
```

```
C:\Users\devex>python --version
Python 3.8.2
```

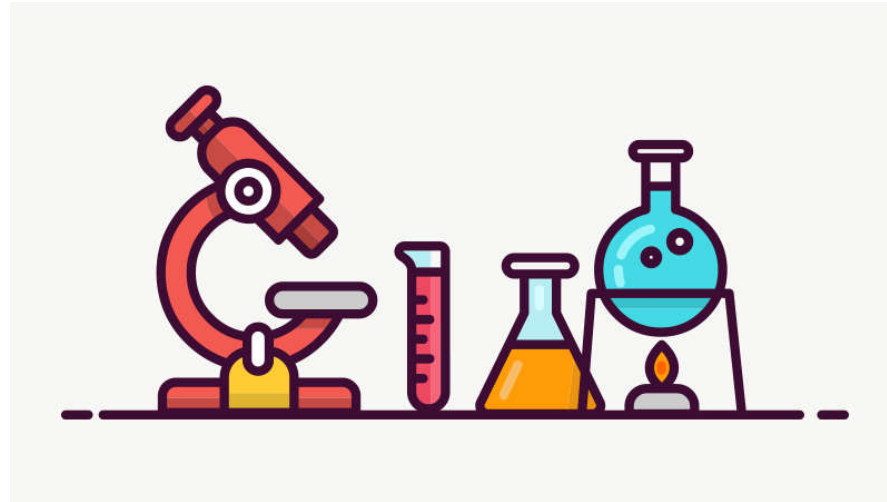
Java vs Python

JAVA	vs	PYTHON
General Purpose Solving Programing Language. Follows - write once, run everywhere		High level programming language. Focus on code readability and shorter syntax.
TYPING		
Java, uses static typing where a user is forced to choose a variable type which cannot be changed later in the programming language.		Python uses dynamic typing, allowing users to change the variable type. Thus making it easier to learn.
SPEED		
Java can be used for desktop programs and web applications. Java is faster than Python		It is not as fast as Java.
PRODUCTIVITY		
Less productive than Python, as each variable has to be defined. Thus, it is time consuming.		Writes code in fewer lines, allowing developers to be more productive.

Atom



- Is a hackable text editor.
- Built on Electron, and based on everything we love about our favorite editors.
- Designed to be deeply customizable, but still approachable using the default configuration.

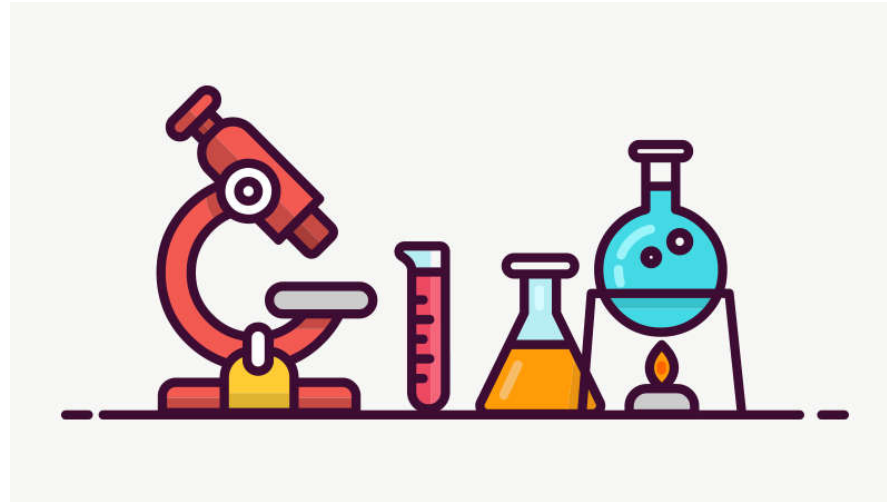


Lab. Atom 설치 및 환경설정

Microsoft Visual Studio Code



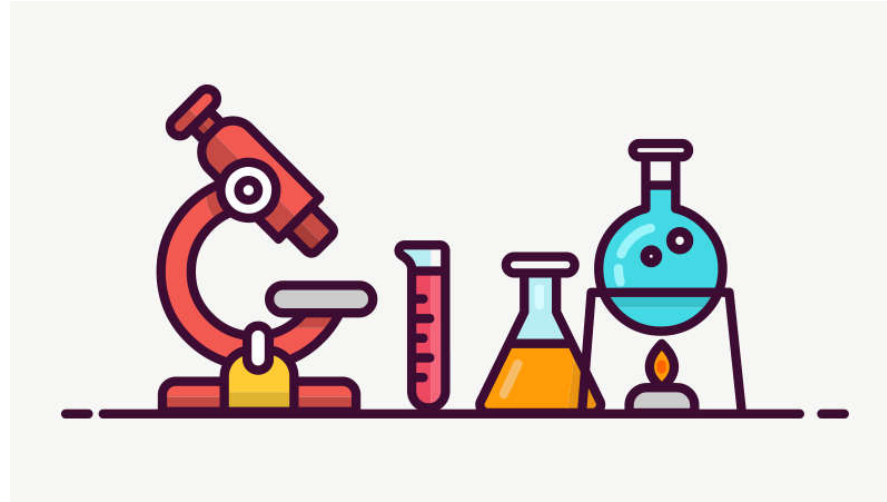
- Is a lightweight but powerful source code editor.
- Runs on your desktop.
- Is available for Windows, macOS and Linux.
- Has a rich ecosystem of extensions for other languages (such as C++, C#, Java, Python, PHP, Go) and runtimes (such as .NET and Unity).



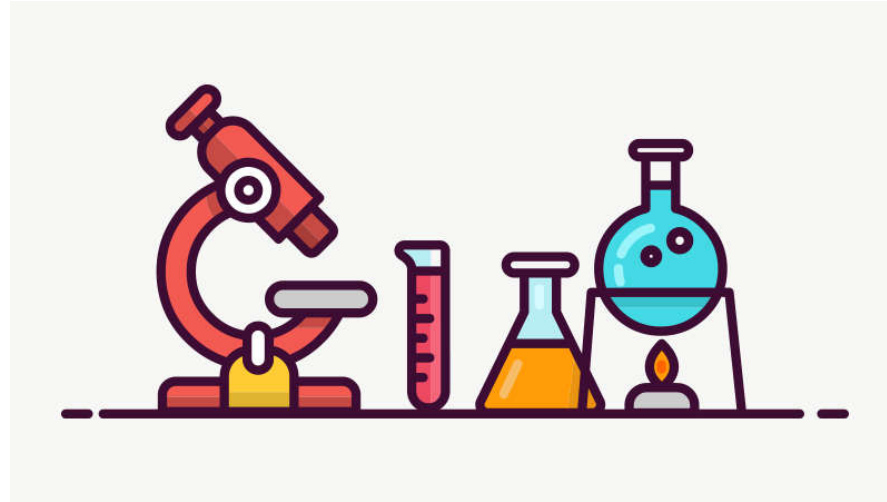
Lab. Visual Studio Code Installation and Configuration for Python

Jupyter Notebook

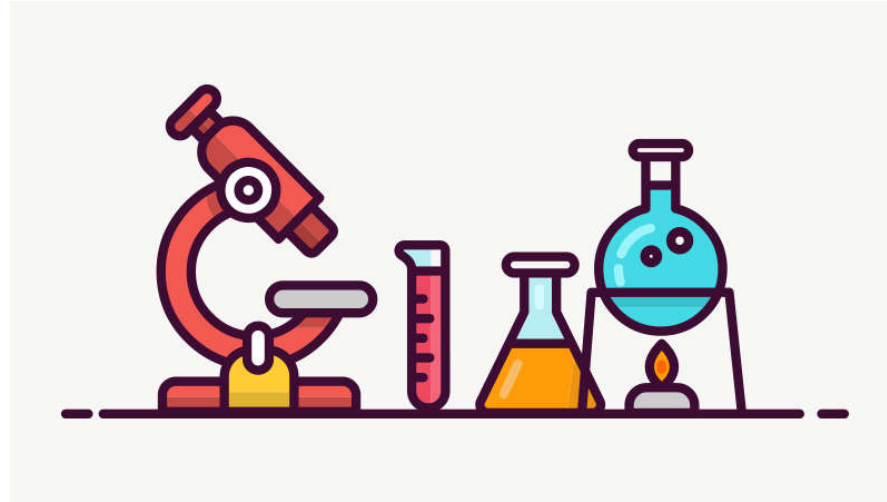
- Web Browser
 - Either Mozilla Firefox or Google Chrome will work well.
 - Try to avoid MS Explorer.
- Prerequisite : Python 3.3 or higher
- Highly recommend installing [Anaconda Distribution](#)
- Alternative for experienced Python users : Installing Jupyter with *pip*
 - `$ pip3 install --upgrade pip`
 - `$ pip3 install jupyter`



Lab. Python virtualenv에서 Jupyter Notebook 설치하기

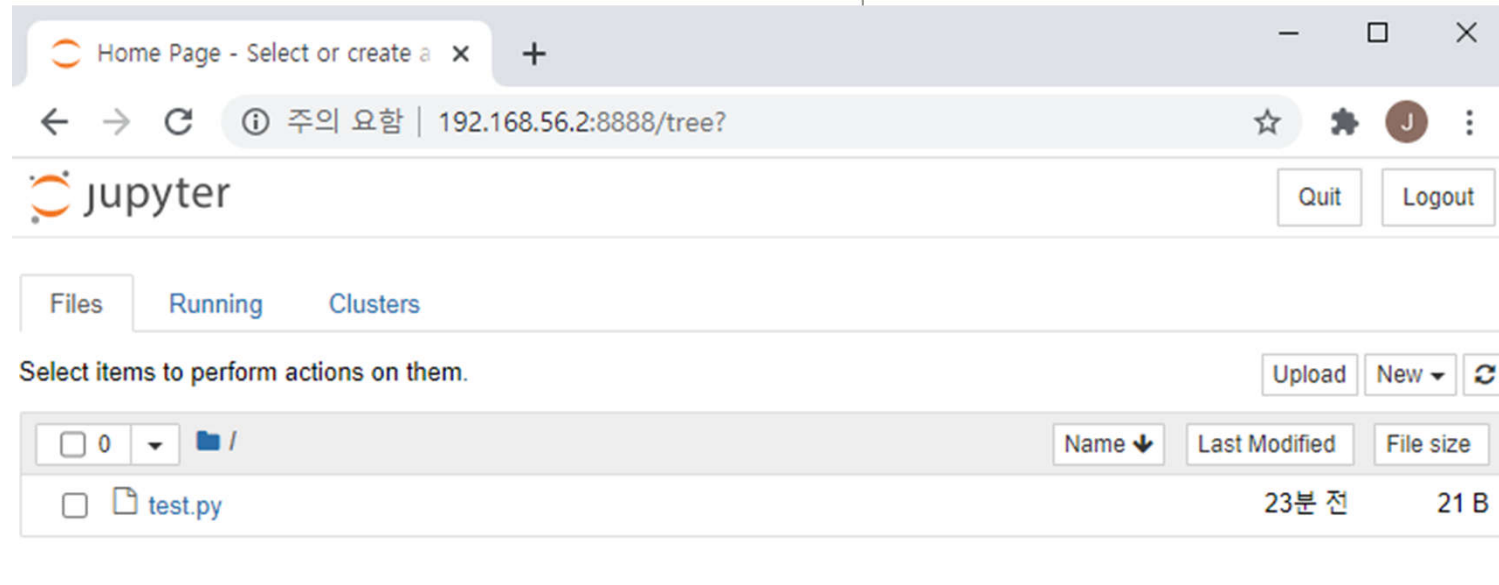
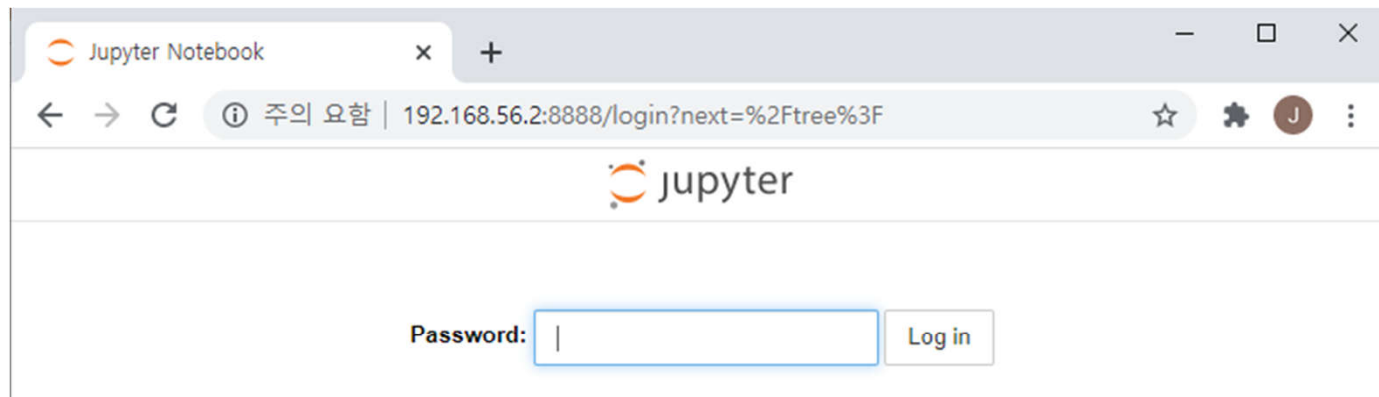


Lab. Jupyter Notebook 사용하기






Lab. Installation Jupyter Notebook on Ubuntu

Jupyter Notebook on Ubuntu

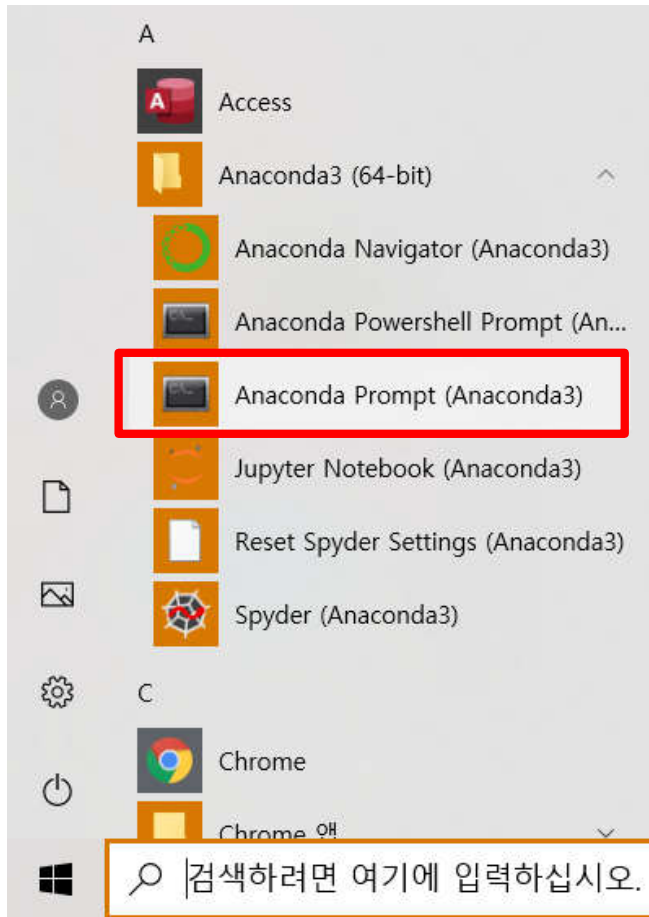


Anaconda

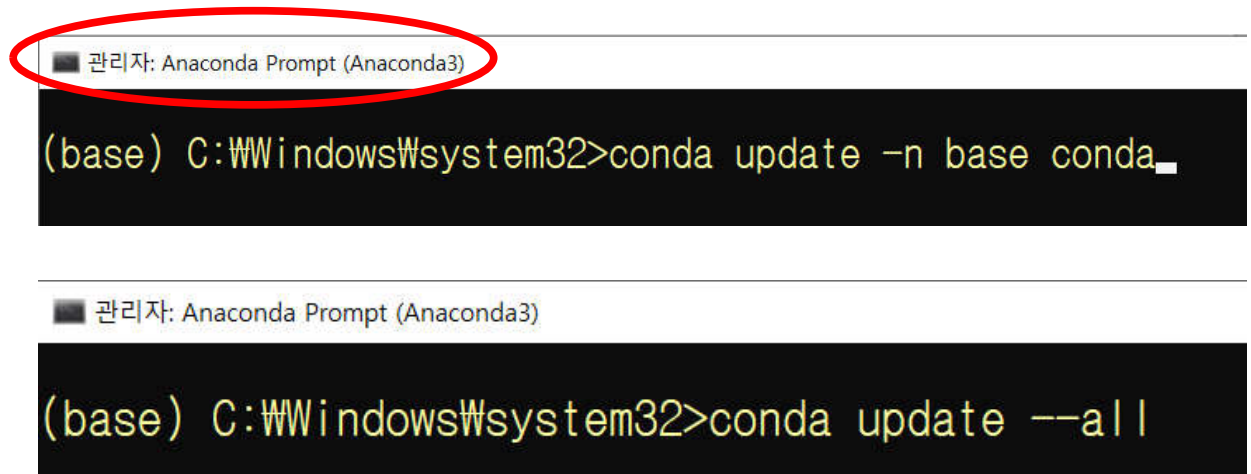
- <https://www.anaconda.com/products/individual>

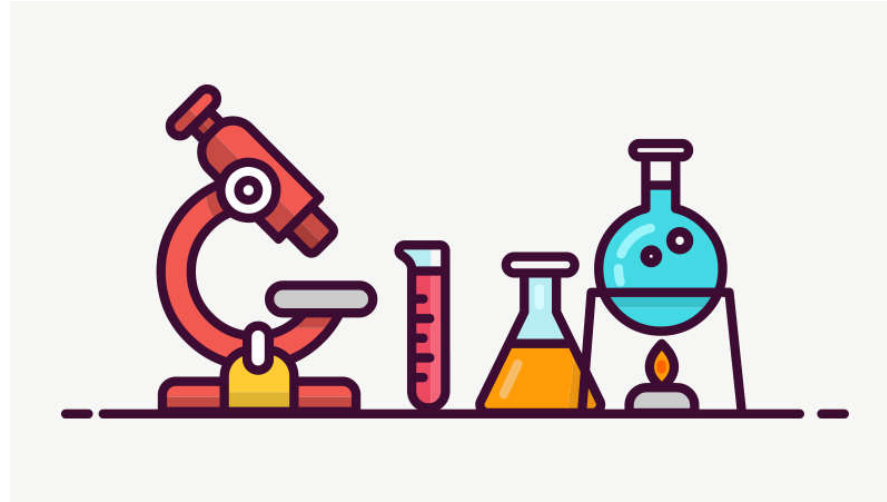
Anaconda Installers		
Windows 	MacOS 	Linux 
Python 3.7	Python 3.7	Python 3.7
64-Bit Graphical Installer (466 MB)	64-Bit Graphical Installer (442 MB)	64-Bit (x86) Installer (522 MB)
32-Bit Graphical Installer (423 MB)	64-Bit Command Line Installer (430 MB)	64-Bit (Power8 and Power9) Installer (276 MB)
Python 2.7	Python 2.7	Python 2.7
64-Bit Graphical Installer (413 MB)	64-Bit Graphical Installer (637 MB)	64-Bit (x86) Installer (477 MB)
32-Bit Graphical Installer (356 MB)	64-Bit Command Line Installer (409 MB)	64-Bit (Power8 and Power9) Installer (295 MB)

Anaconda (Cont.)



- 설치 후 Anaconda Prompt에서 Python Packages를 최신 버전으로 Update





Lab. Anaconda를 이용한 Jupyter Notebook 설치하기

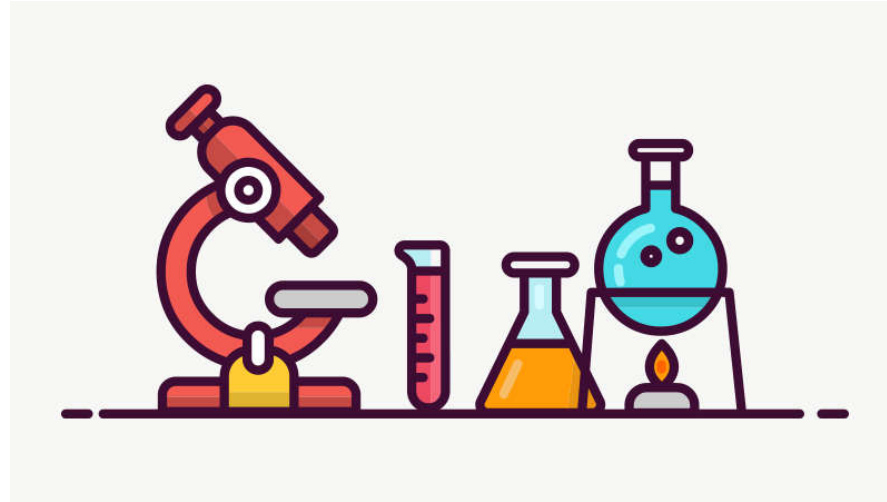
Google Colab



- Google Colaboratory
- Google Drive + Jupyter Notebook
- <https://colab.research.google.com/>
- Computer 사양(2020년 7월 기준)
 - Ubuntu 18.04.3 LTS
 - Intel Xeon CPU 2.30GHz, 6 cpu, cache 46MB
 - 메모리 13.3GB
 - GPU(Tesla T4)
 - TPU도 사용 가능
 - Python 3.6.9

Google Colab (Cont.)

- Internet을 통해 Jupyter Notebook 형식 개발 환경.
- <https://colab.research.google.com/>
- 사용법
 - 동영상 : <https://youtu.be/inN8seMm7UI>
- 김태영님 Blog :
 - https://tykimos.github.io/2019/01/22/colab_getting_started/
- Colab에서 File Upload/Download 방법
 - <http://www.dreamy.pe.kr/zbxe/CodeClip/3769485>



Lab. Google Colab 사용하기

Libraries

- Anaconda를 설치하면 기본적으로 **Numpy, Pandas, Jupyter notebook, scikit-learn, matplotlib, seaborn, nltk** 등이 이미 설치되어 있다.
- 하지만 Anaconda 설치하지 않고 단순히 Python만 설치된 상태이면 위에서 언급한 모든 패키지를 *pip*로 설치해야 한다.

Deep Learning Framework

- TensorFlow
- Keras
- PyTorch (Option)



TensorFlow



Keras



PyTorch

Libraries (Cont.)

■ Tensorflow

```
instructor@Ubuntu-Desktop:~$ ipython
Python 3.7.6 (default, Jan  8 2020, 19:59:22)
Type 'copyright', 'credits' or 'license' for more information
IPython 7.12.0 -- An enhanced Interactive Python. Type '?' for help.

In [1]: import tensorflow as tf

In [2]: tf.__version__
Out[2]: '2.2.0'
```

■ Keras

```
instructor@Ubuntu-Desktop:~$ ipython
Python 3.7.6 (default, Jan  8 2020, 19:59:22)
Type 'copyright', 'credits' or 'license' for more information
IPython 7.12.0 -- An enhanced Interactive Python. Type '?' for help.

In [1]: import keras

In [2]: keras.__version__
Out[2]: '2.4.3'
```

Libraries (Cont.)

■ Gensim

```
instructor@Ubuntu-Desktop:~$ ipython
Python 3.7.6 (default, Jan  8 2020, 19:59:22)
Type 'copyright', 'credits' or 'license' for more information
IPython 7.12.0 -- An enhanced Interactive Python. Type '?' for help.

In [1]: import gensim

In [2]: gensim.__version__
Out[2]: '3.8.0'
```

■ Scikit-learn

```
instructor@Ubuntu-Desktop:~$ ipython
Python 3.7.6 (default, Jan  8 2020, 19:59:22)
Type 'copyright', 'credits' or 'license' for more information
IPython 7.12.0 -- An enhanced Interactive Python. Type '?' for help.

In [1]: import sklearn

In [2]: sklearn.__version__
Out[2]: '0.22.1'
```

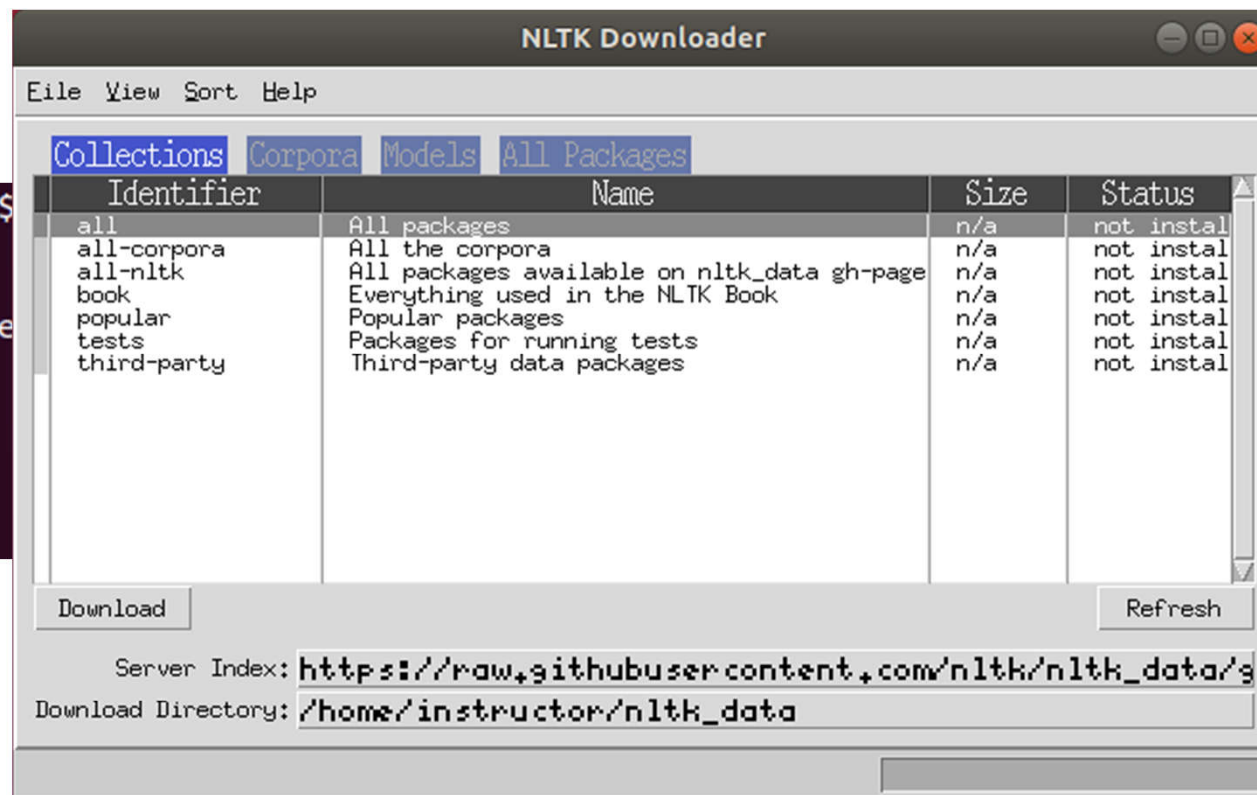
Libraries (Cont.)

■ nltk

```
instructor@Ubuntu-Desktop:~$  
Python 3.7.6 (default, Jan  
Type 'copyright', 'credits'  
IPython 7.12.0 -- An enhance
```

```
In [1]: import nltk
```

```
In [2]: nltk.__version__  
Out[2]: '3.4.5'
```



```
In [1]: import nltk
```

```
In [2]: nltk.download()  
showing info https://raw.githubusercontent.com/nltk/nltk_data/gh-pages/index.xml
```

Libraries (Cont.)

■ konlpy

```
instructor@Ubuntu-Desktop:~$ conda install konlpy
Collecting package metadata (current repodata.json): done
Solving environment: failed with initial frozen solve. Retrying with flexible solve.
Collecting package metadata (repodata.json): done
Solving environment: failed with initial frozen solve. Retrying with flexible solve.

PackagesNotFoundError: The following packages are not available from current channels:

- konlpy

Current channels:

- https://repo.anaconda.com/pkgs/main/linux-64
- https://repo.anaconda.com/pkgs/main/noarch
- https://repo.anaconda.com/pkgs/r/linux-64
- https://repo.anaconda.com/pkgs/r/noarch

To search for alternate channels that may provide the conda package you're
looking for, navigate to

    https://anaconda.org

and use the search bar at the top of the page.
```

Libraries (Cont.)

■ Konlpy

- Installation OpenJDK 8 higher
- \$JAVA_HOME
- Installation Jpype1

```
instructor@Ubuntu-Desktop:~$ conda -V
conda 4.8.3
instructor@Ubuntu-Desktop:~$ conda install -c conda-forge jpype1
Collecting package metadata (current_repodata.json): done
Solving environment: done

## Package Plan ##

  environment location: /home/instructor/anaconda3

  added / updated specs:
    - jpype1

The following packages will be downloaded:
```

package	build		
conda-4.8.3	py37hc8dfbb8_1	3.0 MB	conda-forge
jpype1-0.7.5	py37h99015e2_0	2.8 MB	conda-forge
python_abi-3.7	1_cp37m	4 KB	conda-forge
Total:		5.8 MB	

Libraries (Cont.)

■ Konlpy

```
instructor@Ubuntu-Desktop:~$ sudo pip3 install konlpy
[sudo] password for instructor:
WARNING: The directory '/home/instructor/.cache/pip/http' or its parent director
y is not owned by the current user and the cache has been disabled. Please check
the permissions and owner of that directory. If executing pip with sudo, you ma
y want sudo's -H flag.
WARNING: The directory '/home/instructor/.cache/pip' or its parent directory is
not owned by the current user and caching wheels has been disabled. check the pe
rmissions and owner of that directory. If executing pip with sudo, you may want
sudo's -H flag.
Collecting konlpy
  Downloading https://files.pythonhosted.org/packages/85/0e/f385566fec837c0b83f2
16b2da65db9997b35dd675e107752005b7d392b1/konlpy-0.5.2-py2.py3-none-any.whl (19.4
MB)
    | ████████████████████████████████████████████████████████████████████████████ | 19.4MB 4.0MB/s
Collecting beautifulsoup4==4.6.0 (from konlpy)
  Downloading https://files.pythonhosted.org/packages/9e/d4/10f46e5cfac773e22707
237bfcd51bbffeaf0a576b0a847ec7ab15bd7ace/beautifulsoup4-4.6.0-py3-none-any.whl (
86kB)
```