**FreeCodeCamp: Python With: Dr Duck**

**From page :** [**https://www.freecodecamp.org/learn/?messages=success%5B0%5D%3Dflash.signin-success**](https://www.freecodecamp.org/learn/?messages=success%5B0%5D%3Dflash.signin-success)

**User:** [**JuanCarlos.dergan@gmail.com**](mailto:JuanCarlos.dergan@gmail.com)

**Pwd: FreeCodeCamp1973**

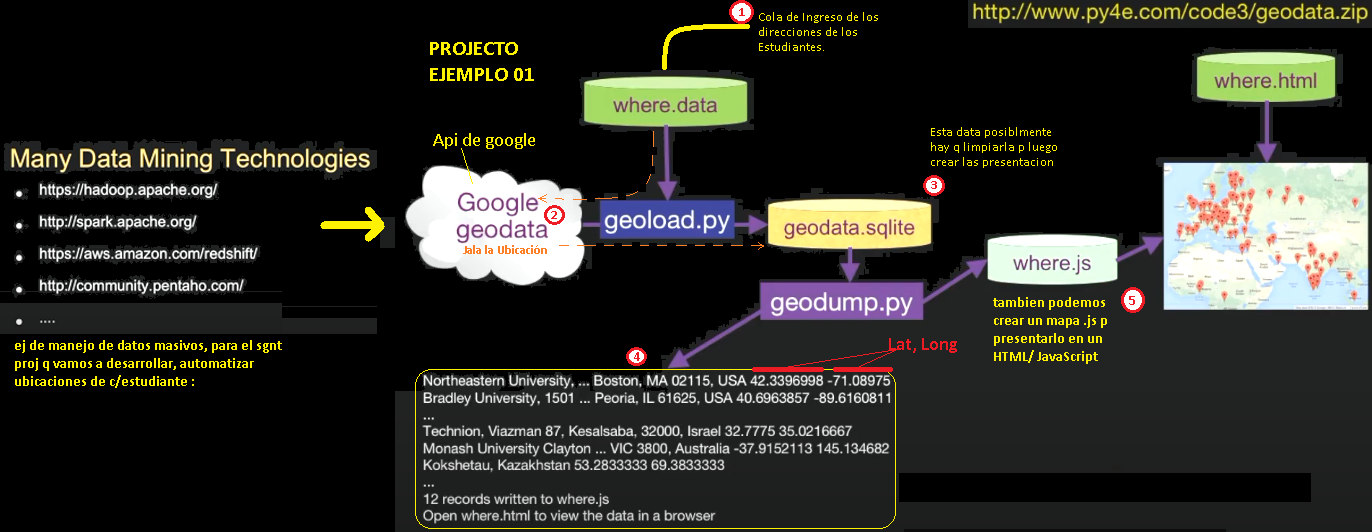
**Subject: Scientific Computing w Python Certification**

**Resumen de Projectos del Dr Chuck:**

**Proj Final : GeoData , Multi-Step Data Analysis. Link =** [**https://www.youtube.com/watch?v=e3lydkH0prw&t=8s**](https://www.youtube.com/watch?v=e3lydkH0prw&t=8s)

**En el sgnt proj q usa el algoritmo se llama : “Geodata”**

El nombre de univ donde estudian los alumnos, no tienen una direccion en mapa, y lo que hace estr prj es consultar a GoogleApi geodata p sustraer un JSON donde se especifica las Latitudes, direccion de ubicación etc. Ver a continuacion, paso **1**, **2** ,**3**, **4** y **5.**



**Nota: Las fuentes para este curso estan en :** [**https://www.py4e.com/code3/**](https://www.py4e.com/code3/)

Where.Data .- es el file q contiene el nombre de las universidades.

Geodata.py **.-** es el programa q Lee el file where.data, lo alamaena en la BD sino existe

Y luego lo prepara con una consulta API rest al GoogleApi p sustraer el JSON con los

Datos de ubicación , direccion exacta, etc y luego lo almacena en la BD, como JSON.

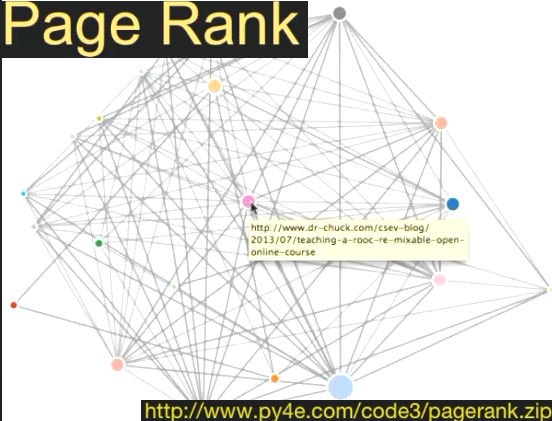
BD.- SQLite.

Luego esta la proyeccion de los datos bajados del googleapi a una pagina web q proyecta el mapa.

Esto lo hace con Where.html y Where.js

**Proj Final : Page Rank, Web Crawler. Link = https://www.youtube.com/watch?v=6-w\_qIUwaxU&t=3s**

**Para iniciar el proj necesitamos entender algunos conceptos de Algoritmo Page Rank:**



Usa WebCrawlers o Spider Web.

Un **Web crawler**, o bot:

Es un algoritmo o bot usado para analizar el código de un sitio **web** en busca de informacion,

para después usarlas y generar insights(comportamientos) o clasificar los datos encontrados.

Un ejemplo muy clásico de **web crawler** está en los sitios de búsqueda como

* **GoogleBot (Google)**
* **Bingbot (Bing)**
* **Slurpbot (Yahoo)**
* **DuckDuckBot (DuckDuckGo)**
* **Baiduspider (Baidu)**
* **Yandex Bot (Yandex)**
* **Sogou Spider (Sogou)**
* **Exabot (Exalead)**
* **Facebot (Facebook)**
* **Alexa Crawler (Amazon)**

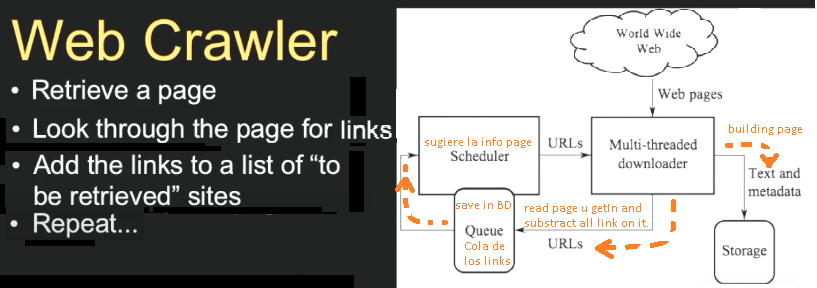
Además, las arañas bots web recopilan datos de contacto y perfiles con fines de marketing. El wecrawler mas antiguo es : **https://www.webcrawler.com/**

¿Un **spider(web)** **crawler** en programación?

Los web crawlers son bots que buscan datos en Internet, analizan el contenido y guardan la información en índices y bases de datos para mejorar el rendimiento de los motores de busqeda.

Una de sus funciones más utilizadas es crear una copia de todo los lugares vistados, para luego procesarlas, para una mejor engine, indexando en una BD las paginas usadas y poder proporcionar una busqeda mas rapida. (fast Searcher)

Lo que vamos a crear es una BD en SQLite para crear este Search Engine.



**Algunas Politicas basicas q tiene un WebCrawler:**

-una reglas según el state q paginas debo registrar.

-según el state,when to check for changes to pages.

-segun el state, cuando evitar en lo posible sobrecargas

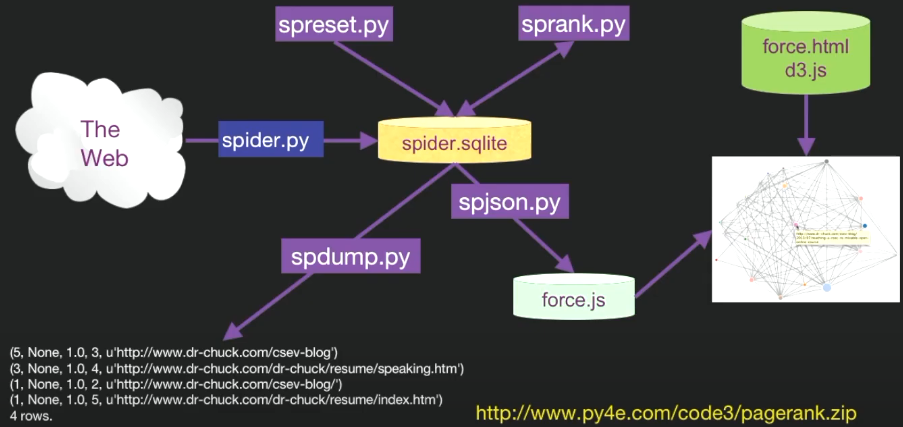
de la web page registrada.(cuidado con ancho de banda)

-policy según states to coordinate of parallelization to

Distributed WebCrawlers.

Hay un file **robot.txt**, q nos indica q debemos ver o no

Ver desde el bot, en la webs.



More resources:

- [Exercise: Geodata](https://www.youtube.com/watch?v=KfhslNzopxo)

- [Exercise: Gmane Model](https://www.youtube.com/watch?v=wSpl1-7afAk)

- [Exercise: Gmane Spider](https://www.youtube.com/watch?v=H3w4lOFBUOI)

- [Exercise: Gmane Viz](https://www.youtube.com/watch?v=LRqVPMEXByw)

- [Exercise: Page Rank](https://www.youtube.com/watch?v=yFRAZBkBDBs)

- [Exercise: Page Spider](https://www.youtube.com/watch?v=sXedPQ_AnWA)

- [Exercise: Page Viz](https://www.youtube.com/watch?v=Fm0hpkxsZoo)

**Proj Final : Mainling Lists - Gname. Link = https://www.youtube.com/watch?v=RYdW660KkaQ**