


## TP2 : Accès aux données Sémantiques via SPARQL

*G. Sanou et K. Todorov*

Dans un premier temps, télécharger et décompresser les fichiers se trouvant sur ce lien :  SPARQLTP . Une fois décompressé, vous aurez deux dossiers :

- Le premier dossier apache-jena-fuseki-4.4.0, contient, le fichier de lancement du serveur Jena fuseki, permettant de servir des graphes RDF ou des knowledge graphs.
- Le second dossier MABKG contient les informations du graphe de la Partie I du TP.

### Partie 0. Requêtes sur le graphe rdf construit dans le TP1

Dans un premier temps, **vous allez vous entraîner sur le jeu de données RDF que vous avez construit au TP1**. Pour cela, vous pouvez lister tous les triplets du graphe. Ensuite, affichez toutes les connaissances liées à Mozart. Qui est le père de Mozart ? son épouse ? Quelles sont ces compositions ?

- Vous pouvez lancer le serveur Fuseki sur votre jeu de données RDF avec : `apache-jena-fuseki-4.4.0/fuseki-server -file RDFFile /nameTripleStore` ou `java -jar apache-jena-fuseki-4.4.0/fuseki-server.jar -file RDFFile /nameTripleStore`

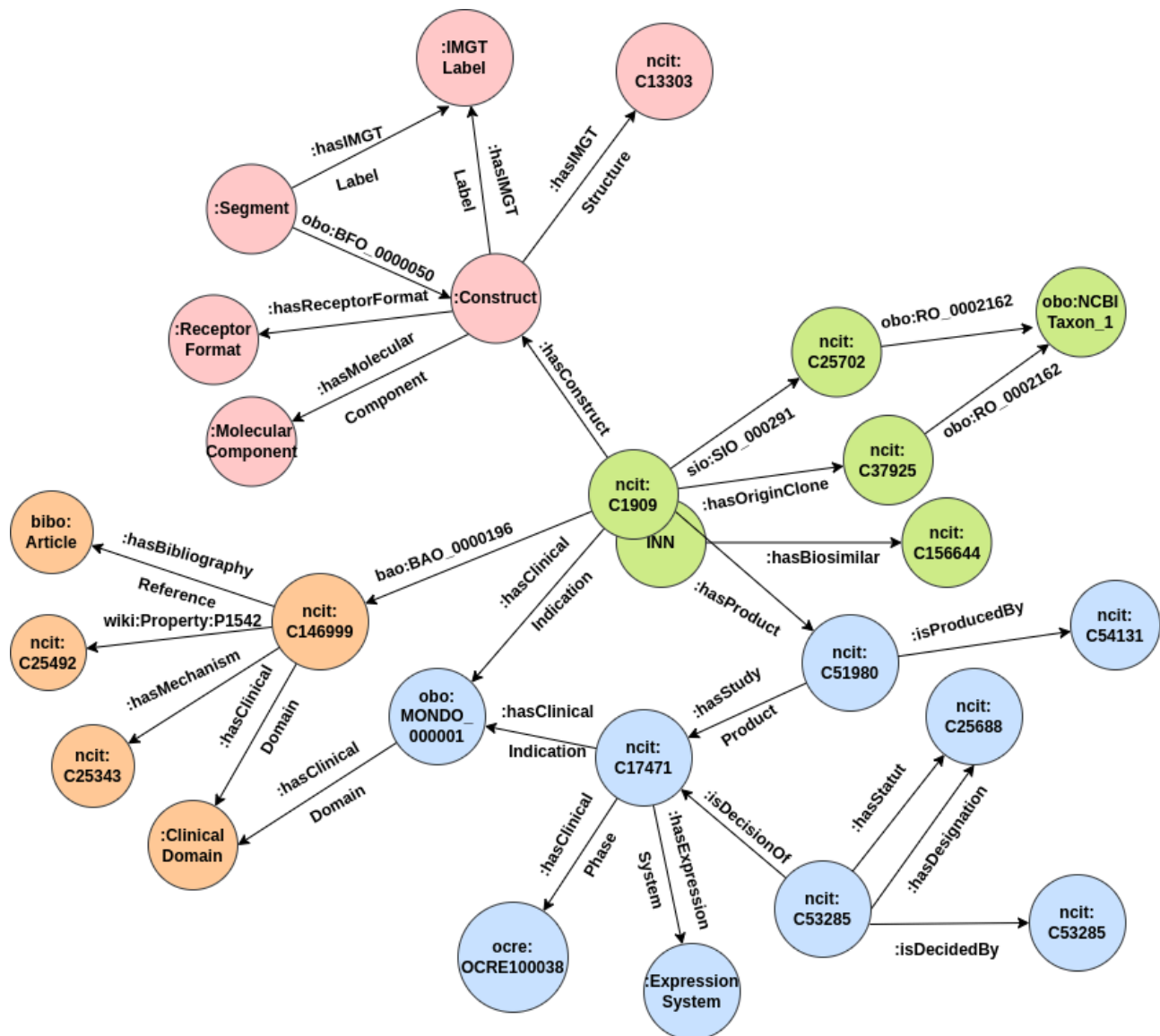
Avec l'interface dédiée de requête <http://localhost:3030/> , effectuer les différentes requêtes demandée.

## **Partie I. Requêtes avec l'interface dédiée : Jena Fuseki**

Considérons le modèle de données de IMGT/mAb-KG voir figure ci-dessous, décrivant les anticorps monoclonaux thérapeutiques (mAbs). Ce modèle donne une idée d'une partie de la représentation de connaissance en immunogénétique. En vert clair, nous avons les connaissances décrivant le mAbs, en rose celles décrivant la construction des mAbs, en bleu, celles qui décrivent le produit du mAb et en orange, les connaissances décrivant les mode d'action des mAbs. Pour résumer :

- Un mAb est une (rdf:type) substance pharmacologique (ncit:C1909) et/ou un INN, il cible (sio:SIO\_000291) une molécule donnée (ncit:C25702), a une construction et un produit associé, enfin, il peut être associé à un clone cellulaire (ncit:C37925). La cible ou target et le clone cellulaire appartiennent à un taxon (obo:NCBITaxon\_1). Enfin, le mAb peut avoir un biosimilaire (ncit:C156644), et avoir un mode d'action (ncit:C146999)
- Une construction fait référence à la composition de divers éléments utilisés pour composer le mAb. Il peut avoir un format de récepteur et un composant moléculaire. Chaque construction possède une étiquette IMGT et au moins un segment avec sa propre étiquette IMGT ; à l'inverse, le segment fait partie d'une construction (obo:BFO\_0000050). De plus, une construction peut être associée à une structure protéique (ncit:C13303), qui reflète la structure protéique utilisée dans IMGT-KG, reliant ainsi les deux KG.
- Un produit est le résultat de la production d'une entreprise (ncit:C54131) et fait invariablement l'objet d'au moins une évaluation ou étude (ncit:C17471). De telles évaluations de produits sont entreprises pour des indications cliniques spécifiques, principalement des maladies (obo:MONDO\_000001). L'état de développement d'un produit est communiqué par sa phase clinique (ocre:OCRE1000038) et peut être exprimé dans un système d'expression. Chaque produit d'étude peut faire l'objet d'une décision (ncit:C53285) qui aboutit à une désignation ou un statut spécifique (ncit:C25688) pris par une organisation (ncit:C53285).
- Le mécanisme d'action (MOA) d'un mAb englobe un domaine clinique, un mécanisme défini (ncit:C25343) via la relation (bao:BAO\_0000196), un ou plusieurs effets (ncit:C25492) et est étayé par des citations bibliographiques

(bibo:article). La relation entre le concept de MOA et ses effets est définie par la relation a un effet (wiki:Property:P1542).



**Model de donnée d'IMGT/mAb-KG: IMGT/mAb-KG: the knowledge graph for therapeutic monoclonal antibodies**

***Dans le schéma, lorsqu'on n'a pas de namespace sur les noms d'entités, on utilise implicitement le namespace imgt.***

obo: <http://purl.obolibrary.org/obo/>

sio: <http://semanticscience.org/resource/>

*bao*: <http://www.bioassayontology.org/bao#>



Avec l'interface dédiée de requête <http://localhost:3030/> , effectuer ces différentes requêtes :

1. Lister 20 mAbs qui sont des INN
2. Lister 20 mAbs ainsi que leur cibles
3. Lister 10 mAbs et les indications cliniques associées
4. Lister 20 mAbs qui sont associées à un mode d'action
5. Associer 30 mAbs avec leur produit d'étude
6. Rechercher des mAbs en utilisant leur nom INN, avec la propriété `imgt:inn_name`
  - a. Suivant la même procédure, rechercher les mAb avec les noms suivants: belimumab, oxelumab, relatlimab.
  - b. Pour les mêmes mAbs de la question a), associer-les à leur construction ainsi que les connaissances de la construction et aussi à leur cible
  - c. Pour les mêmes mAbs de la question a), associé leur au produit, et toutes les connaissances du produit jusqu'à phases cliniques.
  - d. Reproduire le graphe des instances de la figure pour le relatlimab
7. Afficher tous les mAbs indiqué (`imgt:hasClinicalIndication`) dans le traitement des tumeurs solides (`imgt:Solid_tumors`)
8. Lister les mAbs qui sont en phase 3 et M

## **Partie II. Requêtes avec l'API SPARQLWrapper**

Réaliser ces même requêtes en utilisant l'API SPARQLWrapper. Le point d'entrée SPARQL est <http://localhost:3030/mAbKG/sparql> . Sauvegarder chaque résultat dans un fichier csv/tsv.

Lien utile pour prendre en main SPARQLWrapper : [SPARQLWrapper](#)