Anaconda, NLP Environment, pip

STEP1 Install Anaconda & Python

Nearly all NLP routines are written in Python. You will need Anaconda and Python installed in your machine for the NLP Suite to run.

STEP1 will do that for you, downloading and installing the latest Anaconda and Python. **No hard done if you already have Anaconda and Python installed on your machine and you run STEP1. The script will simply exit.**

STEP2: Install all Python (and Java) packages

The NLP Suite scripts rely on a wide range of specialized Python packages. These packages need to be installed in the NLP environment.

STEP2 will do that automatically for you installing all Python packages required by the NLP Suite to run.

requirements.txt

The script basically runs the following pip command.

pip install – r requirements.txt

requirements.txt, housed in the main NLP Suite directory, contains a complete list of all Python packages required by the NLP Suite to run.

The pip command: What is pip?

pip is a package-management system written in Python used to install and manage software packages. Thus, if you open a command line/prompt and type

pip install pandas

```
(NLP) C:\Program Files (x86)\NLP\src>pip install pandas_
```

The command will install the pandas package.

Install new Python packages via pip

STEP2 will install all Python packages required to run the NLP Suite. But new algorithms are constantly added to the NLP Suite. These new algorithms may require Python packages not originally listed in requirements.txt when you ran STEP2.

When that happens, the NLP Suite detects a missing package and will warn you that you need to use the pip install command in command line/terminal to install that specific package. For instance,

pip install pandas

Anaconda sets up different projects in specific environments. The NLP Suite is automatically setup in an environment called NLP.

pip install Python package version

Python scripts are VERY sensitive to the package version used to create the script. To ensure that the package will work on your computer, use the following command to install the required package version.

To install a specific version of a package run the command

pip install pandas~=1.2.1

Permission error

The installation of some packages (e.g., pdfminer.six) may give you a permission error. In that case, add --user to the pip command, for instance,

pip install pdfminer.six --user

pip command does not work

You have installed the pandas package using the pip install command. But when you run the NLP Suite gives you the error that pandas was not installed. Yet, you have done that a number of times following the instructions given by the NLP Suite. **What is going on?**

All packages in Python are installed in specific Anaconda environments. The NLP Suite is setup in the NLP environment. If you open a command line/terminal most likely it will be open on the *base* environment.

```
(base) C:\Program Files (x86)\NLP\src>
```

If you installed pandas in the base environment, it will not be recognized by the NLP Suite. You need to install it in the NLP environment.

If base appears in command line/terminal, you are in the wrong environment. But how do you get to the right NLP environment?

STEP3: Setup the NLP Anaconda environment for pip

The easiest way to open a command line/terminal in the right NLP environment is by running STEP3. **You need to do that only once (but no harm if you accidentally do that repeatedly).** After running STEP3, whenever you open a command line/terminal and type NLP (caps) you will be placed automatically in the right NLP environment.

```
(NLP) C:\Program Files (x86)\NLP\src>
```

At that point you can successfully pip install any Python package to be used by the NLP Suite.

You can also switch to the right NLP environment by typing

conda activate NLP

But then you will also need to change directory (CD) to the src directory under the main NLP Suite directory.

STEP3 will run automatically both commands for you. Much easier!

pip command STILL does not work: Mac users (bash and zsh)

You are in the right NLP environment, you run the pip install pandas command and the NLP Suite continues to warn you that pandas is not installed. **What is going on?**

Since macOS 10.15 (Catalina) the default macOS shell has been switched from bash to zsh. The NLP-Suite setup script has been optimized for zsh not bash. If you encounter errors with the installation of some Python packages (e.g., Stacy), most likely you are running bash. To ensure that you are running zhs open a command line/terminal and type

chsh -s /bin/zsh

If you are running bash, type

conda init bash

Then run STEP2 again.

Useful conda commands

conda info --envs

> will give you a list of the environments on your machine; you will see at least two environments, base and NLP.

conda list

> will give you a list of all Python packages installed in a specific environment; make sure to be in the right environment by using conda activate.

List all Python packages

Listing all packages installed in an environment

The following command will list all packages installed in the active environment:

conda list

If you run the command conda list from the anaconda prompt, rather than from a specific environment, you will list all installed packages in all environments.

conda activate NLP

> will place you in a specific environment (e.g., NLP)

conda env list

> will list all environments

conda env remove --name NLP Suite

  will delete an environment (i.e., NLP Suite) wrongly created or no longer necessary.

conda create -n NLP -y

       will re-create the NLP environment that you may have accidentally deleted. You will then need to run *conda activate NLP* and *CD* to the NLP folder then *python -m pip install -r requirements.txt* or *python3 -m pip install -r requirements.txt* to reinstall all packages and be back to where you were before deleting the NLP environment.

Useful pip commands

pip uninstall pandas

       will uninstall a package (e.g., pandas) in the selected environment

pip show pandas

       will give you the version of any package (e.g., pandas) installed in the selected environment