# Pràctica 1: Web scraping

1. Context. Explicar en quin context s'ha recol·lectat la informació. Explicar per què el lloc web triat proporciona aquesta informació.
2. Definir un títol pel dataset. Triar un títol que sigui descriptiu.
3. Descripció del dataset. Desenvolupar una descripció breu del conjunt de dades que s'ha extret (és necessari que aquesta descripció tingui sentit amb el títol triat).
4. Representació gràfica. Presentar un esquema o diagrama que identifiqui el dataset visualment i el projecte escollit.
5. Contingut. Explicar els camps que inclou el dataset, el període de temps de les dades i com s'ha recollit.
6. Agraïments. Presentar el propietari del conjunt de dades. És necessari incloure cites d'anàlisis anteriors o, en cas de no haver-les, justificar aquesta cerca amb anàlisis similars.
7. Inspiració. Explicar per què és interessant aquest conjunt de dades i quines preguntes es pretenen respondre. És necessari comparar amb les anàlisis anteriors presentades a l’apartat 6..
8. Llicència. Seleccionar una d'aquestes llicències pel dataset resultant i explicar el motiu de la seva selecció:

* Released Under CC0: Public Domain License
* Released Under CC BY-NC-SA 4.0 License
* Released Under CC BY-SA 4.0 License
* Database released under Open Database License, individual contents under Database Contents License
* Other (specified above)
* Unknown License

1. Codi. Adjuntar el codi amb el qual s'ha generat el dataset, preferiblement en Python o, alternativament, en R.
2. Dataset. Publicar el dataset en format CSV a Zenodo (obtenció del DOI) amb una breu descripció.

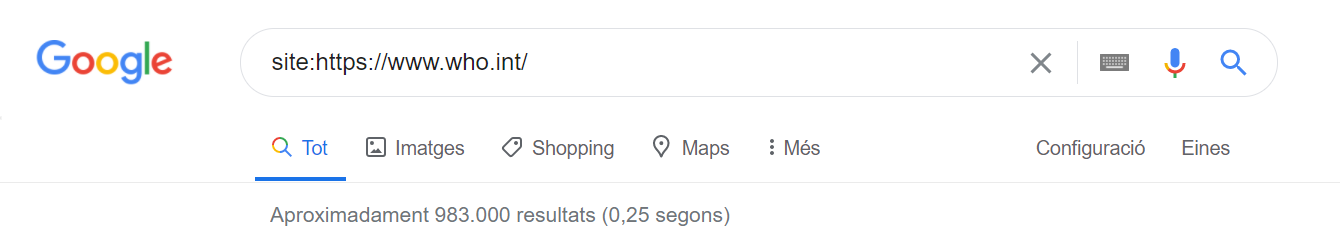
|  |  |
| --- | --- |
| Contribucions | Signa |
| Recerca prèvia | Maria Begoña Felip, Vicenç Pio |
| Redacció de les respostes | Maria Begoña Felip, Vicenç Pio |
| Desenvolupament codi | Maria Begoña Felip, Vicenç Pio |

## Avaluació inicial

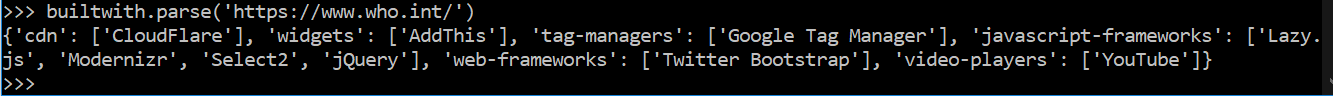
* Arxiu robots.txt
* Mapa del lloc web

He generat el fitxer sitemap.xml amb això: https://www.xml-sitemaps.com/

* La seva grandària



* La tecnologia emprada



* El propietari del lloc web

