

# Introduction to Webscraping with Python - Syllabus

Gabriele Rovigatti

March 1, 2017

## I Syllabus

The lecture is introductory and is meant to provide the audience with a description of what are:

- the key strengths of Python in scraping web pages;
- the most widely tools and modules;
- the best programming practices (secondary).

It is structured to be as interactive as possible, therefore everyone is strongly encouraged to attend with her/his computer: you can find the presentation and the code here. In particular, the latter is in .zip format (direct download here): download it and extract all files in the same directory; Windows users may then call the functions directly from hyperlinks in the presentation, the others will find the relative pieces of code within the /code/ directory.

There are (few) requirements:

- Throughout the lecture I will present many pieces of code. Most of them are fairly general, nonetheless they have been written and tested on Python **2.7**. Hence, I suggest to download and install its last version (available here for any OS). Make sure to install `pip` (or any other module installer available) and the Shell - **Python 2.7 and Python 3.6 are DIFFERENT! Pay attention to it**

- We will use (not so many) modules: these must be downloaded and installed by `pip install` or `easy_install` before calling them. In particular, we will use: `requests`, `selenium`, `time`, `bs4` and `csv`. Please be sure to have them installed before the class start. In case of doubts, follow the instructions provided [here](#) and [here](#)
- A basic knowledge of the concepts of function, for loop, lists and/or arrays is assumed.