Building & Mining Knowledge Graphs

(KEN 4256)

Assignment 1 Part A:

Knowledge graph construction and integration

Due date: 14 February 2022 (upload on student portal before **midnight**)

Description

In this assignment, you will create a knowledge graph to help accommodate guest preferences for a dinner party. Guests follow different diets (vegetarian, vegan etc.), so they would like to know in advance which meals are suitable for their diet. To meet guests' preferences, the knowledge graph should contain information where the following competency questions can be answered:

- How much vitamin C is in "Caprese Stuffed Portobello Mushrooms"?
- Which Mediterranean meals do not include meat?
- Which vegetarian meals are rich in fiber?

In this part of the assignment, you will prepare the knowledge graph by creating and mapping a knowledge graph vocabulary to two data sources (listed below). In the second part of the assignment (to be released later), you will query this knowledge graph with SPARQL to answer the competency questions.

Data sources (attached as zip):
- **Food** dataset, provided in CSV format, contains food nutrition information about various ingredients.
- **The Mediterranean dinner** dataset, provided in text format, contains various recipes from Mediterranean cuisine.

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 essential in your knowledge graph based on the competency questions.

1.2 Map the classes and properties in your knowledge graph to the schema.org or [WhatToMake](#) types and relations.

1.3 Write down a RDF(S) representation of your knowledge graph vocabulary in Turtle syntax.

Task 2: Manually create a knowledge graph from text
Identify individuals (with their types) and relations between them from **The Mediterranean dinner** dataset, and generate RDF triples based on your pre-defined vocabulary from Task 1. You may produce your RDF file by hand or programmatically.

Task 3: Create a knowledge graph from structured source
Write a formal mapping expression (in RML or YARRRML) to convert CSV data to RDF using the KG vocabulary. Generate the RDF from the RML/YARRML definitions with an appropriate RML tool.

Task 4: Integration of two knowledge graphs
Define and execute a strategy to link the Food nutrition and Mediterranean cuisine datasets via ingredients.

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 instructions expressed in the YARRRML/RML mapping file)

- - The linking of the two datasets using LIMES or other methods
  - The number of RDF triples generated by applying the mapping file(s).
  - The type and number of entities linked across the datasets.

Please submit the following files separately (also on Student Portal):
- All YARRRML/RML mapping files used (.yml or .ttl file)
- LIMES configuration file
- Your RDF files as N-triples format with .nt file extension

Questions and comments:

Prof. Michel Dumontier: michel.dumontier@maastrichtuniversity.nl

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