



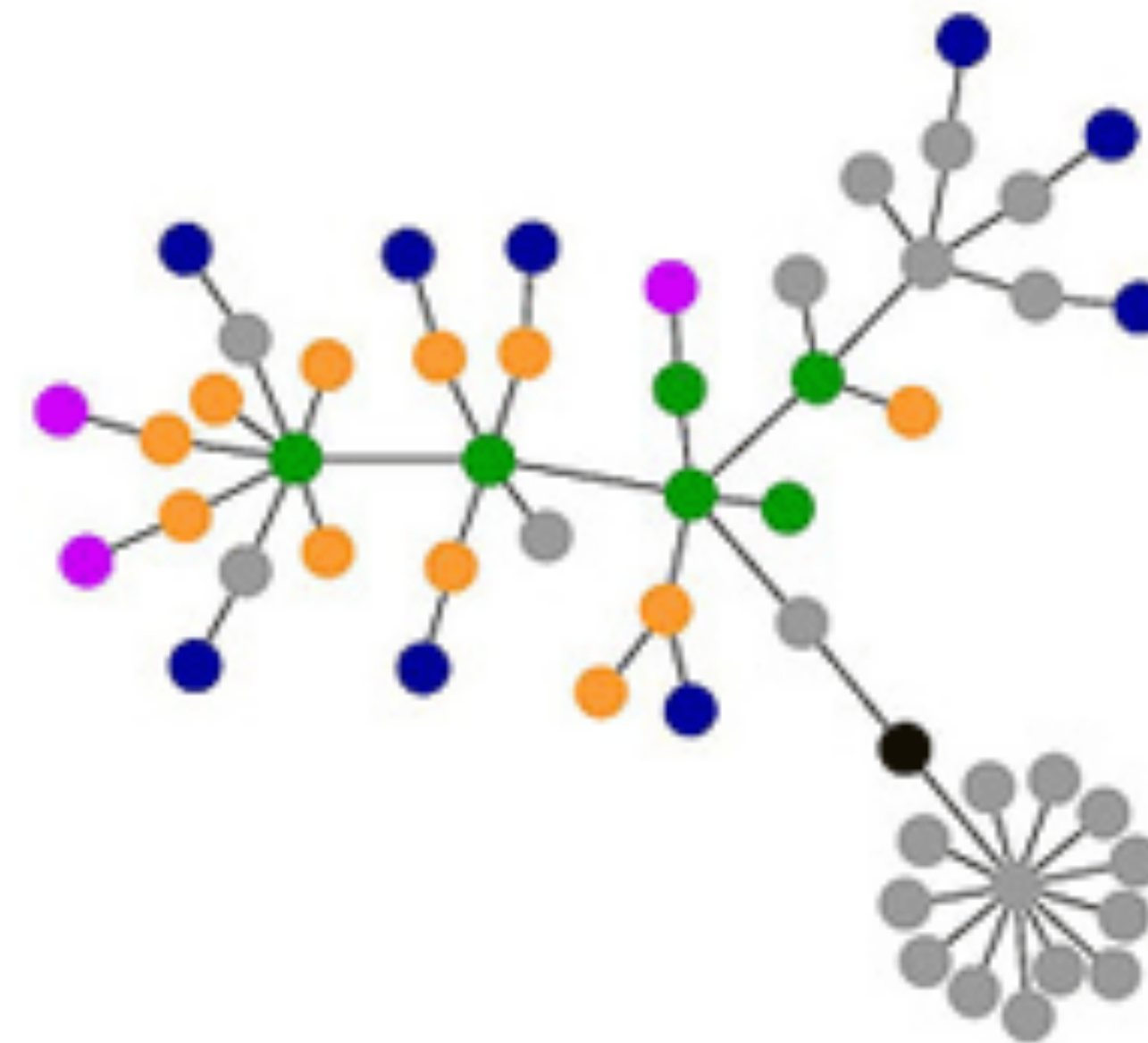
Explorer à l'aide du Web sémantique

EXPLO-SHS, La Rochelle, 15 oct. 2020



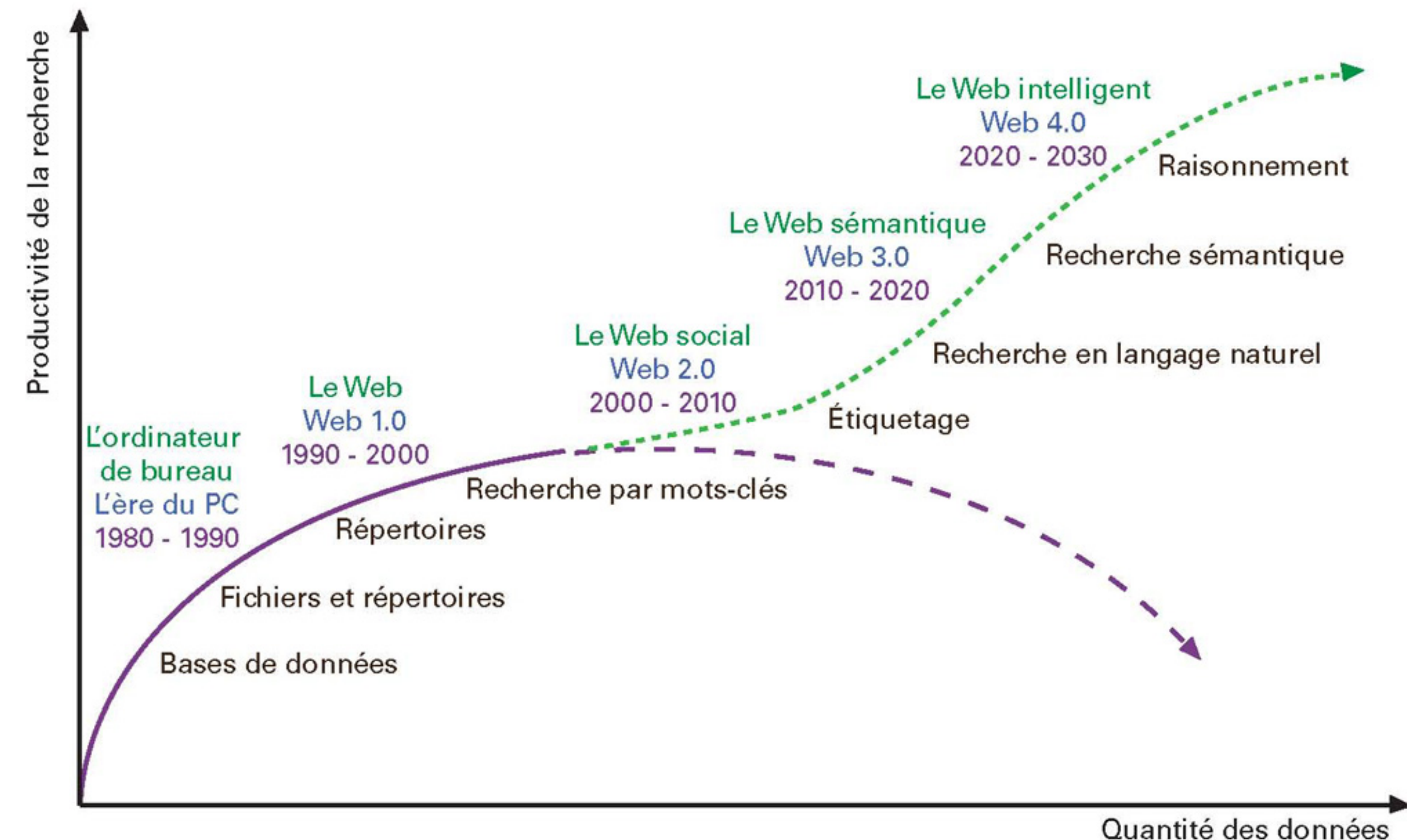
Plan

- Web sémantique : point d'actualité
- Mise en œuvre des principes du Web sémantique
- Champs des possibles pour l'exploration des données



Web sémantique

- Principes du Web : URI, HTTPS, HTML, Hypertextualités
- Utiliser les principes du Web pour structurer de l'information (des données, des documents, etc.)
- Utiliser des outils normalisés (par des instances, des communautés de recherche, etc.) : vocabulaires, ontologie, référentiels...
- Rendre explicite l'information pour la traiter avec des programmes (IA, DL, ML)



Source : François FEYLER, « Web sémantique, Web 3.0, Web de données : trois appellations pour une même notion », <http://www.cndp.fr/ecolenumerique/tous-les-numeros/numero-5-septembre-2010/focus-recherche-documentaire/article/article/le-web-semantique-une-approche-nouvelle-de-lacces-a-linformation-pertinente.html>

Web sémantique

<https://isidore.science/document/10670/1.6g31fu>

The screenshot shows a web browser window with the URL <https://isidore.science/document/10670/1.6g31fu>. The page is titled "Du livre imprimé au Web sémantique : le projet du Dictionnaire des éditeurs français du XIXe siècle".

Fiche du document

Auteurs
Jean-Charles Geslot
Viera Rebolledo-Dhuin

Date
30 juillet 2020

Type de document
Articles

Langue
Français

Mots-clés

histoire histoire du livre
Web sémantique
base de données
prosopographie
constitution de corpus
dictionnaire éditeur
Web history
history of books
semantic Web database
prosopography
corpus building dictionary
publisher Web

Sujets proches

En | Es | Fr

Lexique Vocabulaire
Web WWW
La Toile (Internet) W3
World Wide Web

Citer ce document

Jean-Charles Geslot et al., « Du livre imprimé au Web sémantique : le projet du Dictionnaire des éditeurs français du XIXe siècle »,

Résumé Fr | En

Cet article vise à présenter les enjeux techniques du projet DEF19 (Dictionnaire des éditeurs français du XIXe siècle). Il montre comment les choix scientifiques opérés pour définir le périmètre de ce dictionnaire (définition de ce qu'est un éditeur, recensement de celles et ceux qui entrent dans cette catégorie...) ont en grande partie déterminé les solutions techniques. Il décrit les diverses expériences menées pour mettre au point la base de données, à savoir les logiciels expérimentés et les problèmes qui ont été rencontrés (intégration de données hétérogènes issues de bases différentes, notamment), ainsi que la façon dont ces derniers ont pu être résolus. Il analyse l'intérêt que le Web sémantique, à travers l'usage de la plateforme Omeka S, peut représenter pour une telle entreprise scientifique, en montrant également les diverses potentialités qui se sont ouvertes au fil de la réalisation du projet en termes de collaborations techniques et institutionnelles.

humanités numériques

4 | 2020

Du livre imprimé au Web sémantique : le projet du Dictionnaire des éditeurs français du XIXe siècle

From the Printed Book to the Semantic Web: Towards a Dictionary of 19th-century French Publishers

Jean-Charles Geslot et Viera Rebolledo-Dhuin

FRANÇAIS | ANGLAIS

Cet article vise à présenter les enjeux techniques du projet DEF19 (Dictionnaire des éditeurs français du XIXe siècle). Il montre comment les choix scientifiques opérés pour définir le périmètre de ce dictionnaire (définition de ce qu'est un éditeur, recensement de celles et ceux qui entrent dans cette catégorie...) ont en grande partie déterminé les solutions techniques. Il décrit les diverses expériences menées pour mettre au point la base de données, à savoir les logiciels expérimentés et les problèmes qui ont été rencontrés (intégration de données hétérogènes issues de bases différentes, notamment), ainsi que la façon dont ces derniers ont pu être résolus. Il analyse l'intérêt que le Web sémantique, à travers l'usage de la plateforme Omeka S, peut représenter pour une telle entreprise scientifique, en montrant également les diverses potentialités qui se sont ouvertes au fil de la réalisation du projet en termes de collaborations techniques et institutionnelles.

Web sémantique

<https://isidore.science/document/10670/1.6g31fu>



The screenshot shows a browser window with the URL <https://isidore.science/document/10670/1.6g31fu>. The page displays an RDF table with two columns: the first column contains URIs (predicates) and the second column contains their corresponding values. The table is titled "predicat" and lists various properties such as "description", "publisher", "subject", "type", "title", "date", "language", "provenance", "accessRights", and "subject".

| predicat | |
|---|---|
| http://www.w3.org/1999/02/22-rdf-syntax-ns#type | http://isidore.science/ontology#article |
| http://www.w3.org/1999/02/22-rdf-syntax-ns#type | http://www.openarchives.org/ore/terms/Aggregation |
| http://www.w3.org/1999/02/22-rdf-syntax-ns#type | http://isidore.science/class/publications |
| http://www.w3.org/1999/02/22-rdf-syntax-ns#type | http://isidore.science/class/Document |
| http://purl.org/dc/elements/1.1/description | "This article aims to describe technical issues of DEF19 project" |
| http://purl.org/dc/elements/1.1/description | "Cet article vise à présenter les enjeux techniques du projet DEF" |
| http://purl.org/dc/elements/1.1/publisher | nodeID://b5242301 |
| http://purl.org/dc/elements/1.1/publisher | nodeID://b5242302 |
| http://purl.org/dc/elements/1.1/subject | "Web"@fr |
| http://purl.org/dc/elements/1.1/subject | "base de données"@fr |
| http://purl.org/dc/elements/1.1/subject | "dictionary"@en |
| http://purl.org/dc/elements/1.1/subject | "dictionnaire"@fr |
| http://purl.org/dc/elements/1.1/subject | "histoire"@fr |
| http://purl.org/dc/elements/1.1/subject | "history"@en |
| http://purl.org/dc/elements/1.1/subject | "Web"@en |
| http://purl.org/dc/elements/1.1/subject | "database"@en |
| http://purl.org/dc/elements/1.1/subject | "histoire du livre"@fr |
| http://purl.org/dc/elements/1.1/subject | "prosopographie"@fr |
| http://purl.org/dc/elements/1.1/subject | "prosopography"@en |
| http://purl.org/dc/elements/1.1/subject | "Web sémantique"@fr |
| http://purl.org/dc/elements/1.1/subject | "publisher"@en |
| http://purl.org/dc/elements/1.1/subject | "éditeur"@fr |
| http://purl.org/dc/elements/1.1/subject | "constitution de corpus"@fr |
| http://purl.org/dc/elements/1.1/subject | "semantic Web"@en |
| http://purl.org/dc/elements/1.1/subject | "corpus building"@en |
| http://purl.org/dc/elements/1.1/subject | "history of books"@en |
| http://purl.org/dc/elements/1.1/type | "article" |
| http://purl.org/dc/terms/alternative | "From the Printed Book to the Semantic Web: Towards a Dictionary" |
| http://purl.org/dc/terms/identifier | "10670/1.6g31fu" |
| http://purl.org/dc/terms/identifier | " http://journals.openedition.org/revuehn/426 " |
| http://purl.org/dc/terms/title | "Du livre imprimé au Web sémantique : le projet du Dictionnaire d" |
| http://www.openarchives.org/ore/terms/aggregates | http://journals.openedition.org/revuehn/426 |
| http://www.openarchives.org/ore/terms/aggregates | http://journals.openedition.org/revuehn/tei/426 |
| http://purl.org/dc/terms/coverage | http://sws.geonames.org/3017382/ |
| http://purl.org/dc/elements/1.1/date | "2020-07-30" |
| http://purl.org/dc/terms/creator | http://isidore.science/a/geslot_jean_charles |
| http://purl.org/dc/terms/creator | http://isidore.science/a/rebolledo_dhuin_viera |
| http://purl.org/dc/terms/date | 2020-07-30T00:00:00 |
| http://purl.org/dc/terms/language | http://lexvo.org/id/iso639-3/fra |
| http://purl.org/dc/terms/language | "fr" |
| http://purl.org/dc/terms/provenance | http://journals.openedition.org/revuehn |
| http://www.openarchives.org/ore/terms/isAggregatedBy | http://isidore.science/collection/10670/2.xlh1rj |
| http://purl.org/dc/terms/accessRights | "info:eu-repo/semantics/openAccess" |
| http://purl.org/dc/terms/subject | http://id.loc.gov/authorities/subjects/sh2002006425 |
| http://purl.org/dc/terms/subject | http://id.loc.gov/authorities/subjects/sh00006807 |
| http://purl.org/dc/terms/subject | http://id.loc.gov/authorities/subjects/sh2002006426 |
| http://purl.org/dc/terms/subject | http://id.loc.gov/authorities/subjects/sh85149022 |
| http://purl.org/dc/terms/subject | http://id.loc.gov/authorities/subjects/sh99002428 |
| http://purl.org/dc/terms/subject | http://id.loc.gov/authorities/subjects/sh85055484 |
| http://purl.org/dc/terms/subject | http://id.loc.gov/authorities/subjects/sh85091984 |

Web sémantique

| predicat | |
|---|---|
| http://www.w3.org/1999/02/22-rdf-syntax-ns#type | http://isidore.science/ontology#article |
| http://www.w3.org/1999/02/22-rdf-syntax-ns#type | http://www.openarchives.org/ore/terms/Aggregation |
| http://www.w3.org/1999/02/22-rdf-syntax-ns#type | http://isidore.science/class/publications |
| http://www.w3.org/1999/02/22-rdf-syntax-ns#type | http://isidore.science/class/Document |
| http://purl.org/dc/elements/1.1/description | "This article aims to describe technical issues of DEF19 project |
| http://purl.org/dc/elements/1.1/description | "Cet article vise à présenter les enjeux techniques du projet DEI |
| http://purl.org/dc/elements/1.1/publisher | nodeID://b5242301 |
| http://purl.org/dc/elements/1.1/publisher | nodeID://b5242302 |
| http://purl.org/dc/elements/1.1/subject | "Web"@fr |
| http://purl.org/dc/elements/1.1/subject | "base de données"@fr |
| http://purl.org/dc/elements/1.1/subject | "dictionary"@en |
| http://purl.org/dc/elements/1.1/subject | "dictionnaire"@fr |
| http://purl.org/dc/elements/1.1/subject | "histoire"@fr |
| http://purl.org/dc/elements/1.1/subject | "history"@en |
| http://purl.org/dc/elements/1.1/subject | "Web"@en |
| http://purl.org/dc/elements/1.1/subject | "database"@en |
| http://purl.org/dc/elements/1.1/subject | "histoire du livre"@fr |
| http://purl.org/dc/elements/1.1/subject | " |

Web sémantique

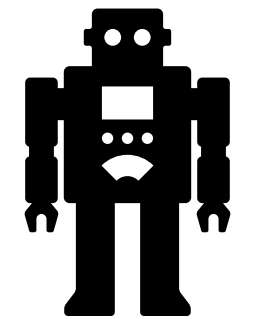
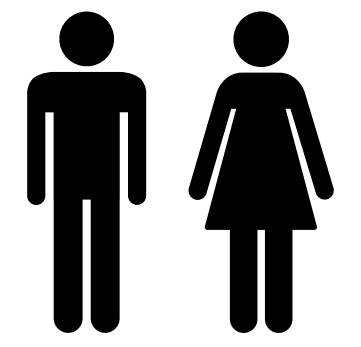
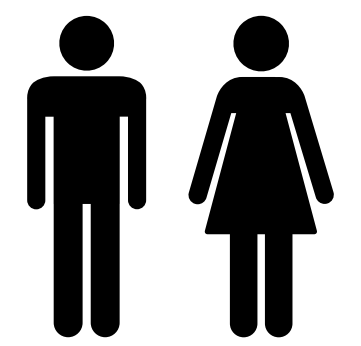
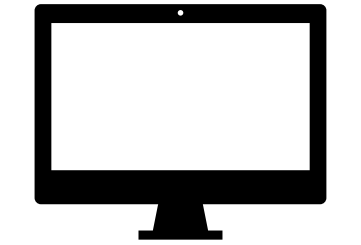
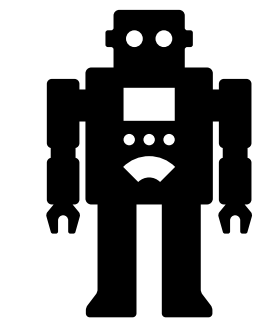
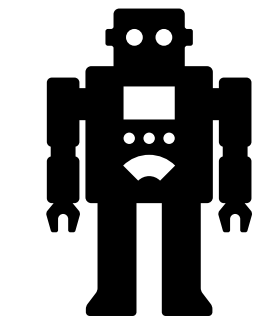
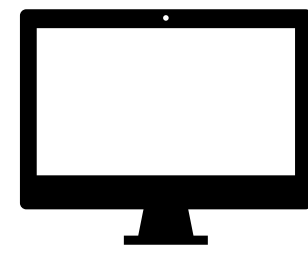
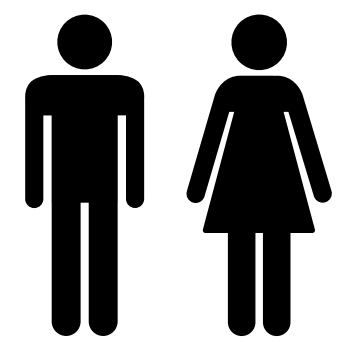
```
@prefix res: <http://www.w3.org/2005/sparql-results#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
_:i a res:ResultSet .
  _:i res:resultVariable "predicat" , "objet" .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix ns1: <http://isidore.science/ontology#> .
_:i res:solution [
  res:binding [ res:variable "predicat" ; res:value rdf:type ] ;
  res:binding [ res:variable "objet" ; res:value ns1:article ] ] .
@prefix ore: <http://www.openarchives.org/ore/terms/> .
_:i res:solution [
  res:binding [ res:variable "predicat" ; res:value rdf:type ] ;
  res:binding [ res:variable "objet" ; res:value ore:Aggregation ] ] .
@prefix isidore: <http://isidore.science/class/> .
_:i res:solution [
  res:binding [ res:variable "predicat" ; res:value rdf:type ] ;
  res:binding [ res:variable "objet" ; res:value isidore:Document ] ] .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
_:i res:solution [
  res:binding [ res:variable "predicat" ; res:value dc:description ] ;
  res:binding [ res:variable "objet" ; res:value "This article aims to describe project (Dictionnaire des \u00E9diteurs fran\u00E7ais du xixe si\u00E8cle, i.e. Publishers Dictionary). Scientific choices have been made (what is a publisher and century France?), which have widely affected technical answers. Several software solutions to build the database, each of them posing specific problems (and resulting in different solutions), particularly in integrating data coming from various preexisting databases. Our final choice, the Omeka platform, offers many development and integration possibilities, together with cooperations allowed by this tool and gradually unveiled since the beginning of the project." ] ] .
_:i res:solution [
  res:binding [ res:variable "predicat" ; res:value dc:description ] ;
  res:binding [ res:variable "objet" ; res:value "Cet article vise \u00E0 pr\u00E9senter les enjeux techniques du projet DEF19 (Dictionnaire des \u00E9diteurs fran\u00E7ais du xixe si\u00E8cle, i.e. comment les choix scientifiques pour d\u00E9finir le projet ont pos\u00E9 des probl\u00E8mes techniques divers, notamment en ce qui concerne l'int\u00E9gration de donn\u00E9es provenant de bases de donn\u00E9es existantes, et les solutions techniques mises en \u00E9vidence pour mettre au point la base de donn\u00E9es du projet, ainsi que les probl\u00E8mes rencontr\u00E9s lors de la r\u00E9alisation du projet en termes de collaborations techniques et institutionnelles." ] ] .
_:i res:solution [
  res:binding [ res:variable "predicat" ; res:value dc:publisher ] ;
  res:binding [ res:variable "objet" ; res:value "_:b5242301" ] ] .
_:i res:solution [
  res:binding [ res:variable "predicat" ; res:value dc:publisher ] ;
  res:binding [ res:variable "objet" ; res:value "_:b5242302" ] ] .
_:i res:solution [
  res:binding [ res:variable "predicat" ; res:value dc:subject ] ;
  res:binding [ res:variable "objet" ; res:value "Web"@fr ] ] .
_:i res:solution [
  res:binding [ res:variable "predicat" ; res:value dc:subject ] ;
  res:binding [ res:variable "objet" ; res:value "base de donn\u00E9es"@fr ] ] .
_:i res:solution [
  res:binding [ res:variable "predicat" ; res:value dc:subject ] ;
  res:binding [ res:variable "objet" ; res:value "dictionary"@en ] ] .
_:i res:solution [
  res:binding [ res:variable "predicat" ; res:value dc:subject ] ;
  res:binding [ res:variable "objet" ; res:value "dictionnaire"@fr ] ] .
_:i res:solution [
  res:binding [ res:variable "predicat" ; res:value dc:subject ] ;
  res:binding [ res:variable "objet" ; res:value "histoire"@fr ] ] .
_:i res:solution [
  res:binding [ res:variable "predicat" ; res:value dc:subject ] ;
  res:binding [ res:variable "objet" ; res:value "history"@en ] ] .
_:i res:solution [
  res:binding [ res:variable "predicat" ; res:value dc:subject ] ;
  res:binding [ res:variable "objet" ; res:value "Web"@en ] ] .
_:i res:solution [
  res:binding [ res:variable "predicat" ; res:value dc:subject ] ;
  res:binding [ res:variable "objet" ; res:value "database"@en ] ] .
_:i res:solution [
  res:binding [ res:variable "predicat" ; res:value dc:subject ] ;
  res:binding [ res:variable "objet" ; res:value "histoire du livre"@fr ] ] .
_:i res:solution [
  res:binding [ res:variable "predicat" ; res:value dc:subject ] ;
```

```
_:ResultSet2053 rdf:type <http://www.w3.org/1999/02/22-rdf-syntax-ns#res:ResultSet> .
_:ResultSet2053 <http://www.w3.org/2005/sparql-results#resultVariable> "predicat" .
_:ResultSet2053 <http://www.w3.org/2005/sparql-results#resultVariable> "objet" .
_:ResultSet2053 <http://www.w3.org/2005/sparql-results#solution> _:ResultSet2053r0 .
_:ResultSet2053r0 <http://www.w3.org/2005/sparql-results#binding> _:ResultSet2053r0c0 .
_:ResultSet2053r0c0 <http://www.w3.org/2005/sparql-results#variable> "predicat" .
_:ResultSet2053r0c0 <http://www.w3.org/2005/sparql-results#value> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> .
_:ResultSet2053r1 <http://www.w3.org/2005/sparql-results#binding> _:ResultSet2053r1c0 .
_:ResultSet2053r1c0 <http://www.w3.org/2005/sparql-results#variable> "objet" .
_:ResultSet2053r1c0 <http://www.w3.org/2005/sparql-results#value> <http://isidore.science/ontology#article> .
_:ResultSet2053r2 <http://www.w3.org/2005/sparql-results#binding> _:ResultSet2053r2c0 .
_:ResultSet2053r2c0 <http://www.w3.org/2005/sparql-results#variable> "predicat" .
_:ResultSet2053r2c0 <http://www.w3.org/2005/sparql-results#value> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> .
_:ResultSet2053r3 <http://www.w3.org/2005/sparql-results#binding> _:ResultSet2053r3c0 .
_:ResultSet2053r3c0 <http://www.w3.org/2005/sparql-results#variable> "objet" .
_:ResultSet2053r3c0 <http://www.w3.org/2005/sparql-results#value> <http://www.openarchives.org/ore/terms/Aggregation> .
_:ResultSet2053r4 <http://www.w3.org/2005/sparql-results#binding> _:ResultSet2053r4c0 .
_:ResultSet2053r4c0 <http://www.w3.org/2005/sparql-results#variable> "predicat" .
_:ResultSet2053r4c0 <http://www.w3.org/2005/sparql-results#value> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> .
_:ResultSet2053r5 <http://www.w3.org/2005/sparql-results#binding> _:ResultSet2053r5c0 .
_:ResultSet2053r5c0 <http://www.w3.org/2005/sparql-results#variable> "objet" .
_:ResultSet2053r5c0 <http://www.w3.org/2005/sparql-results#value> <http://purl.org/dc/elements/1.1/description> .
_:ResultSet2053r6 <http://www.w3.org/2005/sparql-results#binding> _:ResultSet2053r6c0 .
_:ResultSet2053r6c0 <http://www.w3.org/2005/sparql-results#variable> "predicat" .
_:ResultSet2053r6c0 <http://www.w3.org/2005/sparql-results#value> <http://purl.org/dc/elements/1.1/publisher> .
_:ResultSet2053r7 <http://www.w3.org/2005/sparql-results#binding> _:ResultSet2053r7c0 .
_:ResultSet2053r7c0 <http://www.w3.org/2005/sparql-results#variable> "objet" .
_:ResultSet2053r7c0 <http://www.w3.org/2005/sparql-results#value> _:b5242301 .
```

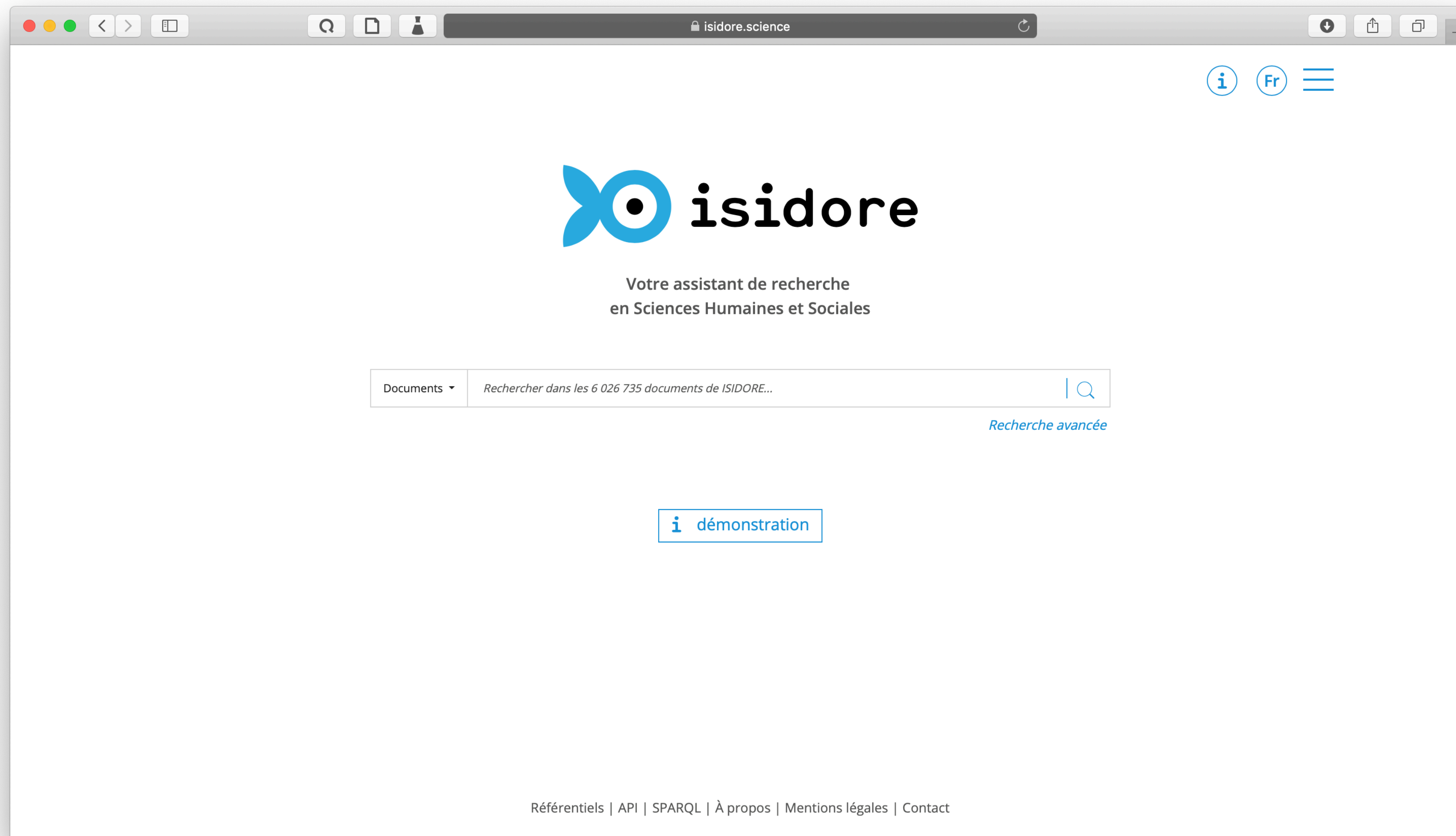
| | A | B | C |
|---|---|---|---|
| 1 | predicat | objet | |
| 2 | http://www.w3.org/1999/02/22-rdf-syntax-ns#type | http://isidore.science/ontology#article | |
| 3 | http://www.w3.org/1999/02/22-rdf-syntax-ns#type | http://www.openarchives.org/ore/terms/Aggregation | |
| 4 | http://www.w3.org/1999/02/22-rdf-syntax-ns#type | http://isidore.science/class/publications | |

```
sparql.xml
1 <rdf:RDF xmlns:res="http://www.w3.org/2005/sparql-results#" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
2 <rdf:Description rdf:nodeID="rset">
3 <rdf:type rdf:resource="http://www.w3.org/2005/sparql-results#ResultSet" />
4 <res:resultVariable>predicat</res:resultVariable>
5 <res:resultVariable>objet</res:resultVariable>
6 <res:solution rdf:nodeID="r0">
7 <res:binding rdf:nodeID="r0c0"><res:variable>predicat</res:variable><res:value rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#type"/></res:binding>
8 <res:binding rdf:nodeID="r0c1"><res:variable>objet</res:variable><res:value rdf:resource="http://isidore.science/ontology#article"/></res:binding>
9 </res:solution>
10 <res:solution rdf:nodeID="r1">
11 <res:binding rdf:nodeID="r1c0"><res:variable>predicat</res:variable><res:value rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#type"/></res:binding>
12 <res:binding rdf:nodeID="r1c1"><res:variable>objet</res:variable><res:value rdf:resource="http://www.openarchives.org/ore/terms/Aggregation"/></res:binding>
13 </res:solution>
14 <res:solution rdf:nodeID="r2">
15 <res:binding rdf:nodeID="r2c0"><res:variable>predicat</res:variable><res:value rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#type"/></res:binding>
16 <res:binding rdf:nodeID="r2c1"><res:variable>objet</res:variable><res:value rdf:resource="http://isidore.science/class/publications"/></res:binding>
17 </res:solution>
18 <res:solution rdf:nodeID="r3">
19 <res:binding rdf:nodeID="r3c0"><res:variable>predicat</res:variable><res:value rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#type"/></res:binding>
20 <res:binding rdf:nodeID="r3c1"><res:variable>objet</res:variable><res:value rdf:resource="http://isidore.science/class/Document"/></res:binding>
21 </res:solution>
22 <res:solution rdf:nodeID="r4">
23 <res:binding rdf:nodeID="r4c0"><res:variable>predicat</res:variable><res:value rdf:resource="http://purl.org/dc/elements/1.1/description"/></res:binding>
24 <res:binding rdf:nodeID="r4c1"><res:variable>objet</res:variable><res:value xml:lang="en">This article aims to describe technical issues of DEF19 project (Dictionnaire des \u00E9diteurs fran\u00E7ais du xixe si\u00E8cle, i.e. Publishers Dictionary). Scientific choices have been made (what is a publisher and century France?), which have widely affected technical answers. Several software solutions to build the database, each of them posing specific problems (and resulting in different solutions), particularly in integrating data coming from various preexisting databases. Our final choice, the Omeka platform, offers many development and integration possibilities, together with cooperations allowed by this tool and gradually unveiled since the beginning of the project.</res:binding>
25 </res:solution>
26 <res:solution rdf:nodeID="r5">
27 <res:binding rdf:nodeID="r5c0"><res:variable>predicat</res:variable><res:value rdf:resource="http://purl.org/dc/elements/1.1/description"/></res:binding>
28 <res:binding rdf:nodeID="r5c1"><res:variable>objet</res:variable><res:value xml:lang="fr">Cet article vise \u00E0 pr\u00E9senter les enjeux techniques du projet DEF19 (Dictionnaire des \u00E9diteurs fran\u00E7ais du xixe si\u00E8cle, i.e. comment les choix scientifiques pour d\u00E9finir le projet ont pos\u00E9 des probl\u00E8mes techniques divers, notamment en ce qui concerne l'int\u00E9gration de donn\u00E9es provenant de bases de donn\u00E9es existantes, et les solutions techniques mises en \u00E9vidence pour mettre au point la base de donn\u00E9es du projet, ainsi que les probl\u00E8mes rencontr\u00E9s lors de la r\u00E9alisation du projet en termes de collaborations techniques et institutionnelles.</res:binding>
29 </res:solution>
30 <res:solution rdf:nodeID="r6">
31 <res:binding rdf:nodeID="r6c0"><res:variable>predicat</res:variable><res:value rdf:resource="http://purl.org/dc/elements/1.1/publisher"/></res:binding>
32 <res:binding rdf:nodeID="r6c1"><res:variable>objet</res:variable><res:value rdf:nodeID="bb5242301"/></res:binding>
33 </res:solution>
```

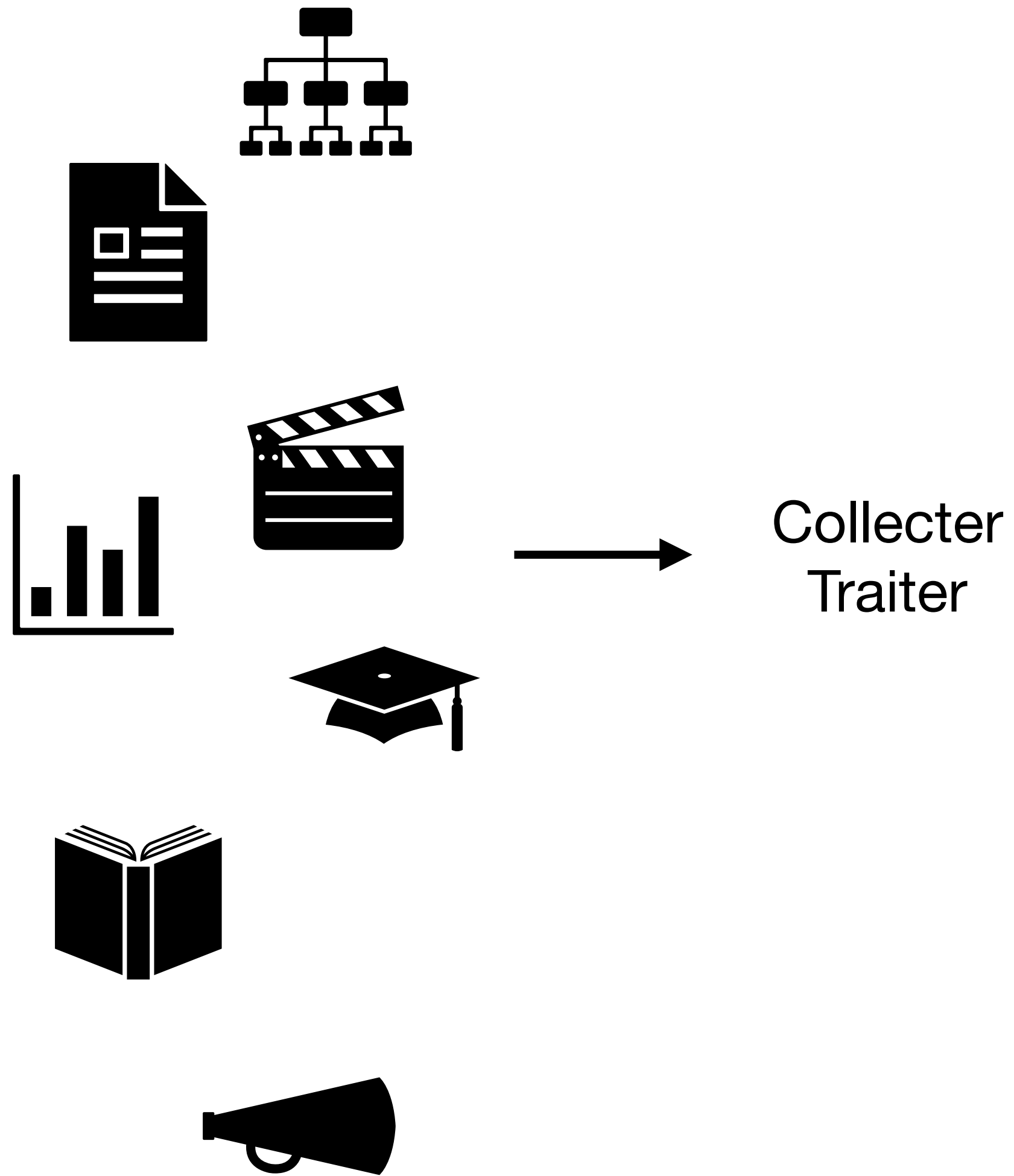
Web sémantique



Mise en œuvre des principes du Web sémantique



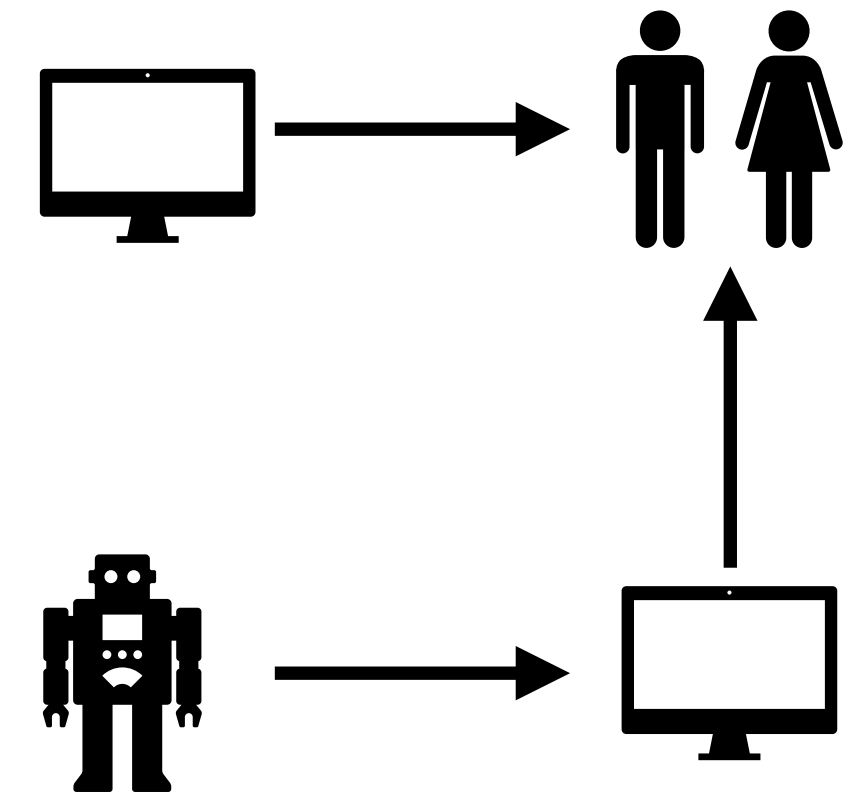
Mise en œuvre des principes du Web sémantique



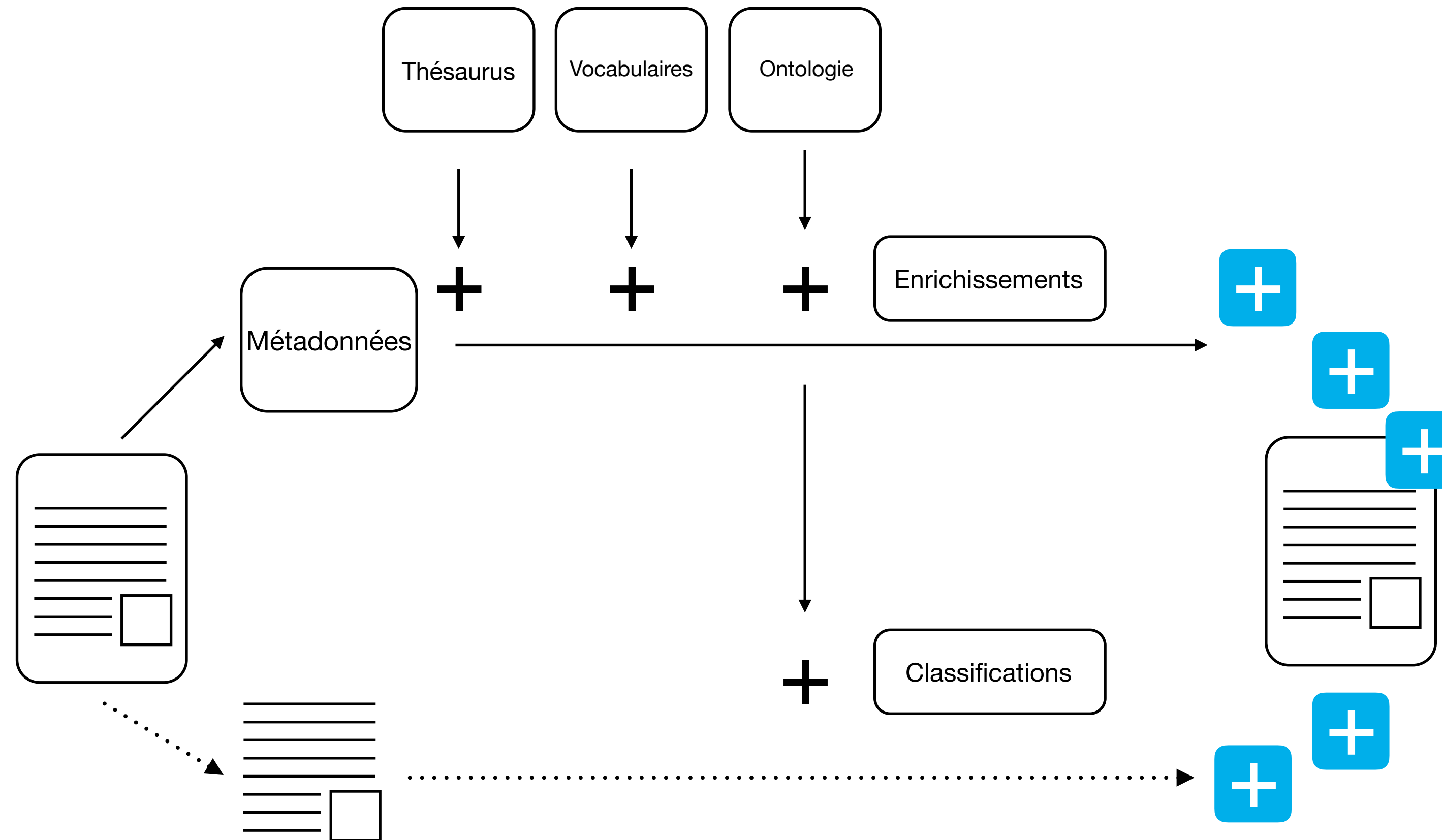
Enrichir
Classer
Catégoriser



Permettre l'accès



Mise en œuvre des principes du Web sémantique



Mise en œuvre des principes du Web sémantique

Enrichir, classer, catégoriser



Mise en œuvre des principes du Web sémantique

The screenshot shows the isidore.science website interface. The search bar contains 'Sujets' and 'Fr'. The main content area displays search results for the subject 'Web'. On the left, there is a sidebar with 'Fiche du sujet' (Subject card) containing synonyms like 'WWW', 'La Toile (Internet)', 'W3', and 'World Wide Web'. The main list shows 11629 documents, including 'Be-Patient : une innovation de rupture centrée sur le patient' by Myriam Le Goff-Pronost et al. (2020), '[S1] Les notions' by Benoît Epron (2020), and 'Perceptions des membres de l'équipe interprofessionnelle face à la présence de la famille lors de la réanimation d'un proche' by Quentin Bevilard-Charrière (2020).

The screenshot shows a browser window displaying a table of RDF triples. The table has two columns: 'predicat' and 'object'. The first row shows a triple with the predicate 'http://www.w3.org/1999/02/22-rdf-syntax-ns#type' and the object 'http://www.w3.org/2004/02/skos/core#Concept'. Other rows include dates like '1995-09-30' and '2001-06-03', and various URIs from BNF and other sources.

| predicat | object |
|---|---|
| http://www.w3.org/1999/02/22-rdf-syntax-ns#type | http://www.w3.org/2004/02/skos/core#Concept |
| http://purl.org/dc/terms/created | "1995-09-30" |
| http://purl.org/dc/terms/modified | "2001-06-03" |
| http://www.w3.org/2000/01/rdf-schema#seeAlso | http://catalogue.bnf.fr/ark:/12148/cb13319953j |
| http://www.w3.org/2002/07/owl#sameAs | http://stitch.cs.vu.nl/vocabularies/rameau/ark:/12148/cb13319953j |
| http://www.w3.org/2004/02/skos/core#note | "Outil de navigation hypermédia sur Internet"@fr |
| http://www.w3.org/2004/02/skos/core#broader | http://data.bnf.fr/ark:/12148/cb12337059x |
| http://www.w3.org/2004/02/skos/core#broader | http://data.bnf.fr/ark:/12148/cb135052753 |
| http://www.w3.org/2004/02/skos/core#altLabel | "La Toile (Internet)"@fr |
| http://www.w3.org/2004/02/skos/core#altLabel | "W3"@fr |
| http://www.w3.org/2004/02/skos/core#altLabel | "WWW"@fr |
| http://www.w3.org/2004/02/skos/core#altLabel | "World Wide Web"@fr |
| http://www.w3.org/2004/02/skos/core#inScheme | http://rameau.bnf.fr/ |
| http://www.w3.org/2004/02/skos/core#narrower | http://data.bnf.fr/ark:/12148/cb13556047k |
| http://www.w3.org/2004/02/skos/core#narrower | http://data.bnf.fr/ark:/12148/cb124996255 |
| http://www.w3.org/2004/02/skos/core#narrower | http://data.bnf.fr/ark:/12148/cb12545061j |
| http://www.w3.org/2004/02/skos/core#narrower | http://data.bnf.fr/ark:/12148/cb12552998v |
| http://www.w3.org/2004/02/skos/core#narrower | http://data.bnf.fr/ark:/12148/cb125428495 |
| http://www.w3.org/2004/02/skos/core#narrower | http://data.bnf.fr/ark:/12148/cb131690783 |
| http://www.w3.org/2004/02/skos/core#narrower | http://data.bnf.fr/ark:/12148/cb12492025t |
| http://www.w3.org/2004/02/skos/core#narrower | http://data.bnf.fr/ark:/12148/cb13191246q |
| http://www.w3.org/2004/02/skos/core#narrower | http://data.bnf.fr/ark:/12148/cb13191909f |
| http://www.w3.org/2004/02/skos/core#narrower | http://data.bnf.fr/ark:/12148/cb125450815 |
| http://www.w3.org/2004/02/skos/core#narrower | http://data.bnf.fr/ark:/12148/cb125614466 |
| http://www.w3.org/2004/02/skos/core#narrower | http://data.bnf.fr/ark:/12148/cb13182541v |
| http://www.w3.org/2004/02/skos/core#narrower | http://data.bnf.fr/ark:/12148/cb131716541 |
| http://www.w3.org/2004/02/skos/core#narrower | http://data.bnf.fr/ark:/12148/cb144109034 |
| http://www.w3.org/2004/02/skos/core#narrower | http://data.bnf.fr/ark:/12148/cb14521343b |
| http://www.w3.org/2004/02/skos/core#narrower | http://data.bnf.fr/ark:/12148/cb146511814 |
| http://www.w3.org/2004/02/skos/core#narrower | http://data.bnf.fr/ark:/12148/cb15592351s |
| http://www.w3.org/2004/02/skos/core#prefLabel | "Web"@fr |
| http://www.w3.org/2004/02/skos/core#related | http://data.bnf.fr/ark:/12148/cb133199588 |
| http://www.w3.org/2004/02/skos/core#closeMatch | http://datos.bne.es/resource/XX4576804 |
| http://www.w3.org/2004/02/skos/core#closeMatch | http://id.loc.gov/authorities/subjects/sh95000541 |
| http://www.w3.org/2004/02/skos/core#closeMatch | http://dewey.info/class/621/ |
| http://purl.org/dc/terms/isPartOf | http://data.bnf.fr/vocabulary/scheme/r166 |
| http://data.bnf.fr/ontology/bnf-onto/FRBNF | 13319953 |

<http://data.bnf.fr/ark:/12148/cb13319953j>

Mise en œuvre des principes du Web sémantique

Virtuoso SPARQL Query Editor

Default Data Set Name (Graph IRI)

Query Text

```
SELECT *  
WHERE {  
<http://data.bnf.fr/ark:/12148/cb13319953j> ?predicat ?object  
} LIMIT 100
```

(Security restrictions of this server do not allow you to retrieve remote RDF data, see [details](#).)

Results Format: HTML (The CXML output is disabled, see [details](#))

Execution timeout: 0 milliseconds (values less than 1000 are ignored)

Options: Strict checking of void variables
 Log debug info at the end of output (has no effect on some queries and output formats)

(The result can only be sent back to browser, not saved on the server, see [details](#))

Run Query Reset

Copyright © 2020 OpenLink Software
Virtuoso version 07.20.3217 on Linux (x64_64-generic-linux-glibc2.17-64), Single Server Edition

BnF Data

Télécharger Les Données | Le Modèle De Données | Contact

Exemple de requêtes

Espace de nom (namespace)

Query

https://data.bnf.fr/sparql

```
1 SELECT *  
2 WHERE {  
3 <http://data.bnf.fr/ark:/12148/cb13319953j> ?predicat ?object  
4 } LIMIT 100
```

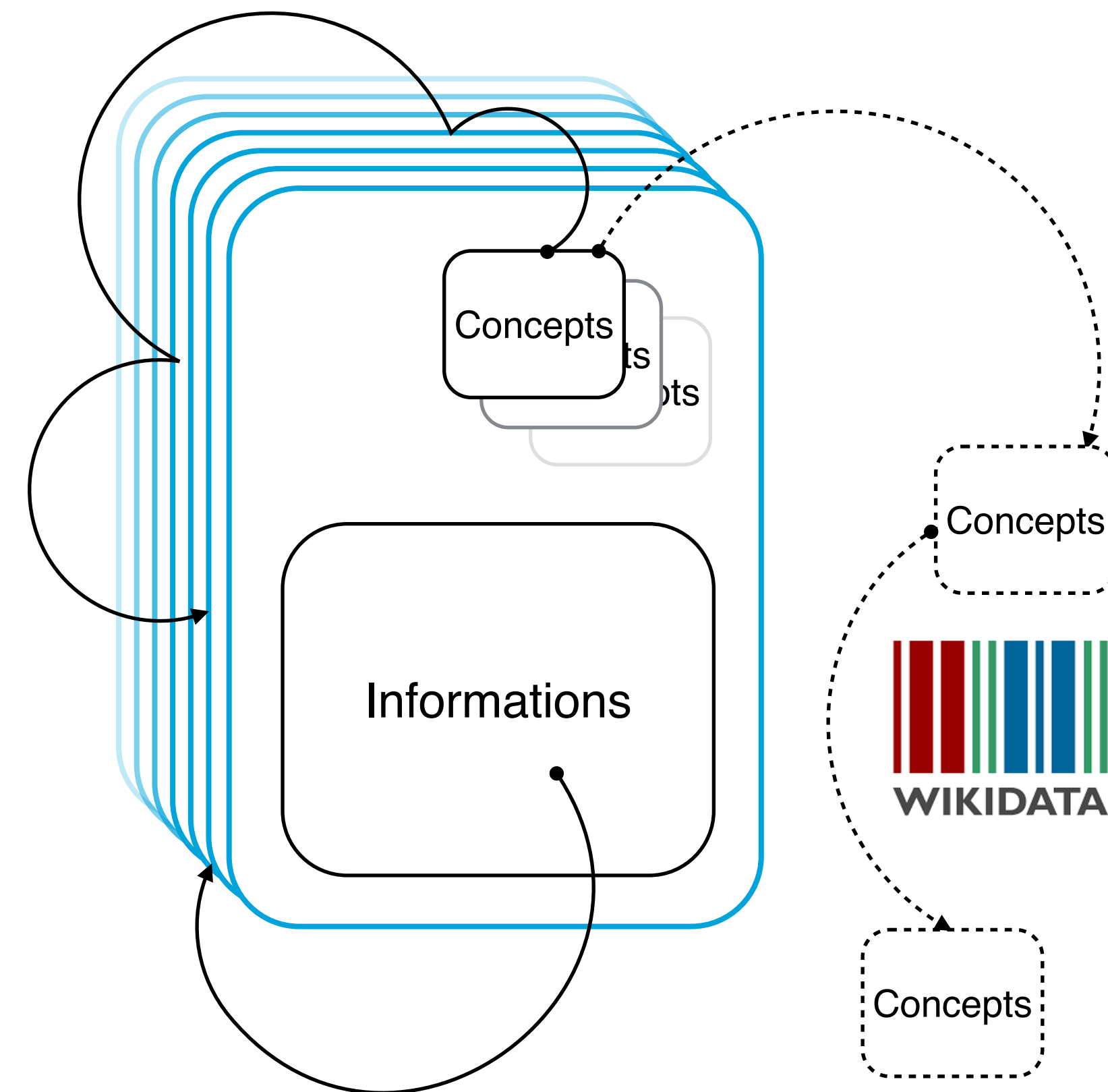
Table Response 41 results in 0.153 seconds Filter query results Page size: 50

| | predicat | object |
|----|---|---|
| 1 | <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> | <http://www.w3.org/2004/02/skos/core#Concept> |
| 2 | <http://purl.org/dc/terms/created> | 1995-09-30 |
| 3 | <http://purl.org/dc/terms/modified> | 2018-12-07 |
| 4 | <http://www.w3.org/2000/01/rdf-schema#seeAlso> | <https://catalogue.bnf.fr/ark:/12148/cb13319953j> |
| 5 | <http://www.w3.org/2002/07/owl#sameAs> | <http://datos.bne.es/resource/XX4576804> |
| 6 | <http://www.w3.org/2002/07/owl#sameAs> | <http://stitch.cs.vu.nl/vocabularies/rameau/ark:/12148/cb13319953j> |
| 7 | <http://www.w3.org/2004/02/skos/core#altLabel> | "W3"@fr |
| 8 | <http://www.w3.org/2004/02/skos/core#altLabel> | "WWW"@fr |
| 9 | <http://www.w3.org/2004/02/skos/core#altLabel> | "La Toile (Internet)"@fr |
| 10 | <http://www.w3.org/2004/02/skos/core#altLabel> | "W3"@fr |
| 11 | <http://www.w3.org/2004/02/skos/core#altLabel> | "WWW"@fr |
| 12 | <http://www.w3.org/2004/02/skos/core#altLabel> | "La Toile (Internet)"@fr |
| 13 | <http://xmlns.com/foaf/0.1/page> | <http://data.bnf.fr/13319953/web/> |
| 14 | <http://data.bnf.fr/ontology/bnf-onto/FRBNF> | "13319953"*<http://www.w3.org/2001/XMLSchema#integer> |

SELECT * WHERE { <http://data.bnf.fr/ark:/12148/cb13319953j> ?predicat ?object }

Mise en œuvre des principes du Web sémantique

Relier les informations avec des concepts scientifiques, donc les documents et/ou les auteurs entre eux.



Mise en œuvre des principes du Web sémantique

The screenshot shows the profile page for Stéphane Pouyllau on the isidore.science platform. The page is divided into several sections:

- Documents:** A list of documents with filters for 'Rapports', 'Articles', and 'Prépublication'. Examples include 'We: a Proposal for the TRIPLE platform' and 'A classifier using ISIDORE, the social and humanities search engine and Keras API for Deep Learning'.
- Auteurs suivis (4):** A list of authors followed, including Robert Vergnieux, Delphine Battistelli, Jean-Luc Minel, and Antoine Courtin.
- Bibliothèques partagées (6):** A list of shared libraries, such as 'À propos d'ISIDORE' (8 documents) and 'Outils SHS - Jupyter' (8 documents).
- Requêtes partagées (1):** A list of shared queries, including 'DL - Keras' with 84 results.

The profile also includes a bio: 'Ingénieur de recherche au CNRS (Centre National de la Recherche Scientifique), Stéphane Pouyllau est spécialisé depuis 1999 en humanités numériques (digital humanities), en information scientifique et technique et en informatisation des données de la recherche en sciences humaines et sociales. Il a co-créé en 2005 et développé le centre national pour la numérisation de sources visuelle jusqu'en 2009.'

The screenshot shows the 'Identifiant' page for Stéphane Pouyllau, providing various identifiers and disciplinary information:

- Identifiant:** IdRef : 197349692, ORCID : 0000-0002-9619-1002, ISNI : 0000000416375245, IdHAL : stephanepouyllau.
- Discipline:** Archéologie et Préhistoire (46) - Sciences de l'information et de la communication (43) - Histoire (11) - Géographie (10) - Histoire, Philosophie et Sociologie des sciences (2) - Art et histoire de l'art (2).
- Type de document:** Autres (69) - Photos et images (61) - Prépublication (15) - Articles (12) - Colloques et conférences (10) - Rapports (5) - Livres et chapitres d'ouvrages (3) - Recensions (2) - Textes imprimés (1) - Documents audiovisuels (1) - Matériels pédagogiques (1) - Fonds d'archives (1).
- Co-auteur:** Stetan Sinclair (2), Michael E. Sinatra (2), Laurent Schmitt (2), Paula Ricaurte (2), Mike Priddy (2), Ernesto Priani Saiso (2), Céline Poudat (2).

The screenshot shows a GitLab notebook titled 'Computing co-authors and co-publications from an author'. It contains a SPARQL query and its results, including a pie chart visualization.

```
In [4]:
from SPARQLWrapper import SPARQLWrapper, JSON
import pandas as pd

sparql = SPARQLWrapper("https://isidore.science/sparql")
sparql.setQuery("""
SELECT ?coauteurs (count(?documents) as ?count) WHERE {
?documents dct:creator <http://isidore.science/a/mounier_pierre>.
?documents dct:creator ?coauteurs
FILTER(?coauteurs != <http://isidore.science/a/mounier_pierre>)
} GROUP BY ?coauteurs
ORDER BY DESC(?count) LIMIT 10
""")
sparql.setReturnFormat(JSON)
results = sparql.query().convert()
results_df = pd.pandas.json_normalize(results['results']['bindings'])
results_df[['coauteurs.value', 'count.value']].head(10)
```

| | coauteurs.value | count.value |
|---|---|-------------|
| 0 | http://isidore.science/a/dacos_marin | 20 |
| 1 | http://isidore.science/a/boutier_jean | 4 |
| 2 | http://isidore.science/a/fabiani_jean_louis | 4 |
| 3 | http://isidore.science/a/wieviorka_michel | 4 |
| 4 | http://isidore.science/a/prochasson_christophe | 4 |
| 5 | http://isidore.science/a/perrot_martyne | 4 |
| 6 | http://isidore.science/a/ozouf_marignier_marie... | 3 |
| 7 | http://isidore.science/a/popovic_alexandre | 3 |
| 8 | http://isidore.science/a/grenier_jean_yves | 3 |
| 9 | http://isidore.science/a/aymard_maurice | 3 |

```
In [5]:
labels = results_df['coauteurs.value']
data = results_df['count.value']
plt.pie(data, labels=labels, explode=None, autopct='%1.2f%%', shadow=False, radius=0.9)
plt.axis('equal')
plt.show()
plt.close()
```

Out [5]:

The pie chart visualizes the distribution of co-authors based on the query results. The largest share is for 'http://isidore.science/a/dacos_marin' at 38.46%.

| coauteurs.value | count.value | Percentage |
|---|-------------|------------|
| http://isidore.science/a/dacos_marin | 20 | 38.46% |
| http://isidore.science/a/boutier_jean | 4 | 7.69% |
| http://isidore.science/a/fabiani_jean_louis | 4 | 7.69% |
| http://isidore.science/a/wieviorka_michel | 4 | 7.69% |
| http://isidore.science/a/prochasson_christophe | 4 | 7.69% |
| http://isidore.science/a/perrot_martyne | 4 | 7.69% |
| http://isidore.science/a/ozouf_marignier_marie... | 3 | 5.77% |
| http://isidore.science/a/popovic_alexandre | 3 | 5.77% |
| http://isidore.science/a/grenier_jean_yves | 3 | 5.77% |
| http://isidore.science/a/aymard_maurice | 3 | 5.77% |

Champs des possibles pour l'exploration des données

The screenshot shows a JupyterLab interface with a notebook titled "Computing co-authors and co-publications from an author". The notebook contains two cells of code and their respective outputs.

In [4]:

```
from SPARQLWrapper import SPARQLWrapper, JSON
import pandas as pd

sparql = SPARQLWrapper("https://isidore.science/sparql")
sparql.setQuery("""
SELECT ?coauteurs (count(?documents) as ?count) WHERE {
?documents dct:creator <http://isidore.science/a/mounier_pierre>.
?documents dct:creator ?coauteurs
FILTER(?coauteurs != <http://isidore.science/a/mounier_pierre>)
} GROUP BY ?coauteurs
ORDER BY DESC(?count) LIMIT 10
""")
sparql.setReturnFormat(JSON)
results = sparql.query().convert()
results_df = pd.pandas.json_normalize(results['results']['bindings'])
results_df[['coauteurs.value', 'count.value']].head(10)
```

Out [4]:

| | coauteurs.value | count.value |
|---|---|-------------|
| 0 | http://isidore.science/a/dacos_marin | 20 |
| 1 | http://isidore.science/a/boutier_jean | 4 |
| 2 | http://isidore.science/a/fabiani_jean_louis | 4 |
| 3 | http://isidore.science/a/wieviorka_michel | 4 |
| 4 | http://isidore.science/a/prochasson_christophe | 4 |
| 5 | http://isidore.science/a/perrot_martyne | 4 |
| 6 | http://isidore.science/a/ozouf_marignier_marie... | 3 |
| 7 | http://isidore.science/a/popovic_alexandre | 3 |
| 8 | http://isidore.science/a/grenier_jean_yves | 3 |
| 9 | http://isidore.science/a/aymard_maurice | 3 |

In [5]:

```
labels = results_df['coauteurs.value']
data = results_df['count.value']
plt.pie(data, labels=labels, explode=None, autopct='%1.2f%%', shadow=False, radius=0.9)
plt.axis('equal')
plt.show()
plt.close()
```

out [5]:

| coauteurs.value | count.value | Percentage |
|---|-------------|------------|
| http://isidore.science/a/dacos_marin | 20 | 38.46% |
| http://isidore.science/a/boutier_jean | 4 | 7.69% |
| http://isidore.science/a/fabiani_jean_louis | 4 | 7.69% |
| http://isidore.science/a/wieviorka_michel | 4 | 7.69% |
| http://isidore.science/a/prochasson_christophe | 4 | 7.69% |
| http://isidore.science/a/perrot_martyne | 4 | 7.69% |
| http://isidore.science/a/ozouf_marignier_marie... | 3 | 5.77% |
| http://isidore.science/a/popovic_alexandre | 3 | 5.77% |
| http://isidore.science/a/grenier_jean_yves | 3 | 5.77% |
| http://isidore.science/a/aymard_maurice | 3 | 5.77% |

Champs des possibles pour l'exploration des données

Virtuoso SPARQL Query Editor

Default Data Set Name (Graph IRI)

Query Text

```
SELECT *  
WHERE {  
<http://data.bnf.fr/ark:/12148/cb13319953j> ?predicat ?object  
} LIMIT 100
```

(Security restrictions of this server do not allow you to retrieve remote RDF data, see [details](#).)

Results Format: HTML (The CXML output is disabled, see [details](#))

Execution timeout: 0 milliseconds (values less than 1000 are ignored)

Options: Strict checking of void variables
 Log debug info at the end of output (has no effect on some queries and output formats)

(The result can only be sent back to browser, not saved on the server, see [details](#))

Run Query Reset

Copyright © 2020 OpenLink Software
Virtuoso version 07.20.3217 on Linux (x64_64-generic-linux-glibc2.17-64), Single Server Edition

SELECT * WHERE { <http://data.bnf.fr/ark:/12148/cb13319953j> ?predicat ?object }

BnF Data

Exemple de requêtes Espace de nom (namespace)

Query [+](#)

https://data.bnf.fr/sparql

```
1 SELECT *  
2 WHERE {  
3 <http://data.bnf.fr/ark:/12148/cb13319953j> ?predicat ?object  
4 } LIMIT 100
```

Table Response 41 results in 0.153 seconds Filter query results Page size: 50

| | predicat | object |
|----|---|---|
| 1 | <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> | <http://www.w3.org/2004/02/skos/core#Concept> |
| 2 | <http://purl.org/dc/terms/created> | 1995-09-30 |
| 3 | <http://purl.org/dc/terms/modified> | 2018-12-07 |
| 4 | <http://www.w3.org/2000/01/rdf-schema#seeAlso> | <https://catalogue.bnf.fr/ark:/12148/cb13319953j> |
| 5 | <http://www.w3.org/2002/07/owl#sameAs> | <http://datos.bne.es/resource/XX4576804> |
| 6 | <http://www.w3.org/2002/07/owl#sameAs> | <http://stitch.cs.vu.nl/vocabularies/rameau/ark:/12148/cb13319953j> |
| 7 | <http://www.w3.org/2004/02/skos/core#altLabel> | "W3"@fr |
| 8 | <http://www.w3.org/2004/02/skos/core#altLabel> | "WWW"@fr |
| 9 | <http://www.w3.org/2004/02/skos/core#altLabel> | "La Toile (Internet)"@fr |
| 10 | <http://www.w3.org/2004/02/skos/core#altLabel> | "World Wide Web"@fr |
| 11 | <http://www.w3.org/2004/02/skos/core#altLabel> | "World Wide Web"@it |
| 12 | <http://www.w3.org/2004/02/skos/core#prefLabel> | "Web"@fr |
| 13 | <http://xmlns.com/foaf/0.1/page> | <http://data.bnf.fr/13319953/web/> |
| 14 | <http://data.bnf.fr/ontology/bnf-onto/FRBNF> | "13319953"^^<http://www.w3.org/2001/XMLSchema#integer> |

Champs des possibles pour l'exploration des données

The screenshot shows a web application for visualizing SPARQL queries. The interface is split into two main sections: a visual query builder on the left and a code editor on the right.

Visual Query Builder:

- It features a flow-based interface with orange arrows and boxes representing query components.
- The first row shows a selection of "Musée" (Museum) with the property "expose" (exhibits) leading to "Oeuvre" (Work of Art).
- The second row shows "Oeuvre" with the property "auteur" (author) leading to "Personne" (Person).
- A search box for "Personne" contains the text "Le Caravage" and a dropdown menu shows "Le Caravage" as a suggestion.
- Other options like "Où" (Where) and "Et" (And) are visible.

Code Editor:

```
1 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
2 SELECT DISTINCT (STR(?label) AS ?nom) ?wikipedia ?this WHERE {
3   ?this <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
4     <http://dbpedia.org/ontology/Museum>;
5     ^<http://dbpedia.org/ontology/museum> ?Artwork1.
6   ?Artwork1 <http://dbpedia.org/ontology/author>
7     <http://fr.dbpedia.org/resource/Le_Caravage>.
8   ?this rdfs:label ?label FILTER(lang(?label) = 'fr')
9   ?this <http://xmlns.com/foaf/0.1/isPrimaryTopicOf> ?wikipedia
10 }
11 ORDER BY ?label LIMIT 5000
```

<https://github.com/sparna-git/Sparnatural>

Champs des possibles pour l'exploration des données

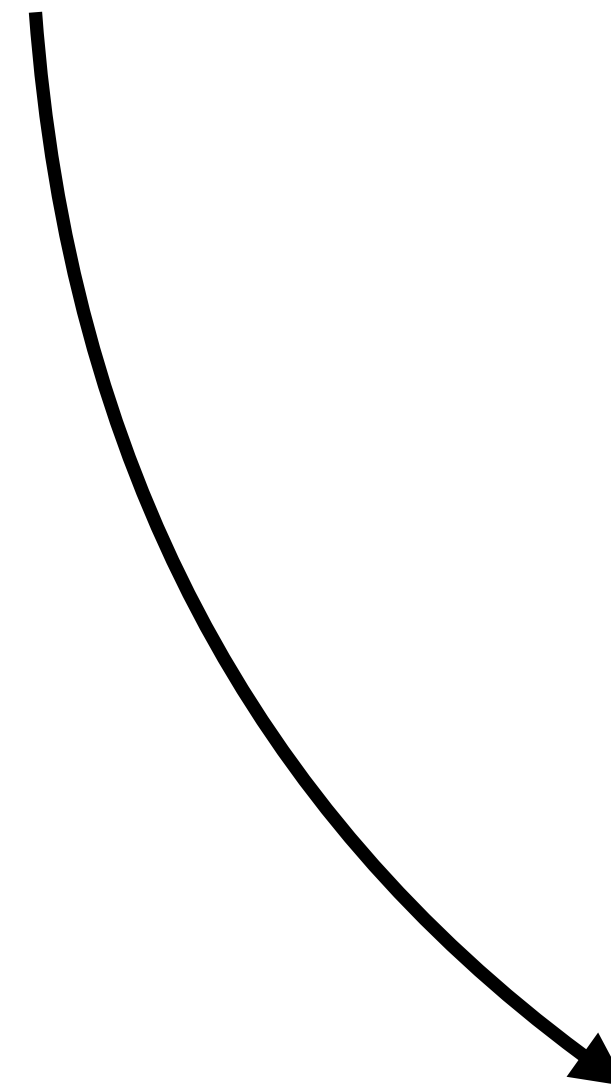
The screenshot shows a JupyterLab notebook titled "Computing co-authors and co-publications from an author". It contains a SPARQL query and a pie chart visualization of the results.

```
[4]: from SPARQLWrapper import SPARQLWrapper, JSON
import pandas as pd

sparql = SPARQLWrapper("https://isidore.science/sparql")
sparql.setQuery("""
SELECT ?coauteurs (count(?documents) as ?count) WHERE {
?documents dcterms:creator <http://isidore.science/a/mounier_pierre>.
?documents dcterms:creator ?coauteurs
FILTER(?coauteurs != <http://isidore.science/a/mounier_pierre>)
} GROUP BY ?coauteurs
ORDER BY DESC(?count) LIMIT 10
""")
sparql.setReturnFormat(JSON)
results = sparql.query().convert()
results_df = pd.pandas.json_normalize(results['results']['bindings'])
results_df[['coauteurs.value', 'count.value']].head(10)
```

| | coauteurs.value | count.value |
|---|---|-------------|
| 0 | http://isidore.science/a/dacos_marin | 20 |
| 1 | http://isidore.science/a/boutier_jean | 4 |
| 2 | http://isidore.science/a/fabiani_jean_louis | 4 |
| 3 | http://isidore.science/a/wieviorka_michel | 4 |
| 4 | http://isidore.science/a/prochasson_christophe | 4 |
| 5 | http://isidore.science/a/perrot_martyne | 4 |
| 6 | http://isidore.science/a/ozouf_marignier_marie... | 3 |
| 7 | http://isidore.science/a/popovic_alexandre | 3 |
| 8 | http://isidore.science/a/grenier_jean_yves | 3 |
| 9 | http://isidore.science/a/aymard_maurice | 3 |

```
[5]: labels = results_df['coauteurs.value']
data = results_df['count.value']
plt.pie(data, labels=labels, explode=None, autopct='%1.2f%%', shadow=False, radius=0.9)
plt.axis('equal')
plt.show()
plt.close()
```



<https://doi.org/10.5281/zenodo.4032622>

The screenshot shows the Zenodo record page for the paper "We: a Proposal for the TRIPLE platform". The page includes the title, authors (POUYLLAU, Stéphane; BUNEL, Mélanie; CAPELLI, Laurent; MINEL, Jean-Luc), a DOI (10.5281/zenodo.4032622), and a list of keywords (Web platform, Discovery platform, Search engine, Social sciences and Humanities, Digital Humanities, Information management, Deep Learning). The page also shows the publication date (September 16, 2020), the number of views (143) and downloads (64), and a list of versions (Version 1, 2, 3).

September 16, 2020

"We": a Proposal for the TRIPLE platform

POUYLLAU, Stéphane; BUNEL, Mélanie; CAPELLI, Laurent; MINEL, Jean-Luc

This research report is a proposal to define a platform, called We, for the TRIPLE project. First, the We platform is situated in the ecosystem of academic tools. Then its functionalities are described. This document proposes a conceptual representation of an innovative platform based on the discovery of experts, topics and projects using the analysis, classification, linking and enrichment of data. It is not a proposal of a Human-Machine Interface (HMI) which is under the responsibility of WP3 and WP5 as well as the editorialization of the components, but there are ideas and may constitute some areas of work to be discussed. Also, the different diagrams aim at illustrating the functionalities and not the design of the screens. These suggestions include elements of the proposal (experts), new public (Python library) and technological advances (Deep Learning) which were not so advanced when the proposal has been elaborated.

143 views, 64 downloads

Indexed in OpenAIRE

Publication date: September 16, 2020

DOI: 10.5281/zenodo.4032622

Keyword(s): Web platform, Discovery platform, Search engine, Social sciences and Humanities, Digital Humanities, Information management, Deep Learning

Subject(s): machine learning, search engine

Related identifiers: References, <https://doi.org/10.5281/zenodo.3965604> (Report)

Communities: Huma-Num Lab

License (for files): Creative Commons Attribution 4.0 International

Files (24.4 MB)

| Name | Size | Preview | Download |
|---------------------|---------|--------------------------|--------------------------|
| WeDesign-TRIPLE.pdf | 24.4 MB | <input type="checkbox"/> | <input type="checkbox"/> |

md5:0b34b62ecd979ae86149fd460f2e08d


Citations 0

Show only: Literature (0), Dataset (0), Software (0), Unknown (0), Citations to this version

Versions

| Version | Date |
|-----------|--------------|
| Version 3 | Sep 30, 2020 |
| Version 2 | Sep 16, 2020 |
| Version 1 | Sep 16, 2020 |

Champs des possibles pour l'e

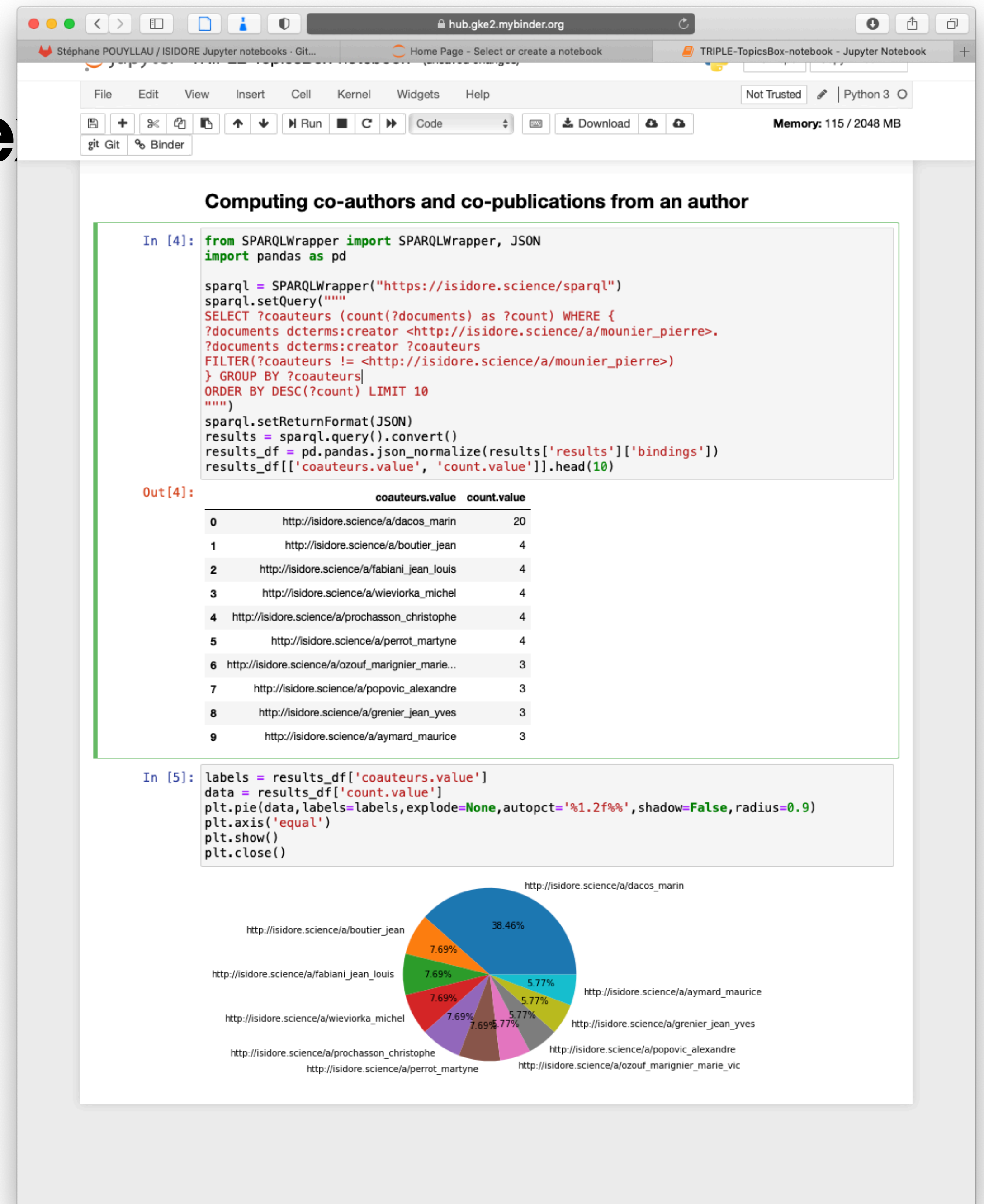


Starting repository: <https://gitlab.huma-num.fr/spouyllau/isidore-jupyter/master>

Take a look at our [gallery of example repositories](#).

```
Build logs
```

```
Downloading rdflib-5.0.0-py3-none-any.whl (231 kB)
Collecting SPARQLWrapper==1.8.5
  Downloading SPARQLWrapper-1.8.5-py3-none-any.whl (26 kB)
Requirement already satisfied: python-dateutil>=2.1 in /srv/conda/envs/notebook/lib/python3.7/site-packages (from matplotlib==3.2.1->-r requirements.txt (line 4)) (2.8.1)
Collecting kiwisolver>=1.0.1
  Downloading kiwisolver-1.2.0-cp37-cp37m-manylinux1_x86_64.whl (88 kB)
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in /srv/conda/envs/notebook/lib/python3.7/site-packages (from matplotlib==3.2.1->-r requirements.txt (line 4)) (2.4.7)
Collecting cyclor>=0.10
  Downloading cyclor-0.10.0-py2.py3-none-any.whl (6.5 kB)
Collecting pytz>=2017.2
  Downloading pytz-2020.1-py2.py3-none-any.whl (510 kB)
Collecting isodate
  Downloading isodate-0.6.0-py2.py3-none-any.whl (45 kB)
Requirement already satisfied: six in /srv/conda/envs/notebook/lib/python3.7/site-packages (from rdflib==5.0.0->-r requirements.txt (line 7)) (1.15.0)
Installing collected packages: numpy, kiwisolver, cyclor, matplotlib, pytz, pandas, isodate, rdflib, SPARQLWrapper
```



hub.gke2.mybinder.org

Stéphane POUYLLAU / ISIDORE Jupyter notebooks - Git...

Home Page - Select or create a notebook

TRIPLE-TopicsBox-notebook - Jupyter Notebook

File Edit View Insert Cell Kernel Widgets Help

Not Trusted Python 3

Memory: 115 / 2048 MB

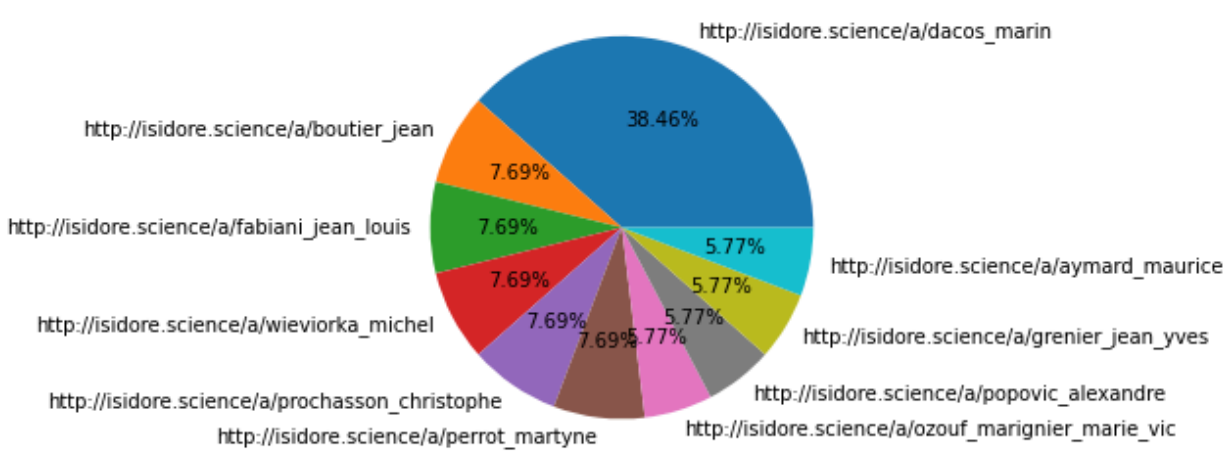
Computing co-authors and co-publications from an author

```
In [4]: from SPARQLWrapper import SPARQLWrapper, JSON
import pandas as pd

sparql = SPARQLWrapper("https://isidore.science/sparql")
sparql.setQuery("""
SELECT ?coauteurs (count(?documents) as ?count) WHERE {
?documents dcterms:creator <http://isidore.science/a/mounier_pierre>.
?documents dcterms:creator ?coauteurs
FILTER(?coauteurs != <http://isidore.science/a/mounier_pierre>)
} GROUP BY ?coauteurs
ORDER BY DESC(?count) LIMIT 10
""")
sparql.setReturnFormat(JSON)
results = sparql.query().convert()
results_df = pd.pandas.json_normalize(results['results']['bindings'])
results_df[['coauteurs.value', 'count.value']].head(10)
```

| | coauteurs.value | count.value |
|---|---|-------------|
| 0 | http://isidore.science/a/dacos_marin | 20 |
| 1 | http://isidore.science/a/boutier_jean | 4 |
| 2 | http://isidore.science/a/fabiani_jean_louis | 4 |
| 3 | http://isidore.science/a/wieviorka_michel | 4 |
| 4 | http://isidore.science/a/prochasson_christophe | 4 |
| 5 | http://isidore.science/a/perrot_martyne | 4 |
| 6 | http://isidore.science/a/ozouf_marignier_marie... | 3 |
| 7 | http://isidore.science/a/popovic_alexandre | 3 |
| 8 | http://isidore.science/a/grenier_jean_yves | 3 |
| 9 | http://isidore.science/a/aymard_maurice | 3 |

```
In [5]: labels = results_df['coauteurs.value']
data = results_df['count.value']
plt.pie(data, labels=labels, explode=None, autopct='%1.2f%%', shadow=False, radius=0.9)
plt.axis('equal')
plt.show()
plt.close()
```



| coauteurs.value | count.value | Percentage |
|---|-------------|------------|
| http://isidore.science/a/dacos_marin | 20 | 38.46% |
| http://isidore.science/a/boutier_jean | 4 | 7.69% |
| http://isidore.science/a/fabiani_jean_louis | 4 | 7.69% |
| http://isidore.science/a/wieviorka_michel | 4 | 7.69% |
| http://isidore.science/a/prochasson_christophe | 4 | 7.69% |
| http://isidore.science/a/perrot_martyne | 4 | 7.69% |
| http://isidore.science/a/ozouf_marignier_marie... | 3 | 5.77% |
| http://isidore.science/a/popovic_alexandre | 3 | 5.77% |
| http://isidore.science/a/grenier_jean_yves | 3 | 5.77% |
| http://isidore.science/a/aymard_maurice | 3 | 5.77% |

Champs des possibles pour l'exploration

Mise en œuvre d'un classifieur de texte en Deep Learning avec Keras/Tensorflow

Mise en œuvre d'un classifieur de texte en Deep Learning avec Keras

Nous allons utiliser des titres de documents, issus du moteur de recherche SHS isidore.science pour entraîner un classifieur à réseau de neurones à l'aide de Keras. Keras est une API pour le Deep Learning qui permet de créer, d'entraîner et d'utiliser des réseaux de neurones (voir les ressources bibliographiques en infra).

Pour entraîner le classifieur, nous constituons un corpus d'entraînement à l'aide de la SPARQL dans le SPARQL endpoint d'isidore.science :

```
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX sioc: <http://rdfs.org/sioc/ns#>
SELECT ?uri ?title where {
  ?uri sioc:topic <http://aurehal.archives-ouvertes.fr/subject/shs.XXXX>.
  ?uri dcterms:provenance <https://halshs.archives-ouvertes.fr>.
  ?uri dcterms:title ?title.
  FILTER (lang(?title)='fr')
} LIMIT 2000
```

Nous collectons 4000 titres :

- 2000 titres en histoire (XXXX = shs.hist), étiqueté 1
- 2000 titres en qui ne sont pas de l'histoire (XXXX = shs.info + shs.hist), étiqueté 0

Nous éliminons les mots vides, signes de ponctuation, etc. pour ne garder que les mots utiles.

Exemple :

```
Vivre travailler capotteri grecs Venise ;1
Lutter ennemi interne longue histoire obsession
```

Préparation des données et entraînement

On charge le corpus d'entraînement :

```
[1]: import pandas as pd
filepath_dict = {'histgeo': 'MotsISIDORE-histoire-info-soci...'
df_list = []
for source, filepath in filepath_dict.items():
    df = pd.read_csv(filepath, names=['words', 'label'], sep=';', encoding='utf-8')
    df['source'] = source
    df_list.append(df)
df = pd.concat(df_list)
print(df.iloc[0])
words    poèmes Lu You végétarisme
label    1
source   histgeo
Name: 0, dtype: object
```

Préparation des données et création des trains de données. Il s'agit de la vectorisation du corpus d'entraînement (et on utilise sklearn pour le machine learning).

```
[3]: from sklearn.model_selection import train_test_split
df_hist = df[df['source'] == 'histgeo']
sentencesA = df_hist['words'].values
y = df_hist['label'].values
sentences_train, sentences_test, y_train, y_test = train_test_split(
    sentencesA, y, test_size=0.25, random_state=100)
from sklearn.feature_extraction.text import CountVectorizer
```

Extracteur de titres d'ISIDORE

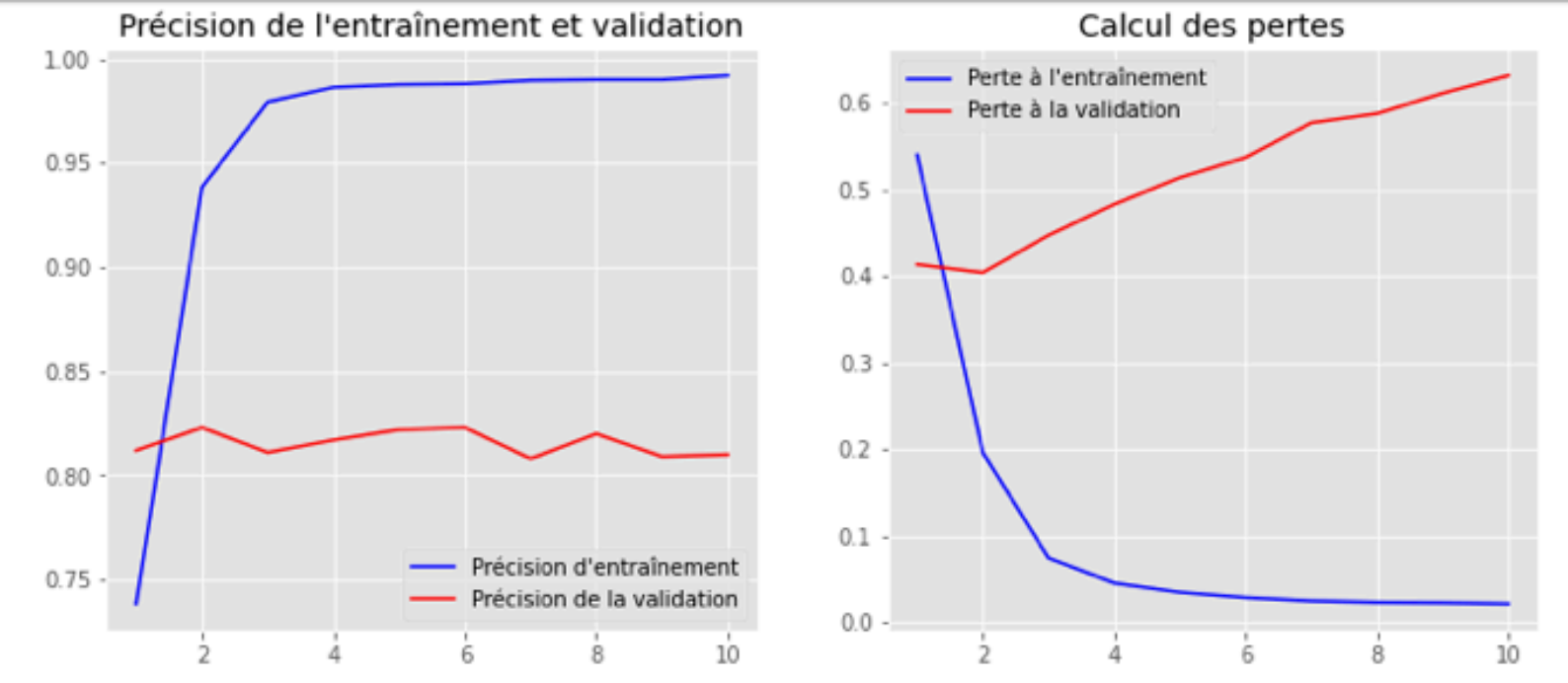
Extraire 100 titres d'ISIDORE (via son SPARQL endpoint) suivant une discipline SHS (référentiel MOSA).
But : préparer les jeux d'entraînement pour Keras.

Les disciplines SHS dans ISIDORE :

```
from SPARQLWrapper import SPARQLWrapper, JSON
import pandas as pd

sparql = SPARQLWrapper("http://isidore.science/sparql")
sparql.setQuery("""
    SELECT DISTINCT ?discipline WHERE {
      ?i <http://rdfs.org/sioc/ns#topic> ?discipline.
    } LIMIT 100
""")
sparql.setReturnFormat(JSON)
results = sparql.query().convert()
results_df = pd.json_normalize(results['results']['bindings'])
results_df.head(100)
```

<https://isidore.science/document/10670/1.rs01x9>



Utilisation du classifieur pour prédire si des titres d'articles sont de l'histoire

On définit des titres d'articles à classifier et on les vectorise pour les classifier avec le classifieur :

```
[11]: sentences_apredire = ['Stratégies éditoriales des musées. Une approche de la médiation par
    'Le monde karstique',
    'Au plus près des âmes et des corps. Une histoire intime des catholiqu
    'Cuba: pour une géographie du socialisme',
    'Migrations en Turquie',
    'Les autoroutes et informations',
    'Les seigneuries et baronies au Moyen-âge',
    'La guerre civile espagnole']
X_apredire = vectorizer.transform(sentences_apredire)
X_apredire = X_apredire.todense()
vectorizer.transform(sentences_apredire).toarray()
```

```
[11]: array([[0, 0, 0, ..., 0, 0, 0],
            [0, 0, 0, ..., 0, 0, 0],
            [0, 0, 0, ..., 0, 0, 0],
            ...,
            [0, 0, 0, ..., 0, 0, 0],
            [0, 0, 0, ..., 0, 0, 0],
            [0, 0, 0, ..., 0, 0, 0]])
```

On utilise l'entraînement avec `model.predict_classes` de Keras pour prédire les articles qui sont des articles d'histoire. Les articles = 1 sont sans doute des articles en histoire.

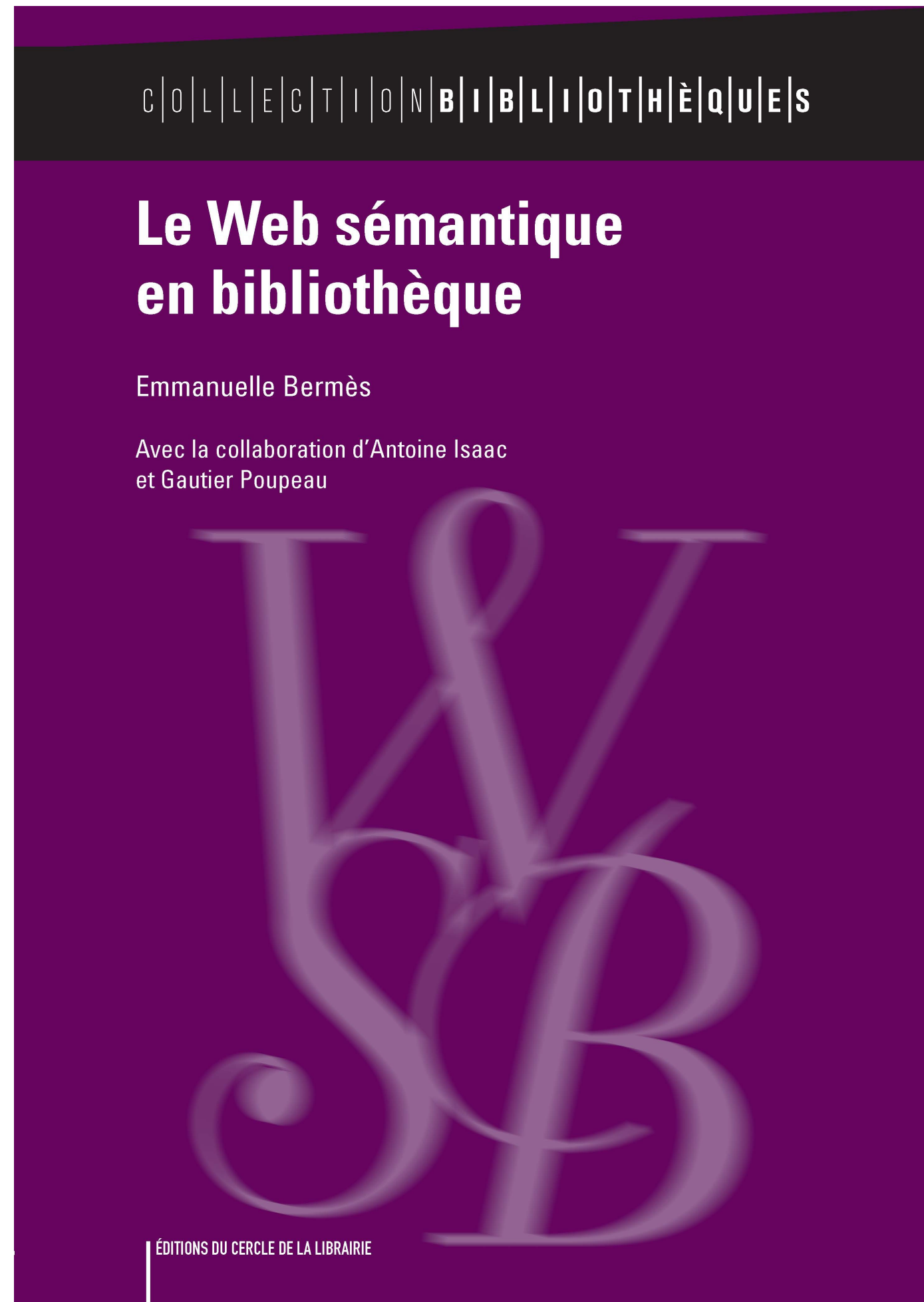
```
[13]: predictions = model.predict_classes(X_apredire, verbose=1)
print()
print("Affichage des résultats :")
print()
for i in range(len(X_apredire)):
    if predictions[i] == 0:
        print("\'%s\' = %s ==> pas de l'histoire" % (sentences_apredire [i], predictions [i]))
    else:
        print("\'%s\' = %s ==> sans doute de l'histoire" % (sentences_apredire [i], predictions [i]))
```

```
1/1 [=====] - 0s 1ms/step
```

Affichage des résultats :

```
"Stratégies éditoriales des musées. Une approche de la médiation par l'accès ouvert aux do
nnées numérisées" = [0] ==> pas de l'histoire
"Le monde karstique" = [0] ==> pas de l'histoire
"Au plus près des âmes et des corps. Une histoire intime des catholiques au xixe siècle" =
[1] ==> sans doute de l'histoire
"Cuba: pour une géographie du socialisme" = [1] ==> sans doute de l'histoire
"Migrations en Turquie" = [0] ==> pas de l'histoire
"Les autoroutes et informations" = [0] ==> pas de l'histoire
"Les seigneuries et baronies au Moyen-âge" = [1] ==> sans doute de l'histoire
"La guerre civile espagnole" = [1] ==> sans doute de l'histoire
```

Lectures...



Bermès, E., Le Web sémantique en bibliothèque, (2013).

Janakiev, N., Practical Text Classification With Python and Keras, realpython.com (2020).

Géron. A., Deep Learning avec Keras et TensorFlow, DUNOD, 2e ed. (2020)

Pouyllau, S., Web de données, big data, open data, quels rôles pour les documentalistes ? DocSI (2013).<http://doi.org/10.3917/docsi.503.0026>

Pouyllau, S. A classifier using ISIDORE, the social and humanities search engine and Keras API for Deep Learning (Version V1), (2020). <http://doi.org/10.5281/zenodo.3994126>



Explorer à l'aide du Web sémantique

EXPLO-SHS, La Rochelle, 15 oct. 2020

