

remote-course-setup

September 17, 2020

1 Setup for the Text Mining course (outside of UoE)

1.0.1 Prior Knowledge

This course is aimed at participants without any prior experience with text analysis of data sets. You will benefit from knowing how to start a terminal on your computer, how to run commands on it and how to do basic coding in the programming language python but we will go over each step required in the course for anyone with no previous knowledge of these things.

The following instruction are for how to get the course setup on you local laptop, external to the University of Edinburgh's system.

1.0.2 Terminal Setup

Linux: Look for Terminal in your applications and click on it to open it.

macOS: Bash is the default shell in all versions of macOS, you do not need to install anything. Open the Terminal application from `/Applications/Utilities` or using Spotlight search.

Windows: Open the Start Menu, type `command` or `cmd` in the search field, and then press Enter or click on the Command Prompt shortcut.

1.0.3 Software Installs for Day 1

For this course, we ask you to install Miniconda and as well as a series of software packages with it (jupyter notebook, nltk, numpy, matplotlib and wordcloud). Please take time to install these different things, ideally before the lesson if you can. You will need to know what operating system you are using before you begin and whether you are on a 32-bit or a 64-bit machine: this differs from machine to machine but some googling should bring you the answer.

1. Install Miniconda

Got to <https://docs.conda.io/en/latest/miniconda.html> and download the correct version. We will be working with Python version 3.7, so you need to select the appropriate version of Miniconda for your operating system with Python 3.7. Most operating systems are 64 bit these days. So unless you have a very old computer 64 bit likely going to be the download you need.

Windows: Click on Miniconda3 Windows 64-bit to download miniconda and once that's done you should see the exe (executable) at the bottom left of your browser window. Click on it and go through the installation process. Once installed, go to the Start menu and start up the Anaconda Powershell Prompt. You can find it by typing "Anaconda".

macOs: Download the Miniconda3 MacOSX 64-bit pkg (package) file. It will be saved in your Downloads directory. Once downloaded, double-click on it and go through the installation.

Linux: Download the appropriate Miniconda package and install it.

For all of the software packages to be installed next (using conda) you will need to proceed through the terminal for Mac and Linux, or if on Windows then enter the commands below in the Anaconda Powershell prompt. Some packages might take a while to install and you have to type “y” and press Enter when asked to proceed. A package is finished installing once it says “Executing transaction: done”.

To make it easier and avoid typos, just copy/paste the commands below across.

2. Install Jupyter Notebook

Windows: install jupyter notebook via the Anaconda Powershell prompt. To open it, go to the Start menu, then find and start up the Anaconda Powershell Prompt. You can find it by typing “Anaconda”. macOS/Linux: go to a terminal window and run the following command:

```
conda install -c anaconda jupyter
```

3. Install NLTK

Windows: install NLTK via the Anaconda Powershell prompt
macOS/Linux: run the following command in your terminal:

```
conda install -c anaconda nltk
```

4. Install numpy

Windows: install numpy via the Anaconda Powershell prompt
macOS/Linux: run the following command in your terminal:

```
conda install numpy
```

5. Install matplotlib

Windows: install matplotlib via the Anaconda Powershell prompt
macOS/Linux: run the following command in your terminal

```
conda install -c conda-forge matplotlib
```

6. Install wordcloud

Windows: install wordcloud via the Anaconda Powershell prompt
macOS/Linux: run the following command in your terminal:

```
conda install -c conda-forge wordcloud
```

1.0.4 Software Installs for Day 2

7. Installing the spaCy and downloading a model

Windows: install spacy via the Anaconda Powershell prompt
macOS/Linux: run the following command in your terminal:

```
conda install -c conda-forge spacy
```

Download a spacy model in your terminal:

```
python -m spacy download en_core_web_sm
```

8. Install pandas

Windows: install pandas via the Anaconda Powershell prompt
macOS/Linux: run the following command in your terminal:

```
conda install pandas
```

9. Install networkx

Windows: install networkx via the Anaconda Powershell prompt
macOS/Linux: run the following command in your terminal:

```
conda install networkx
```

1.0.5 Introduction to Jupyter Notebook

You can run the course notebooks using [Jupyter Notebook](#). It is an open-source web application that allows you to create and share documents that contain live code, equations, visualisations and narrative text. You interact with it via a browser window rather than a separate stand-alone application.

You can run the notebooks within a Jupyter Notebook interface so please type `jupyter notebook` into the command window on your machine. Ideally you would do that where you have saved all the course notebooks. You may have to navigate to that directory in your terminal window. For example the following command will take you to your Downloads folder on a Mac or Linux:

```
cd ~/Downloads
```

On Windows you can also use the `cd` command but you need to remember to use backslashes in your directory path.

Once you are in the directory containing the course notebooks (on the terminal) you can start a Jupyter Notebook server. To do this you need to:

Windows: open the Anaconda Powershell prompt and type: **macOS/Linux:** open a terminal window and type:

```
jupyter notebook
```

This will open up a browser window or tab containing the base directory of where you can store your new notebook on your computer. It may take a while for the browser window to open.

To create a new jupyter notebook you need to select a location for it on your computer via the browser window that opened up and click on “New” in the top right corner of the browser. You need to select Python 3. Once the new notebook opens you can give it a name by changing the word “Untitled” in the first line of the notebook.

You can see the first cell in your new notebook. You can enter Python code into this cell and press “Run” as long as it is marked as “Code” in the menu at the top of your notebook. This will run your code and you will see any output created by the code immediately below it. If you simply press Enter, this won’t run the code: instead, use the keyboard shortcut for Run which is Command-Enter. (To run the cell you are currently on and insert a new cell below, use Option-Enter on a Mac, and Alt-Enter on a PC.)

In this course we provide all the notebooks which you can adapt, but if you want to create a new empty one then you know how to do that now.