

# Pràctica 1 (35% nota final)

## Descripció de la Pràctica a realitzar

L'objectiu d'aquesta activitat serà la creació d'un dataset a partir de les dades contingudes en una web. Per a la seva realització, s'han de complir els següents punts:

1. Context. Explicar en quin context s'ha recollit la informació. Explicar per què el lloc web triat proporciona aquesta informació.

Hem pensat en la necessitat que tenim els que ens agrada anar a fer passejades per la natura d'aconseguir informació de rutes fetes per altres persones anònimes.

Generar aquest dataset ens donarà la possibilitat de poder consultar una gran quantitat de rutes sense haver-hi d'anar navegant d'una a una altra. L'objectiu d'aquest site és poder compartir rutes a l'aire lliure entre els diferents usuaris tant si estan enregistrats com si no. Els usuaris enregistrats fan rutes i posteriorment les penjen al site, d'aquesta manera la resta pot veure el tipus de ruta i decidir si li agradaria fer-la o no.

2. Definir un títol pel dataset. Triar un títol que sigui descriptiu.

`rutes_{activitat}_{nom_pais}_{nom_regio1}_{nom_regio2}`

Si no s'introdueix activitat, aquesta per defecte serà "outdoor". Si no s'introdueix nom\_pais, nom\_regio1 o nom\_regio2, aquests no apareixeran al nom. Per exemple si només introduïm activitat i nom\_pais el nom del dataset serà: `rutes_{activitat}_{nom_pais}`.

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).

El dataset resultant contindrà informació de totes les rutes que han estat valorades per algun altre usuari, en funció d'un país, regió1 i regió2 específica, o bé en funció d'una activitat i una regió.

Les dades que hi podrem trobar són: tipus, puntuació de la ruta, puntuació del tracker, distància, alçada, desnivell, circular, votació dels usuaris, etc.

4. Representació gràfica. Presentar una imatge o esquema que identifiqui el dataset visualment



5. Contingut. Explicar els camps que inclou el dataset, el període de temps de les dades i com s'ha recollit.

El dataset inclou els següents camps:

- TypeTrack: tipus de ruta
- TrailRank: rànkung de la pista
- UserRank: rànkung de l'usuari que ha pujat la informació.
- Distància: longitud de la ruta en km.
- Loop: indica Si la ruta és circular o No.
- Desnivell acumulat pujant en metres.
- Altitud màxima en metres.
- Desnivell acumulat baixant en metres.
- Altitud mínima en metres.
- Dificultat Tècnica : Difícil, Moderat o Fàcil.
- Data en la qual es va fer.

- Valoració sobre 5 punts.
- Vots (quantitat)

SI trobem o NO els següents elements a la ruta:

- Lake: SI/NO.
- Fountain: SI/NO.
- Panoramic: SI/NO.
- River: SI/NO.
- Parking: SI/NO.
- Bridge: SI/NO.
- WaterFall: SI/NO.
- Refuge: SI/NO.
- Shelter: SI/NO.
- Url: l'adreça de la pàgina on trobem la informació.

Pel que fa a la durada de les dades hem de dir que aquestes poden arribar a estar emmagatzemades durant uns quants anys, ja que ens hem trobat amb rutes de l'any 2006. De totes maneres l'antiguitat de la informació en el nostre cas no és gaire important pel fet que possiblement el contingut i l'explicació de la ruta poc variarà encara que passin uns quants anys.

Aquest site està estructurat en tres nivells tal com s'observa en el path de l'url quan volem arribar a consultar una ruta, per exemple una ruta que es trobi a Andorra, Canillo, el Tarter tindrà un path com <https://ca.wikiloc.com/rutes/outdoor/andorra/canillo/el-tarter> on podem observar que les diferents parts del path indiquen les zones geogràfiques que acotem. Un cop estem situats a la zona desitjada ja podem escollir la ruta i passar a obtenir la informació sobre aquesta. És un cop ja dins d'aquest últim nivell quan començarem a fer Web Scrapping i a obtenir totes les dades que s'han comentat amb anterioritat.

6. **Agraïments. Presentar el propietari del conjunt de dades. És necessari incloure cites de recerca o anàlisis anteriors (si n'hi ha).**

Voldríem agrair al propietari de les dades (Wikiloc) per permetre que es realitzi web scraping a la majoria del seu domini, perquè ens permetrà a nosaltres que compartim la mateixa filosofia que ells: "Creure en un futur on la humanitat passa més temps gaudint i preservant la natura". Poder recollir les dades de les rutes que ens interessin de forma automàtica i després tractar aquestes dades per a escollir les rutes que més ens atraguin. D'aquesta manera podrem seguir gaudint

de la natura sense perdre tant de temps cercant una ruta per la interfície gràfica de Wikiloc.

7. Inspiració. Explicar per què és interessant aquest conjunt de dades i quines preguntes es pretenen respondre.

A part de la utilitat funcional que ja he comentat prèviament també li podem trobar una part d'interès de cara a l'estudi, per exemple podries arribar a saber:

- Quin tipus de rutes són les més valorades
- Si les rutes que pugen els usuaris són circulars normalment
- Quines són les zones amb més rutes
- Quines són les zones que tenen rutes específiques en una modalitat
- Quins són els usuaris més fiables

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

La llicència que s'ha escollit per a la publicació d'aquest projecte és **CC BY-NC-SA 4.0 License**. Els motius pel qual s'ha portat a terme l'elecció d'aquesta llicència són:

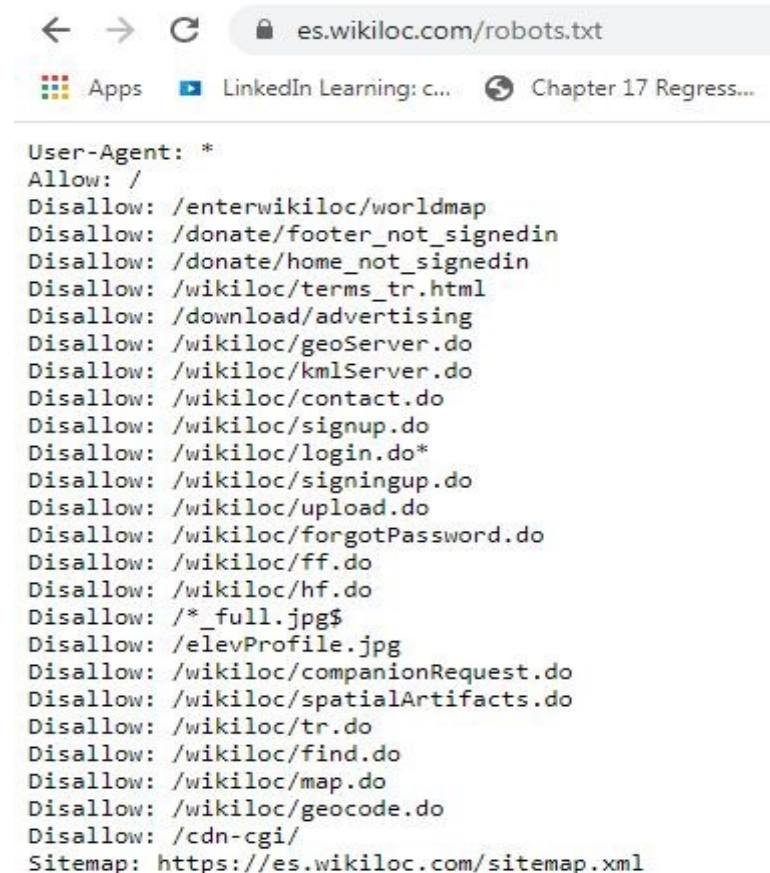
- Es permet copiar i distribuir el material en qualsevol mitjà o format.
- Es permet adaptar, transformar i construir sobre del projecte.
- S'ha de donar crèdit als propietaris de la llicència, és a dir, del projecte i indicar si es realitzen canvis.
- No es permeten usos comercials del projecte. WikiLoc no permet fer un ús comercial del contingut del seu domini.
- Si es realitza alguna transformació o construcció a sobre del material, s'ha de distribuir sobre la mateixa llicència que l'original.

9. Codi. Adjuntar el codi amb el qual s'ha generat el dataset, preferiblement en Python o, alternativament, en R.

Aprofitem aquesta secció per explicar una mica tot l'estudi previ que hem fet del site i com hem arribat a extreure informació de les url's desitjades així com informació pròpia del site.

## Estructura del Site

En aquest portal podem trobar 2 tipus de fitxers que ens poden ajudar en gran manera a desenvolupar una actuació de Web Scrapping el màxim d'ètica possible, per una part tenim el fitxer [robots.txt], del qual mostrem el seu contingut:



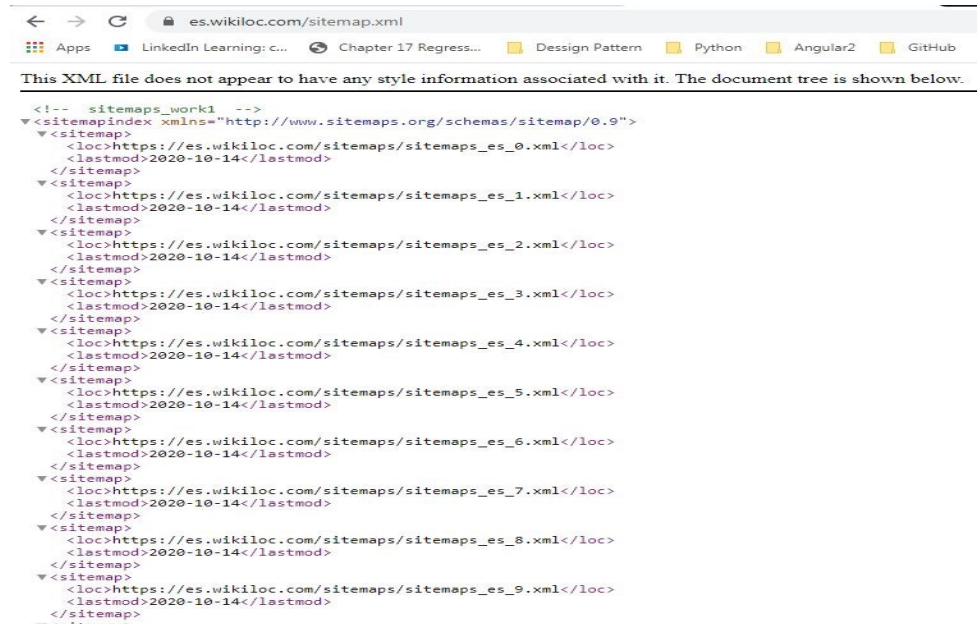
```
User-Agent: *
Allow: /
Disallow: /enterwikiloc/worldmap
Disallow: /donate/footer_not_signedin
Disallow: /donate/home_not_signedin
Disallow: /wikiloc/terms_tr.html
Disallow: /download/advertising
Disallow: /wikiloc/geoServer.do
Disallow: /wikiloc/kmlServer.do
Disallow: /wikiloc/contact.do
Disallow: /wikiloc/signup.do
Disallow: /wikiloc/login.do*
Disallow: /wikiloc/signingup.do
Disallow: /wikiloc/upload.do
Disallow: /wikiloc/forgotPassword.do
Disallow: /wikiloc/ff.do
Disallow: /wikiloc/hf.do
Disallow: /*_full.jpg$
Disallow: /elevProfile.jpg
Disallow: /wikiloc/companionRequest.do
Disallow: /wikiloc/spatialArtifacts.do
Disallow: /wikiloc/tr.do
Disallow: /wikiloc/find.do
Disallow: /wikiloc/map.do
Disallow: /wikiloc/geocode.do
Disallow: /cdn-cgi/
Sitemap: https://es.wikiloc.com/sitemap.xml
```

Que observem en aquest fitxer?

- Totes les rutes són aplicables a qualsevol “user.agent”
- Trobem directives “allow” i “disallow” les quals tenen significats diferents i a més tenen interpretacions diferents:



- Standard Implementation: la 1ª directiva que coincideixi amb la url per la qual es pregunta s'aplicarà.
- Google Implementation: si el patró indicat en la directiva "allow" té una longitud igual o superior a la de la directiva "disallow" llavors s'aplicarà la directiva "allow".
- Bing Implementation: si la url preguntada coincideix amb més d'una directiva, independentment de que sigui "allow" o "disallow" s'aplicarà la que tingui una longitud major.
- Es recomana no accedir a algunes pàgines amb extensió ".do"
- Es recomana no accedir a url's que comencin amb "/enterwikiloc/workmap"
- Es recomana no accedir a url's que acabin amb "\_full.jpg"
- Es recomana no accedir a la carpeta "/cdn-cgi/"
- A l'última línia es troba una referència al fitxer SiteMaps, del qual mostrem una part del seu contingut.



```

<!-- sitemaps_work1 -->
<?xml version="1.0" encoding="UTF-8" ?>
<sitemapindex xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
  <sitemap>
    <loc>https://es.wikiloc.com/sitemaps/sitemaps_es_0.xml</loc>
    <lastmod>2020-10-14</lastmod>
  </sitemap>
  <sitemap>
    <loc>https://es.wikiloc.com/sitemaps/sitemaps_es_1.xml</loc>
    <lastmod>2020-10-14</lastmod>
  </sitemap>
  <sitemap>
    <loc>https://es.wikiloc.com/sitemaps/sitemaps_es_2.xml</loc>
    <lastmod>2020-10-14</lastmod>
  </sitemap>
  <sitemap>
    <loc>https://es.wikiloc.com/sitemaps/sitemaps_es_3.xml</loc>
    <lastmod>2020-10-14</lastmod>
  </sitemap>
  <sitemap>
    <loc>https://es.wikiloc.com/sitemaps/sitemaps_es_4.xml</loc>
    <lastmod>2020-10-14</lastmod>
  </sitemap>
  <sitemap>
    <loc>https://es.wikiloc.com/sitemaps/sitemaps_es_5.xml</loc>
    <lastmod>2020-10-14</lastmod>
  </sitemap>
  <sitemap>
    <loc>https://es.wikiloc.com/sitemaps/sitemaps_es_6.xml</loc>
    <lastmod>2020-10-14</lastmod>
  </sitemap>
  <sitemap>
    <loc>https://es.wikiloc.com/sitemaps/sitemaps_es_7.xml</loc>
    <lastmod>2020-10-14</lastmod>
  </sitemap>
  <sitemap>
    <loc>https://es.wikiloc.com/sitemaps/sitemaps_es_8.xml</loc>
    <lastmod>2020-10-14</lastmod>
  </sitemap>
  <sitemap>
    <loc>https://es.wikiloc.com/sitemaps/sitemaps_es_9.xml</loc>
    <lastmod>2020-10-14</lastmod>
  </sitemap>
</sitemapindex>

```

Aquest fitxer no és exactament un fitxer que segueixi el format del protocol SiteMap. ja que no conté algun dels tags necessaris especificats en aquest protocol, com poden ser <urlset> o <url>.

En la següent imatge es pot observar el format d'aquest tipus de fitxers:

Atributo		Descripción
<sitemapindex>	obligatorio	Encapsula información acerca de todos los Sitemaps del archivo.
<sitemap>	obligatorio	Encapsula información acerca de un Sitemap concreto.
<loc>	obligatorio	Identifica la ubicación del Sitemap.  Esta ubicación puede ser un Sitemap, un archivo Atom, un archivo RSS o un archivo de texto.
<lastmod>	opcional	Identifica la hora a la que se modificó el Sitemap correspondiente. No es la hora de modificación de alguna de las páginas incluidas en el Sitemap. El valor de la etiqueta lastmod debe encontrarse en formato Fecha y hora de W3C.  Al proporcionar la marca horaria de la última modificación, permite que los rastreadores de motores de búsqueda obtengan únicamente un subconjunto de los Sitemaps del índice, es decir, el rastreador sólo podrá obtener Sitemaps que han sido modificados a partir de una fecha concreta. Este mecanismo de obtención de Sitemap incremental permite un rápido descubrimiento de nuevas URL en sitios de gran tamaño.

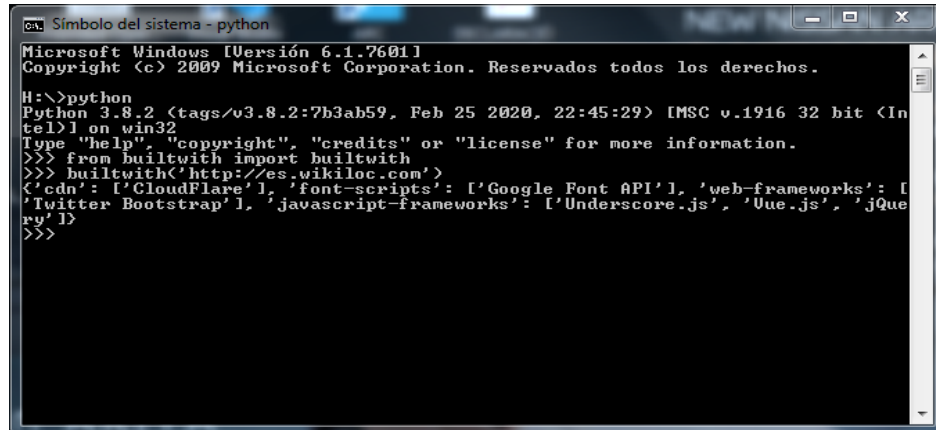
En el seu lloc ens trobem amb un arxiu índex de SiteMap, que ens permet agrupar diversos fitxers SiteMap. Aquest tipus de fitxers als quals es fa referència dins de l'arxiu índex de SiteMap han de complir unes especificacions:

- El nombre màxim d'urls no ha de superar les 50.000.
- Els fitxers no han de superar una mida de 50 MB.

En aquest cas particular l'arxiu índex de SiteMap fa referència a un total de 160 fitxers SiteMap, amb aquesta quantitat d'informació ja podem veure que es tracta d'un site amb moltes urls i que per tant hem de ser molt curiosos a l'hora de realitzar accions.

## Tecnologia utilitzada en el Site

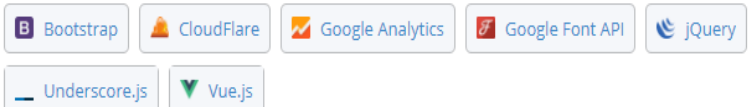
Per aconseguir aquesta informació utilitzarem el package [builtwith].



Un cop executat, a l'imatge superior podem observar les diferents tecnologies que s'utilitzen en aquest Site.

IP del servidor 104.26.5.38

Tecnologies



Herramientas de analítica

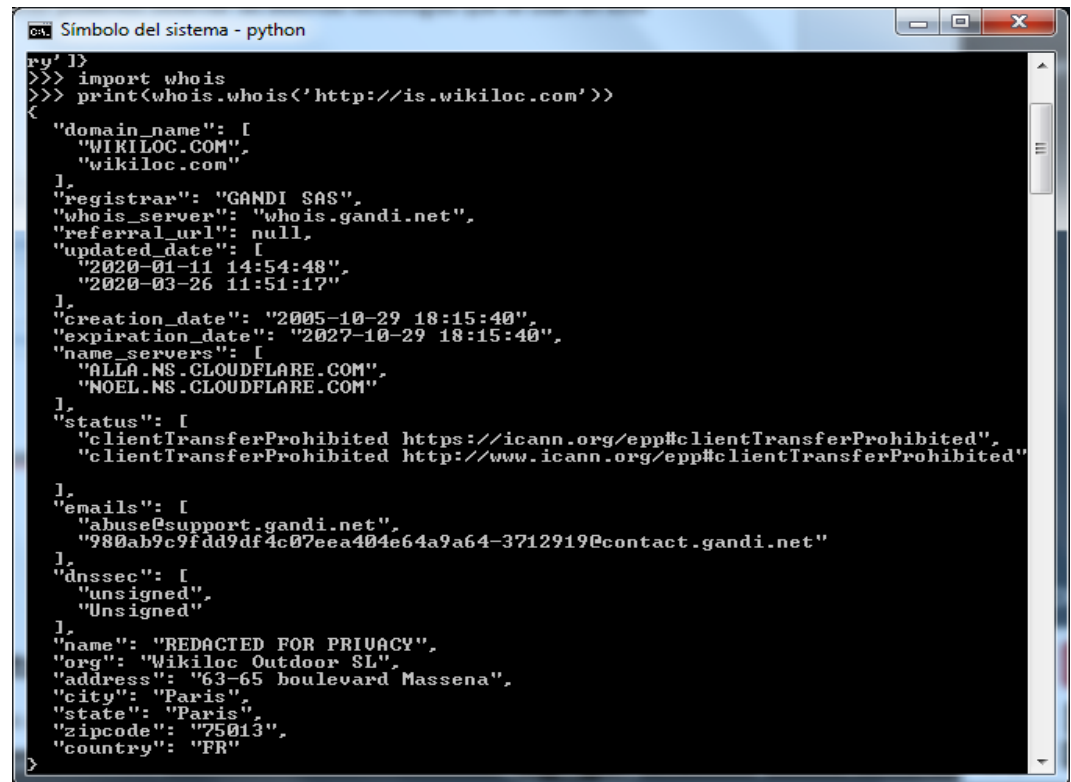


Google Analytics



## Propietari del Site

Per aconseguir informació del propietari del site utilitzarem el package [whois], el qual ens tornarà la següent informació:



```
Símbolo del sistema - python
>>> import whois
>>> print(whois.whois('http://is.wikiloc.com'))
{
  "domain_name": [
    "WIKILOC.COM",
    "wikiloc.com"
  ],
  "registrar": "GANDI SAS",
  "whois_server": "whois.gandi.net",
  "referral_url": null,
  "updated_date": [
    "2020-01-11 14:54:48",
    "2020-03-26 11:51:17"
  ],
  "creation_date": "2005-10-29 18:15:40",
  "expiration_date": "2027-10-29 18:15:40",
  "name_servers": [
    "ALLA.NS.CLOUDFLARE.COM",
    "NOEL.NS.CLOUDFLARE.COM"
  ],
  "status": [
    "clientTransferProhibited https://icann.org/epp#clientTransferProhibited",
    "clientTransferProhibited http://www.icann.org/epp#clientTransferProhibited"
  ],
  "emails": [
    "abuse@support.gandi.net",
    "980ab9c9fdd9df4c07eea404e64a9a64-3712919@contact.gandi.net"
  ],
  "dnssec": [
    "unsigned",
    "Unsigned"
  ],
  "name": "REDACTED FOR PRIVACY",
  "org": "Wikiloc Outdoor SL",
  "address": "63-65 boulevard Massena",
  "city": "Paris",
  "state": "Paris",
  "zipcode": "75013",
  "country": "FR"
}
```

## Estudi de l' Estructura del Site

El site al qual volem accedir és el: <https://ca.wikiloc.com>.

Si accedim a l'url "<https://ca.wikiloc.com/rutes>" se'ns renderitza una pàgina HTML que ens agrupa la informació en 2 grups ben diferenciats, tenim les "Activitats" i els "Països".



## Activitats

Rutes <a href="#">BTT</a>	Rutes <a href="#">Senderisme</a>	Rutes <a href="#">Bicicleta de Carretera</a>	Rutes <a href="#">Running</a>
Rutes <a href="#">Trail Running</a>	Rutes <a href="#">Alpinisme</a>	Rutes <a href="#">Cicloturisme</a>	Rutes <a href="#">Passeig</a>
Rutes <a href="#">Moto de Carretera</a>	Rutes <a href="#">Esquí de Muntanya</a>	Rutes <a href="#">Moto Trail</a>	Rutes <a href="#">Quad</a>
Rutes <a href="#">Caiaç/ Canoa</a>	Rutes <a href="#">Veler</a>	Rutes <a href="#">Raquetes de Neu</a>	Rutes <a href="#">Esquí de Fons</a>
Rutes <a href="#">Esquí Alpí</a>	Rutes <a href="#">Avió</a>	Rutes <a href="#">Passeig a Cavall</a>	Rutes <a href="#">Trineu de Gossos</a>
Rutes <a href="#">Escalada</a>	Rutes <a href="#">Patinatge en Línia</a>	Rutes <a href="#">Patinatge</a>	Rutes <a href="#">Tren</a>
Rutes <a href="#">Barranquisme</a>	Rutes <a href="#">Submarinisme</a>	Rutes <a href="#">Espeleologia</a>	Rutes <a href="#">Ala Delta</a>
Rutes <a href="#">Globus Aerostàtic</a>	Rutes <a href="#">Snowboard</a>	Rutes <a href="#">Escalada en Gel</a>	Rutes <a href="#">Moto de Neu</a>
Rutes <a href="#">Mobilitat Reduïda</a>	Rutes <a href="#">Tot Terreny</a>	Rutes <a href="#">Rem</a>	Rutes <a href="#">Cotxe</a>
Rutes <a href="#">Kitesurf</a>	Rutes <a href="#">Kite Skiing</a>	Rutes <a href="#">Trineu</a>	Rutes <a href="#">Kickbike</a>
Rutes <a href="#">Parapent</a>	Rutes <a href="#">Persones Invidents</a>	Rutes <a href="#">Marxa Nòrdica</a>	Rutes <a href="#">Moto Trial</a>
Rutes <a href="#">Moto Enduro</a>	Rutes <a href="#">Via Ferrada</a>	Rutes <a href="#">Natació</a>	Rutes <a href="#">Orientació</a>
Rutes <a href="#">Multiesport</a>	Rutes <a href="#">Paddle Surf</a>	Rutes <a href="#">Descalç</a>	Rutes <a href="#">Canicross</a>
Rutes <a href="#">Rollerski</a>	Rutes <a href="#">Longboard/ Monopati</a>	Rutes <a href="#">Monocicle de Muntanya</a>	Rutes <a href="#">Golf</a>
Rutes <a href="#">Autocaravana</a>	Rutes <a href="#">Aerolliscador</a>	Rutes <a href="#">Segway</a>	Rutes <a href="#">Camell</a>
Rutes <a href="#">Esquí Freeride</a>	Rutes <a href="#">Drone</a>	Rutes <a href="#">Vaixell Motor</a>	Rutes <a href="#">Observació d'Aus</a>
Rutes <a href="#">Bicicleta amb Remolc</a>	Rutes <a href="#">Moto d'Aigua</a>	Rutes <a href="#">Bicicleta de Mà</a>	Rutes <a href="#">Ràfting</a>
Rutes <a href="#">Descens BTT</a>	Rutes <a href="#">eBike</a>	Rutes <a href="#">Salt BASE</a>	Rutes <a href="#">Joëlette</a>
Rutes <a href="#">Cotxet de Bebé</a>	Rutes <a href="#">Splitboard</a>	Rutes <a href="#">Bicicleta Gravel</a>	

## Països

Rutes a <a href="#">Afganistan</a>	Rutes a <a href="#">Albània</a>	Rutes a <a href="#">Alemanya</a>	Rutes a <a href="#">Algèria</a>
Rutes a <a href="#">Andorra</a>	Rutes a <a href="#">Angola</a>	Rutes a <a href="#">Anguilla</a>	Rutes a <a href="#">Antigua i Barbuda</a>
Rutes a <a href="#">Antilles Holandeses</a>	Rutes a <a href="#">Aràbia Saudita</a>	Rutes a <a href="#">Argentina</a>	Rutes a <a href="#">Armènia</a>
Rutes a <a href="#">Austràlia</a>	Rutes a <a href="#">Àustria</a>	Rutes a <a href="#">Bahames</a>	Rutes a <a href="#">Bahrain</a>
Rutes a <a href="#">Bangla Desh</a>	Rutes a <a href="#">Barbados</a>	Rutes a <a href="#">Batllia de Guernsey</a>	Rutes a <a href="#">Batllia de Jersey</a>
Rutes a <a href="#">Bèlgica</a>	Rutes a <a href="#">Belize</a>	Rutes a <a href="#">Bermudes</a>	Rutes a <a href="#">Bhutan</a>
Rutes a <a href="#">Bielorússia</a>	Rutes a <a href="#">Birmània</a>	Rutes a <a href="#">Bolívia</a>	Rutes a <a href="#">Bonaire, Saint Eustatius and Saba</a>
Rutes a <a href="#">Bòsnia i Hercegovina</a>	Rutes a <a href="#">Botswana</a>	Rutes a <a href="#">Brasil</a>	Rutes a <a href="#">Brunei</a>
Rutes a <a href="#">Bulgària</a>	Rutes a <a href="#">Burkina Faso</a>	Rutes a <a href="#">Burundi</a>	Rutes a <a href="#">Cambodja</a>
Rutes a <a href="#">Camerun</a>	Rutes a <a href="#">Canadà</a>	Rutes a <a href="#">Cap Verd</a>	Rutes a <a href="#">Colòmbia</a>
Rutes a <a href="#">Comores</a>	Rutes a <a href="#">Congo - Brazzaville</a>	Rutes a <a href="#">Congo - Kinshasa</a>	Rutes a <a href="#">Corea del Nord</a>
Rutes a <a href="#">Corea del Sud</a>	Rutes a <a href="#">Costa d'Ivori</a>	Rutes a <a href="#">Costa Rica</a>	Rutes a <a href="#">Croàcia</a>
Rutes a <a href="#">Cuba</a>	Rutes a <a href="#">Dinamarca</a>	Rutes a <a href="#">Djibouti</a>	Rutes a <a href="#">Dominica</a>
Rutes a <a href="#">Egipte</a>	Rutes a <a href="#">El Benín</a>	Rutes a <a href="#">El Kazakhstan</a>	Rutes a <a href="#">El Salvador</a>
Rutes a <a href="#">El Sudan</a>	Rutes a <a href="#">El Tadjikistan</a>	Rutes a <a href="#">Emirats Àrabs Units</a>	Rutes a <a href="#">Equador</a>
Rutes a <a href="#">Eritrea</a>	Rutes a <a href="#">Eslovàquia</a>	Rutes a <a href="#">Eslovènia</a>	Rutes a <a href="#">Espanya</a>
Rutes a <a href="#">Estats Units</a>	Rutes a <a href="#">Estònia</a>	Rutes a <a href="#">Etiòpia</a>	Rutes a <a href="#">Fiji</a>
Rutes a <a href="#">Filipines</a>	Rutes a <a href="#">Finlàndia</a>	Rutes a <a href="#">França</a>	Rutes a <a href="#">Gabon</a>
Rutes a <a href="#">Gàmbia</a>	Rutes a <a href="#">Geòrgia</a>	Rutes a <a href="#">Ghana</a>	Rutes a <a href="#">Gibraltar</a>
Rutes a <a href="#">Grècia</a>	Rutes a <a href="#">Grenada</a>	Rutes a <a href="#">Groenlàndia</a>	Rutes a <a href="#">Guadeloupe</a>
Rutes a <a href="#">Guam</a>	Rutes a <a href="#">Guatemala</a>	Rutes a <a href="#">Guinea</a>	Rutes a <a href="#">Guinea Bissau</a>
Rutes a <a href="#">Guinea Equatorial</a>	Rutes a <a href="#">Guyana</a>	Rutes a <a href="#">Haití</a>	Rutes a <a href="#">Hondures</a>
Rutes a <a href="#">Hong Kong</a>	Rutes a <a href="#">Hongria</a>	Rutes a <a href="#">Iemen</a>	Rutes a <a href="#">Illa de Man</a>

Si procedim a observar l'estructura d'aquesta pàgina observem l'ús de:

- `<ul>`: mostra una llista sense ensenyar l'ordinal de la llista.
- `<li>`: mostra cadascun dels ítems de la llista.
- `<a>`: link a l' url indicada.

[illegible]

Si realitzem un clic a sobre “Andorra” se’ns dirigeix a un altre url “[ca.wikiloc.com/rutes/outdoor/andorra](https://ca.wikiloc.com/rutes/outdoor/andorra)” on observem:

[← Rutes en altres països](#)

## Rutes a Andorra

53.100 rutes ★★★★★ (1543)

Ordena

[Rutes a Andorra](#)  
[Rutes a Andorra \(general\)](#)  
[Rutes a Andorra la Vella](#)  
[Rutes a Canillo](#)  
[Rutes a Encamp](#)  
[Rutes a Escaldes-Engordany](#)  
[Rutes a la Massana](#)  
[Rutes a Ordino](#)  
[Rutes a Sant Julià](#)


**141021 ANDORRA. Fonts i bosc del Campeà.**
**6,01 quilòmetres** - Fàcil - *per GERCA*

a prop de Les Bons, Parròquia d'Encamp (Andorra)

 amb 

★★★★★ (17)


**180830 100 Cims. Andorra. ( n° 205 ) 'Pic de Tristaina' .  
Circ i estanys des de la estació de Ordino Arcalís**
**7,28 quilòmetres** - Moderat - *per GERCA*

a prop de El Santí, Ordino (Andorra)

Si obtenim l'estructura de la pàgina, observem que segueix el mateix format que anteriorment

```

▼<div id="content" class="row" style="height: auto !important;">
  ::before
  ▶<div id="trails" class="col-xs-12 col-sm-9 col-sm-push-3" style="height: auto
!important; min-height: 0px !important;">...</div>
▼<div id="filters" class="col-xs-12 col-sm-3 col-sm-pull-9">
  ▼<ul class="limited-list" data-limit="20">
    ▼<li>
      ***
      ▶<a href="https://ca.wikiloc.com/rutes/outdoor/andorra/andorra" title="Rutes
a Andorra">...</a> == $0
    </li>
    ▼<li>
      ▶<a href="https://ca.wikiloc.com/rutes/outdoor/andorra/andorra-general"
title="Rutes a Andorra (general)">...</a>
    </li>
    ▼<li>
      ▶<a href="https://ca.wikiloc.com/rutes/outdoor/andorra/andorra-la-vella"
title="Rutes a Andorra la Vella">...</a>
    </li>
    ▶<li>...</li>
    ▶<li>...</li>
    ▶<li>...</li>
    ▶<li>...</li>
    ▶<li>...</li>
    ▶<li>...</li>
  </ul>
  <br>

```

Realitzem un clic sobre “Rutes en Canillo” i se’ns dirigeix a l’url “[ca.wikiloc.com/rutes/outdoor/andorra/canillo](https://ca.wikiloc.com/rutes/outdoor/andorra/canillo)”

[← Rutes a Andorra](#)

## Rutes a Canillo (Andorra)

17.400 rutes ★★★★★ (616)

Ordena

### Llocs a Canillo

[Rutes a Bordes d'Envalira](#)  
[Rutes a Canillo](#)  
[Rutes a El Tarter](#)  
[Rutes a El Vilar](#)  
[Rutes a Els Plans](#)  
[Rutes a Incles](#)  
[Rutes a La Costa](#)  
[Rutes a l'Aldosa](#)  
[Rutes a Meritxell](#)  
[Rutes a Molleres](#)  
[Rutes a Prats](#)



**120706 ANDORRA. Camí Interparroquial ( C.I. ) . TOMB de les NERES**

**6,87 quilòmetres** - Fàcil - *per GERCA*

a prop de Molleres, Parròquia de Canillo (Andorra)

★★★★★ (14)



**Coll d'Ordino – Tomb per les Neres**

**7,97 quilòmetres** - Fàcil - *per erill*

a prop de Meritxell, Canillo (Andorra)

★★★★★ (5)



**130928 100 Cims (100). ANDORRA. 'Casamanya Nord (2752 m)' . El meu '100 Cims' N° 100**

Tornem a observar l’estructura i segueix el mateix patró que les vistes anteriorment.

```

<!-- google_ad_section_start -->
▶<div class="topad topad--landing adsearch col-xs-12">...</div>
<!-- google_ad_section_end -->
▶<div class="row lang-anchor">...</div>
▼<div id="content" class="row" style="height: auto !important;">
  ::before
  ▶<div id="trails" class="col-xs-12 col-sm-9 col-sm-push-3" style="height: auto
  !important; min-height: 0px !important;">...</div>
  ▼<div id="filters" class="col-xs-12 col-sm-3 col-sm-pull-9">
    <h4>Llocs a Canillo</h4>
    ▼<ul class="limited-list" data-limit="20">
      ▶<li>...</li>
      ▼<li>
        ***
        ▶<a href="https://ca.wikiloc.com/rutes/outdoor/andorra/canillo/canillo"
        title="Rutes a Canillo">...</a> == $0
        </li>
        ▼<li>
          ▶<a href="https://ca.wikiloc.com/rutes/outdoor/andorra/canillo/el-tarter"
          title="Rutes a El Tarter">...</a>
          </li>
          ▼<li>
            ▶<a href="https://ca.wikiloc.com/rutes/outdoor/andorra/canillo/el-vilar"
            title="Rutes a El Vilar">...</a>
            </li>
            ▶<li>...</li>
            ▶<li>...</li>
            ▶<li>...</li>

```

Realitzem un clic sobre “Rutes al Tarter”, que ens portarà a l’url  
“ca.wikiloc.com/rutes/outdoor/andorra/canillo/el-tarter”

← Rutes a Canillo

## Rutes a El Tarter, Canillo (Andorra)

31 rutes ★★★★★ (1)

### Activitats a El Tarter, Canillo (Andorra)

Rutes Senderisme a El Tarter, Canillo (Andorra)  
Rutes Trail Running a El Tarter, Canillo (Andorra)  
Rutes Esquí de Muntanya a El Tarter, Canillo (Andorra)  
Rutes Esquí Alpí a El Tarter, Canillo (Andorra)  
Rutes Snowboard a El Tarter, Canillo (Andorra)



#### pic de les pales de la cabana sorda

**9,19 quilòmetres** - Moderat - *per passamuntanyes*  
a prop de El Tarter, Parròquia de Canillo (Andorra)  
★★★★★ (1)



#### la vall del riu

**16,78 quilòmetres** - Díficil - *per Muntanya Activa*  
a prop de El Tarter, Parròquia de Canillo (Andorra)



#### Pic de la coma de varilles per la serra d'Entor

**12,65 quilòmetres** - Moderat - *per bouvier*  
a prop de El Tarter, Parròquia de Canillo (Andorra)

On ja ens apareixeran totes les rutes i el tipus de ruta que podem realitzar a la zona seleccionada.

L'estructura d'aquesta pàgina és diferent de les anteriors:



- <h3>: heading de tipus 3
- <a>: link a la url indicada.

```

    <div class="tools col-xs-12 display">...</div>
    <!-- /tools -->
    <ul class="trail-list">
      <li>...</li>
      <li>...</li>
      <li>
        <div class="row">
          ::before
          <div class="col-sm-1 col-xs-2">...</div>
          <div class="info col-md-7 col-sm-6 col-xs-10">
            <h3>
***      <a href="https://ca.wikiloc.com/rutes-senderisme/pic-de-la-coma-de-
          varilles-per-la-serra-dentor-2086042" title="ruta Pic de la coma de
          varilles per la serra d'Entor - El Tarter, Parròquia de Canillo
          (Andorra)" class="trail-title dont-break-out">Pic de la coma de
          varilles per la serra d'Entor</a> == $0
            </h3>
            <p class="meta">...</p>
            <p class="description">
              a prop de&nbsp;El Tarter,
              Parròquia de Canillo (Andorra)</p>
            </div>
          <div class="col-xs-10 col-xs-push-2 col-sm-push-0 col-sm-5 col-md-4">...
          </div>
          ::after
        </div>
      </li>
    </ul>
  </div>

```

I ara sí que ja només ens queda seleccionar la ruta desitjada.



Senderisme (Rutes Senderisme a Andorra → Canillo → Canillo)

Canillo-Refugi i Estanys Vall de Riu-  
Vall de Montaup 87

Descarrega

Afegeix a la llista Fes la teva valoració

Comparteix



Autor  
Terricat

15.352 504 11210

He fet aquesta ruta

→ 11,85 km ↺ Sí

↗ 850 m ↗ 2.605 m

↘ 850 m ↘ 1.798 m

△ Moderat

## Temps

6 hores 18 minuts

## Coordenades

1547

## Data de pujada

23 / de setembre / 2018

## Data de realització

de setembre 2018

## Valoració



informació



Fàcil de seguir



Entorn



On ja podem obtenir la informació específica de la ruta:

- Tipus: Senderisme
- Distància: 11,85 km
- Desnivell acumulat pujant: 850 m

- Desnivell acumulat baixant: 850 m
- Dificultat Tècnica: moderat
- Ruta circular: si
- Altitud màxima: 2605 m
- Altitud mínima: 1798 m

Informació addicional:

- Valoració: de 1 a 5 estrelles
- Vots: 2 vots
- Data realització: Setembre 2018
- Característiques de la ruta, indicant si existeix algun d'aquests elements:
  - Llac
  - Font
  - Riu
  - Panoràmica
  - Pàrquing
  - Pont
  - Cataracta
  - Refugi
  - Refugi Lliure

Un cop vist l'estudi previ ara el que farem serà analitzar breument el contingut del fitxer .log que genera la nostra aplicació.

```
2020-11-03 18:17:13.899032 ===== Start The Web Scrapping =====
Paràmetre activitat = outdoor
Paràmetre pais = espanya
Paràmetre regio1 = catalunya
Paràmetre regio2 = -1 (tots)
                                Paràmetres Inicials

ALLOWED URL: https://ca.wikiloc.com/rutes/outdoor/espanya/catalunya  Url inicial, país + regió1

FILTER ----> https://ca.wikiloc.com/rutes/outdoor/espanya/catalunya/abella
FILTER ----> https://ca.wikiloc.com/rutes/outdoor/espanya/catalunya/abella-de-la-conca
FILTER ----> https://ca.wikiloc.com/rutes/outdoor/espanya/catalunya/abrera
FILTER ----> https://ca.wikiloc.com/rutes/outdoor/espanya/catalunya/adrall
FILTER ----> https://ca.wikiloc.com/rutes/outdoor/espanya/catalunya/adri
FILTER ----> https://ca.wikiloc.com/rutes/outdoor/espanya/catalunya/age
FILTER ----> https://ca.wikiloc.com/rutes/outdoor/espanya/catalunya/agell
FILTER ----> https://ca.wikiloc.com/rutes/outdoor/espanya/catalunya/ager
FILTER ----> https://ca.wikiloc.com/rutes/outdoor/espanya/catalunya/agramunt
FILTER ----> https://ca.wikiloc.com/rutes/outdoor/espanya/catalunya/aquilar-de-segarra
FILTER ----> https://ca.wikiloc.com/rutes/outdoor/espanya/catalunya/aquilo
FILTER ----> https://ca.wikiloc.com/rutes/outdoor/espanya/catalunya/aquiro
FILTER ----> https://ca.wikiloc.com/rutes/outdoor/espanya/catalunya/aquillana
                                Url's descendents
```

Les primeres línies ens mostren els paràmetres que volem utilitzar a l'aplicació, en aquest cas volem obtenir totes les rutes del país [Espanya] i regió1 [Catalunya].

A continuació ens mostra que l'url resultant està permesa segons les directives del fitxer [Robots.txt] del site i tot seguit totes les urls descendents de la 1<sup>a</sup>.

```
FILTER ----> https://ca.wikiloc.com/rutes/outdoor/espanya/catalunva/zona-franca
FILTER ----> https://ca.wikiloc.com/rutes/outdoor/espanya/catalunva/zona-hoteleria
FILTER ----> https://ca.wikiloc.com/rutes/outdoor/espanya/catalunva/zona-universitaria
```

```
ALLOWED URL: https://ca.wikiloc.com/rutes/outdoor/espanya/catalunva/abella
ALLOWED URL: https://ca.wikiloc.com/rutes/outdoor/espanya/catalunva/abella?from=24&to=48
ALLOWED URL: https://ca.wikiloc.com/rutes/outdoor/espanya/catalunva/abella?from=48&to=72
ALLOWED URL: https://ca.wikiloc.com/rutes/outdoor/espanya/catalunva/abella?from=72&to=96
ALLOWED URL: https://ca.wikiloc.com/rutes/outdoor/espanya/catalunva/abella?from=96&to=120
ALLOWED URL: https://ca.wikiloc.com/rutes/outdoor/espanya/catalunva/abella?from=120&to=144
ALLOWED URL: https://ca.wikiloc.com/rutes/outdoor/espanya/catalunva/abella?from=144&to=168
ALLOWED URL: https://ca.wikiloc.com/rutes/outdoor/espanya/catalunva/abella?from=168&to=192
ALLOWED URL: https://ca.wikiloc.com/rutes/outdoor/espanya/catalunva/abella?from=192&to=216
ALLOWED URL: https://ca.wikiloc.com/rutes/outdoor/espanya/catalunva/abella?from=216&to=240
```

Per cada url obtinguda hi trobem 24 rutes

En funció de les rutes que hi hagin emmagatzemades en cada url haurem de continuar fent "gets" d'aquesta zona o no. Cada "get" ens retorna un total de 24 rutes. Hi ha moltes url's que arriben a tenir fins a 10000 rutes per visitar. Cada "allowed" associat a una url ens indica que aquesta està permesa segons les directives del fitxer "Robots.txt".

```
ALLOWED URL: https://ca.wikiloc.com/rutes/outdoor/espanya/catalunva/abella?from=768&to=792
ALLOWED URL: https://ca.wikiloc.com/rutes/outdoor/espanya/catalunva/abella?from=792&to=816
ALLOWED URL: https://ca.wikiloc.com/rutes/outdoor/espanya/catalunva/abella?from=816&to=840
ALLOWED URL: https://ca.wikiloc.com/rutes/outdoor/espanya/catalunva/abella?from=840&to=864
ALLOWED URL: https://ca.wikiloc.com/rutes/outdoor/espanya/catalunva/abella?from=864&to=888
```

```
Num rutes valorades a https://ca.wikiloc.com/rutes/outdoor/espanya/catalunva/abella -> 35
```

```
ALLOWED URL: https://ca.wikiloc.com/rutes/outdoor/espanya/catalunva/abella-de-la-conca
ALLOWED URL: https://ca.wikiloc.com/rutes/outdoor/espanya/catalunva/abella-de-la-conca?from=24&to=48
ALLOWED URL: https://ca.wikiloc.com/rutes/outdoor/espanya/catalunva/abella-de-la-conca?from=48&to=72
ALLOWED URL: https://ca.wikiloc.com/rutes/outdoor/espanya/catalunva/abella-de-la-conca?from=72&to=96
ALLOWED URL: https://ca.wikiloc.com/rutes/outdoor/espanya/catalunva/abella-de-la-conca?from=96&to=120
ALLOWED URL: https://ca.wikiloc.com/rutes/outdoor/espanya/catalunva/abella-de-la-conca?from=120&to=144
```

Un cop hem visitat totes les rutes disponibles en aquella url indiquem les rutes valorades que hi hem trobat.

```
ALLOWED URL: https://ca.wikiloc.com/rutes-a-peu/abella-serrat-de-la-varguera-abella-7876143
Warning -> La url: https://ca.wikiloc.com/rutes-a-peu/abella-serrat-de-la-varguera-abella-7876143 no té data de creació
Warning: Method getDateCreation(): Error: 'NoneType' object has no attribute 'text'
Warning -> La url: https://ca.wikiloc.com/rutes-a-peu/abella-serrat-de-la-varguera-abella-7876143 no té informació adicional
Method getCards(): Error: 'NoneType' object has no attribute 'find_all_next'
```

Com ja hem comentat anteriorment hi ha un conjunt de dades que no sempre existeixen al DOM de les url's analitzades, quan ens trobem amb aquests casos traurem missatges informatius com es pot veure a la imatge anterior.



```
Rutes trobades = 60057
Ruta de les dades: ./rutes_outdoor_espanya_catalunya.csv
2020-11-04 10:19:39.228638 ===== Elapsed Time: 57745.329606 segons =====
2020-11-04 10:19:39.228638 ===== End The Web Scrapping =====
```

Un cop finalitzi ens indicarà quantes rutes ha analitzat, el nom del fitxer on emmagatzemarà les dades i el temps que ha dedicat per obtenir tota la informació, que en aquest cas s'aproxima a les 16 h de treball.

10. Dataset. Publicar el dataset en format CSV a Zenodo (obtenció del DOI) amb una breu descripció.

DOI: 10.5281/zenodo.4246805

Target URL: <https://doi.org/10.5281/zenodo.4246805>

## Taula contribucions

Contribucions	Signa
Recerca prèvia	P.M.B, X.M.
Redacció de les respostes	P.M.B, X.M.
Desenvolupament del codi	P.M.B, X.M.