

Projet SBC : visualisation des Linked Open Data

2017 – 2018

Travail demandé

La visualisation des *Linked Open Data* est essentielle pour permettre à un utilisateur d’appréhender la structure et les différents liens dans un jeu de données qui lui est inconnu. Une visualisation proposée [1] est construite en détectant les éléments de la *TBox* les plus fréquemment instanciés (classes et prédicats entre les instances des classes) à partir de requêtes SPARQL sur le SPARQL *endpoint* d’un jeu de données.

Néanmoins, cette visualisation n’affiche que les éléments de la *TBox* les plus fréquents. Il pourrait être intéressant d’afficher toute la *TBox* puis de mettre cette *TBox* en perspective avec ce qui existe fréquemment dans la *ABox*. Par exemple, supposons l’existence d’une propriété **partOf** dont le **domain** est défini comme étant la classe **GeneVariant** et le **range** comme étant la classe **Gene**. Il serait utile d’afficher cette définition de la *TBox* puis d’afficher (d’une manière élégante) les véritables classes fréquemment appliquées en **domain** et en **range** de cette propriété. Il ne s’agit bien évidemment pas de l’unique possibilité. Il serait intéressant d’exploiter l’ensemble des connaissances proposées par les ontologies : hiérarchies de classes, de propriétés, mappings **owl:sameAs**, liens de *cross-references*... L’outil présenté propose également un unique mode d’exploration du jeu de données considéré : à partir de la fréquence d’instanciation des éléments de la *TBox*. On pourrait également réfléchir à d’autres approches : à partir d’une instance particulière, à partir d’un groupe d’instances, à partir d’une propriété, d’un groupe de propriétés, en interaction avec les souhaits d’expansion de l’utilisateur...

Il vous est donc demandé dans ce projet de :

- Proposer diverses stratégies affichant la *TBox* d’un jeu de données *Linked Open Data* en exploitant conjointement la structure de la *TBox* (hiérarchies de classes et de propriétés, mappings **owl:sameAs**, *cross-references*...) et la fréquence des instances dans la *ABox*
- Proposer diverses stratégies d’exploration du jeu de données *Linked Open Data*

Vous pourrez expérimenter vos propositions sur les jeux de données cités dans [1]. Vous pourrez également construire votre propre jeu de données de test (cf. TP2, exercice 3).

Évaluation

Votre travail sera évalué lors d’une **soutenance** :

- Date : 15/03/2018 de 8h à 12h
- Salle : E1.16 – TELECOM Nancy
- Membres du jury : Pierre Monnin et Malika Smaïl-Tabbone
- Durée d’une soutenance : 15 minutes de présentation et 5 minutes de questions

Les critères d’évaluation sont les suivants :

1. La compréhension et la reformulation du problème
2. La proposition argumentée de plusieurs stratégies améliorant la visualisation d’un jeu de données en exploitant conjointement *TBox* et *ABox*
3. L’expérimentation de ces stratégies
4. La bonne répartition du travail
5. La qualité de la soutenance (support, présentation, réponses aux questions)

Vous déposerez sur Arche avant le 14/03/2018 23:59 une archive contenant le code source de l’outil développé et votre support de présentation.

Références

- [1] Steffen Lohmann, Vincent Link, Eduard Marbach, and Stefan Negru. Extraction and visualization of *tbox* information from SPARQL endpoints. In *Proceedings of the 20th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2016)*, volume 10024 of *LNAI*, pages 713–728. Springer, 2016.