**Assignment 1**

**Simon Jasansky**

**Gergely Parádi**

**Ákos Engelmann**

**İpek Çakın**

A knowledge graph is a type of graph data structure that represents entities and their relationships as build with types and properties. It provides a visual representation of the relationships between entities, enabling users to understand the connections between different pieces of information. They are used at different application like information retrieval, recommendation systems, semantic search, and data management, they are very helpful for dealing with large amounts of structured and unstructured data, as they provide a scalable and flexible way to represent and manage complex data. Therefore, building knowledge graphs is important because it provides a flexible and scalable way to represent and manage complex data, leading to improved decision-making, data management, and discovery.

RDF (Resource Description Framework) is a standard data model for representing, exchanging data on the web, and also providing the data model for implementing knowledge graphs. Herewith, the aim of project is the understanding basic structure of knowledge graphs and clarify types and properties with using RDFs. For this purpose, we first defined types and properties, assigned classes them from existing vocabularies, and then prepared RDFs suitable for the Turtle syntax.

RDFs are based on the idea of a triples which is consist of a subject, predicate and object. Each triple represents a single relationship between two resources, where the subject is the resource being described, the predicate is the relationship between the subject and object, and the object is the resource that the subject is related to.

In our project, we tried to built triple to graph. At first step, we crate a graph and interated over all rows and added triples to the graph. Then we bound prefix to namespaces, and applied this process to both the artist and album types.

We converted the dataset into an KG, however with the countries we ran into the problem with using real URI-s. Since the exact countynames are not always in the same vocabulary, we instead specified them as string literals. We also used ID-s for the graph to overcome the problem of potential duplictes and spaces, but we added both the artist and the album names with the "names" property.

**Tasks**

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

* 1. **Make a list of types and properties.**

Types:

* Artist
* Album
* Country
* Genre
* Literal

Properties:

* country from
* is genre
* amount sold
* rating
* created by
* released date

First, we determined the "types" and "properties" to create the Knowledge Graph. We used the information in competency questions for these.

To explain in more detail, in first question "Which albums/artists have achieved a rating of more than 4.5 from MTV?", albums and artists are types, while ranting (than 4.5) is property. In second question, album, artist and country are types because we need to find sales number of “album” in Netherlands and meanwhile country from and amount sold are properties. And for the third question, country, album and genre are types, on the other hand released date, created by and (is) genre are properties.

Moreover, to explain the literal part separately, we used this type for constant parts in questions, for example "4.5 from MTV" in the first question, "from the Netherlands" in the second question, and "2012" in the third question.

1.2 **Map the classes and properties in your knowledge graph to the existing vocabularies.**

Then we matched the types and properties selected for the chart with the existing vocabularies.

Types:

* Genre: <http://purl.org/ontology/mo/Genre>
* Artist: <http://purl.org/ontology/mo/MusicArtist>
* Album: <https://schema.org/MusicAlbum>
* Country: <https://schema.org/Country>

Properties:

* born in: <https://schema.org/birthPlace>
* is genre: <https://schema.org/genre>
* released in: <https://schema.org/releaseDate>
* number of sales: <https://dbpedia.org/ontology/numberSold>
* created by: <https://dbpedia.org/property/createdBy>
* MTV rating: <https://schema.org/starRating>

Then we matched the types and properties selected for the chart with the existing vocabularies.

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

@prefix purl: <http://purl.org/ontology/mo/>

@prefix db: <https://dbpedia.org/>

@prefix sch: <https://schema.org/>

* purl:ontology/mo/MusicArtist sch:birthPlace sch:Country.
* sch:MusicAlbum db:property/createdBy purl:MusicArtist ;
* sch:releaseDate "..."^^xsd:date ;
* db:ontology/numberSold "..."^^xsd:integer ;
* sch:genre purl:Genre ;
* sch:starRating "..."^^xsd:double .

**Task 2:** Programmatically using library create a knowledge graph from structured data.