Compte rendu Webmapping des restaurants de Sortir Télérama.

Compte rendu Webmapping des restaurants de Sortir Télérama (scrapping des données, structuration, carte interactive...).



Figure restaurant le rond (image credit: Télérama)

1. Introduction

Le scrapping de donnée consiste à extraire les données de site web via un script ou un programme.

Le scrapping peut aussi peut permettre d’étudier les résultats d’une recherche par mot clef dans un moteur de recherche.

Mon projet s’est déroulé en deux temps. Dans un premier temps il faut scrapper les données. Dans un deuxième temps les données scrappées serviront. Il existe plusieurs façons de scrappées les données d’un site internet.

Je ne détaillerai pas l’étape de géocodage entre l’étape de scrapping et l’étape de webmaping. Il s’agit d’un simple géocodage avec le géocodeur de data.gouv.

1. Le scrapping de données
2. Avec le plugin web scrapper (piste suivis)

Web scrapper est un plugin de google très utile. Il permet comme pourrait le faire un programme du scrapping.

Tous d’abord il faut savoir que la manière de scrapper ou d’extraire du contenu de site web diffère pour chaque site web. La manière de scrapper diffère par exemple suivant la pagination d’un site web et de comment il est organisé.

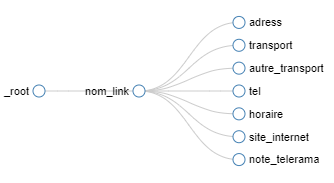
J’ai privilégié web scrapper car c’est la plus rapide à mettre en place pour le projet. Pour extraire ou scrapper avec webscrapper il faut commencer par créer une sitemap. Dans la sitemap d’indique l’url des pages ou je veux extraire du contenu.



https://sortir.telerama.fr/listes/restos?page=[1-120]

J’indique avec page=[1-120] que je voudrais scrapper de la page 1 à la page 120.

Il faut ensuite créer des sélectors.



Web scrappeur permet de faire des graph de sélector. Ce système de sélector est assez intelligent. Il est possible de mettre un sélector nom\_link pour entrer dans la page de chaque restaurant. Dans le graph je lis ensuite ce premier sélector à un deuxième sélector qui permet de récupérer du texte sur les adresses, transport….

1. Avec python (tentative)

Il est possible de scrapper en python avec la librairie Beautiful Soup.

1. Avec R (tentative)

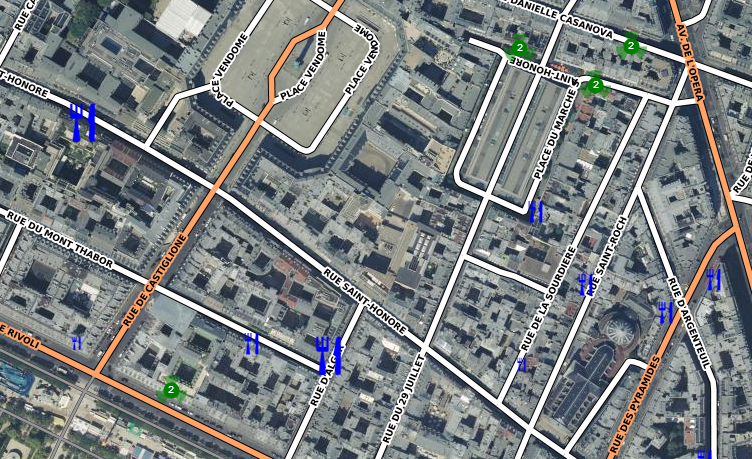
Il est aussi possible avec R de faire de scrapping.

1. Le webmapping des restaurants télérama
2. Avec un lien iframe de l’IGN (piste suivis)

J’ai opté pour une solution sans programmation. En effet créer ma page web avec une carte sans script en javascript.



Pour faire cela il faut sur le site ma carte de l’IGN importé le dataset issue du scrapping. Il est ensuite possible de faire un travail de représentation avec des icônes de différentes taille en fonction de la note de télérama. Il est aussi possible de faire des clusters pour un meilleur rendu sur les différentes échelles.



Le lien iframe est générer sur le site ma carte de l’IGN. Il n’y a plus qu’à mettre ce lien dans une div.

<div id="mapid" class="mx-auto">  
 <iframe width="853" height="480" src="https://macarte.ign.fr/carte/77f50640ef57fddcfacbd51ecd2b43d9/restaurant\_telerama" frameborder="0" marginheight="0" marginwidth="0"></iframe>  
</div>

Cela me donne ceci.



1. En programmant et en important un geojson (piste non suivis mais réalisable)

Il est aussi possible de faire des cartes en programmant en HTML, CSS, javascript avec les librairies leaflet ou openlayer.