

【初心者向け】 Python スクレイピング実 践入門

== MENTA教材 ==
written by bayashi





目次

1. この教材の前提知識
2. この教材の対象者
3. 対応バージョン
4. スクレイピングとは
5. スクレイピングの注意事項
6. 環境構築(Anaconda・Jupyter notebook)
7. Pythonの簡単な基礎文法のおさらい
8. スクレイピングをやってみよう
9. MENTAのサイトから名前とプラントタイトルを取得する
10. ページ遷移にも挑戦
11. ログインを自動化してみよう



この教材の前提知識

- ・ProgateやドットインストールなどでHTML/CSSの基礎知識を身につけている
- ・プログラミングの基本構文を知っている(for文、if文)
(何となくのレベルでも構いません)



この教材の対象者

- ・Progateをやったけど、次に何をすればいいかわからない
- ・プログラミングを使って何かしてみたい
- ・データ分析のためにweb上にあるデータを収集したい



対応バージョン

python == 3.6

selenium 3.13.0

Mac

windows

Google Chrome



スクレイピングとは

- ・ネット上にあるwebサイトの情報を実行されたプログラムによって情報を取得し、その情報を加工したり分析したりすることです。

- ・大量のデータを収集することができるので業務効率化やデータ分析に有効な手段です。

- ・スクレイピング の利用例

株価の変動を確かめる、ネットオークションの価格の変動を確かめる、
マーケティングに使用するデータを収集する



スクレイピング の注意事項①

・スクレイピングを行う際にはサイトの規約などをしっかり読みましょう。取得したデータを再利用することを利用規約で禁じているサイトもあります。スクレイピングを禁止しているサイトだと法的に訴えられることもあるため、注意が必要です。

環境構築編



環境構築

早速ですが、**Python**を手軽に実行できるための環境である、「**Jupyter Notebook**」をインストールしましょう。手順としては

① **Anaconda**のダウンロード、インストール

② **Anaconda**に搭載されている「**Jupyter Notebook**」を起動する

というプロセスになります。

インストールのやり方は次ページにある記事を参考にしてみてください。



Anacondaのインストールをする

- ・ダウンロードしたファイルをどこに配置するかなどは、添付先URLの記事を参考にしてみてください。
- ・なにかお困りごとや確認したいことがあればお気軽にお尋ねください。

【Windows用】

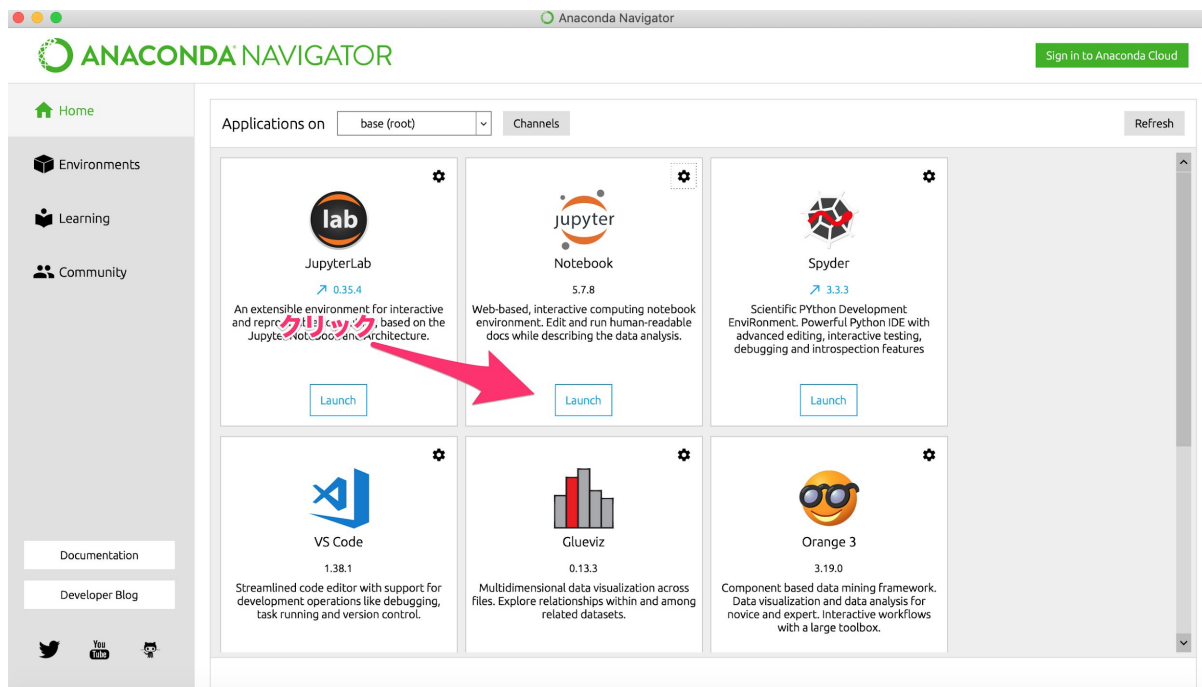
<https://did2memo.net/2017/07/09/python-anaconda-jupyter-notebook-windows/>

【Mac用】

<https://qiita.com/turmericN/items/37ea643439c40c556aa4>

Jupyter Notebook使い方編

①Anacondaを開きJupyter notebookをクリックする





②Jupyter notebookが開たら、以下の場所をクリックして「Python3」を選択する

[Quit](#)[Logout](#)[Files](#)[Running](#)[Clusters](#)

Select items to perform actions on them.

[Upload](#)[New ▾](#)☐ 0 ▾ /[Name ▾](#)[Last Modified](#)[File size](#)☐ [dir_name]

9ヶ月前

☐ AnacondaProjects

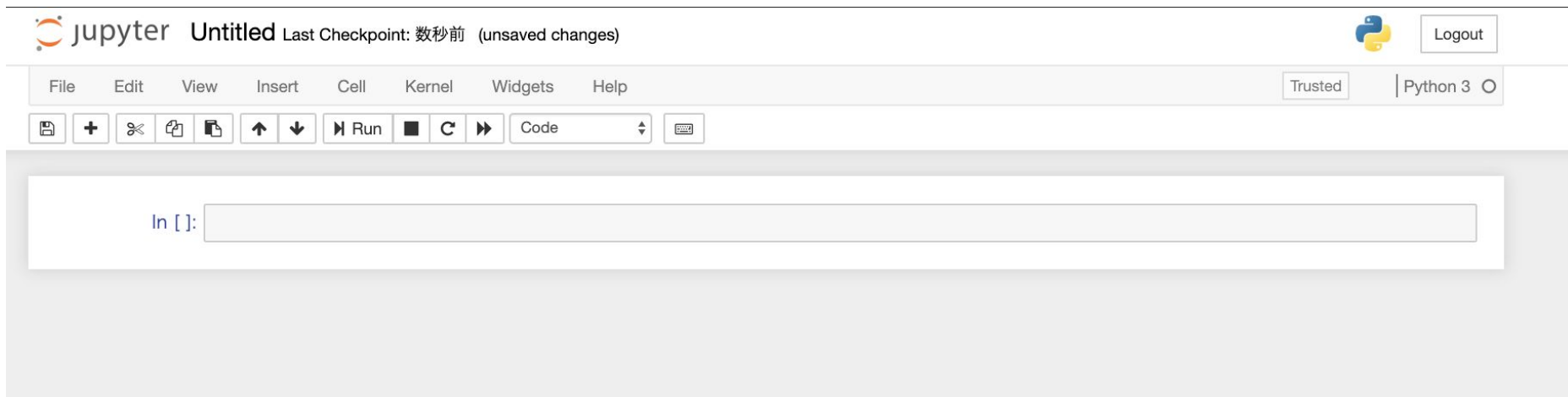
6ヶ月前

クリック

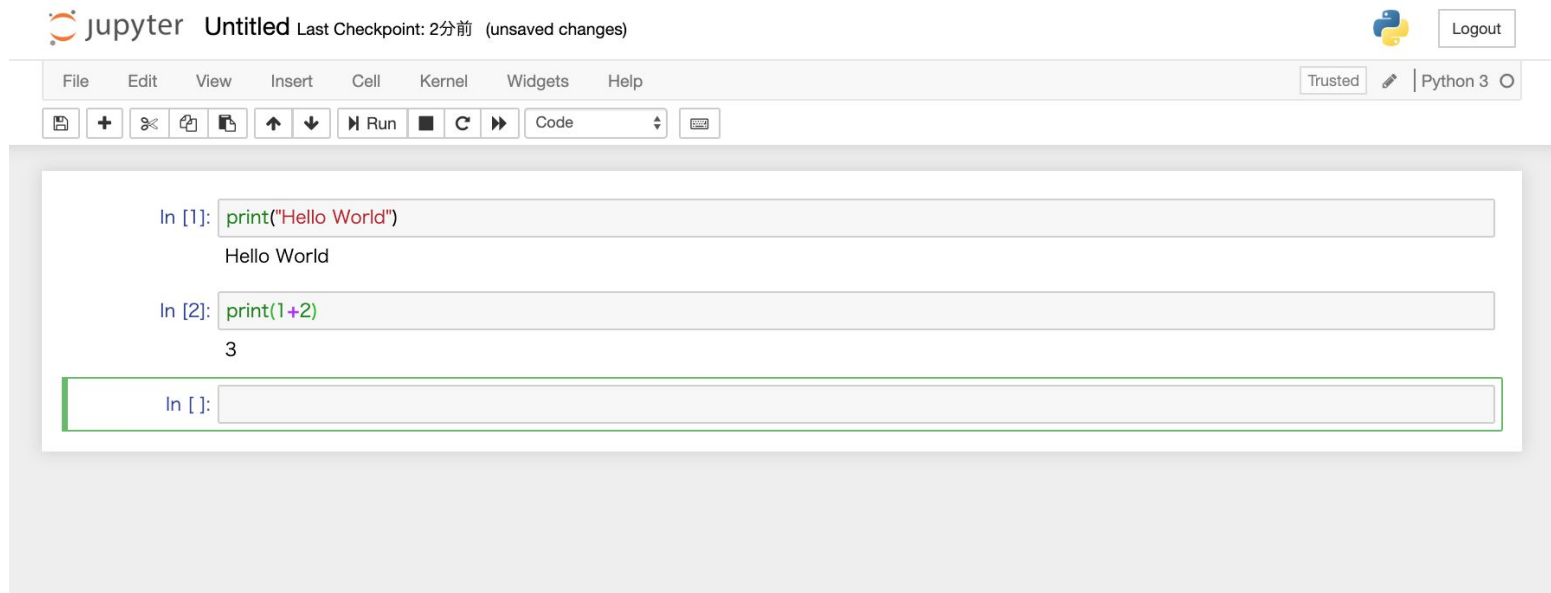




③セルの中にPythonのプログラムを書いてみましょう。
実行は「Shift + Enter」です。



実際のプログラム出力画面(参考)



The screenshot displays the Jupyter Notebook interface. At the top, the header shows the Jupyter logo, the text "jupyter", and the document title "Untitled". To the right of the title, it indicates "Last Checkpoint: 2分前 (unsaved changes)". Further right are a Python logo icon and a "Logout" button. Below the header is a menu bar with options: File, Edit, View, Insert, Cell, Kernel, Widgets, and Help. To the right of the menu bar is a "Trusted" status indicator, an edit icon, and the text "Python 3". Below the menu bar is a toolbar with icons for saving, adding cells, undo, redo, copy, paste, scroll up, scroll down, run, interrupt kernel, and restart kernel. A dropdown menu is set to "Code". The main area contains three code cells. The first cell has the input `In [1]: print("Hello World")` and the output `Hello World`. The second cell has the input `In [2]: print(1+2)` and the output `3`. The third cell is currently empty, showing `In []:` followed by a text input field.

jupyter Untitled Last Checkpoint: 2分前 (unsaved changes) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

In [1]: `print("Hello World")`
Hello World

In [2]: `print(1+2)`
3

In []: