

Initiation au web sémantique avec Wikidata

1. Le web sémantique – définition

« **Le web sémantique permet aux données d'être partagées et réutilisées entre plusieurs applications, entreprises et groupes d'ordinateurs** ». Voici comment le World Wide Web Consortium définit le web sémantique. Il s'agit d'un **réseau de données qui peuvent être traitées directement ou indirectement pour aider les utilisateurs à créer de nouvelles connaissances**. En d'autres termes, il s'agit de **faire émerger des connaissances à partir de masses de données retrouvées sur internet**.

Le web sémantique, qu'on appelle aussi le web des métadonnées, repose sur un ensemble de technologies censées rendre le contenu des ressources (images, textes, vidéos, etc) du web plus accessibles et utilisables par les humains. Ce système permet de relier entre elles les informations qui jusque-là étaient traitées séparément. Le fonctionnement du web sémantique intègre donc au web actuel la possibilité d'agréger plusieurs données liées entre elles soit sémantiquement, soit par des attributs qui les déterminent.

« **Le web sémantique, c'est le web qu'on veut faire parler** ».

Afin de fonctionner, il faut mettre en lien toutes les données existantes. C'est-à-dire **structurer au préalable les sources d'informations pertinentes en leur attribuant des métadonnées lisibles par l'Homme et la machine et surtout standardisées**. Concrètement une page web doit être annotée en fonction des informations qu'elle contient, et ce, selon un vocabulaire standardisé (syntaxe RDF, vocabulaire de Schema.org et identifiant Wikidata).

Une recherche grâce au web sémantique nous permettra de répondre à des questions comme :

- Quelles sont les plus grandes villes du monde qui ont une femme maire ?
- Quels aéroports sont situés dans un rayon de 100 km de Berlin ?
- Qui sont tous les descendants connus de Gengis Khan ?

2. Les champs d'application du web sémantique

L'ensemble du web est impacté par le web sémantique. Les réseaux sociaux, les blogs, la recherche scientifique, la recherche d'information sur internet, le e-commerce...

3. Les limites

Le problème c'est qu'il faut structurer le web et cela prend du temps ! En effet, chaque page, chaque élément (texte, photo, vidéo, lien hypertexte...) doit être référencé. Chaque élément doit avoir un attribut, une valeur. Le second problème est qu'il faut que ces métadonnées soient standardisées pour que cela soit le plus efficace possible.

4. Wikidata

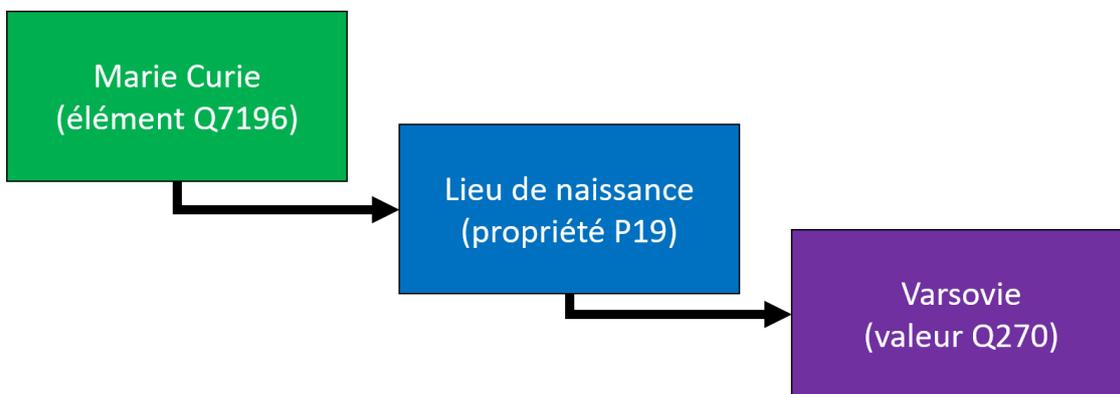
Wikidata est **une base de données liées libre** qui peut être lue et modifiée par les êtres humains et les machines. Wikidata sert de dépôt central à des **données structurées** utilisées pour les projets frères de Wikimedia, parmi lesquels Wikipédia, Wikivoyage, Wikisource et d'autres.

Dans Wikidata, un concept, un sujet ou un objet est représenté par un **élément**. Chaque élément dispose d'une page spécifique commençant par **Q**. Par exemple, pour l'humain « Marie Curie » on retrouvera la page Q7186. Pour l'objet « Continent Amérique du Sud », la page concernée est Q18. Un volcan (édifice géologique) aura la page Q8072.

Dans Wikidata, les **déclarations** permettent de stocker des informations à propos d'un élément. Cela se fait en formant une paire : une **propriété** avec au moins une **valeur**. Cette paire est le cœur de la déclaration. Les déclarations peuvent être complétées ou contextualisées à l'aide de valeurs additionnelles, comme les qualificatifs, les références ou les rangs. Elles permettent de relier les éléments les uns aux autres, formant ainsi un graphe d'éléments : **les données sont structurées**.

Par exemple, Marie Curie (élément Q7186), possède un lieu de naissance (propriété P19) qui se trouve être la ville de Varsovie (valeur Q270).

Graphiquement, cela peut être représenté ainsi :



Chaque donnée est donc structurée ainsi dans Wikidata.

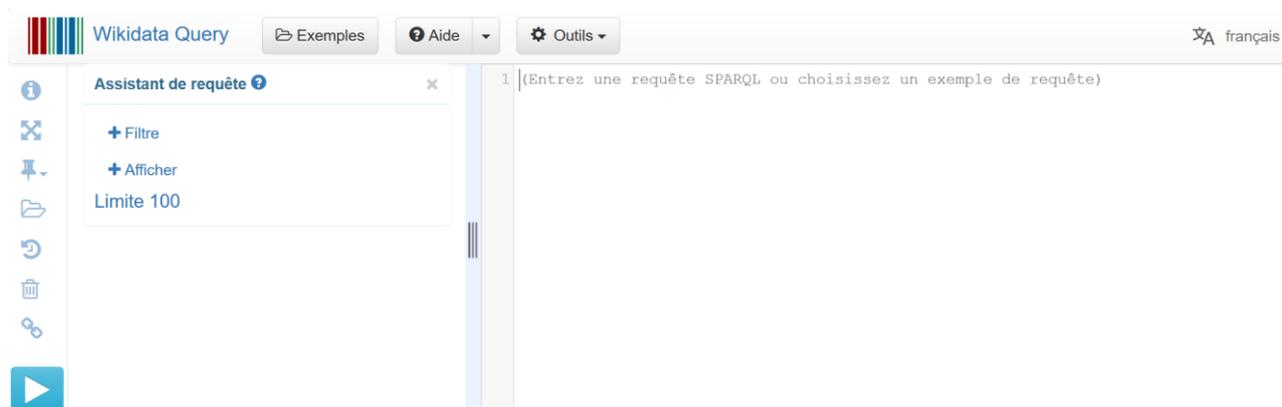
5. Rechercher sur Wikidata

Faire une recherche sur Wikidata consiste à requêter (c'est-à-dire interroger) la base de données de Wikidata. Pour ce faire, on peut utiliser l'assistant de requêtes de Wikidata ou bien interroger la base de données en lui parlant le SPARQL...

Le langage SPARQL (SPARQL Protocol and RDF Query Language) est un langage de requête et un protocole qui permet de rechercher, d'ajouter, de modifier ou de supprimer des données disponibles à travers internet. Ce langage se rapproche du SQL (pour travailler sur certaines bases de données).

L'assistant de requêtes de Wikidata est bien plus convivial et bien plus intuitif !

Pour lancer une requête, il faut se rendre sur <https://query.wikidata.org> puis sélectionner, à l'aide des filtres, les éléments et les propriétés recherchées. L'intérêt avec l'assistant de requêtes, c'est qu'il propose des autocomplétions (qui permettent de proposer des éléments de recherche par intuition). On peut aussi voir dans le volet de droite la requête s'afficher en langage SPARQL.



Wikidata propose une série d'exemples de recherches, pratique pour se lancer au début et pour comprendre le langage SPARQL si on le souhaite.

Les résultats s'affichent en bas de la page. Une multitude de formats est disponible pour l'affichage des résultats (tableau, grilles d'images, cartes, graphiques...).

Il faut noter que les résultats obtenus sont uniquement ceux référencés dans Wikidata.

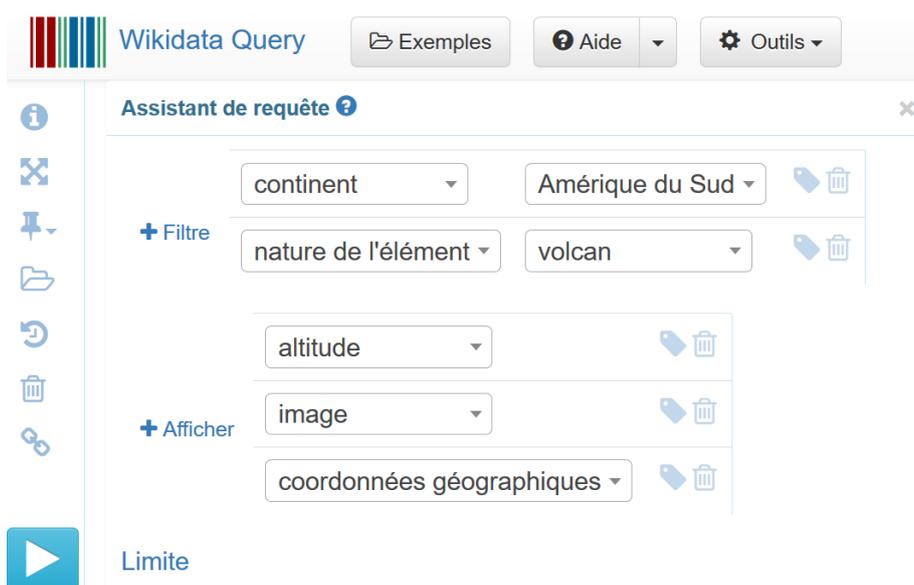
6. Un exemple de recherche

Je souhaite chercher les volcans présents sur le continent Amérique du Sud et je voudrai que Wikidata m'affiche leurs coordonnées géographiques, leur altitude et qu'il m'affiche une photo de l'édifice volcanique. Je souhaite aussi que Wikidata m'affiche les résultats sous forme de grille d'images avec les informations recherchées.

Dans la zone « Filtre », j'écris « Amérique du Sud ». Je m'assure que l'assistant de requête affiche la propriété « continent » puis j'écris « volcan » en 2^e élément et je m'assure qu'il propose bien « nature de l'élément » comme propriété.

Dans la zone « Afficher », je rentre les informations suivantes : « altitude », « image » et « coordonnées géographiques ».

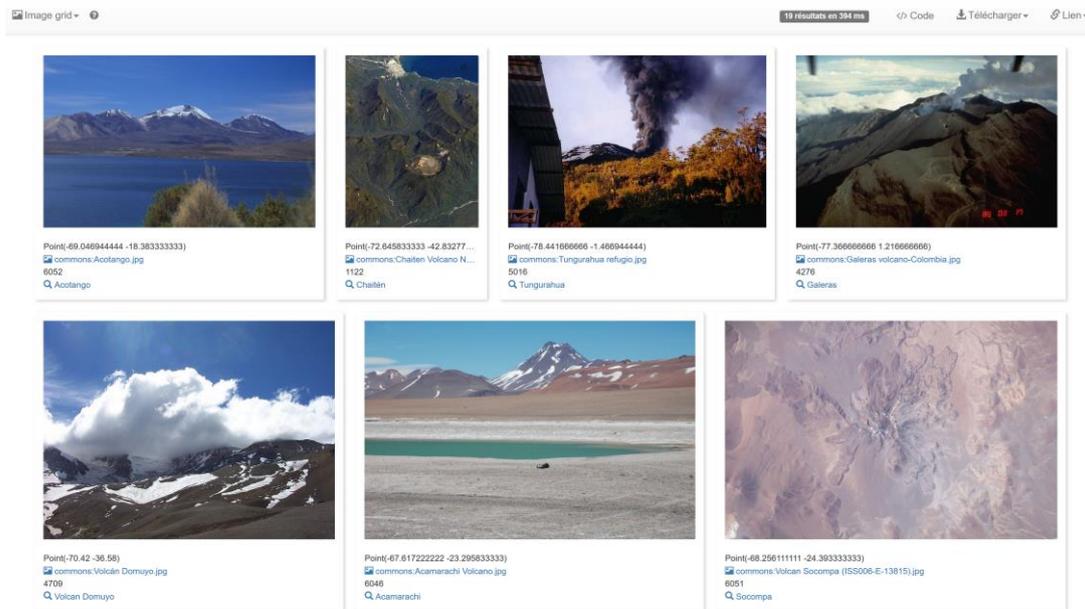
Je mets la limite à zéro (initialement définie à 100) pour m'assurer d'avoir tous les résultats disponibles sur Wikidata.



On peut observer le code SPARQL s'afficher en même temps qu'on définit notre requête. Pour les plus téméraires, cela donne une petite idée de comment cela est organisé. Notez qu'on y retrouve les noms des éléments au format *Q+nombre* (exemple : Amérique du Sud = Q18) et les propriétés au format *P+nombre* (exemple : coordonnées géographiques = P625).

```
1 SELECT ?Am_rique_du_Sud ?Am_rique_du_SudLabel ?altitude ?image ?coordonn_es_g_ographiques
2 WHERE {
3   SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }
4   ?Am_rique_du_Sud wdt:P30 wd:Q18.
5   ?Am_rique_du_Sud wdt:P31 wd:Q8072.
6   OPTIONAL { ?Am_rique_du_Sud wdt:P2044 ?altitude. }
7   OPTIONAL { ?Am_rique_du_Sud wdt:P18 ?image. }
8   OPTIONAL { ?Am_rique_du_Sud wdt:P625 ?coordonn_es_g_ographiques. }
9 }
```

Je choisis ensuite le type d'affichage des résultats. Ici, étant donné que je demande à voir des images, je choisis l'affichage « Image grid » puis je clique sur la flèche. Les résultats s'affichent alors :



7. Applications en SVT

Cf. article sur svt.ac-creteil.fr

8. Pour aller plus loin :

Introduction à Wikidata : <https://www.wikidata.org/wiki/Wikidata:Introduction/fr>

Les déclarations : <https://www.wikidata.org/wiki/Help:Statements/fr>

Service de requête SPARQL/Une introduction en douceur au service de requête de Wikidata :

https://www.wikidata.org/wiki/Wikidata:SPARQL_query_service/A_gentle_introduction_to_the_Wikidata_Query_Service/fr

Bien sûr, Wikidata est un projet collaboratif. Si vous souhaitez y contribuer, n'hésitez pas !