KG Improvement: Embeddings

Credit

This assignment is identical to the one in the City uni course.

Ontology Embeddings (Task Vector)

This task consists in creating embeddings capturing the rich semantics of the created ontology and generated data. We will use the tool OWL2Vec*. These embeddings can be used in a subsequent Machine Learning model that requires as input the encoding of the ontology entities. In this coursework we are computing clusters of the ontology entities.

Subtask Vector.1 Run OWL2Vec* with the created ontology and generated data. Test three different configurations. Save the generated vectors in both binary and textual format (20%).

Subtask Vector.2 Select 5 pairs of entities (for any of the tested configurations) and discuss the similarity of their vectors (e.g., compare the vectors of the concept pizza:Margherita and the word “pizza”). (20%).

Subtask Vector.3 Compute clusters (e.g., using K-means) for the embeddings of the ontology concepts (i.e., the URI embeddings). Test the algorithm with different number of clusters, visualize the clusters and discuss the results (50%).

Subtask Vector.4 Correctness and documentation of the codes (10%).