Document Analysis Assignment 1: Setup instructions

Setting up a personal machine:

It is recommended that you install the anaconda3 python environment from here:

https://www.anaconda.com/products/individual

We will be using python 3 in this course; the lab machines use python 3.8.10. The code has also been tested on python 3.7.9. Please do not use a version of python less than 3.7 (in previous years the fact that python 3.6 dictionaries do not remember insertion order, while python 3.7 dictionaries do remember insertion order, confused several students).

Once installed you will also need to install the nltk package:

pip install --user -U nltk

You may also need to download the stopword dictionary. To do this open a python interactive prompt and type in the following:

import nltk
nltk.download('stopwords')

You will also need the trectools package:

pip install trectools

Google Colab:

(Google Colab is not recommended for Assignment 1, but might be useful for the labs and Assignment 2)

You might like to use Google Colab if your personal machine is slow and you don't want to use the VDI infrastructure (described below). Details on how to setup Colab can be found here:
https://cloudstor.aarnet.edu.au/plus/s/tqqb8VfcM8IpBTx

Setting up lab machines:

We will be using anacoda3 on the lab machines. Most things should already be setup, you just need to enable the anaconda environment. In the terminal, **from your home directory** type *anacoda3* and press enter. You will need to do this every time you want to work with anaconda3.

You can then run your code by typing `python <your python file here>`.

Both *trectools* and *nltk* are already installed on the lab machines.

VDI Remote access:

If you are remote but would still like to use the ANU infrastructure, you can use the VDI infrastructure by following the instructions here:

https://cs.anu.edu.au/docs/student-computing-environment/linuxlabs/VDI/