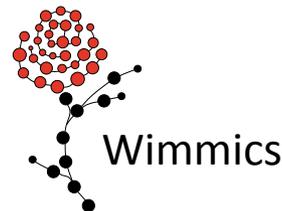




Wikipédia pour les machines

Julien Cojan



CaféIn – 27 juin 2013

Sémantisation des ressources francophones de Wikimedia



Signature de la
convention
Sémanticpédia,



Données contenues dans les articles de Wikipédia

The image shows a screenshot of the French Wikipedia article for "Musée du Louvre". Several elements are highlighted with red boxes and labeled with yellow boxes:

- titre**: Points to the article title "Musée du Louvre".
- liens**: Points to internal links within the text, such as "musées", "Paris", and "Seine".
- image**: Points to the main image of the Louvre Museum.
- champs infobox**: Points to the infobox section, which contains:
 - Informations géographiques**:
 - Pays**: France
 - Ville**: Paris
 - Adresse**: 99, rue de Rivoli, 75001 Paris
 - Coordonnées**: 48° 51' 40" Nord, 2° 20' 09" Est
 - Informations générales**:
 - Date d'inauguration**: 10 août 1793
 - Collections**: Antiquités orientales, Antiquités égyptiennes, Antiquités grecques, étrusques et romaines du Louvre, Arts de l'islam, Sculptures, Objets d'art, Peintures, Arts graphiques
 - Nombre d'œuvres**: 35 000 en exposition, 445 000 au total (dont 140 000 œuvres graphiques)
- liens interlangues**: Points to the language selection menu on the left, specifically "Deutsch" and "English".
- categories**: Points to the "Catégories" section at the bottom, which includes: Musée national (France), Musée archéologique en France, Musée gréco-romain, Musée égyptologique, Musée d'art islamique, Musée fondé en 1793, Sculpture en France, Monument parisien, Patrimoine mondial à Paris, 1er arrondissement de Paris, Louvre, and [+].

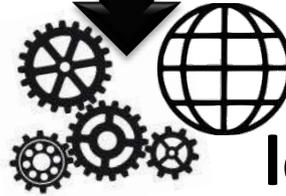
The article text itself is partially visible, starting with: "Le musée du Louvre est l'un des plus grands musées du monde et le plus grand musée de Paris par sa surface de 210 000 m² dont 60 600 consacrés aux expositions². Situé au cœur de la ville, entre la rive droite de la Seine et la rue de Rivoli, dans le 1^{er} arrondissement, le bâtiment est un ancien palais royal, le palais du Louvre. La statue équestre de Louis XIV..."



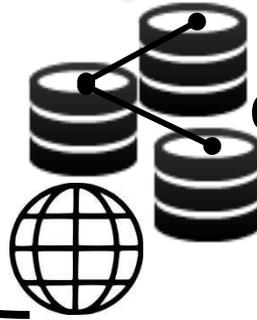
wikipedia.fr



dbpedia.fr

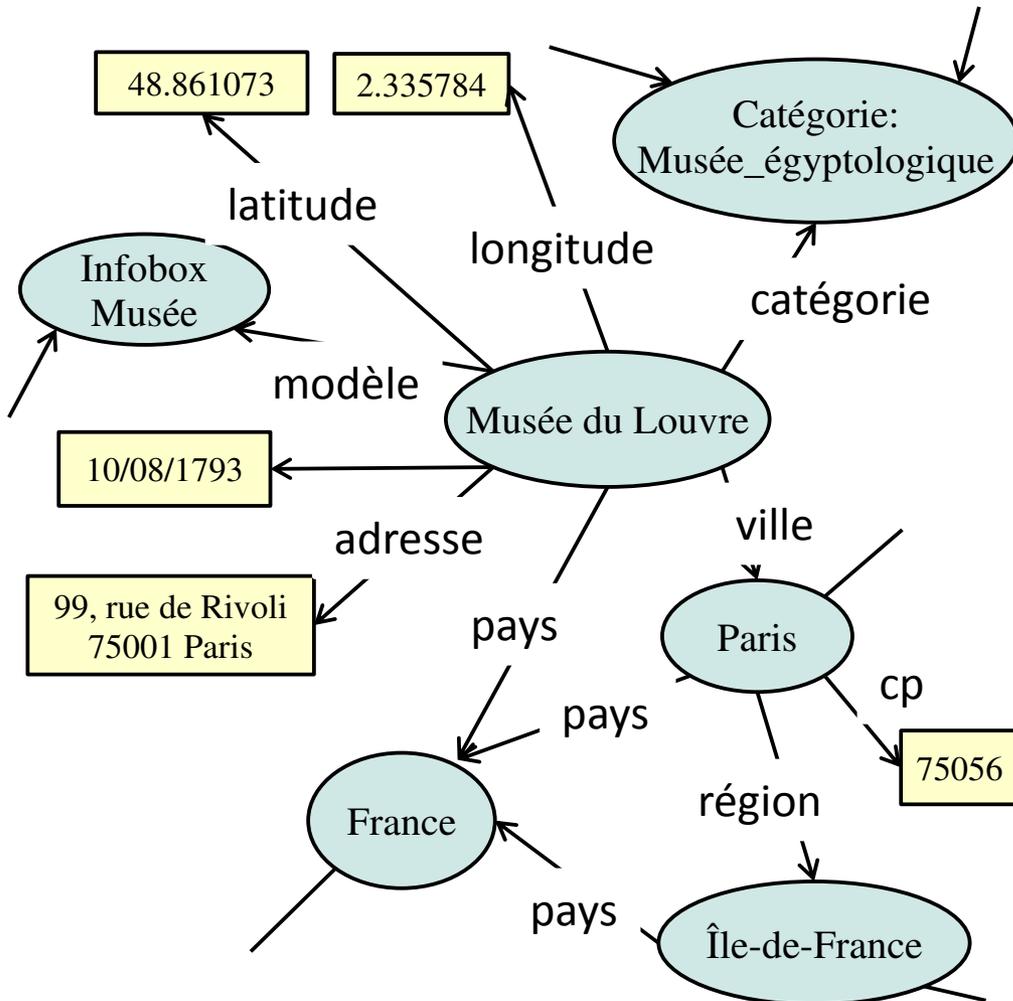


logiciel d'extraction



données encyclopédiques

Données extraites de Wikipédia



Musée du Louvre



Informations géographiques

Pays	France
Ville	Paris
Adresse	99, rue de Rivoli 75001 Paris
Coordonnées	48° 51' 40" Nord 2° 20' 09" Est

Informations générales

Date d'inauguration	10 août 1793
Collections	Antiquités orientales Antiquités égyptiennes Antiquités grecques, étrusques et romaines du Louvre Arts de l'islam Sculptures Objets d'art Peintures Arts graphiques
Nombre d'œuvres	35 000 en exposition 445 000 au total (dont 140 000 œuvres graphiques)

le gréco-romain | Musée égyptologique
 ent parisien | Patrimoine mondial à Paris

Publication sur le Web de Données



<http://linkeddata.org/>

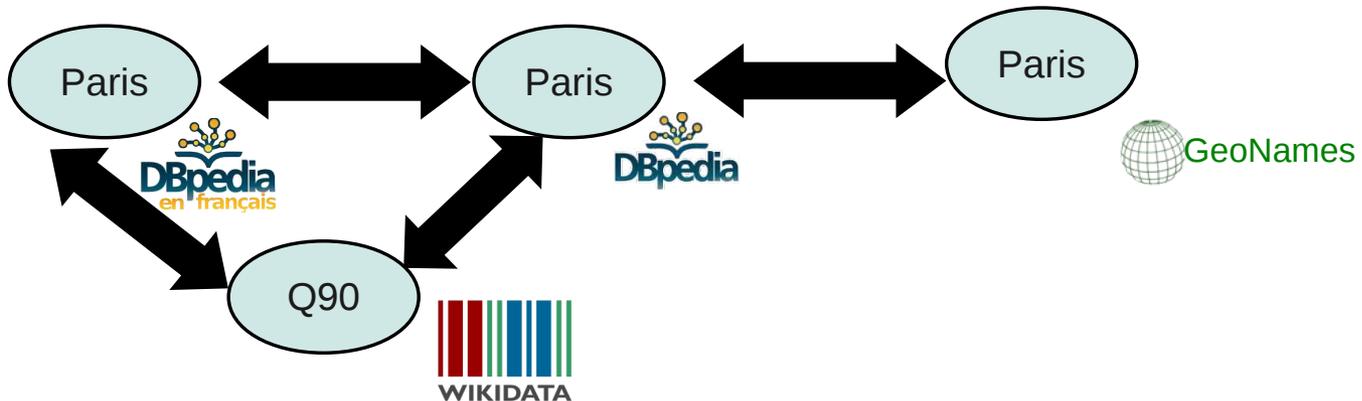
- Données ouvertes



- Données web



- Données liées



Publication sur le Web Sémantique

- Schémas ouverts



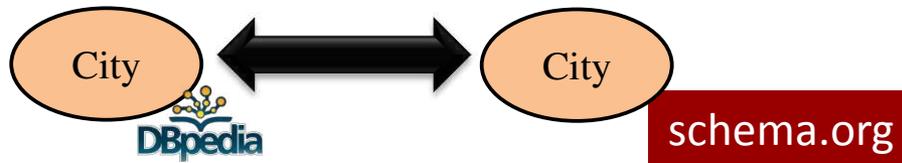
```
owl:Thing
[...]
Place (edit)
  PopulatedPlace (edit)
    Continent (edit)
    Country (edit)
    Island (edit)
    Region (edit)
    Settlement (edit)
      City (edit)
      CityDistrict (edit)
      Town (edit)
      Village (edit)
[...]
```

Extrait de la hiérarchie de classes de l'ontologie de DBpedia

- Schémas web



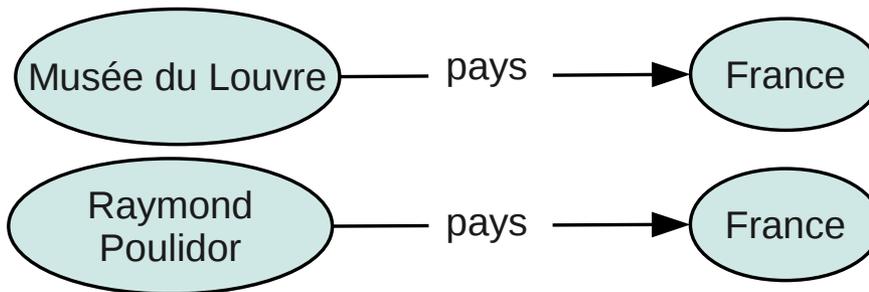
- Schémas liés



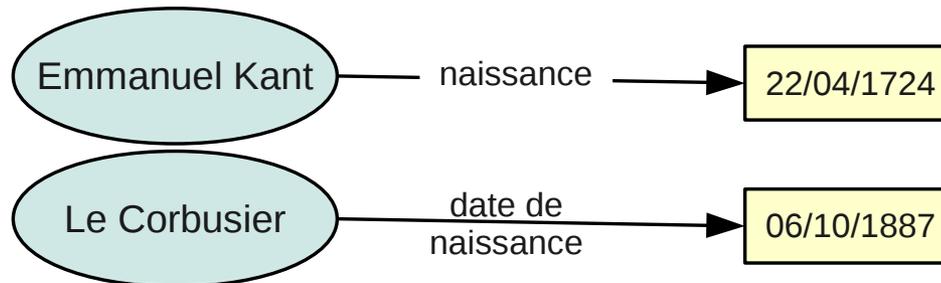
Limites de l'extraction générique

Problèmes soulevés par l'extraction textuelle

- Mêmes termes, plusieurs sens



- Différents termes avec le même sens



Vocabulaire défini collaborativement sur
<http://mappings.dbpedia.org>

Ontologie :

- Hiérarchie de classes

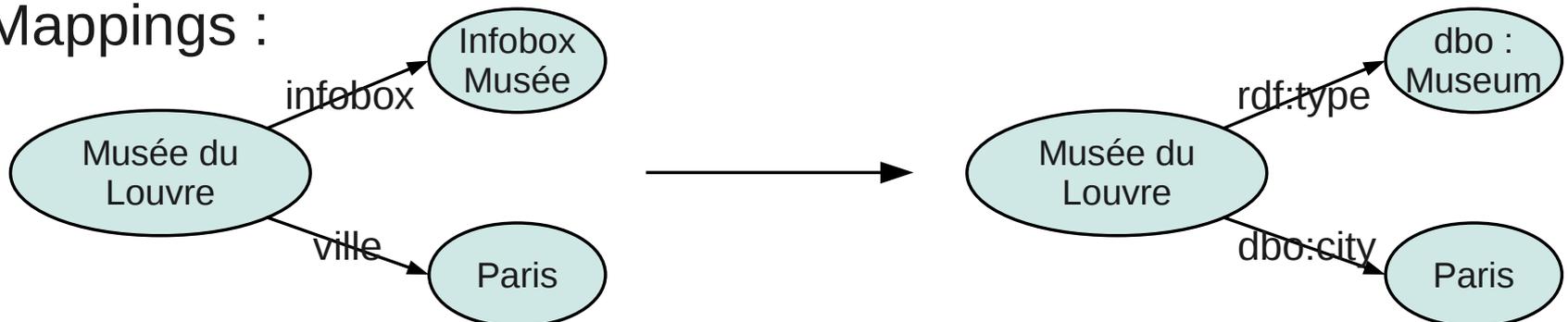
<http://dbpedia.org/ontology/Museum>

- Propriétés

<http://dbpedia.org/ontology/city>

dbo:

Mappings :



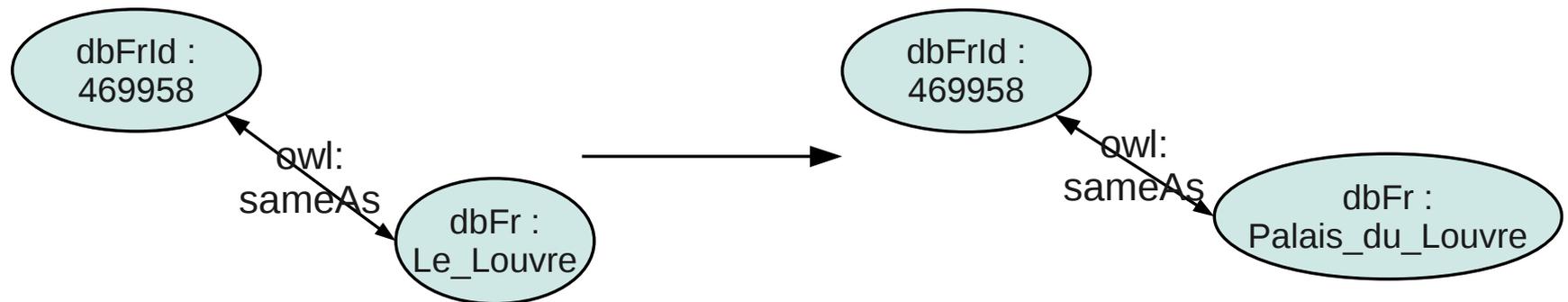
- Lorsqu'une page est renommée

« Le Louvre » → « Palais du Louvre »

~~dbFr :Le_Louvre~~ → dbFr :Palais_du_Louvre

- Introduction d'URI ne indépendante du titre

<http://fr.dbpedia.org/resourceById/469958>





Interroger les données

Accès aux données

- Téléchargement de l'ensemble des données
<http://fr.dbpedia.org/download/>
(licence: CC – BY – SA)

- Par déréférencement



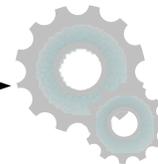
http →



- SPARQL
<http://fr.dbpedia.org/sparql>



SPARQL →



Déréférencement

http://fr.dbpedia.org/resource/Musée_du_Louvre



About: [Musée du Louvre](#)

An Entity of Type : [architectural structure](#), from Named Graph : <http://fr.dbpedia.org>, within Data Space : fr.dbpedia.org

Property	Value
dbpedia-owl:abstract	<ul style="list-style-type: none"> Le musée du Louvre est l'un des plus consacrés aux expositions. Situé a bâtiment est un ancien palais royal mais le palais n'est pas aligné sur c depuis les rois capétiens jusqu'à n
dbpedia-owl:address	<ul style="list-style-type: none"> 99, rue de Rivoli 75001 Paris
dbpedia-owl:city	<ul style="list-style-type: none"> dbpedia-fr:Paris
dbpedia-owl:collection	<ul style="list-style-type: none"> Arts de l'Islam Peintures Sculptures Arts graphiques Antiquités grecques, étrusques et Objets d'art Antiquités orientales Antiquités égyptiennes
dbpedia-owl:country	<ul style="list-style-type: none"> dbpedia-fr:France

```

<rdf:RDF>
- <rdf:Description rdf:about="http://fr.dbpedia.org/resource/Musée_du
  <rdf:type rdf:resource="http://schema.org/Place"/>
  <rdf:type rdf:resource="http://dbpedia.org/ontology/Place"/>
  <rdf:type rdf:resource="http://www.opengis.net/gml/_Feature"/>
  <rdf:type rdf:resource="http://dbpedia.org/ontology/Building"/>
  <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Thing"/>
  <rdf:type rdf:resource="http://dbpedia.org/ontology/ArchitecturalS
  <rdf:type rdf:resource="http://dbpedia.org/ontology/Museum"/>
  <rdf:type rdf:resource="http://schema.org/Museum"/>
  <rdfs:label xml:lang="fr">Musée du Louvre</rdfs:label>
  <geo:lat rdf:datatype="http://www.w3.org/2001/XMLSchema#float"
  <geo:long rdf:datatype="http://www.w3.org/2001/XMLSchema#floa
  <dbpedia-owl:country rdf:resource="http://fr.dbpedia.org/resource
  <dbpedia-owl:city rdf:resource="http://fr.dbpedia.org/resource/Par
  <dbpedia-owl:address xml:lang="fr">99, rue de Rivoli</dbpedia-owl:
  <dbpedia-owl:address xml:lang="fr">75001 Paris</dbpedia-owl:
  <dbpedia-owl:openingDate rdf:datatype="http://www.w3.org/200
  
```

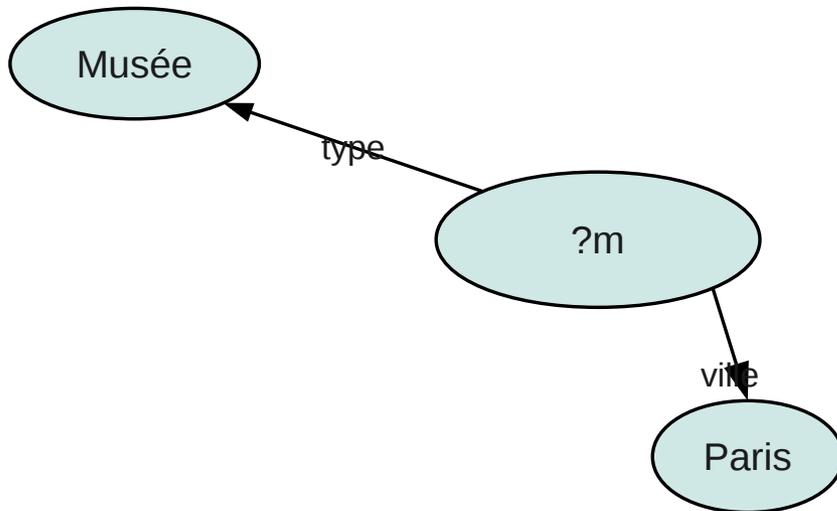


SPARQL

- Langage de requête sur des données RDF
- Standard du W3C
- Accessible à <http://fr.dbpedia.org/sparql>
- Voir présentation d'Olivier Corby :
http://devlog.cnrs.fr/_media/jdev2013/sparql-jdev.pdf

Exemple SPARQL 1

Musées à Paris



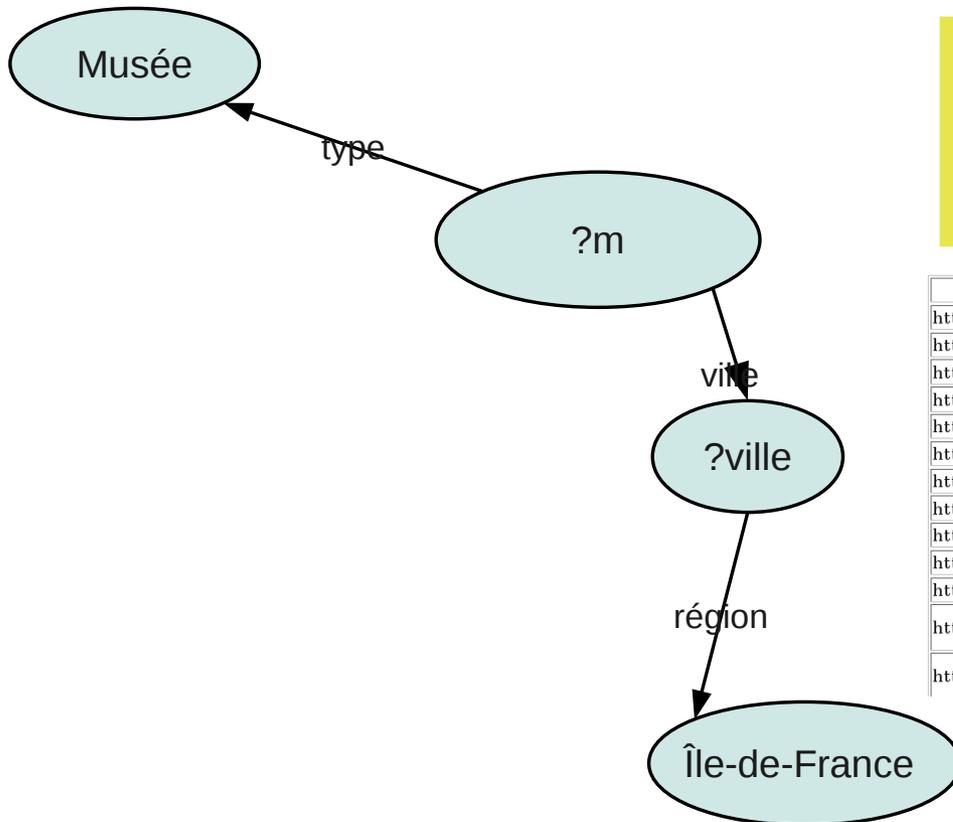
```

SELECT DISTINCT ?m WHERE {
  ?m rdf:type dbpedia-owl:Museum .
  ?m dbpedia-owl:city dbpedia-fr:Paris .
}
  
```

m
http://fr.dbpedia.org/resource/Catacombes_de_Paris
http://fr.dbpedia.org/resource/Bibliothèque_nationale_de_France
http://fr.dbpedia.org/resource/Musée_de_Cluny
http://fr.dbpedia.org/resource/Musée_Picasso_(Paris)
http://fr.dbpedia.org/resource/Musée_du_Louvre
http://fr.dbpedia.org/resource/Musée_Grévin
http://fr.dbpedia.org/resource/Musée_d'Orsay
http://fr.dbpedia.org/resource/Centre_national_d'art_et_de_culture_Georges-Pompidou
http://fr.dbpedia.org/resource/Musée_Nissim-de-Camondo
http://fr.dbpedia.org/resource/Muséum_national_d'histoire_naturelle
http://fr.dbpedia.org/resource/Palais_de_la_découverte
http://fr.dbpedia.org/resource/Musée_du_quai_Branly
http://fr.dbpedia.org/resource/Bibliothèque_historique_de_la_Ville_de_Paris
http://fr.dbpedia.org/resource/Cité_des_sciences_et_de_l'industrie
http://fr.dbpedia.org/resource/Musée_des_arts_et_métiers
http://fr.dbpedia.org/resource/Musée_national_d'art_moderne
http://fr.dbpedia.org/resource/Musée_Rodin
http://fr.dbpedia.org/resource/Musée_national_Eugène-Delacroix
http://fr.dbpedia.org/resource/Grand_Palais_(Paris)
http://fr.dbpedia.org/resource/Petit_Palais
http://fr.dbpedia.org/resource/Musée_Bourdelle
http://fr.dbpedia.org/resource/Jeu_de_Paume_(musée)

Exemple SPARQL 2

Musées dans une ville d'Île de France

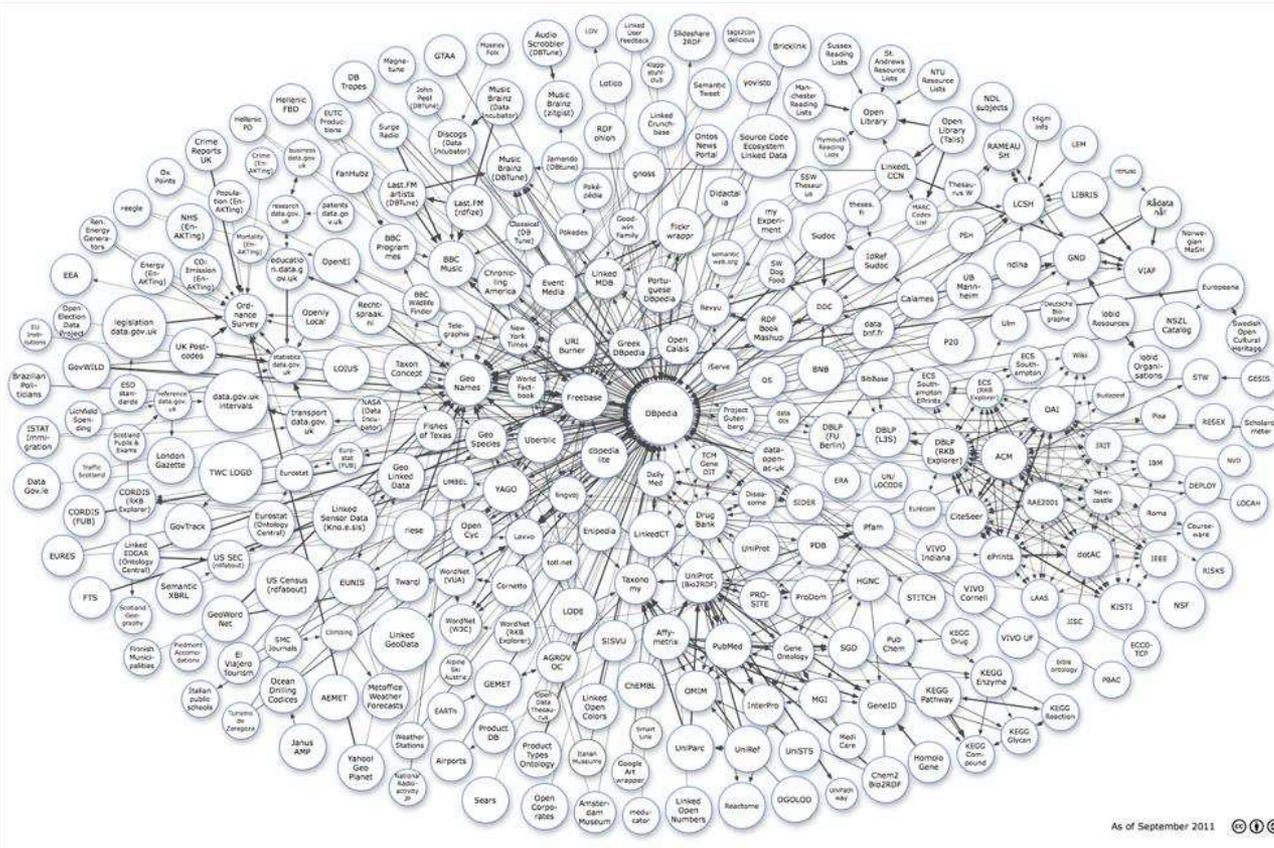


```

SELECT DISTINCT ?m?ville WHERE {
  ?m rdf:type dbpedia-owl:Museum .
  ?m dbpedia-owl:city ?ville .
  ?ville dbpedia-owl:region dbpedia-fr:Île-de-France
}
  
```

m	ville
http://fr.dbpedia.org/resource/Catacombes_de_Paris	http://fr.dbpedia.org/resource/Paris
http://fr.dbpedia.org/resource/Bibliothèque_nationale_de_France	http://fr.dbpedia.org/resource/Paris
http://fr.dbpedia.org/resource/Musée_de_Cluny	http://fr.dbpedia.org/resource/Paris
http://fr.dbpedia.org/resource/Musée_Picasso_(Paris)	http://fr.dbpedia.org/resource/Paris
http://fr.dbpedia.org/resource/Musée_du_Louvre	http://fr.dbpedia.org/resource/Paris
http://fr.dbpedia.org/resource/Musée_Grévin	http://fr.dbpedia.org/resource/Paris
http://fr.dbpedia.org/resource/Musée_d'Orsay	http://fr.dbpedia.org/resource/Paris
http://fr.dbpedia.org/resource/Musée_Nissim-de-Camondo	http://fr.dbpedia.org/resource/Paris
http://fr.dbpedia.org/resource/Muséum_national_d'histoire_naturelle	http://fr.dbpedia.org/resource/Paris
http://fr.dbpedia.org/resource/Palais_de_la_découverte	http://fr.dbpedia.org/resource/Paris
http://fr.dbpedia.org/resource/Musée_du_quai_Branly	http://fr.dbpedia.org/resource/Paris
http://fr.dbpedia.org/resource/Musée_d'archéologie_nationale	http://fr.dbpedia.org/resource/Saint-Germain-en-Laye
http://fr.dbpedia.org/resource/Rambolitrain	http://fr.dbpedia.org/resource/Rambouillet

Comment lier les données ?

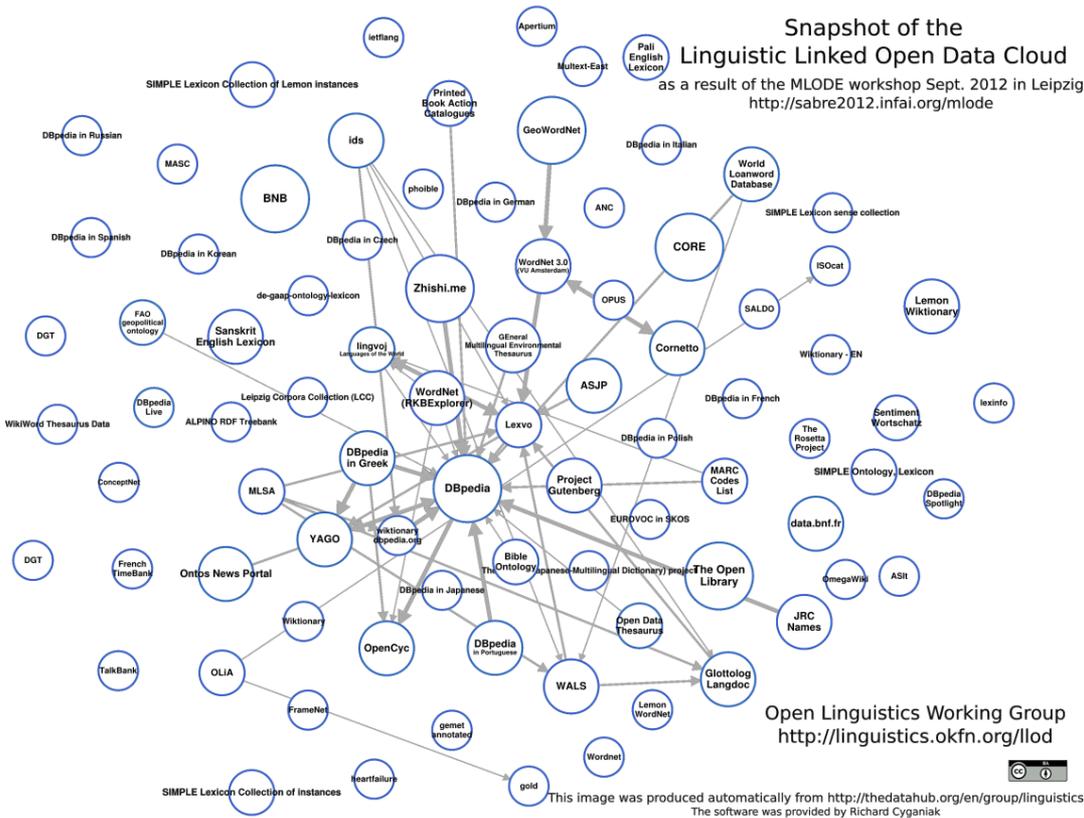


Liens vers :

- Freebase
- CIA factbook
- Geonames
- Project Gutenberg
- DrugBank

...

Web de Données multilingue



- Membre du comité d'internationalisation de DBpedia [1] (14 chapitres)
- Interconnexion
 - Liens d'équivalence entre instances
 - Vocabulaire commun (Ontologie DBpedia)

[1] <http://dbpedia.org/Internationalization>



	400 millions de triplets	130 millions de triplets
	760 000	143 000
	570 000	318 000
	330 000	70 000
...

49 000 hors DBpedia En

180 000 hors DBpedia En

30 000 hors DBpedia En

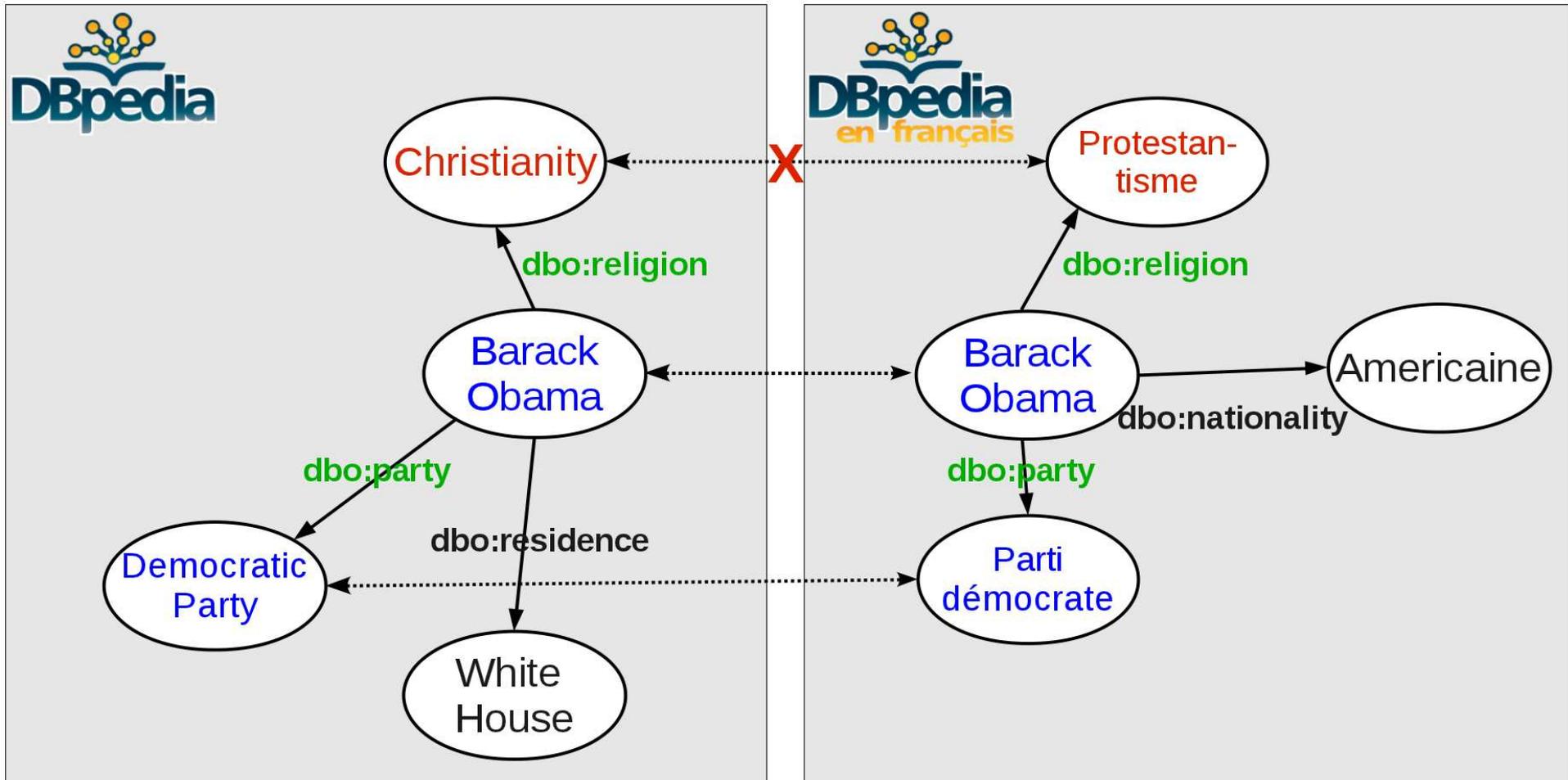
Nombre de valeurs communes 343 583

Nombre de valeurs distinctes 175 227

Nombre de valeurs uniquement dans DBpedia Fr 460 586

Nombre de valeurs uniquement dans DBpedia En 2 022 557

Comment agréger les données de différentes sources ?





<https://www.wikidata.org>

- Données éditées collaborativement
(Wikimedia commons de données)
- Recouvrement avec DBpedia
 - ✓ Phase 1 : liens interlangue
 - 🕒 Phase 2 : infoboxes
- Futur
 - Création de liens entre DBpedia et Wikidata (compatibilité)
 - Réutilisation d'outils de DBpedia



Export RDF de Wikidata

- Dump disponible

<http://semanticweb.org/RDF/Wikidata/>

- Pas chargé dans un serveur SPARQL

- Modèle de données

<https://meta.wikimedia.org/wiki/Wikidata/Development/RDF>



Export RDF de Wikidata

Spain

[\[edit \]](#)

country located mostly in Western Europe

[\[edit \]](#)

Also known as: Kingdom of Spain España Reino de España

[\[edit \]](#)

Statements

capital	Madrid	wd:Q29	wd:P30v	wd:Q2807
		Spain	capital	Madrid

▼ 1 source

[\[edit \]](#)

stated in Spanish Constitution of 1978

quote La capital del Estado es la villa de Madrid.

Wikipedia pages linked to this item (240 entries)

Language	Code	Linked article	
Acèh	acewiki	Seupanyo	[edit]
Afrikaans	afwiki	Spanje	[edit]



Export RDF de Wikidata

Spain

[\[edit \]](#)

country located mostly in Western Europe

[\[edit \]](#)

Also known as:

Kingdom of Spain

España

Reino de España

[\[edit \]](#)

Statements

capital	Madrid	[edit]
	<p>▼ 1 source</p> <p>stated in Spanish Constitution of 1978</p> <p>quote La capital del Estado es la villa de Madrid.</p>	[edit]

Wikipedia pages linked to this item (see also)

Language	Code
Acèh	acew
Afrikaans	afwik

wd:Q29	wd:P30s	wd:Q29SA[...]
wd:Q29SA[...]	a	wo:Statement .
wd:Q29SA[...]	wd:P30v	wd:Q46 .
wd:Q29SA[...]	pv:wasDerivedFrom	w:R694d42ff4494a7e9 .
w:R694d42ff4494a7e9	a	wo:Reference .
w:R694d42ff4494a7e9	w:P248r	w:Q847664 .
w:R694d42ff4494a7e9	w:P387r	"La capital del Estado es la villa de Madrid." .



Que faire avec ces
données ?

Zone 47

Indexation multi-lingue
d'œuvres d'art



index de
collections



international

DBpedia Artworks | Œuvres d'a...

zone47.com/db/

Français 20 ok

DBpedia (en, fr, de, nl, el) Artworks | Œuvres d'art | Gemälde | Schilderijen | Έργα Ζωγραφικής

Page 1 [Suivant](#) [Fin](#)

 Chimère d'Arezzo	 Avukana Buddha statue (en)	 La Joconde Léonard de Vinci, c. 1503–1519	 Le Radeau de La Méduse Théodore Géricault, 1819 - Musée du Louvre	 Tres de Mayo Francisco de Goya, 1814 - Musée du Prado
 Portrait du Dr Gachet avec branche de digitale Vincent van Gogh, 1890 - Musée d'Orsay	 Manneken-Pis	 Spoliarium (en) Juan Luna, circa 1894 - Musée national des Philippines	 Femmes de Tahiti 1891 - Musée d'Orsay	 Rhinocéros de Dürer 1515 - British Museum
 La Laitière Johannes Vermeer, c. 1657–1658 -	 The Music Lesson (en) Johannes Vermeer, 1662 - Royal Collection	 La Madone du Grand-Duc Raphaël (peintre), 1505 - Palais Pitti	 La Reddition de Breda Diego Vélasquez, 1634 - Musée du Prado	 La Vierge aux rochers Léonard de Vinci, 1483 - Musée du Louvre

<http://zone47.com/db/>

Discovery Hub.

Comment découvrir de nouveaux sujets d'intérêt?



index de collections



international



exploration

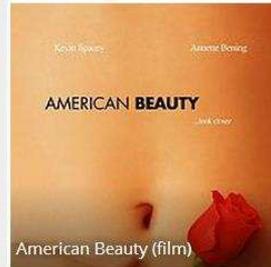
Discovery Hub.

Home Search Profile

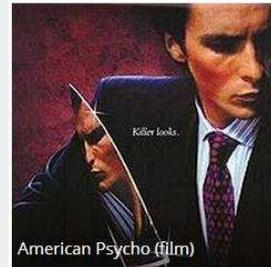
Mosaic
Detailed List

Fight Clu... X Search

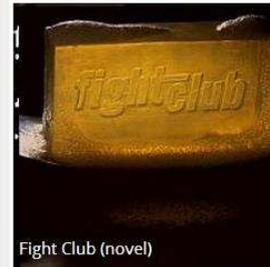
Top



American Beauty (film)



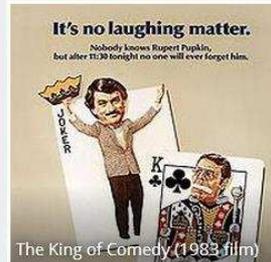
American Psycho (film)



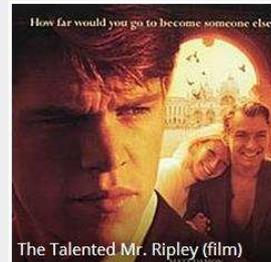
Fight Club (novel)



Magnolia (film)



The King of Comedy (1983 film)



The Talented Mr. Ripley (film)



List of American films of 1999



Unreliable narrator

Search Box

A Clo

The W



Question Answering
wiKiframework-based
System

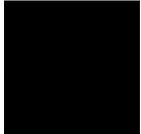
Comment interroger les
données en langue naturelle?



**index de
collections**



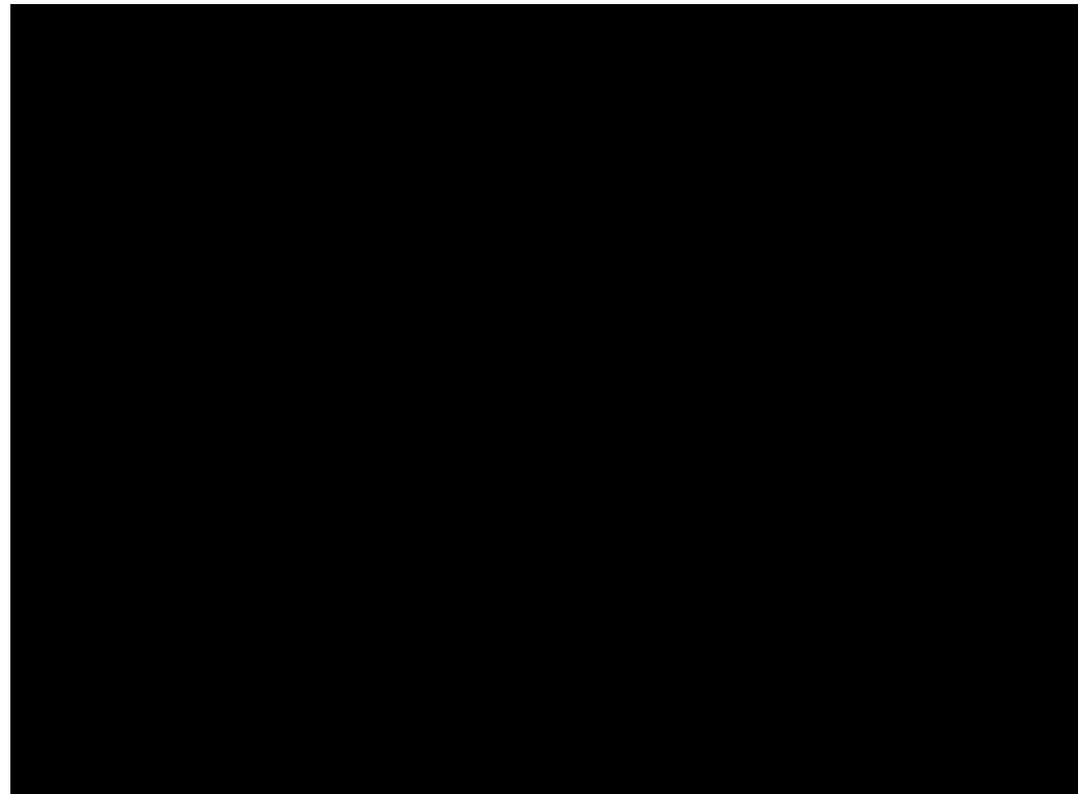
international



exploration



accessibilité



<http://dbpedia.inria.fr/qakis>



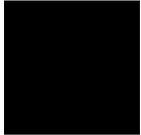
Reador: agrégateur sémantique de news



index de collections



international



exploration



accessibilité



aggregation

Reador - a semantic feed aggregator

Reador Home News Sources Your twitter timeline! Edit account Logout

Follow the project

Reador can help you?

Advanced search Go reader!

More about Reador!

<http://reador.net>

Gephi

+SemanticWebImport

Outil de visualisation de graphes

<http://wiki.gephi.org/index.php/SemanticWebImport>



index de collections



international



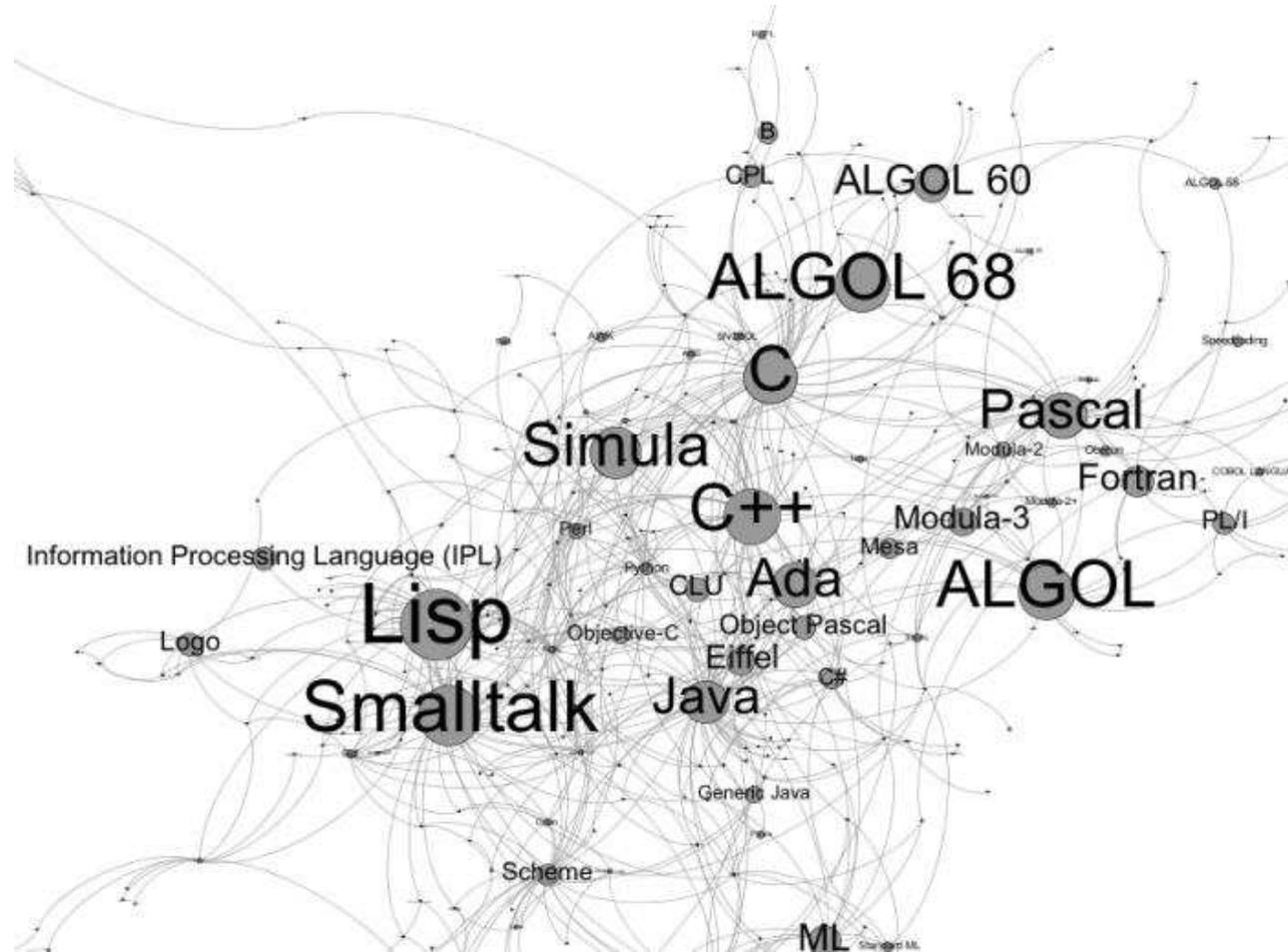
exploration



accessibilité



agrégation



<http://blog.ouseful.info/2012/07/03/mapping-how-programming-languages-influenced-each-other-according-to-wikipedia/>

HdA Lab

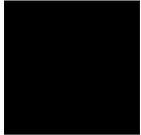
Indexation sémantique de
fiches éducatives



index de
collections



international



exploration



accessibilité



aggregation



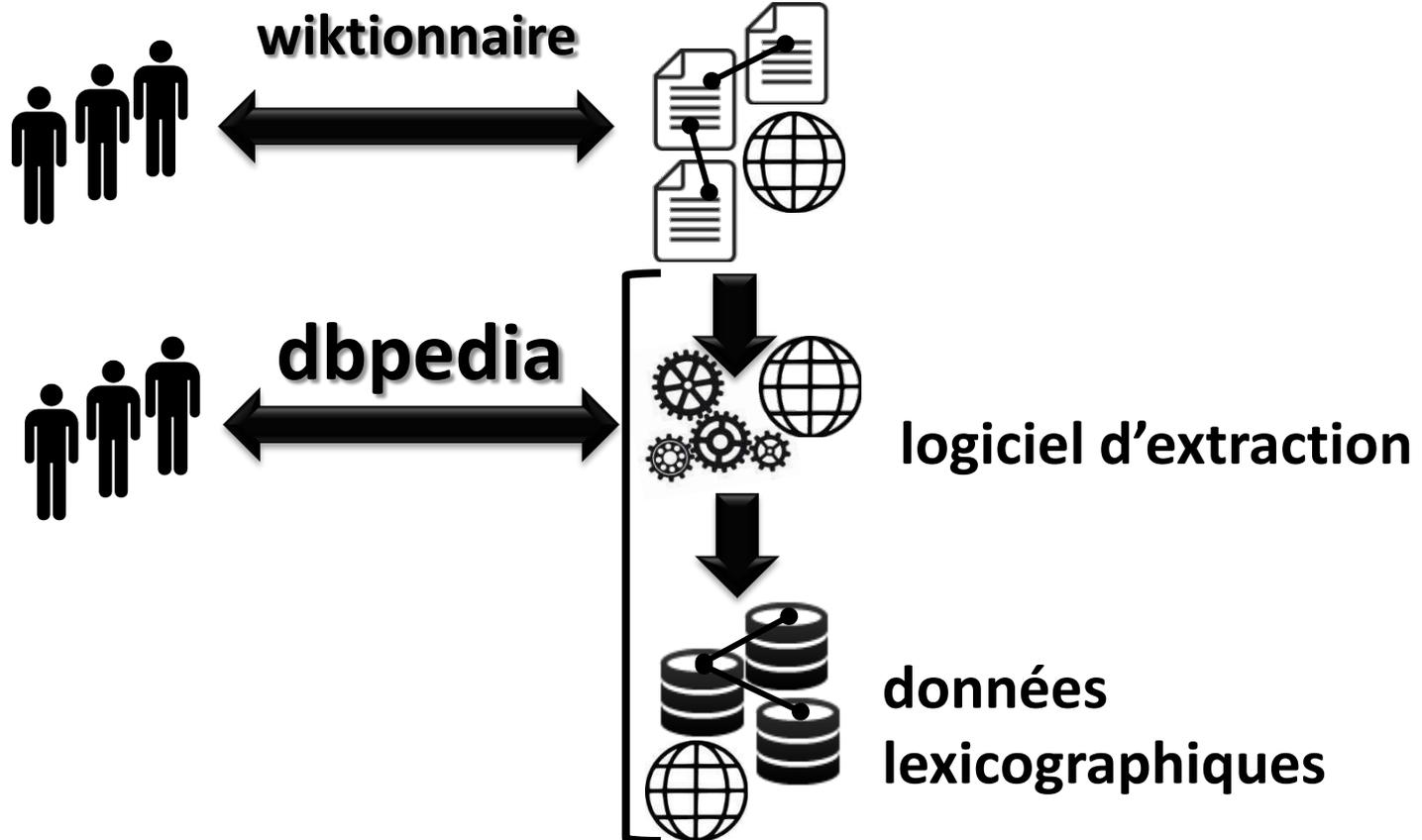
éducation

The screenshot shows the HdA Lab website interface. At the top, there's a navigation bar with "HdA Lab" and "Accueil". Below that, a search bar and a "Recherche par facettes : Nouvelle session" button. A timeline filter is set to "Toutes périodes". The main content area is divided into several sections:

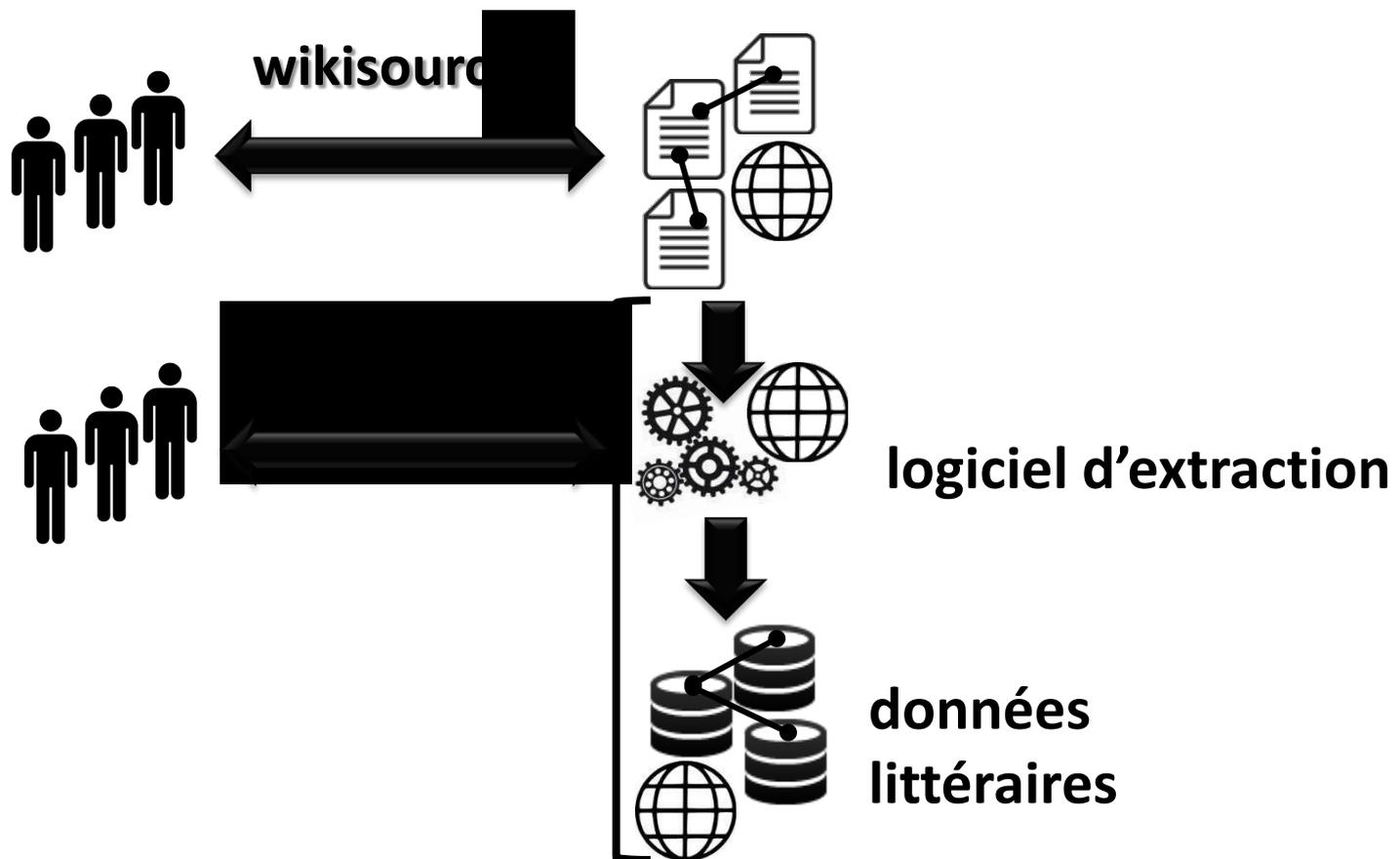
- Gérer la vue "Mes résultats de recherche"**: A world map showing search results by country.
- Notes**: A section for adding notes to the current view.
- Disciplines artistiques**: A horizontal bar chart showing the distribution of search results across various artistic disciplines: Peinture, Sculpture, Musique, Photographie, Mise en scène, Architecture, Cinéma, Théâtre, Littérature, and Dessin.
- Résultats de recherche**: A list of search results, with the first one being "Louise Bourgeois" (4562 notices). The result includes a snippet of text, a map of Paris, and a list of tags: Louise Bourgeois (plasticienne), Installation (art), XXe siècle, 2008, Mémoire (sciences humaines), Sculpture, Peinture, Art contemporain, Enfance, Métamorphose, Inconscient, and Famille.
- Annotations**: A section for adding annotations to the current result.

<http://hdalab.iri-research.org/hdalab/>

Application à d'autres ressources ?



Application à d'autres ressources wiki ?



?Questions