

Projet Jeux Olympiques - Journal de bord

T. Burnel, N. Grim, M. Griveau, M. Mechentel

Janvier 2024

1 Introduction

Nous sommes une équipe de datajournalistes chargée d'étudier l'existence ou non de liens entre le succès d'un pays aux Jeux Olympiques et sa richesse.

2 Jeux de données

3 Traitement des données

3.1 Objectif du traitement

3.2 Chaîne de traitement

Pensez à parler aussi de vos échecs, des enrichissements ou croisements qui n'ont pas marché... et de ce que vous avez appris.

3.2.1 Requête SPARQL

Pour enrichir nos données *via* Wikidata, notre objectif était de mettre au point une requête SPARQL retournant le nombre d’habitants de tous les pays du monde sur trente ans (1993 - 2003).

```
SELECT ?paysLabel ?population ?date
WHERE
{
  ?pays wdt:P31 wd:Q6256.
  ?pays p:P1082 ?populationStatement.
  ?populationStatement ps:P1082 ?population.
  ?populationStatement pq:P585 ?date.
  FILTER(YEAR(?date) >= (YEAR(NOW()) - 30)).
  SERVICE wikibase:label { bd:serviceParam wikibase:language
    "[AUTO_LANGUAGE],fr". }
}
ORDER BY ?paysLabel ?date
```

Être en mesure de requêter la liste de l’ensemble des pays a été la première étape de la construction de notre requête. Le premier triplet utilise la variable inconnue **?pays** dont l’objet est **country** (wd :Q6256) :

```
?pays wdt:P31 wd:Q6256.
```

La deuxième partie de la requête doit retourner la liste du nombre d’habitants de chaque pays en prenant en compte une dimension chronologique. La

complexité de cette demande correspond à un parcours de graph en quatre temps – et non pas en trois.

Nous avons donc créé une nouvelle variable, **?populationStatement**, contenant toutes les propriétés utilisées dans la classe **population** grâce au préfixe **p**¹. Pour obtenir le nombre d’habitants, nous avons utilisé le préfixe **ps** permettant d’obtenir la valeur de la propriété relative à la population :

```
?pays p:P1082 ?populationStatement.  
?populationStatement ps:P1082 ?population.
```

Enfin, nous avons utilisé le préfixe **pq** pour récupérer la valeur chronologique dans la variable **?date**. Un filtre a été appliqué pour exprimer les limites de notre période, soit **NOW** pour 2023 et **-30** pour 1993 :

```
?populationStatement pq:P585 ?date.  
FILTER(YEAR(?date) >= (YEAR(NOW()) - 30)).
```

La toute dernière ligne de la requête – couplée à la toute première – permet de contraindre l’affichage des résultats. Lorsque la requête s’exécute, le service Wikidata renvoie le nom de chaque pays (**?paysLabel**) par ordre alphabétique, le nombre d’habitants (**?population**) et l’année correspondante (**?date**) par ordre croissant :

```
SELECT ?paysLabel ?population ?date  
ORDER BY ?paysLabel ?date
```

¹Pour ce faire, nous avons consulté la liste de préfixes Wikidata : <https://www.wikidata.org/wiki/EntitySchema:E49>.

Nous nous sommes heurtés à plusieurs erreurs avant de rendre la requête fonctionnelle. Nous avons dû nous éloigner du parcours de graphique à trois étapes vu en cours et comprendre la nécessité de recourir à la variable **?paysStatement**. La documentation a été d'un grand secours pour nous aider à saisir l'utilisation des propriétés **population**² et **date**³.

Nous avons également pris en exemple la requête *Population in Europe after 1960*. La consulter a été l'occasion d'un meilleur apprentissage sur les propriétés Wikidata, une porte ouverte à la lecture de la documentation et à l'appréhension des requêtes en deux temps : d'abord la collecte des propriétés d'une classe puis le requêtage de chacune d'elles en fonction de notre besoin.

4 Visualisation des données

4.1 Présentation des visualisations

4.2 Analyse

Qu'apprend-on en regardant les visualisations ? Quels sont les biais ?

²Confer. <https://www.wikidata.org/wiki/Property:P1082>.

³Confer. <https://www.wikidata.org/wiki/Property:P585>.

Table des matières

1	Introduction	1
2	Jeux de données	1
3	Traitement des données	1
3.1	Objectif du traitement	1
3.2	Chaîne de traitement	1
3.2.1	Requête SPARQL	2
4	Visualisation des données	4
4.1	Présentation des visualisations	4
4.2	Analyse	4