Getting started with the Hands-on Assignments

Marieke van Erp Valentina Maccatrozzo Chris Dijkshoorn
Niels Ockeloen Davide Ceolin Cristina Bucur

February 8, 2018

1 Introduction

To give you some hands-on experience with social web data, we will be mining and analysing some during the hands-on sessions. For the majority of you, this will mean that you will be going through the exercises in the hands-on sessions using the provided Python scripts (if you want to use another programming/scripting language, by all means do, but we cannot provide you with as much detailed support).

The hands-on assignments are chosen such that you will gain a first-hand understanding of the good and the bad of social web data. We will prep you for it during the lectures, but nothing beats experience. The assignments during the different sessions also build up to what you need to know in order to do the final assignment, so make sure you understand what is going on in the code.

In this document, we will give some tips and hints on how to get started with Python and what you need to do to run the tools used in the assignments.

2 Python

There are huge differences between different versions of Python. We have tested the assignments on python version 2.7. The computer labs have 2.7.3 installed. Newer is not always better. Not all dependencies have been ported to Python 3 yet, so to play it safe, please install a 2.7.x version.

2.1 Installing Python on your own machine

If you're working on a *NIX system, chances are Python is already up and running. Just open a terminal and type "python" followed by a return and if the python interpreter comes on it will tell you which version you have. If it's not the right version, you can download it here: http://www.python.org/download. For Windows users, ActivePython is recommended, as it automatically adds Python to your path on the Windows Command Prompt. You can find it at ActivePython http://www.activestate.com/activepython. You can download the 'Community Edition' for free at the bottom of the page. If you have persistent problems getting stuff working on your own machine, consider using one in the computer labs.

2.2 Creating a Virtual Environment for Pip

As we are going to be using quite a few packages, a package manager is highly recommended. We'll use pip. On the VU lab computers, we will setup a 'virtual environment' in order to install

packages. When you are working on your own machine, the below instructions might need to be changed to accommodate for alternative locations of your working folder and python installation and/or installed tools such as 7zip on Windows. You can also unpack using a GUI program available to you. To set up the virtual environment on the lab computers, do the following:

2.2.1 Instructions for the Linux labs

Open the terminal (Applications - System Tools - Terminal)
Create a working folder:

```
$ mkdir socialweb
$ cd socialweb
```

Download latest virtual environment from https://pypi.python.org/pypi/virtualenv#downloads
or

```
$ curl -O https://pypi.python.org/packages/source/v/virtualenv/virtualenv-1.11.1.tar.gz
```

Unpack virtual environment (where X corresponds to your downloaded version):

```
$ tar xvfz virtualenv-X.X.X.tar.gz
$ rm virtualenv-X.X.X.tar.gz
```

Create a new environment:

```
$ python virtualenv-1.11.1/virtualenv.py socialEnv
```

Activate the virtual environment:

```
$ source socialEnv/bin/activate
```

NB: Every time you open a new terminal, e.g. at the beginning of a new hands on session, you have to navigate to your working folder and activate the virtual environment.

2.2.2 Instructions for the Windows labs

Open a command prompt (All Programs - Accessories - Command Prompt; Or Win+R, 'cmd')
Go to your personal share and create a working folder:

```
> H:
> mkdir socialweb
> cd socialweb
```

Download latest virtual environment from https://pypi.python.org/pypi/virtualenv#downloads
or

```
> curl -O https://pypi.python.org/packages/source/v/virtualenv/virtualenv-1.11.1.tar.gz
```

Unpack virtual environment (where X corresponds to your downloaded version)::

```
> rename virtualenv-X.X.X.tar.gz ve.tar.gz
> "C:\Program Files\7-Zip\7z" x ve.tar.gz -so | "C:\Program Files\7-Zip\7z" x -aoa -si -ttar
> del ve.tar.gz
```

NB: You can unpack the virtual environment also via the graphical interface of Windows. In this case, make sure that you unpack both the .gz and the .tar archive (so two steps) and that the unpacked files reside in the folder: H:\socialweb\virtualenv-X.X.X .

Create a new environment:

```
> python virtualenv-1.11.1/virtualenv.py socialEnv
> xcopy C:\PYTHON27\tcl /h socialEnv\lib /e
```

Activate the virtual environment:

```
> socialEnv\Scripts\activate
```

NB: Every time you open a new command prompt, e.g. at the beginning of a new hands on session, you have to navigate to your working folder and activate the virtual environment.

2.2.3 Installing required packages with pip

Before installing packages, make sure the current directory is your working folder (socialweb), and that you have activated the virtual environment. Now, whenever you need a new package, you just run 'pip install PACKAGENAME' from the command line. You need to restart your Python interpreter whenever you have installed a new package. For the hands-on sessions, all required packages are listed in the prerequisites, so start every session with installing all of them so you won't have to restart halfway.

Now you're all set. If you're not familiar with Python, have a look at the Official Python tutorial. One of the most important things to know about Python is that whitespace matters. We also recommended the book How to Think Like a Computer Scientist: Learning with Python by Jeffrey Elkner, Allen B. Downey and Chris Meyers. This should provide you with enough background material to do the exercises. Also remember that Google (or Bing, for that matter) is your friend if you get stuck.

3 Working with the command line/prompt

The command line is a powerful tool, if you know how to wield it. If you're wondering why we want you to use that black screen where you have to type in things instead of use a fancy GUI, read In the beginning was the commandline by Neil Stephenson. It is not all about programming or even the command line, but about computers and culture. If you're not going to use your own machine, make sure you know how to move around the file system, start up python, run python scripts and write to files before the first hands-on session. Below are some helpful links.

3.1 *NIX

Here's some first aid (from our colleagues at Bioinformatics) to get you started.

3.2 Windows

Here's some first aid (from our colleagues at Princeton) to get you started.

4 Visualisations

We are also going to be doing some visualisations. For this, we will use the .dot language and some JavaScript libraries. In the *Linux labs*, circo is available from the command line that can convert .dot files to an image file.

In the *Windows labs*, we will use GraphViz (also possible on *NIX machines, if your want to edit your image in a graphical user interface). To use GraphViz in the Windows labs, you have to do the following:

Open a command prompt (All Programs - Accessories - Command Prompt; Or Win+R, 'cmd') Go to your personal share and working folder:

```
> H:
> cd socialweb
```

Download and install GraphViz:

```
> curl -O https://graphviz.gitlab.io/_pages/Download/windows/graphviz-2.38.msi
> msiexec /i graphviz-2.38.msi
--> IN THE GUI INSTALLER, CHANGE THE INSTALL DIRECTORY TO "H:\socialweb\Graphviz2.38\".
```

To start GraphViz:

```
> Graphviz2.38\bin\gvedit
```

For the more complex visualisations, we will use d3.js.

5 Clean up after the course

To free space on your personal VU share, you can delete the 'socialweb' working folder you created. This folder contains all the files needed for the Python virtual environment, as well as the used packages and possibly GraphViz. NB: IF YOU WANT TO SAVE YOUR ASSIGNMENTS OR OTHER MATERIAL, FIRST COPY THEM TO A DIFFERENT LOCATION. Due to rights issues, delete the folder on the type of machine that you used to create it and install its contents.