Data Mining

Données ouvertes liées (Linked Open Data)

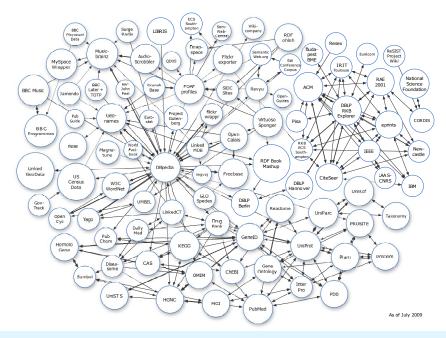
John Samuel CPE Lyon

Year: 2020-2021

Email: john(dot)samuel(at)cpe(dot)fr



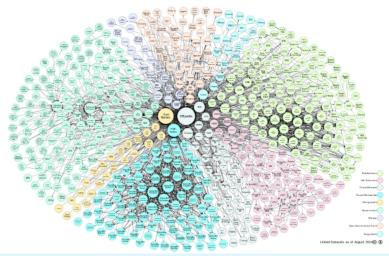
Linked Open Data



Linked Open Data: Représentation des relations (2009)

1. https://commons.wikimedia.org/wiki/File:Lod-datasets_2009-07-14.svg

Linked Open Data



Données ouvertes liées (Linked Open data: LOD), 2010

1. https://commons.wikimedia.org/wiki/File:Lod-datasets_2010-09-22_colored.png

Wikidata

- o Début en 2012
- o Une base de connaissances libre, ouverte, liée, structurée, collaborative et multilingue
- o Des sites Wikipédia multilingues à plusieurs sous-domaines à un site multilingue à un seul domaine



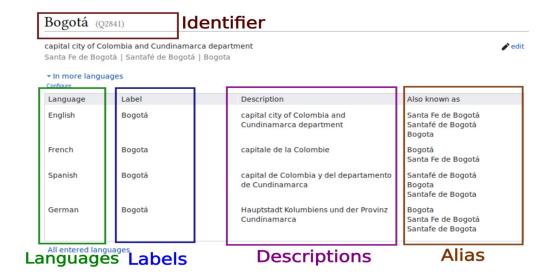
Wikipédia: Articles multilingues



Wikipédia: Articles multilingues

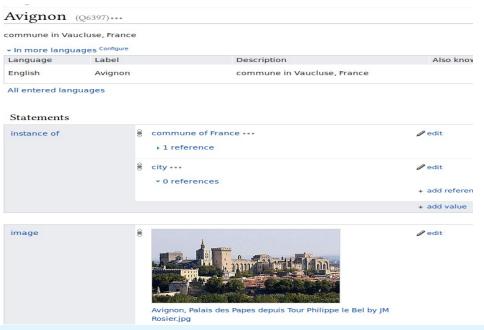
Data Mining | John Samuel 5 < | >

Élément de Wikidata: Étiquettes,



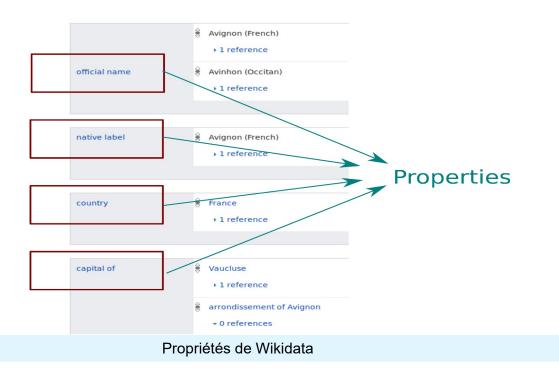
Exemple: Bogotá(Q2841)

Élément (item) de Wikidata: Étiquettes et

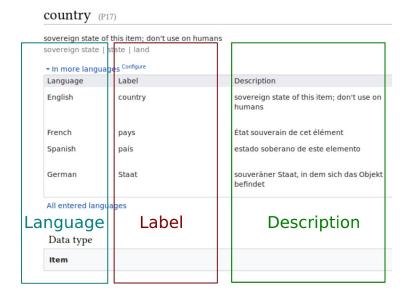


Exemple: Avignon (Q6397)

Propriétés de Wikidata



Propriétés: Étiquettes et descriptions



https://www.wikidata.org/wiki/Property:P17

Exemple: pays (P17)

Propriétés

instance of (P31)

that class of which this subject is a particular example and member (subject typically an individual member with a proper | Pedit label); different from P279; using this property as a qualifier is deprecated—use P2868 or P3831 instead is a | is an | has class | has type | is a particular | is a specific | is an individual | is a unique | is an example of | member of | unique individual of | distinct member of | non-type member of | unsubclassable example of | uninstantiable instance of | unsubclassifiable member of | not a type but is instance of | unsubtypable particular | unitary element of class | distinct element of | distinct individual member of | occurrence of | rdf:type | type

▼ In more languages

Language	Label	Description	Also known as
English	instance of	that class of which this subject is a particular example and member (subject typically an individual member with a proper name label); different from P279; using this property as a qualifier is deprecated—use P2868 or P3831 instead	is a is an has class has type Is a particular is a specific is an individual is a unique is an example of member of unique individual of distinct member of unsubclassable example of unsubclassable example of unsubclassable particular unitary element of closs distinct element of distinct individual member of occurrence of rdf:type type

Étiquettes, descriptions et alias de Property:P31 en anglais

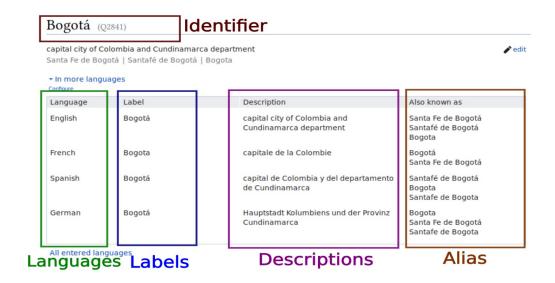
Propriétés de l'Infobox de Wikipedia



Propriétés existantes de l'Infobox de Wikipedia en anglais pour une ville colombienne

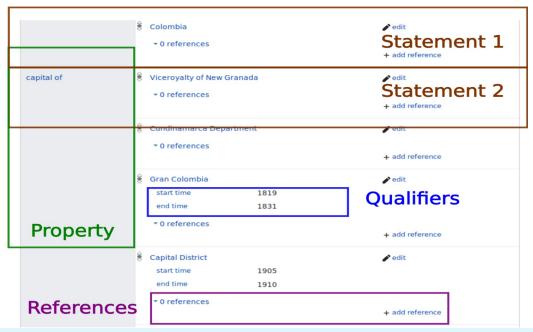
Data Mining | John Samuel 11 < | >

Identifiant, étiquettes, descriptions et alias



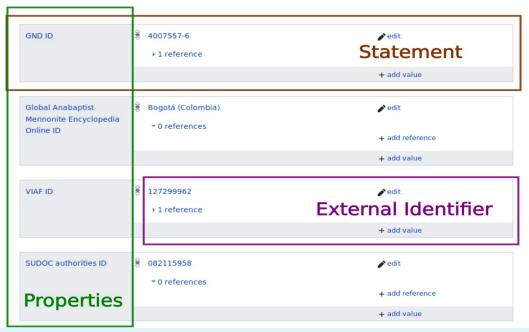
Traductions actuelles: Bogotá (Q2841)

Wikidata: Informations, propriétés,



Informations actuelles: Bogotá (Q2841)

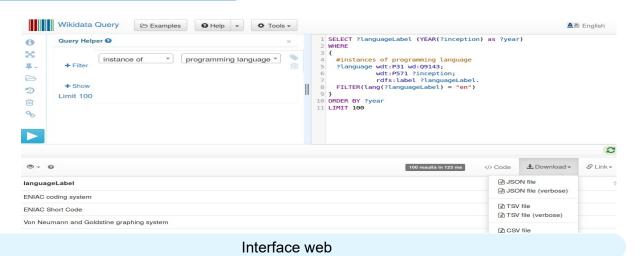
Wikidata: External identifiers



Identifiants externes actuels: Bogotá (Q2841)

Data Mining

Requêtes SPARQL



Requêtes SPARQL

Les noms de 100 langages de programmation.

```
SELECT ?languageLabel (YEAR(?inception) as ?year)
WHERE
{
    #instances of programming language
    ?language wdt:P31 wd:Q9143;
    wdt:P571 ?inception;
    rdfs:label ?languageLabel.
    FILTER(lang(?languageLabel) = "en")
}
ORDER BY ?year
LIMIT 100
```

Requêtes SPARQL

Les noms de 100 langages de programmation et leurs paradigmes.

```
SELECT ?languageLabel ?paradigmLabel (YEAR(?inception) as ?year)
WHERE
{
    #instances of programming language
    ?language wdt:P31 wd:Q9143;
    wdt:P571 ?inception; #inception
    wdt:P3966 ?paradigm; #programming language paradigm
    rdfs:label ?languageLabel. #label
    ?paradigm rdfs:label ?paradigmLabel #label
    FILTER(lang(?languageLabel) = "en" && lang(?paradigmLabel) = "en") #English
}
ORDER BY ?year ?paradigmLabel
LIMIT 100
```

Requêtes SPARQL

Les informations disponibles sur la population des différents pays à différentes périodes.

```
SELECT DISTINCT ?countryLabel (YEAR(?date) as ?year) ?population
WHERE {
   ?country wdt:P31 wd:Q6256; #Country
      p:P1082 ?populationStatement;
   rdfs:label ?countryLabel. #Label
   ?populationStatement ps:P1082 ?population; #population
   pq:P585 ?date. #period in time
   FILTER(lang(?countryLabel)="en") #Label in English
}
ORDER by ?countryLabel ?year
LIMIT 1000
```

Références

Ressources en ligne

- Resource Description Framework (RDF) https://www.w3.org/RDF/
- Resource Description Framework https://en.wikipedia.org/wiki/Resource_Description_Framework

Références

Couleurs

• Color Tool - Material Design

Images

• Wikimedia Commons