Nanyang Technological University
H6751 Text and Web Mining
Lab: Installation and Text Preprocessing (I)

In this lab, you will learn how to set up the environment for Python programming, and explore various text preprocessing techniques.

1. Install Python tools on a Local Machine.

– **Anaconda** (https://www.continuum.io/downloads) has been installed on lab machines.
- Windows 64-bit Graphical Installer: Anaconda3-x.x.x-Windows-x86_64.exe
– If you cannot run Jupyter Notebook, reinstall Anaconda:
- Uninstall anaconda by executing E (or D):\Anaconda3**Uninstall-Anaconda.exe**
- And reinstall anaconda using **Anaconda3-x.x.x-Windows-x86_64.exe** with the following options:
  - Just Me (recommended)
  - Destination folder: E (or D):\Anaconda3

2. Running the Jupyter Notebook
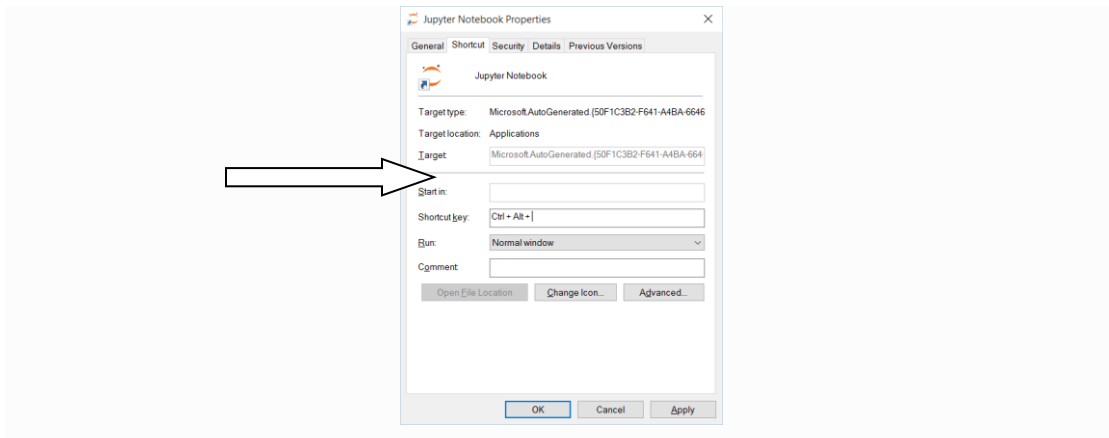
2.1. Launching *Jupyter Notebook App*

The Jupyter Notebook App can be launched by clicking on the *Jupyter Notebook* icon on task bar (or type in "**Jupyter Notebook**" in Window Menu) or by typing in a terminal (*cmd* on Windows): *ipython notebook*

This will launch a new browser window (or a new tab) showing the Notebook Dashboard, a sort of control panel that allows the selection of which notebook to open.
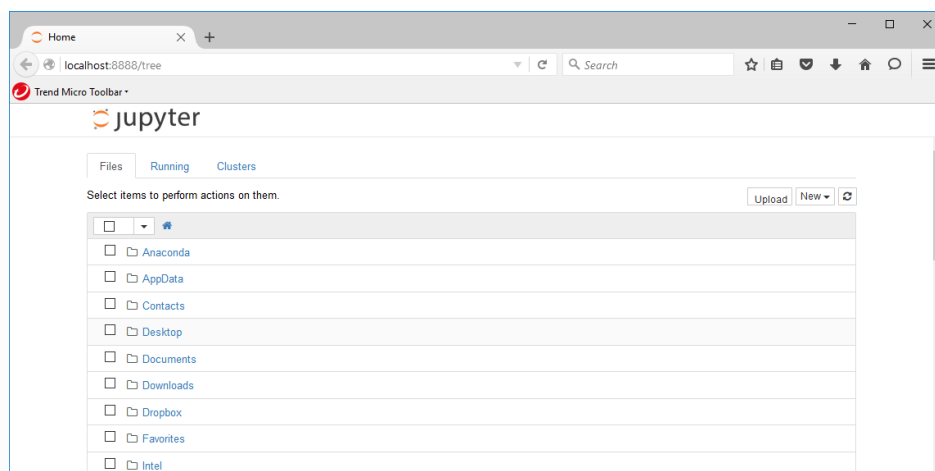
When started, the Jupyter Notebook App can access only files within its start-up folder (including any sub-folder). If you store the notebook documents in a subfolder of your user folder no configuration is necessary. Otherwise, you need to choose a folder which will contain all the notebooks and set this as the Jupyter Notebook App start-up folder. Follow the following steps if you want to change the Jupyter Notebook startup folder:

- Copy the *IPython Notebook* launcher from the menu to the desktop.

- Right click on the new launcher and change the "Start in" field by pasting the full path of the folder which will contain all the notebooks.


- Double-click on the *Jupyter Notebook* desktop launcher to start the Jupyter Notebook App, which will open in a new browser window (or tab). Note also that a secondary terminal window (used only for error logging and for shut down) will be also opened. If only the terminal starts, try opening this address with your browser: **http://localhost:8888/**.

2.2. Shutting down the *Jupyter Notebook App*

The Jupyter Notebook App is a server that appears in your browser at a default address (*http://localhost:8888*). Closing the browser (or the tab) will not shut down the server. You can reopen the previous address and the Jupyter Notebook App will be redisplayed. To completely shut it down you need to **close the associated terminal**.

You can run many copies of the Jupyter Notebook App (double-clicking on the Jupyter Notebook icon) and they will show up at a similar address (only the number after ":", which is the port, will increment for each new copy).
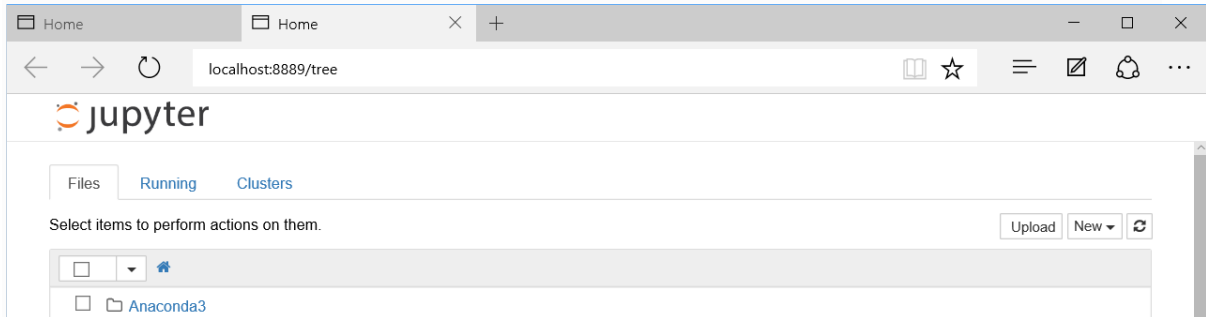

Since with a single Jupyter Notebook App you can already open many notebooks, running multiple copies of Jupyter Notebook App is not recommended.

2.3 Opening and Closing a notebook

You can open a new notebook by clicking on the "**New**" list box and selecting "**Python 3** or **Python [Default]**"
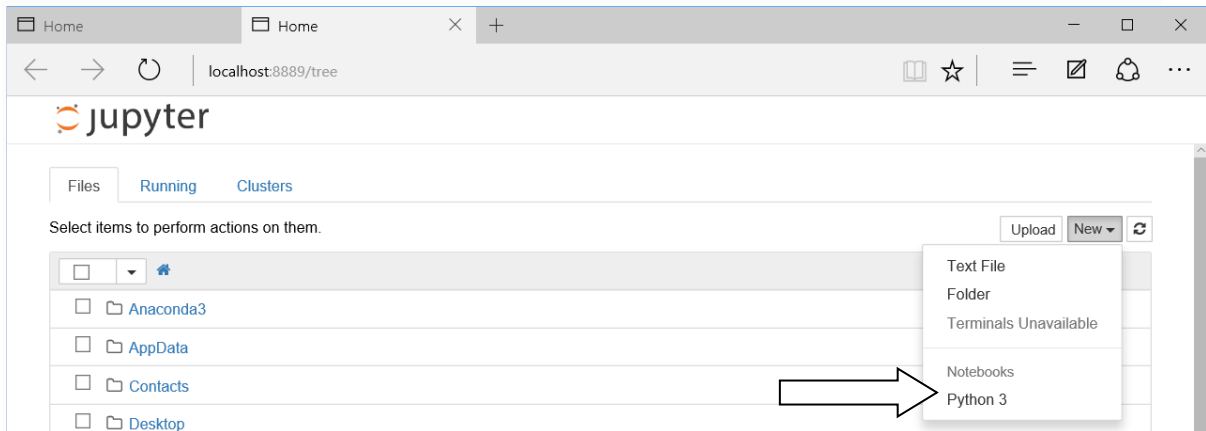
This will open a new notebook, starting the computational engine called the kernel. From here, codes could be written and executed.
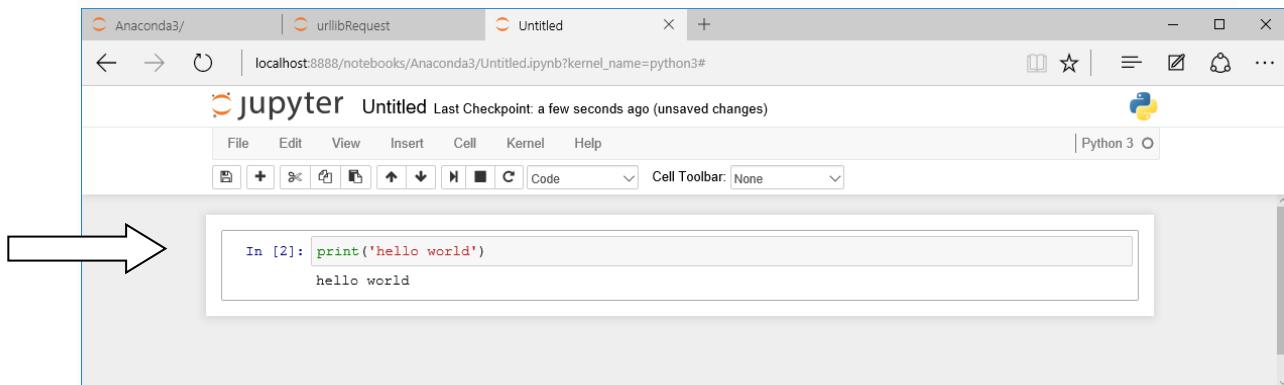
When a notebook is opened, its "computational engine" (called the kernel) is automatically started. Closing the notebook browser tab, will not shut down the kernel, instead the kernel will keep running until is explicitly shut down.

To shut down a kernel, go to the associated notebook and click on menu *File -> Close and Halt*. Alternatively, the Notebook Dashboard has a tab named *Running* that shows all the running notebooks (i.e. kernels) and allows shutting them down (by clicking on a *Shutdown* button).

2.4 Executing a notebook

Download the notebook you want to execute and put it in your notebook folder (or a sub-folder of it).

- In the Notebook Dashboard navigate to find the notebook: clicking on its name will open it in a new browser tab.
- You can run the notebook document step-by-step (one cell a time) by pressing *shift + enter*.
- You can run the whole notebook in a single step by clicking on the menu *Cell -> Run All*.
- To restart the kernel (i.e. the computational engine), click on the menu *Kernel -> Restart*. This can be useful to start over a computation from scratch (e.g. variables are deleted, open files are closed, etc...).

3. Running the Jupyter Notebook – wk2-text-preprocessing-v1.ipynb

3.1. Copy the **notebook files** downloaded from the class website to
C:\Users*youraccountname*\documents\python\H6751

3.2 Open a Jupyter Notebook

- File – Open …, **wk2-text-preprocessing-v1.ipynb in C:\Users***youraccountname* **\documents\python\H6752\wk2-labfiles**
- Run the notebook document step-by-step (one cell a time)