**PDF DOWNLOAD PROJECT**

**Python based PDF extraction tool**

PDF download project is a Web Scraping project used to download PDF from HTML webpage.

Website : <http://parivesh.nic.in/>

* ENVIRONMENT CLEARANCE > DASHBOARD

## **Download & Installation**

**1.** Clone the repository.

git clone github\_url Or you can download the zip file and Extract the files from there.

**2.** Download dependencies (see below)

**3.** Change directory(cd) to /pdf\_project folder and run python3 workingpdf.py from terminal. (make sure you are using python version 3.6.x or greater)

**Installation Guide**

* >= Python 3.6.x : <https://www.python.org/downloads/release/python-360/>

**Installing Dependencies**

**Method 1 :**

**Using requirements.txt ( pip recommended )**

* requirements.txt:

pip install -r requirements.txt

**Method 2 :**

**Using pip ( recommended )**

* BeautifulSoup4: <https://pypi.org/project/beautifulsoup4/>

pip install beautifulsoup4

This allows us to search & extract content from an HTML webpage.

* Requests: <https://pypi.org/project/requests/>

pip install requests

The requests module allows you to send HTTP requests using Python. The HTTP request returns a Response Object with all the response data (content, encoding, status, etc).

* Pyinstaller : <https://pypi.org/project/PyInstaller/>

pip install pyinstaller

This allows us to build executable file.

To build executable file, run :

pyinstaller --onefile working\_pdf.py

Documentation : <https://pyinstaller.readthedocs.io/en/stable/usage.html>