

**CENTRALE
LYON**

ÉCOLE CENTRALE LYON

MOD BIG DATA
RAPPORT

BE Linked Data

Élève :
Corentin TRIBOULET

Tuteurs :
Stephane DERRODE

1 Introduction

Dans ce TP, il est demandé d'écrire une requête originales sur une base linked data. Nous utiliserons la base de donnée de la Bibliothèque Nationale de France. Ainsi le *SparQL endpoint* est le suivant <https://data.bnf.fr/sparql/>.

L'objectif de cette requête est d'obtenir l'année de naissance des auteurs mort cette année et d'identifier les 10 années de naissance les plus représentées.

2 Code de la requête

```
1 PREFIX bnf-onto: <http://data.bnf.fr/ontology/bnf-onto/>
2 PREFIX foaf: <http://xmlns.com/foaf/0.1/>
3 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
4 PREFIX bio: <http://vocab.org/bio/0.1/>
5
6 SELECT (SUBSTR(?naissance, 1, 4) AS ?year_naissance) (COUNT(?mort)
7         AS ?count_mort)
8 WHERE {
9     ?auteur a foaf:Person;
10            bnf-onto:lastYear ?mort;
11            bio:birth ?naissance;
12 FILTER (xsd:integer(?mort) = 2024)
13 }
14 GROUP BY (SUBSTR(?naissance, 1, 4))
15 ORDER BY DESC(?count_mort)
16 LIMIT 10
```

Explications :

- Ligne 1-4 : Récupération des données
- Ligne 6 : `SUBSTR(?naissance, 1, 4)` permet de récupérer uniquement l'année (YYYY) car certaines dates sont au format YYYY-MM-DD
- Ligne 11 : `xsd:integer(?mort)` convertit la valeur de `?mort` en entier pour permettre une comparaison numérique avec 2024.
- Ligne 14 : On groupe selon l'année de naissance pour compter le nombre de morts
- Ligne 15 : On trie le résultat de manière à afficher le TOP 10 des années de naissance les plus représentées

3 Graph RDF de la requête

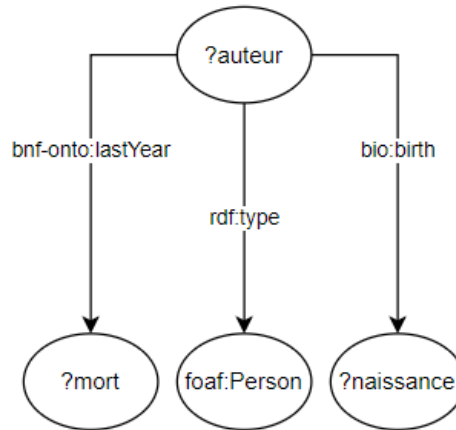


FIGURE 1 – Graph RDF

4 Résultat de la requête

	year_naissance	count_mort
1	1935	"36" ^{^^xsd:integer}
2	1930	"36" ^{^^xsd:integer}
3	1936	"34" ^{^^xsd:integer}
4	1938	"34" ^{^^xsd:integer}
5	1931	"32" ^{^^xsd:integer}
6	1934	"32" ^{^^xsd:integer}
7	1940	"32" ^{^^xsd:integer}
8	1939	"31" ^{^^xsd:integer}
9	1944	"28" ^{^^xsd:integer}
10	1946	"27" ^{^^xsd:integer}

FIGURE 2 – Résultat

Avec l'entièreté des données de la requête, il est possible de tracer l'histogramme du nombre de décès par année de naissance des auteurs morts cette année.

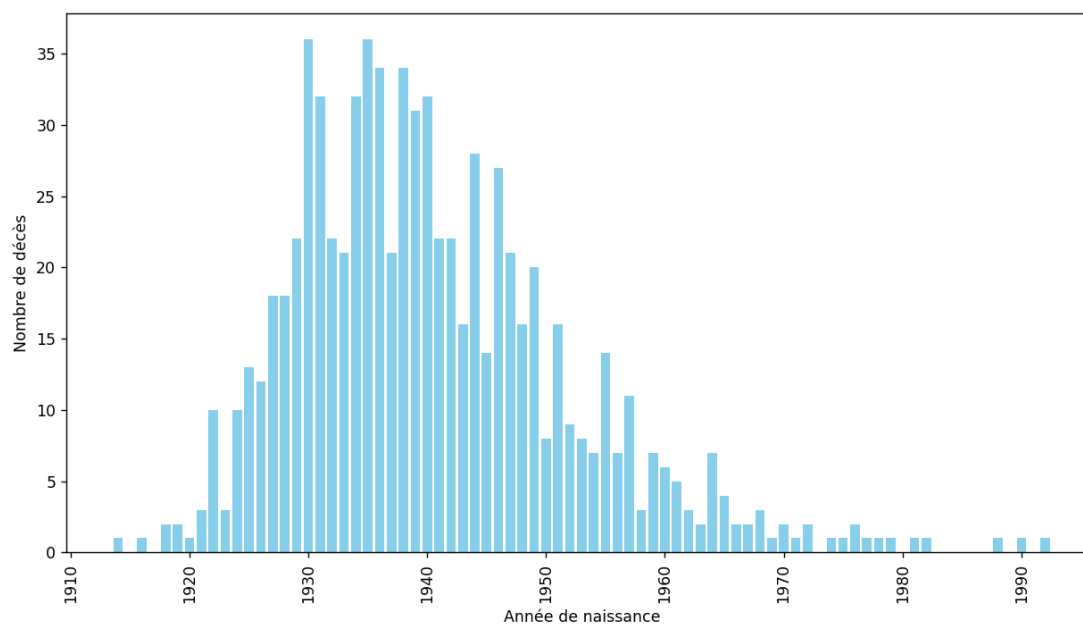


FIGURE 3 – Histogramme du nombre de décès par année de naissance