

Installazione del software del Data Journalist

Angelica Lo Duca
angelica.loduca@iit.cnr.it

Fase	Software
Data Collection	Python + jupyter + librerie specifiche
Data Cleaning	Open Refine + Python + pandas
Data Exploration	Tableau
Data Analysis	Python + jupyter + scikit-learn
Data Visualization	jupyter + D3.js
Story Packaging	XAMPP (opzionale), HTML + js

Python 3 - Windows

- <https://www.python.org/downloads/> (pagina generale)
- <https://www.python.org/downloads/release/python-382/> (pagina specifica)

Version	Operating System	Description	MD5 Sum	File Size	GPG
Gzipped source tarball	Source release		f9f3768f757e34b342dbc06b41cbc844	24007411	SIG
XZ compressed source tarball	Source release		e9d6ebc92183a177b8e8a58cad5b8d67	17869888	SIG
macOS 64-bit installer	Mac OS X	for OS X 10.9 and later	f12203128b5c639dc08e5a43a2812cc7	30023420	SIG
Windows help file	Windows		7506675dcbb9a1569b54e600ae66c9fb	8507261	SIG
Windows x86-64 embeddable zip file	Windows	for AMD64/EM64T/x64	1a98565285491c0ea65450e78afe6f8d	8017771	SIG
Windows x86-64 executable installer	Windows	for AMD64/EM64T/x64	b5df1cbb2bc152cd70c3da9151cb510b	27586384	SIG
Windows x86-64 web-based installer	Windows	for AMD64/EM64T/x64	2586cdad1a363d1a8abb5fc102b2d418	1363760	SIG
Windows x86 embeddable zip file	Windows		1b1f0f0c5ee8601f160cfad5b560e3a7	7147713	SIG
Windows x86 executable installer	Windows		6f0ba59c7dbeba7bb0ee21682fe39748	26481424	SIG
Windows x86 web-based installer	Windows		04d97979534f4bd33752c183fc4ce680	1325416	SIG

Installazione Python 3 su Windows



Selezionare entrambe le caselle con il segno di spunta.

Verifica della correttezza dell'installazione

- Chiudere (se aperto) e Aprire un terminale
 - cercare nella barra di ricerca la parola “terminale”
 - selezionare prompt dei comandi
- Digitare **python --version**
 - come output dovrebbe dare: Python 3.9.2

Python 3 - Mac OS

- <https://www.python.org/downloads/> (pagina generale)
- <https://www.python.org/downloads/release/python-382/> (pagina specifica)

Version	Operating System	Description	MD5 Sum	File Size	GPG
Gzipped source tarball	Source release		f9f3768f757e34b342dbc06b41cbc844	24007411	SIG
XZ compressed source tarball	Source release		e9d6ebc92183a177b8e8a58cad5b8d67	17869888	SIG
macOS 64-bit installer	macOS	Python 3.8.2 and later	f12203128b5c639dc08e5a43a2812cc7	30023420	SIG
Windows help file	Windows		7506675dccb9a1569b54e600ae66c9fb	8507261	SIG
Windows x86-64 embeddable zip file	Windows	for AMD64/EM64T/x64	1a98565285491c0ea65450e78afe6f8d	8017771	SIG
Windows x86-64 executable installer	Windows	for AMD64/EM64T/x64	b5df1cbb2bc152cd70c3da9151cb510b	27586384	SIG
Windows x86-64 web-based installer	Windows	for AMD64/EM64T/x64	2586cdad1a363d1a8abb5fc102b2d418	1363760	SIG
Windows x86 embeddable zip file	Windows		1b1f0f0c5ee8601f160cfad5b560e3a7	7147713	SIG
Windows x86 executable installer	Windows		6f0ba59c7dbeba7bb0ee21682fe39748	26481424	SIG
Windows x86 web-based installer	Windows		04d97979534f4bd33752c183fc4ce680	1325416	SIG

Verifica della correttezza dell'informazione

- Aprire un terminale
- Mac OS di default ha installato python 2.7
 - digitando **python --version** dovrebbe dare come output Python 2.7.10
- Per usare python 3 si deve usare il comando python3
 - digitando **python3 --version** dovrebbe dare come output Python 3.8.2

Python 3 - Ubuntu

- In un terminale scrivere i seguenti comandi:
 - `sudo apt update`
 - `sudo apt install software-properties-common`
 - `sudo add-apt-repository ppa:deadsnakes/ppa`
 - `sudo apt update`
 - `sudo apt install python3.9 python3.9-venv python3.9-dev`
 - `python3.9 -V`
- **Installare pip3**
 - `sudo apt install python3-pip`

Jupyter

- <https://jupyter.org/install> (Windows)
 - pip install jupyterlab
 - pip install notebook
 - pip install voila
- <https://jupyter.org/install> (Mac/Linux)
 - pip3 install jupyterlab
 - pip3 install notebook
 - pip3 install voila

Data Collection

	Librerie
Web API Flickr	<code>pip install flickrapi</code> <code>pip3 install flickrapi (Linux/Mac)</code>
Web API Twitter	<code>pip install python-twitter</code> <code>pip3 install python-twitter</code> (Linux/Mac)
Estrazione da PDF	<code>pip install tabula-py</code> <code>pip3 install tabula-py(Linux/Mac)</code> tabula - https://tabula.technology/
Web Scraping	selenium + chrome driver

Selenium + Chrome Driver per python

- Da terminale digitale
 - `pip install selenium`
 - `pip3 install selenium` (Mac/Linux)
 - <https://selenium-python.readthedocs.io/installation.html>
- Web Driver per python
 - verificare la versione del browser

Chrome:	https://sites.google.com/a/chromium.org/chromedriver/downloads
Edge:	https://developer.microsoft.com/en-us/microsoft-edge/tools/webdriver/
Firefox:	https://github.com/mozilla/geckodriver/releases
Safari:	https://webkit.org/blog/6900/webdriver-support-in-safari-10/

Versione di Chrome

- Per conoscere la versione di Chrome, nella barra degli indirizzi digitare
 - `chrome://settings/help`



Google Chrome

Aggiornamento quasi terminato. Riavvia Google Chrome per completarlo. Le finestre di navigazione in incognito non verranno riaperte.



Versione 80.0.3987.116 (Build ufficiale) (a 64 bit)

Riavvia

Installazione Web Driver

- Dopo aver scaricato il Web driver, posizionarlo in una cartella a piacere del proprio file system
- Impostare la variabile di ambiente \$PATH in modo da contenere il percorso alla cartella prescelta
 - Windows - <https://youtu.be/mHtlBq5cP2Y>
 - Mac OS - modificare il file .bash_profile
 - aprire un terminale
 - nella propria home digitare il comando **vi .profile**
 - in fondo al file scrivere
 - digitare il tasto i (insert)
 - export PATH="`<percorso al web driver>:$PATH`"
 - premere ESC :wq per salvare il file
 - riavviare il computer
 - aprire il terminale e digitare echo \$PATH

Data Cleaning

- Open Refine
 - <https://openrefine.org/download.html>
- Python Pandas
 - `pip install pandas`
 - `pip3 install pandas (Linux/Mac)`

Data Exploration

- Tableau
 - <https://public.tableau.com/en-us/s/download>
 - <https://www.tableau.com/academic/students>

Data Analysis

- Scikit-learn
 - <https://scikit-learn.org/stable/install.html>
 - `pip install -U scikit-learn`
 - `pip3 install -U scikit-learn` (Linux/Mac)

Story Packaging

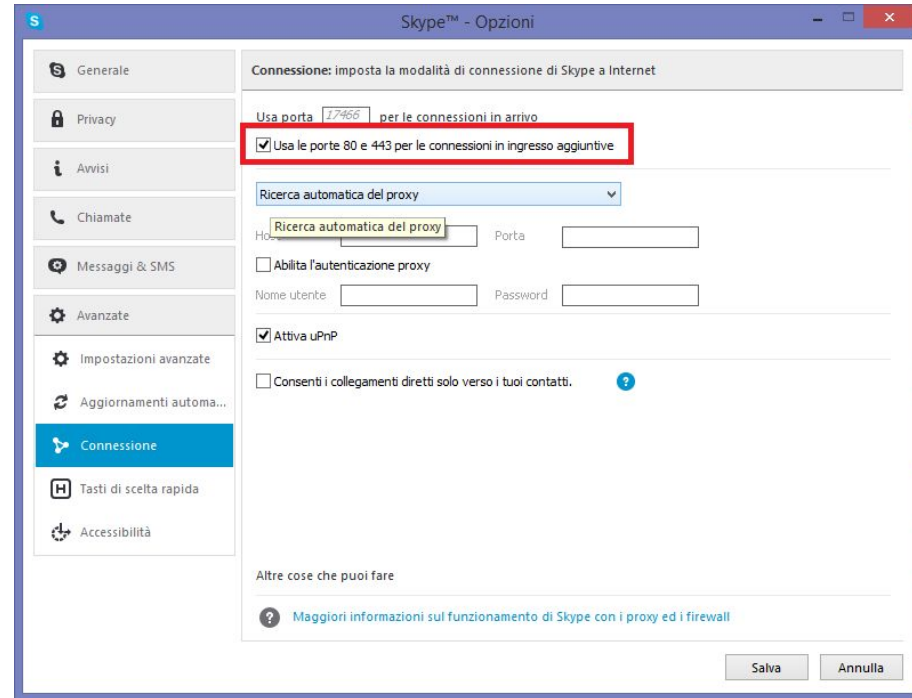
- XAMPP
 - <https://www.apachefriends.org/it/download.html>

Conflitti con altri programmi

- Di default il server Apache sta in ascolto sulla porta 80.
- La porta 80 potrebbe essere utilizzata da altri programmi, come ad esempio Skype e Spotify.
- Per evitare conflitti, chiudere Spotify e i programmi che stanno in ascolto sulla porta 80.

Conflitto con Skype

- Eseguire Skype
- Accedere al menu Strumenti e poi Opzioni
- Selezionare menu Avanzate e poi Connessione
- Deselezionare la voce evidenziata in ROSSO (**Usa le porte 80 e 443 per le connessioni in ingresso aggiuntive**)



Github

Strumento di lavoro per la condivisione del codice

Scaricare l'applicazione:

- Windows/Mac
 - <https://desktop.github.com/>
- Ubuntu
 - `sudo wget https://github.com/shiftkey/desktop/releases/download/release-2.1.0-linux1/GitHubDesktop-linux-2.1.0-linux1.debsudo gdebi`
 - `GitHubDesktop-linux-2.1.0-linux1.deb`

[Tutorial sull'Installazione](#)

Github a linea di comando

Scaricare il seguente software:

<https://git-scm.com/downloads>