

TP2 - Énoncé

POLYTECHNIQUE
MONTREAL

LE GÉNIE
EN PREMIÈRE CLASSE



Manipulation d'une base de connaissances en RDF

INF8410 - Ontologie et web sémantique

Automne 2021

Département de génie informatique
École Polytechnique de Montréal

Chargé de laboratoire :

Jean-Simon Marrocco (jean-simon.marrocco@polymtl.ca)

Date de remise : 7 novembre 2021 à 23h55

1 Objectif

Lors du premier laboratoire, vous avez eu l'occasion d'écrire une base de connaissances en RDF afin de représenter du contenu. L'objectif de ce second laboratoire sera d'écrire des requêtes SPARQL et des contraintes SHACL afin, dans un premier temps, de répondre à des questions sur une base de connaissances donnée, puis, d'exploiter votre base de connaissances et d'en valider le contenu.

Familiarisation avec TopBraid Composer

Dans un premier temps, vous allez vous familiariser avec l'outil que nous utiliserons pour la suite du cours : TopBraid Composer Free Edition, que vous pouvez [télécharger](#) gratuitement. Il s'agit d'un logiciel complet qui vous permet d'éditer des documents en RDF ou OWL, et qui vous permet de faire des requêtes SPARQL et de la validation avec SHACL. Il offre une documentation complète. Avant de poursuivre avec ce travail pratique, nous vous suggérons donc de suivre les tutoriels suivants :

1. [TopBraid Composer - Getting Started Guider](#) (lisez seulement jusqu'à la section 3.4, inclusivement)
2. [SHACL Tutorial: Getting Started](#)

2 Première partie : expérimentations avec une base de connaissances

2.1 Présentation de la base de connaissances

Pour cette première partie du deuxième laboratoire d'INF8410, vous aurez à rédiger des requêtes SPARQL et des contraintes SHACL sur la base de connaissances *academic.ttl*. Cette base de connaissances utilise le vocabulaire schématisé à la figure 1, permet de représenter les notions de cours, sessions, étudiants suivant un cours et enseignant le donnant (elle permet ainsi de répondre à la question 2.3.2 du premier laboratoire). Certaines erreurs ont été volontairement ajoutées lors du peuplement de cette base de connaissances et l'objectif des contraintes SHACL sera de les mettre en évidence.

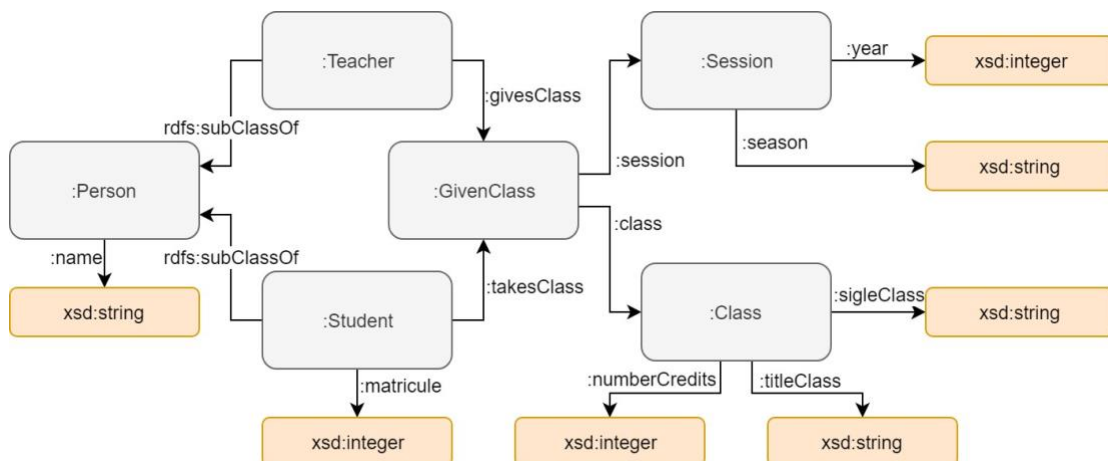


FIGURE 1 - Graphe du vocabulaire utilisé dans *academic.ttl*

2.2 Requêtes SPARQL

Afin de rédiger et tester vos requêtes, utilisez l'outil TopBraid Composer Free Edition.

Pour chacune des questions suivantes, rédigez une requête SPARQL permettant d'y répondre le plus précisément possible et exécutez celle-ci sur la base de connaissances *academic.ttl*.

- Quel est le nom de l'enseignant(e) ayant offert le cours de sigle "INF8410" à l'étudiant nommé "Jesse" et à quelle session ?
- Quel sont les noms des étudiants ayant suivis le cours intitulé "Champs électromagnétiques" avec l'étudiante nommée "Violet"? (Violet peut elle-même se retrouver dans cette liste).
- Quel est l'enseignant(e) ayant offert le plus de cours au total sur toutes les sessions et combien en a-t-il(elle) offert(e) ?
- L'étudiant nommé "Leonard" a-t-il suivi les 11 cours offerts au moins une fois chacun ?
- Quelle est la moyenne du nombre de crédits suivis par les étudiants pour chaque session ?

Pour chaque requête, ajoutez le code SPARQL ainsi que le résultat de la requête (copié-collé ou capture d'écran) à votre rapport.

2.3 Contraintes SHACL

Pour tester vos contraintes, le *SHACL playground* est disponible en ligne, mais pour une validation plus complète, utilisez TopBraid Composer. Pour chaque contrainte, ajoutez le code SHACL ainsi que le résultat de la contrainte (copié-collé ou capture d'écran) à votre rapport.

Pour chacun des requis suivants, rédigez une contrainte SHACL permettant de le valider le plus précisément possible et exécutez celle-ci sur l'ontologie *academic.ttl*.

- Une personne doit avoir exactement un nom qui est un *string*.
- Un cours doit avoir un nombre de crédits qui lui est attribué entre 1 et 3, inclusivement, et son sigle doit avoir une longueur exacte de 7 caractères.
- Un étudiant doit avoir suivi au moins une fois le cours INF3005.
- À chaque session, il ne peut pas y avoir plus d'un professeur qui enseigne le même cours.
- Le cours MTH1110 ne peut pas être donné à la même session que le cours PHS1102.

2.4 Règles SHACL

On désire créer la classe `BusyTeacher`, qui est une sous-classe de `Teacher`. Tout professeur appartient à cette classe s'il enseigne plus de cinq cours. Écrivez une règle en SHACL pour représenter cette classe.

3 Ontologie personnelle

Pour cette seconde partie du laboratoire, vous aurez à imaginer et appliquer sur votre base de connaissance du TP1:

- 5 requêtes SPARQL
- 5 contraintes SHACL
- 2 règles d'inférence SHACL

Vous êtes autorisés à modifier ou ajouter plus d'instances à votre base de connaissances. Les requêtes et contraintes doivent avoir une complexité similaire ou supérieures à celles de la première partie, c'est à dire utiliser les différents outils des langages à votre disposition (par exemple, pour SPARQL : `FILTER`, `GROUP BY`, `COUNT` etc.). Vos requêtes et contraintes doivent donc démontrer votre compréhension des deux langages et utiliser le plus possible les différentes possibilités permises par SHACL et SPARQL. Des précisions vous seront fournies ultérieurement par votre chargé de laboratoire.

4 Outils à votre disposition

Une série d'outils sont à votre disposition :

- Spécifications de SHACL : <https://www.w3.org/TR/shacl/>
- Spécifications de SPARQL 1.1: <https://www.w3.org/TR/sparql11-query/>
- Guide SPARQL : <https://www.w3.org/TR/rdf-sparql-query/>
- Guide de syntaxe de Turtle : <https://www.w3.org/TR/turtle/>
- Valideur SHACL : <https://shacl-play.sparna.fr/play/>
- SHACL Playground : <http://shacl.org/playground/>
- TopBraid Composer Free Edition : <https://www.topquadrant.com/topbraid-composer-install/>

5 Remise

La remise se fait sur Moodle au plus tard le 7 novembre à 23h55. Une pénalité de 10% de la note finale sera appliquée pour chaque jour de retard. Un seul membre par équipe aura à effectuer la remise. Le format de remise est une archive au format INF8410 TP2 [matricule1] [matricule2].zip contenant votre rapport en format word ou pdf .

6 Barème d'évaluation

L'évaluation portera sur les points suivants :

- 2.3 - Base de connaissances commune - SPARQL (20%) (4% par requête)
- 2.3 - Base de connaissances commune - SHACL (20%) (4% par requête)
- 2.4 - Base de connaissances commune - Règles SHACL (4%)
- 3 - Base de connaissances personnelle - SPARQL (25%) (5% par requête)
- 3 - Base de connaissances personnelle - SHACL (25%) (5% par requête)
- 3 - Base de connaissances personnelle - Règles SHACL (6%)

Des points peuvent être déduits si vos requêtes, contraintes ou règles sont trop basiques.