

Integrációs és ellenőrzési technikák

2. gyakorlat

Adat integráció RDF adatbázisok között

Az alábbi feladatokat a virtuális gépre telepített RDF4J adatbázis workbench felületén oldja meg:

<http://localhost:8080/rdf4j-workbench/>

Az OpenRDF Workbench felület angol nyelvű kézikönyvét itt találja:

http://docs.rdf4j.org/server-workbench-console/#_rdf4j_workbench

A SPARQL lekérdező nyelv specifikációja:

<https://www.w3.org/TR/sparql11-query/>

A gyakorlat során több adatbázis együttes lekérdezését fogjuk gyakorolni, ahol az adatbázisok külső kulccsal vannak kapcsolva egymáshoz. Az első gyakorlaton is használt Szépművészeti Múzeum műkincseit tartalmazó adatbázist kapcsoljuk a DBpedia adatbázishoz.

A felhasznált adatbázisok:

- Open linked data from the Museum of Fine Arts Budapest
A Szépművészeti Múzeumban tárolt műkincsekről tartalmaz információt.
<http://data.szepmuveszeti.hu/>
 - SPARQL endpoint: <http://data.szepmuveszeti.hu/sparql>
 - Példa rekord: http://data.szepmuveszeti.hu/doc/collections/museum/E39_Actor/267eb328-1384-38c9-a197-cc123d6e1cdc
- DBpedia
<http://wiki.dbpedia.org/>
 - SPARQL endpoint: <https://dbpedia.org/sparql>
 - Példa rekord: <http://dbpedia.org/page/Rembrandt>

0) Töltse be az OpenRDF Workbench-be a minta adatbázist.

A gyakorlat során használt adatbázist az alábbi címről töltse le:

<http://home.mit.bme.hu/~fandrew/integralt/szepmuveszeti-dbpedia.2017-02-24.trig>

Az adatbázis file típusa TriG!

Nyissa meg böngészőben az RDF4J Workbench felületét:

<http://localhost:8080/rdf4j-workbench/>

A bal oldali menüben válassza a "Repositories / New repository" menüpontot, majd az alábbi beállításokkal hozhatja létre az adatbázist:

- Type: In Memory Store
(A gyakorlat során az adatbázist a teljesítmény kedvéért csak a memóriában tároljuk. Ha meg akarjuk őrizni az adatokat, válasszuk a Native Java Store lehetőséget.)
- ID: szepmuveszeti
- Next
- Create

A fejlécben az "Repository:" után látja mindig az aktuálisan kiválasztott adatbázis azonosítóját. Ha több adatbázisa is van, váltani a "Repositories" menüpontban tud.

Az adatok betöltéséhez a "Modify / Add" menüpontban:

- Kapcsolja ki a "use base URI as context identifier" checkbox-ot.
- RDF Data File: Choose File
Válassza ki a kívánt RDF file-t.
- Data format: TriG
- Upload

Ha később újra akarja kezdeni a munkát, a "Repositories / Delete repository" menüpontban törölheti az adatbázist, és újra létrehozhatja.

1) A Szépművészeti Múzeum adatbázisa mellé bekerült a DBpedia adatbázis egy része, mely művészekről tartalmaz relációkat. Nézze meg, milyen információk szerepelnek Rembrandtról, azonosítója "<http://dbpedia.org/resource/Rembrandt>".

Az új adatok a "<dbpedia:actors>" kontextusba kerültek.

Hasonlítsa össze az adatokat a <https://en.wikipedia.org/wiki/Rembrandt> wikipedia oldallal.

2) A szepmuveszeti adatbázisban az "owl:sameAs" relációkon keresztül külső adatbázisok vannak az entitásokhoz kapcsolva. Hány hivatkozás van a DBpedia adatbázisra?

A külső hivatkozások típusait meg tudja különböztetni egy FILTER szűréssel, melyben az 1. gyakorlat regex() függvényét az entitás azonosítójára használja. Az IRI-ből az str() függvénnyel kaphat string-et:

```
FILTER(regex(str(?id), "dbpedia"))
```

3) Kérdezze le, melyik külső adatbázisra pontosan hány "owl:sameAs" hivatkozás van az adatbázisban. A külső adatbázisokat az azonosító namespace szerint különböztesse meg.

A külső azonosítókból le tudja vágni az utolsó tagot, amivel namespace-eket kap, az alábbi változó értékadással:

```
BIND(replace(str(?id), "/[^/]*$", "") AS ?db)
```

Itt a replace() függvény a regex()-hez hasonló, de reguláris kifejezést cserél le. A BIND segítségével a ?db változóhoz rendeljük a kifejezés értékét.

Végül a hivatkozás-típusok összeszámolásához használhatja a "COUNT" és "GROUP BY" kulcsszavakat, az alábbi minta szerint:

```
SELECT ?type (COUNT(?type) as ?count)
WHERE {
  ?s rdf:type ?type
} GROUP BY ?type
```

4) Kérdezze le a DBpediából azokat a művészeket, akik a 16. század első felében születtek.

A `dbo:birthDate` tulajdonsággal megadott születési évet tekintse `xsd:date` típusúnak, és az alábbi módon szűrhet:

```
FILTER("2000-01-01"^^xsd:date <= ?date)
```

5) Írjon lekérdezést, ami a DBpedia adatbázisban “<<http://dbpedia.org/resource/Rembrandt>>” azonosítójú művész alkotásainak címeit megtalálja a szepmuveszeti adatbázisban.

A szepmuveszeti és a külső adatbázisok kulcsai `owl:sameAs` tulajdonsággal vannak összekapcsolva, például:

```
<.../E39_Actor/267eb328-...> <http://www.w3.org/2002/07/owl#sameAs> <http://dbpedia.org/resource/Rembrandt>
```

Az alkotót az alkotási folyamathoz a `ecrm:P11_had_participant`, a folyamatot a képhez a `ecrm:P12i_was_present_at` tulajdonság kapcsol.

Egy kép címét az alábbi módon tudja lekérdezni:

```
?thing rdfs:label ?title
```

6) Kérdezze le azoknak az alkotóknak a neveit, akikre Rembrandt hatással volt, és a Szépművészeti Múzeum adatbázisában szerepelnek. Használja a DBpedia “`dbo:influencedBy`” tulajdonságát.

7) Kérdezze le az összes kép címét, melyeket olyan alkotók készítettek, akikre Rembrandt hatással volt. (3 kép)

Házi feladat

Kérdezze le azokat a tanítvány-tanító párokat névvel, akik szerepelnek a szepmuveszeti adatbázisban. A DBpedia “`dbo:training`” tulajdonságát használja.