

HAI824 - Traitement Sémantique des Données

TP1 : Modélisation Sémantique des données en RDF

K. Todorov et G. Sanou — 2025

Partie I. Modélisation d'un domaine à partir d'un court texte

Considérons le texte suivant:

Wolfgang Amadeus Mozart, né à Salzbourg le 27 janvier 1756, mort à Vienne, la capitale autrichienne, le 5 décembre 1791, est un compositeur. Il est fils de Léopold Mozart et était marié à Constance Weber. Parmi ses œuvres les plus célèbres, on trouve l'opéra "La flûte enchantée" et la 41 Symphonie en ut majeur, dite "Jupiter". Cette symphonie est composée de 4 parties : 1. Allegro Vivace, 2. Andante Cantabile, 3. Menuetto, 4. Molto Allegro. Elle a été enregistrée par l'orchestre symphonique de Londres sous la direction de Claudio Abbado en 1980.

- 1) Extraire manuellement les connaissances exprimées dans ce texte : penser aux entités (personnes, œuvres, lieux, etc.), aux relations entre ces entités (« fils », « compositeur », « capitale », etc.) et à leur caractéristiques (l'œuvre « Jupiter »; est dans le genre "symphonie", etc.). Présenter cela sous la forme d'une table.
- 2) Dessiner un graphe RDF qui correspond aux informations dans le texte concernant l'œuvre « Jupiter ». Autrement dit, créer un modèle d'une œuvre et l'instancier par les données sur « Jupiter ». Vous pouvez utiliser <https://app.diagrams.net/> pour le dessin.
- 3) Créer manuellement un fichier qui traduit le graphe que vous avez dessiné en RDF ;
- 4) Vérifier si le fichier est bon par un validateur en-ligne et visualiser le graphe (à l'aide par exemple du validateur de W3C <http://www.w3.org/RDF/Validator/>).
- 5) (Bonus) Utiliser l'outil DBpedia Spotlight ou l'outil TagMe pour extraire et ancrer automatiquement à DBpedia des entités à partir de votre texte et comparer le résultat à votre extraction manuelle faite en point 1.
- 6) Visualisation du graphe avec pyViz: [Introduction — pyvis 0.1.3.1 documentation](#)

Remarques :

- Ré-utiliser des vocabulaires existants pour nommer vos propriétés et vos classes. Utiliser le service LOV pour retrouver et réutiliser des vocabulaires pour décrire vos données (classes, propriétés) : <http://lov.okfn.org/dataset/lov/>. Quel critère de choix de vocabulaires ?
- Typier les instances (en utilisant des classes existantes (LOV)) quand cela est nécessaire ;
- Attention à l'identification des ressources (et des propriétés et classes) par des URIs !

Partie II. Modélisation et intégration avec l'API python RDFlib

- Exploiter l'API RDFlib afin de créer le graphe RDF
- Tutoriels : <https://rdflib.readthedocs.io/en/stable/gettingstarted.html>