
TP1 - Énoncé

POLYTECHNIQUE
MONTREAL

LE GÉNIE
EN PREMIÈRE CLASSE



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 : 3 octobre 2021 23h55

Les laboratoires d'INF8410 se font par groupe de 2 ou 3. Envoyez le plus tôt possible par courriel le matricule des membres de votre équipe au chargé de laboratoire.

1 Objectif

Les premières séances du cours d'ontologie et Web sémantique INF8410 vous ont permis de découvrir RDF ainsi que ses différentes formes de sérialisation (N-Triples, RDF/XML et Turtle). L'objectif de ce premier laboratoire est de mettre en pratique ces notions par la création de représentations formelles en RDF.

2 Partie 1

2.1 Analyse de données RDF

Soit la description suivante donnée en Turtle :

```
@prefix : <www.exemple.org/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix schema: <https://schema.org/> .

:B1 a schema:Book ;
    :langue "Anglais" ;
    schema:author (:A1 [ schema:name "Michel Gagnon" ] ) ;
    :derivedFrom [ a schema:Book ;
                  :langue "Francais" ;
                  :title "Les joies du web sémantique" ] ;
    schema:translator :T1 .
:A1 a schema:Person ;
    :occupation "Politician", "writer" ;
    schema:name "Michelle Tremblay" .
[ :attributedTo :B1 ;
  a schema:Rating ;
  schema:ratingValue "9" ] .
```

Décrivez, en vos propres mots et le plus précisément possible, ce qui est décrit dans ce graphe, et discutez de la qualité de cette représentation. Proposez, s'il y a lieu, une version améliorée. Pour vous aider dans votre discussion, prenez en compte les considérations suivantes (notez qu'il peut y en avoir d'autres qui ne sont pas listées ici):

- Fait-on une bonne utilisation des vocabulaires disponibles?
- Y a-t-il une meilleure manière de structurer les informations en RDF?
- Y a-t-il des informations pertinentes qui sont manquantes?
- Sera-t-il facile de maintenir et mettre à jour ces données?
- Quelles sont les conséquences sur le genre de requête qu'on peut faire?

2.2 RDF vers graphe

Fournissez le graphe RDF de la description de la question 2.1. N'utilisez pas un outil automatique de conversion. dessinez le graphe manuellement.

2.3 Description textuelle vers RDF

Soit la situation suivante :

Polytechnique Montréal (anciennement École Polytechnique de Montréal) est un établissement d'enseignement supérieur d'ingénierie affilié à l'Université de Montréal, situé à Montréal (Québec) et fondé en 1873. En plus de ses programmes de baccalauréat en génie, Polytechnique Montréal offre une formation aux deuxième et troisième cycles.

Représentez cette situation en RDF, et fournissez des sérialisations dans les trois notations suivantes: Turtle, RDF/XML et JSON-LD. Assurez-vous de bien structurer votre représentation de manière à ce qu'elle soit facile à comprendre et à maintenir et, surtout, facile à lier à d'autres données provenant d'autres sources du web sémantique.

3 Deuxième partie

L'objectif de cette seconde partie est de présenter la description d'un thème de votre choix, parmi la liste qui vous sera soumise par votre chargé de laboratoire. Le choix de la forme de sérialisation (N-Triples, RDF/XML ou Turtle) est laissé à votre discrétion (il est toutefois conseillé d'utiliser Turtle). Un simple éditeur de texte est suffisant pour l'écriture de votre base de connaissances.

Votre représentation devra inclure les éléments suivants:

- Les définitions de préfixe (il faut les utiliser afin de faciliter la compréhension).
- Les classes nécessaires à vos instances. Vous devez le plus possible réutiliser des classes déjà existantes dans des vocabulaires externes. Limitez-vous toutefois aux vocabulaires reconnus, comme ceux dont les liens se trouvent dans la section « Outils à votre disposition ».
- Les déclarations de propriétés entre objets ou vers des littéraux (ici aussi vous devez maximiser l'utilisation de vocabulaires externes).
- Les déclarations de domaine des propriétés.
- Les déclarations de portée (range) des propriétés.
- Les instances.
- Les étiquettes (rdfs:label) des classes, instances et propriétés.
- Les commentaires (rdfs:comment) des classes, instances et propriétés.
- Les relations entre instances.
- Les littéraux.

Une partie de l'évaluation portera sur la couverture de votre ontologie. Il est donc nécessaire que celle-ci contienne suffisamment de classes (au moins une dizaine) et de propriétés entre ces classes pour présenter un intérêt. Il est également demandé de peupler votre ontologie de quelques instances (2 ou 3 par classe environ) afin de démontrer des cas d'utilisation. Il vous sera également nécessaire de faire des choix conceptuels lors de l'élaboration de votre base de connaissances (tel que décider ce qui doit être une classe ou une instance). Il vous est demandé d'inclure un court rapport décrivant ces choix conceptuels ainsi que d'ajouter quelques cas d'utilisation textuels afin de prouver l'intérêt de cette base de connaissances (pour le domaine du cinéma, on pourrait, par exemple, avoir le cas d'utilisation: « nous pouvons facilement identifier les films réalisés par Quentin Tarantino »). Il vous est également demandé d'ajouter à votre rapport au moins un graphe RDF, plusieurs si nécessaire, qui illustrent votre base de connaissances.

4 Outils à votre disposition

Une série d'outils sont à votre disposition afin de faciliter l'écriture et la validation de votre rdf :

- Outil pour dessin de graphe RDF : <https://www.draw.io/>
- Guide de syntaxe de N-Triples : <https://www.w3.org/TR/n-triples/>
- Guide de syntaxe de Turtle : <https://www.w3.org/TR/turtle/>
- Guide de syntaxe de RDF/XML : <https://www.w3.org/TR/rdf-syntax-grammar/>
- Convertisseur de syntaxes : <http://rdf-translator.appspot.com/>
- Valideur de RDF de W3C : <https://www.w3.org/RDF/Validator/>
- Valideur de RDF : <http://www.macs.hw.ac.uk/~ajg33/validata/>
- Visualisateur de RDF : <http://rhizomik.net/html/redefer/rdf2svg-form/>
- Vocabulaire de FOAF : <http://xmlns.com/foaf/spec/>
- Vocabulaire de schema : <https://schema.org/docs/schemas.html>
- Vocabulaire de dublin core : <http://dublincore.org/documents/dcmi-terms/>
- Vocabulaire de event : <http://motools.sourceforge.net/event/event.html>
- Vocabulaire de time : <https://www.w3.org/TR/owl-time/>
- Vocabulaire geoname : <https://www.geonames.org/ontology/documentation.html>
- Vocabulaire dbpedia : <http://dbpedia.org/>

5 Remise

La remise est à effectuer par Moodle au plus tard le 3 octobre 2021 à 23h55. La politique appliquée pour les retards est de 10% de la note finale en moins par jour de retard. (Ainsi un étudiant ayant théoriquement 92% mais rendant son travail avec 2 jours de retard aura 72%.)

Un seul membre par équipe aura à effectuer la remise. Le format de remise est une archive au format INF8410 TP1 [matricule1] [matricule2].zip contenant un rapport avec vos réponses pour chaque partie (un fichier LaTeX, Word, ou texte avec des images jointes pour le(s) graphe(s)). D'autres détails concernant la remise vous seront fournis ultérieurement par le chargé de laboratoire.

6 Évaluation

Le barème pour l'évaluation est le suivant:

- 2.1 - Analyse de données RDF (15%)
- 2.2 - RDF vers graphe (5%)
- 2.3 - Description textuelle vers RDF (10%)
- 3 - Couverture du thème choisi (20%)
- 3 - Qualité des choix conceptuels et graphe (20%)
- 3 - Qualité des cas d'utilisation (10%)
- 3 - Bonne documentation (domaine et portée des propriétés, étiquettes et commentaires) (10%)
- Qualité générale du travail remis (10%)