



Web des Données et Web Sémantique

Mini-projet

Bilal ZARAKET

Enseignant :

M. Sebastien FERRE

Introduction :

Introduction :

Le Web des Données et le Web Sémantique ont ouvert la voie à une approche plus intelligente et interconnectée de la gestion des informations en ligne. Dans le cadre de ce mini-projet, mené sous la direction de M. Sebastien FERRE, j'ai exploré les possibilités offertes par ces technologies émergentes en utilisant un jeu de données extrait du "Crime Data from 2020 to Present - Catalog". Ce jeu de données documente les incidents criminels survenus dans la ville de Los Angeles depuis 2020, offrant une plongée approfondie dans les différents aspects de la criminalité.

Les données, transcrites à partir de rapports originaux, introduisent la complexité inhérente à la nature des crimes et aux variations potentielles dans la déclaration. L'objectif principal de ce projet est d'explorer, représenter et analyser ces données à l'aide du Web Sémantique, en convertissant le jeu de données brut en une représentation RDF (Resource Description Framework). Cette transformation a été réalisée à l'aide de l'outil OpenRefine, où un schéma spécifique a été élaboré pour connecter et organiser les différentes entités du jeu de données.

Le schéma RDF ainsi créé a été ensuite intégré dans une base de données Apache Jena Fuseki, permettant ainsi l'interrogation des données à l'aide du langage de requête SPARQL. Cette démarche offre une manière plus flexible et puissante d'explorer les relations entre les entités criminelles, les victimes, les lieux et les circonstances, grâce à la sémantique sous-jacente du Web des Données.

Dans la suite de ce rapport, nous présenterons les différentes entités du jeu de données, les relations établies dans le schéma RDF, ainsi que le processus d'interrogation à l'aide de SPARQL. En outre, nous détaillerons la visualisation des données obtenues, mettant en avant plusieurs aspects significatifs des crimes à Los Angeles. Enfin, nous concluons avec une analyse critique des résultats obtenus et des perspectives d'amélioration de cette approche pour une compréhension plus approfondie des données criminelles.

The github repository of this project is found in: <https://github.com/bilalzaraket/semantic-web.git>

Dans ce projet j'ai utilisé le dataset extrait du « [Crime Data from 2020 to Present - Catalog](#) », Cet ensemble de données reflète les incidents criminels survenus dans la ville de Los Angeles remontant à 2020. Ces données sont transcrites à partir de rapports de criminalité originaux qui sont tapés sur papier et il peut donc y avoir certaines inexactitudes dans les données.

Ce dataset étudie plusieurs aspects des crimes survenus dans la ville de Los Angeles, incluent :

1. DR_NO:
 - a. Signification : Numéro d'annuaire
 - b. Explication : un identifiant unique/ un numéro de référence attribué au rapport de crime.
2. Date Rptd:
 - a. Signification : Date de déclaration
 - b. Explication : La date à laquelle le crime a été signalé aux autorités.
3. DATE OCC:
 - a. Signification : Date d'apparition
 - b. Explication : la date et l'heure auxquelles le crime a eu lieu
4. TIME OCC:
 - a. Signification : le temps s'est produit
 - b. Explication : L'heure à laquelle le crime a eu lieu.
5. AREA:
 - a. Signification : indicatif régional
 - b. Explication : Un code indiquant la zone où le crime a eu lieu.
6. AREA NAME:
 - a. Signification : Nom de la zone
 - b. Explication : le nom associé à la zone où le crime a eu lieu (par exemple, Wilshire, Central, Southwest).
7. Rpt Dist No:
 - a. Signification : Numéro de district déclarant
 - b. Explication : Un numéro distinct lié au signalement du crime.
8. Part 1-2:
 - a. Signification : Classification des crimes (partie 1 ou partie 2)
 - b. Explication : Indique si le crime est classé dans la partie 1 ou dans la partie 2 selon le système de déclaration uniforme de la criminalité (DUC).
9. Crm Cd:
 - a. Signification : Code criminel
 - b. Explication : Un code représentant le type de crime.
10. Crm Cd Desc:
 - a. Signification : Description du code criminel
 - b. Explication : Une description correspondant au code criminel.
11. Mocodes 1 to Mocodes 10:



























































- a. Signification : codes de modus operandi
 - b. Explication : Ces colonnes peuvent contenir des codes ou des descriptions liés à la méthode ou à la caractéristique du crime.
- 12. Vict Age:
 - a. Signification : victim age
 - b. Explication : L'âge de la victime.
- 13. Vict Sex:
 - a. Signification: victim sex
 - b. Le sexe de la victime (M pour Homme, F pour Femme).
- 14. Vict Descent:
 - a. Signification: Victim Descent
 - b. L'origine ethnique ou raciale de la victime.
- 15. Premis Cd:
 - a. Premise code
 - b. Un code indiquant le type d'endroit où le crime a eu lieu
- 16. Premis Desc:
 - a. Premise Description
 - b. Explication : Une description correspondant au code du local.
- 17. Weapon Used Cd:
 - a. Weapon used code
 - b. Explication : Un code représentant le type d'arme utilisé dans le crime.
- 18. Weapon Desc:
 - a. Weapon Description
 - b. Explication : Une description correspondant au code de l'arme.
- 19. Status:
 - a. Criminal status code
 - b. Un code indiquant le statut du crime (par exemple, AA pour l'arrestation d'un adulte, IC pour la poursuite de l'enquête).
- 20. Status Desc:
 - a. Description of the criminal status
 - b. Explication : Une description correspondant au code de statut criminel.
- 21. Crm Cd 1 to Crm Cd 4:
 - a. Supplementary Criminal Codes
 - b. Explication : Ces colonnes peuvent contenir des codes criminels supplémentaires liés à l'incident.
- 22. Location:
 - a. Crime Location
 - b. Explication: l'adresse ou le lieu précis où le crime a eu lieu.
- 23. Cross Street:
 - a. Cross street of the location
 - b. Explication : Le nom de la rue transversale liée au lieu du crime.
- 24. LAT:
 - a. Latitude
 - b. Explication : la coordonnée de latitude géographique du lieu du crime.

25. LON:

- Longitude
- Explication : La coordonnée de longitude géographique du lieu du crime.

On a converti ce dataset à une représentation RDF utilisant l'outil OpenRefine avec ce schéma (on a pris un sample de ce dataset (2000 lignes)):



 > Premis →	<div> <div>Add object...</div> <div>  R: Premis Cd Add type... </div> </div>	<div> <div>  > Premis_Description →  > rdfs:label → Add property... </div> <div>  L: Premis Desc Add object...  L: Premis Cd Add object... </div> </div>
 > Weapon_Used →	<div> <div>Add object...</div> <div>  R: Weapon Used Cd Add type... </div> </div>	<div> <div>  > rdfs:label →  > Weapon_Desc → Add property... </div> <div>  L: Weapon Used Cd Add object...  L: Weapon Desc Add object... </div> </div>
 > Status →	<div> <div>Add object...</div> <div>  R: Status Add type... </div> </div>	<div> <div>  > status_description →  > rdfs:label → Add property... </div> <div>  L: Status Desc Add object...  L: Status Add object... </div> </div>
 > Crm_Cd_2 →	<div> <div>Add object...</div> <div>  L: Crm Cd 2 Add object... </div> </div>	
 > Crm_Cd_3 →	<div> <div>Add object...</div> <div>  L: Crm Cd 3 Add object... </div> </div>	
 > Crm_Cd_4 →	<div> <div>Add object...</div> <div>  L: Crm Cd 4 Add object... </div> </div>	
 > victim_info →	<div> <div>Add object...</div> <div>  B: Blank Add type... </div> </div>	<div> <div>  > Victim_age →  > Victim_sex → </div> <div>  L: Vict Age Add object...  L: Vict Sex Add object... </div> </div>
 > victim_info →	<div> <div>Add object...</div> <div>  B: Blank Add type... </div> </div>	<div> <div>  > Victim_age →  > Victim_sex →  > Victim_descent → Add property... </div> <div>  L: Vict Age Add object...  L: Vict Sex Add object...  L: Vict Descent Add object... </div> </div>
 > location_data →	<div> <div>Add object...</div> <div>  B: Blank Add type... </div> </div>	<div> <div>  > location_name →  > cross_street →  > Latitude →  > Longitude → Add property... </div> <div>  L: LOCATION Add object...  L: Cross Street Add object...  L: LAT Add object...  L: LON Add object... </div> </div>
 > Mocodes2 →	<div> <div>Add object...</div> <div>  L: Mocodes 2 Add object... </div> </div>	
 > Mocodes3 →	<div> <div>Add object...</div> <div>  L: Mocodes 3 Add object... </div> </div>	
 > Mocodes4 →	<div> <div>Add object...</div> <div>  L: Mocodes 4 Add object... </div> </div>	
 > Mocodes5 →	<div> <div>Add object...</div> <div>  L: Mocodes 5 Add object... </div> </div>	
 > Mocodes6 →	<div> <div>Add object...</div> <div>  L: Mocodes 6 Add object... </div> </div>	



Dans ce schéma :

- J'ai connecté LOCATION, Cross Street, LAT, LON à une Blank Node (les locations ne répètent pas)
- J'ai connecté Victime âge, Victime sex, Victime Descend à une Blank Node (les victimes ne répètent pas)
- J'ai créé des IRI : Status (connecté à Status Desc, Status), Weapon_Used (connecté à Weapon Used Cd, Weapon Desc), et premis_cd (connecté à Premis cd, Premis desc), Crm Cd (connecté à crm cd, crm cd desc), Area (connecté à AREA, AREA_NAME)

Après avoir exporté le schéma rdf résultant, nous avons utilisé la base de données Apache Jena Fuseki pour stocker les données et interagir avec elles à l'aide de SPARQL.

SPARQL Endpoint

Content Type (SELECT)

Content Type (GRAPH)

/crime/query

JSON

Turtle

```

1 PREFIX : <http://crime-register/>
2 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
4 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
5 SELECT (?lab AS ?label) (str(count(?lab)) AS ?count) WHERE {
6   ?s :status ?stat .
7   ?stat rdfs:label ?lab
8 }
9 GROUP BY ?lab
10
11
12 #SELECT DISTINCT ?domain
13 #WHERE {
14 #  ?s :AREA ?a .

```

Table

Response

5 results in 0.084 seconds

Simple view

Ellipse

Filter query res

	label	count
1	JD	5
2	IC	1735
3	JA	5
4	AA	83
5	AO	172

Après avoir démarré le serveur fuseki, nous pouvons maintenant envoyer des requêtes sparql à l'API du serveur et extraire de nos données rdf stockées.

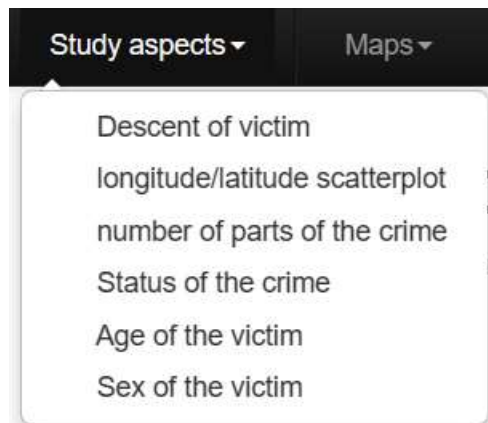
J'ai ensuite utilisé une bibliothèque appelée [d3sparql](#), une extension de la bibliothèque javascript d3 pour visualiser nos données. (Utilisant un serveur Node.js)



Welcome to the Crime Study App

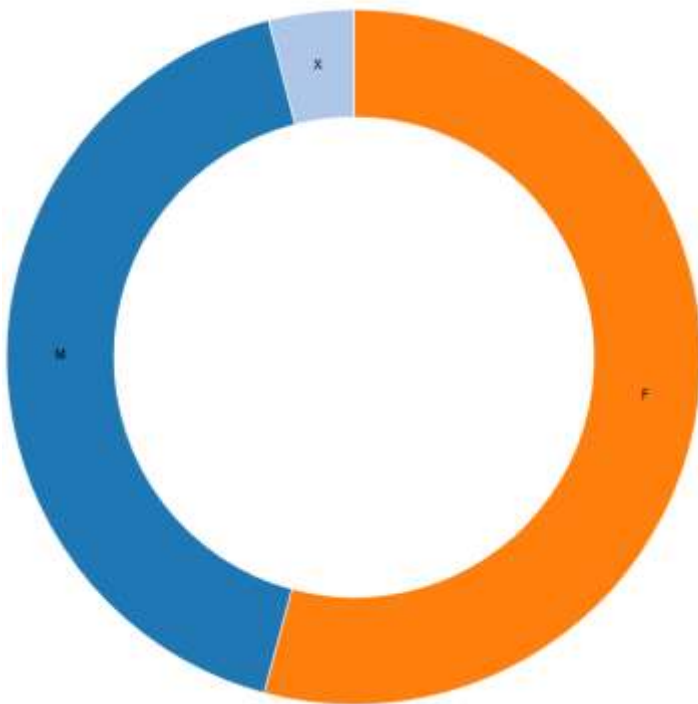
Explore and analyze crime data in the USA.

Si on choisit «Study Aspects », on va avoir une liste des diagrammes :



Chaque page a un diagramme différent qui représente des aspects différents de notre donnée.

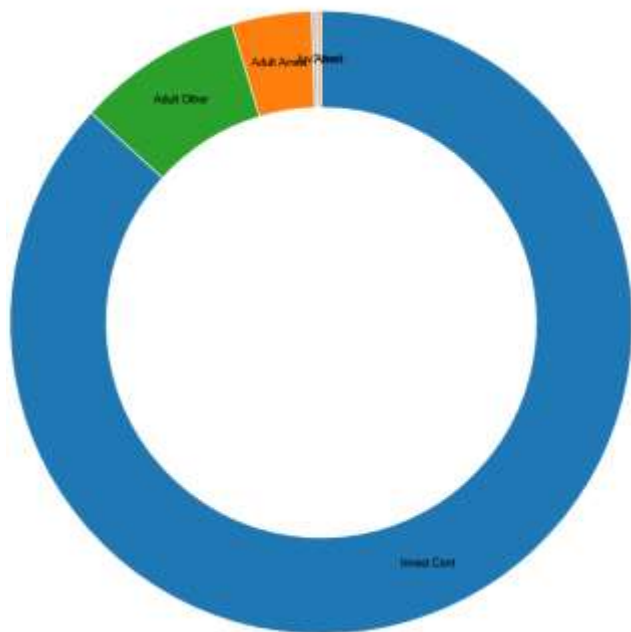
Par exemple, si on choisit Sex of the victim, on va avoir un piechart qui représentes les sexes des victimes



La proportion de femmes victimes de crimes est plus élevée que celle des hommes à Los Angeles

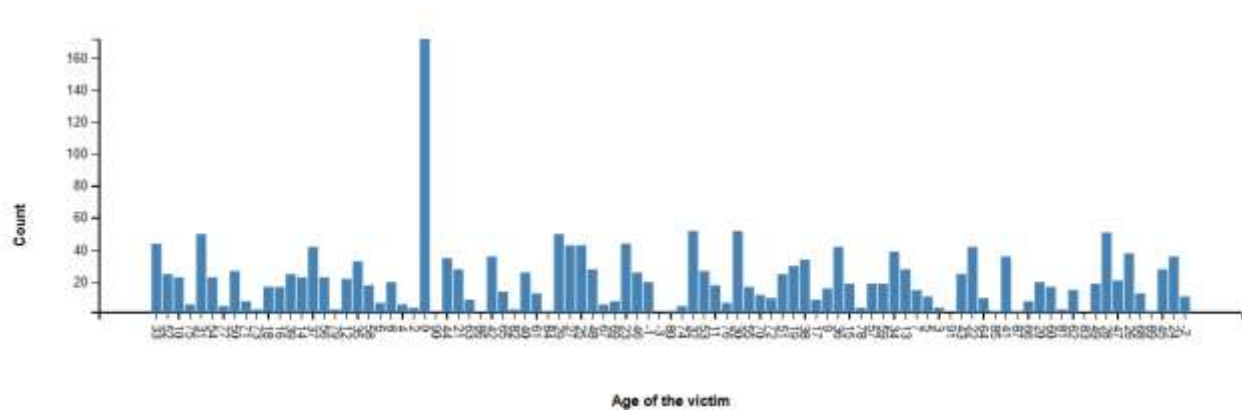
Status of the Crime :

La proportion des crimes de invest Cont sont plus élevée que les autres crimes



Victim age :

On voit ici que les crimes contre l'âge «0» sont élevées, qui signifie qu'il y a un problème avec le dataset.



```

PREFIX : <http://crime-register/>

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

SELECT (?Victim_age AS ?vage) (str(COUNT(?Victim_age)) AS ?count) WHERE {

    ?s :victim_info ?vi .

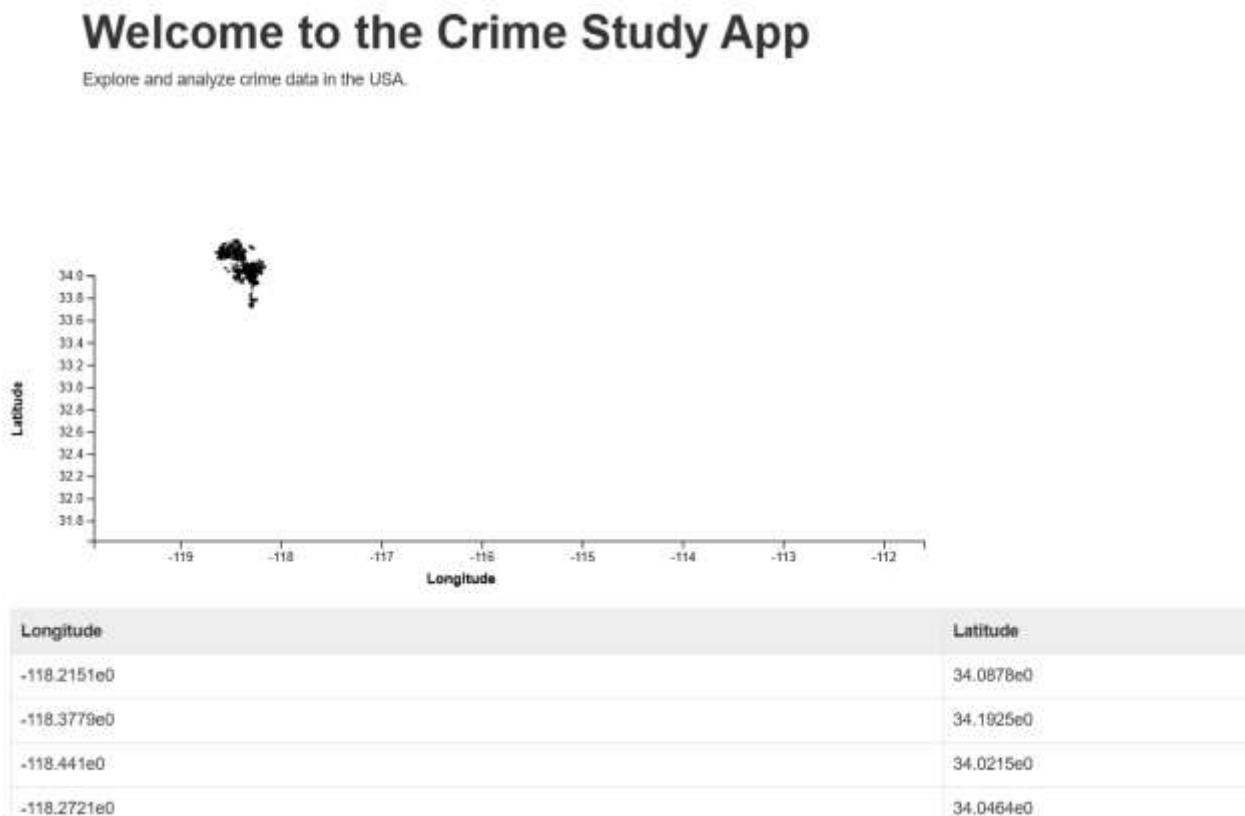
    ?vi :Victim_age ?Victim_age

}

GROUP BY ?Victim_age

```

Représentation des endroits des crimes :



```
PREFIX : <http://crime-register/>

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT DISTINCT (str(?lon) as ?Longitude) (str(?lat) as ?Latitude)
WHERE {
    ?s :location_data ?b1 .
    ?b1 :Latitude ?lat .
    ?b1 :Longitude ?lon .
}
```

Maps :

Pour représenter les endroits des crimes, j'ai implémenté une fonction pour représenter les crimes sur une carte (dans d3sparql.js : crimeLocationsMap), en utilisant des fichiers écrit le format topojson : [037.json](#) (pour représenter les crimes sur la carte de Los Angeles), [countries-10m.json](#) (la carte de USA).

