Building & Mining Knowledge Graphs

(KEN 4256)

Assignment 1 part 1

Knowledge graph construction with RDFLIB

Due date: 13 February 2023 (upload on student portal before **18:00 CET**)

Description

In this assignment, you will use RDFLIB (python package) to create a knowledge graph. Source datasets will be music data (link) which contains information of artists and albums. After creating the knowledge graph, the knowledge graph will be used to retrieve the answers of following competency questions:
- Which albums/artists have achieved a rating of more than 4.5 from MTV?
- What is the number of album sales for all artists from the Netherlands?
- How many albums with genre "country" were released in the year of 2012? Which one was the best-selling?

Data sources (link or attached):
- **Artists** dataset, provided in CSV format.
- **Albums** dataset, provided in CSV format.

Tasks

Task 1: Develop a vocabulary for your knowledge graph (conceptual design and representation of the types and properties**).**

1.1. Make a list of types and properties that are needed in your knowledge graph based on the competency questions.

1.2 Map the classes and properties in your knowledge graph to the existing vocabularies such as schema.org or LOD Vocabularies. Or, you can use MusicOntology.

1.3 Write down a RDFS representation of your knowledge graph vocabulary in Turtle syntax.

Task 2: Programmatically using library create a knowledge graph from structured data
Identify individuals (with their types) between them from **Artists & Albums** dataset, and generate RDF triples based on your pre-defined vocabulary from Task 1. Assign unique URI to each individual. You may produce your RDF file using different conversion tools.

Deliverables

One student per group will deliver a written technical report (max 5 pages), which contains the following information:
- A description, in your own words, of the **methodology used** and the **motivation for any choices made** in the following steps of the assignment:
  - The conversion of the datasets to RDF (including a description in your words of the conversion script)
- A diagram showing your conceptual design

Please submit the following files separately (also on Student Portal):
- Your RDF(S) files as N-triples/Turtle format (also accept any kind of RDF format)

Reading Materials

- RDFLib documentation, https://rdflib.readthedocs.io

- TRANSFORMATION FROM SEMANTIC DATA MODEL TO RDF
  https://media.neliti.com/media/publications/144976-EN-transformation-from-semantic-data-model.pdf
- Data modelling with RDF: a tutorial
  https://sinaahmadi.github.io/posts/rdf-data-modelling-a-tutorial.html
- Build a medium size KG from a CSV dataset https://derwen.ai/docs/kgl/ex2_0/
- Lecture 2 slides (Generate RDF graph from structured data)

Questions and comments:

Prof. Dr. Christopher Brewster: christopher.brewster@maastrichtuniversity.nl

Dr. Remzi Çelebi: remzi.celebi@maastrichtuniversity.nl