

2. Projektni zadatak iz predmeta

Laboratorij programskog inženjerstva i informacijskih sustava 1

u okviru predmeta Napredni modeli i baze podataka

Tema: Semantički web

akademska godina 2016/2017

Priprema

1. Proučiti slajdove s predavanja i vlastite zabilješke

2. U web pregledniku otvorite url:

<http://www.linkedmdb.org/snorg1/>

Dočekat će vas sučelje za interaktivno postavljanje upita nad grafom podataka iz **Linked movie database(LMDB)** izvora.

Alternativno, možete koristiti Fuseki server na svom virtualnom stroju.

Fuseki se nalazi u `/usr/fuseki-server/` i nije stalno pokrenut, nego se po potrebi može ručno pokrenuti iz konzole, sa:

```
cd /usr/fuseki-server/  
./fuseki-server --loc=LMDB/ /LMDB
```

Ili, ako želimo da nam Fuseki prihvaća i SPARQL UPDATE naredbe:

```
./fuseki-server --update --loc=LMDB/ /LMDB
```

Nakon toga je web sučelje dostupno na http://IP_RAČUNALA:3030/

I, kao treća opcija, upite možete postavljati i nad lokalnom kopijom LMDB grafa:

a) Sa stranica predmeta preuzmite datoteku (zbog veličine razrezanu u dva dijela):

```
jena-fuseki+LMDBdump.7z
```

b) Raspakirajte 7zip datoteku i iz komandne linije pokrenite Fuseki server:

```
fuseki-server.bat --loc=LMDB /ds
```

(proučite i ostale parametre pri pokretanju servera, npr. `--update`)

c) U web pregledniku otvorite url:

<http://localhost:3030>

na poveznici "control panel" potrebno je izabrati jedini ponuđeni set podataka (dataset) - `/ds`. Nakon toga otvara se interaktivno sučelje gdje je moguće postavljati upite.

3. Isprobajte nekoliko jednostavnih upita, npr.

a. **ukupni broj trojki u grafu**

```
SELECT (COUNT(*) AS ?ukupnoTrojki)  
WHERE {  
    ?s ?p ?o  
}
```

b. **popis svih klasa**

```
SELECT DISTINCT ?class  
WHERE {  
    [] a ?class  
}  
ORDER BY ?class
```

- c. **(samo na virtualnom računalu ili na lokalnoj kopiji!)** postoji li u grafu film koji je režirao **Quentin Tarantino**?

```
PREFIX movie: http://data.linkedmdb.org/resource/movie/
ASK {
    ?movie movie:director ?director .
    ?director movie:director_name "Quentin Tarantino"
}
```

- d. **U koliko drama i akcijskih filmova je glumio Will Smith?**

```
PREFIX movie: <http://data.linkedmdb.org/resource/movie/>
SELECT (COUNT(DISTINCT ?movie) AS ?brojFilmova)
WHERE {
    ?movie movie:actor ?actor .
    ?actor movie:actor_name "Will Smith" .
    ?movie movie:genre ?genre .
    ?genre movie:film_genre_name ?gname .
    VALUES ?gname {"Action" "Drama"}
}
```

- e. **U kojim filmovima, snimljenim nakon 2000. godine, je glumila Catherine Zeta-Jones?**
Filmove poredati po datumu.

```
PREFIX movie: <http://data.linkedmdb.org/resource/movie/>
PREFIX dc: <http://purl.org/dc/terms/>
SELECT ?title ?date
WHERE {
    ?movie movie:actor ?actor .
    ?movie dc:date ?date .
    ?movie dc:title ?title .
    ?actor movie:actor_name "Catherine Zeta-Jones"
    FILTER(?date>"2000-12-31")
}
ORDER BY ASC(?date)
```

Pokušajte sami napisati nekoliko upita:

- f. Koliko filmova u grafu je režirao Quentin Tarantino?
- g. Koje filmove je režirao Quentin Tarantino?
- h. Tko je sve glumio u filmu „Reservoir Dogs“?
- i. Napišite upit iz d) dijela pomoću ključne riječi UNION
- j. Koje su to drame i akcijski filmovi u kojima je glumio Will Smith?
- k. U kojem filmu u grafu je glumilo najviše glumaca?

Potrebna programska podrška

Mogući odabiri tehnologija:

JAVA	PHP
<ul style="list-style-type: none">• Apache Jena (apache-jena-2.11.0.zip)• Razvojno okruženje po želji (npr. Eclipse)	<ul style="list-style-type: none">• Arc2 library (arc2-master.zip)• Graphite library (Graphite.php)• Graphite sparqllib (sparqllib.php)

U oba slučaja potreban je Web poslužitelj, za Javu npr. **Apache Tomcat** ili **Jetty**, a ako koristite PHP, npr. **Apache** ili **IIS**.

Alternativno, za rješavanje zadatka možete koristiti i neku drugu kombinaciju tehnologija, no morate biti sigurni da s njima možete uspješno riješiti sve zadatke.

Napomena #1: Ako za vrijeme testiranja upite želite izvršavati nad lokalnom kopijom podataka (npr. zbog brzine ili nedostupnosti mreže/servera i sl.), preuzmite datoteku „**jena-fuseki+LMDBDump.7z**“, koju čini Fuseki server sa „utočenim“ podacima iz LMDb grafa.

Napomena #2: Za izvršavanje „Federated“ upita potrebna je SPARQL pristupna točka koja „zna“ obaviti raspodjelu upita i udruživanje rezultata, tzv. „mediator endpoint“. Ako koristite **Apache Jena** programski okvir, ugrađeni ARQ engine će obaviti medijaciju. U slučaju upotrebe nekog drugog programskog okvira, za medijaciju možete iskoristiti lokalni **Fuseki** server. U tom slučaju vam ne trebaju podaci, nego samo „čista instalacija“ - **jena-fuseki-1.0.0-distribution.zip**

Napomena #3: Cilj ovog projekta je upoznati se sa tehnologijama Semantičkog weba, naučiti osnove baratanja RDF-om i pisanja SPARQL upita. Zbog toga nije dozvoljeno korištenje tehnologija kao što je LINQ.

Zadatak

Potrebno je ostvariti jednostavan filmski web portal, koji kao izvor podataka koristi semantičke izvore **Linked movie database** i **Dbpedia**. Po želji, može se uključiti i dodatni semantički izvor (npr. Musicbrainz, Geonames i sl.). Umjesto **Linked movie database** izvora možete koristiti i lokalni SPARQL endpoint sa kopijom podataka.

Portal treba sadržavati:

- Jednostavnu naslovnicu sa **poljem za pretraživanje po naslovu filma ili imenu glumaca**. Na dnu naslovnice potrebno je ispisati **barem 4 statistička podataka o izvorima koji se koriste** (npr. ukupni broj trojki, trenutni broj filmova u bazi ili sl. – naravno brojke ne smiju biti „hardcodirane“ nego dohvaćene odgovarajućim upitom).
Dodatno, na stranici prikazite i sljedeće informacije
 - 3 slučajno odabrana filma
 - link na stranicu o glumcu/glumicama koji danas slave rođendan
- Stranicu za **prikaz rezultata pretraživanja**. Upit mora pronalaziti sve glumce i filmove koji bilo gdje u imenu ili nazivu sadrže zadani niz znakova. Rezultate je potrebno sortirati abecedno. Rezultati moraju biti prikazani tako da se klikom na odgovarajući link otvara stranica koja prikazuje podatke o odabranom glumcu ili filmu.
- Stranicu za **prikaz podataka o filmu** (npr. naslov, popis glumaca, datum, redatelj i scenarist, kratki opis...). Potrebno je prikazati najmanje 5 različitih podataka. **Obavezno** dohvatiti i podatke sa Dbpedije (ako u LMDb za taj film postoji owl:sameAs URI prema Dbpediji)
- Stranicu za **prikaz podataka o glumcu** (npr. ime i prezime, datum i mjesto rođenja, fotografija, wiki stranica, kratki životopis...). Podatke o glumcima potrebno je dohvatiti kroz združeni (en. *federated*) SPARQL upit. Potrebno je prikazati najmanje 5 različitih podataka. Ako ne postoji owl:sameAs link na Dbpediju, pokušajte napraviti združeni upit preko imena glumca. Uočite da LMDb nema oznaku jezika, dok Dbpedija ima (hint: FILTER (STR(?imeDBP) = ?imeLMDb)).
- Potrebno je omogućiti **dohvat podataka o filmovima i glumcima na različitim jezicima**. Jezik se može izabrati npr. preko padajućeg izbornika na samoj stranici.

Linked movie database

Linked movie database je otvorena baza filmova, povezana s raznim drugim semantičkim izvorima, kao što su Dbpedia, Geonames i sl.

WEB: <http://www.linkedmdb.org/>

SPARQL endpoint: <http://data.linkedmdb.org/sparql>

HTML verzija: npr. <http://data.linkedmdb.org/page/film/2014>

LOKALNA KOPIJA PODATAKA, interaktivno sučelje: <http://localhost:3030/>

LOKALNA KOPIJA PODATAKA, SPARQL endpoint: <http://localhost:3030/ds/sparql>

Dbpedia

Dbpedia je, prema definiciji s njihovih web stranica, „community effort“ tj. otvoreni zajednički projekt izdvajanja strukturiranih informacija iz Wikipedije. Izdvojeni podaci dostupni su u formi RDF grafa na Webu i omogućavaju postavljanje kompliciranih upita i pretraživanje Wikipedije na nove i prije

Dbpedije nemoguće načine.

WEB: <http://dbpedia.org/>

SPARQL endpoint: <http://dbpedia.org/sparql>

HTML verzija: npr. [http://dbpedia.org/page/The Lord of the Rings](http://dbpedia.org/page/The_Lord_of_the_Rings)

Primjeri SPARQL upita

OPĆENITO: Upit koji određuje ukupan broj trojki u grafu

```
SELECT (COUNT(*) AS ?ukupnoTrojki)
WHERE {
    ?s ?p ?o
}
```

LMDB: Filmovi u kojima je glumio ili ih je režirao Clint Eastwood

```
PREFIX imdb: <http://data.linkedmdb.org/resource/movie/>
SELECT DISTINCT ?movie
WHERE {
    {
        ?movie imdb:actor ?personURI .
        ?personURI imdb:actor_name "Clint Eastwood"
    }
    UNION
    {
        ?movie imdb:director ?personURI .
        ?personURI imdb:director_name "Clint Eastwood"
    }
}
```

LMDB: Popis glumaca koji su glumili s Kevinom Baconom

```
PREFIX imdb: <http://data.linkedmdb.org/resource/movie/>
SELECT DISTINCT ?actorName
WHERE {
    ?kb imdb:actor_name "Kevin Bacon" .
    ?movie imdb:actor ?kb .
    ?movie imdb:actor ?actor .
    ?actor imdb:actor_name ?actorName .
    FILTER (?kb != ?actor)
}
ORDER BY ASC(?actorName)
```

LMDB+DBPEDIA: Podaci o filmovima Arnolda Schwarzeneggera

```
PREFIX imdb: <http://data.linkedmdb.org/resource/movie/>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX dbpo: <http://dbpedia.org/ontology/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?dbpediaUrl ?title ?date ?abstract ?runtime ?wikiPage ?picURL
WHERE {
    ?actor imdb:actor_name "Arnold Schwarzenegger".
    ?movie imdb:actor ?actor .
}
```

```

?movie dcterms:title ?title .
?movie dcterms:date ?date .
?movie owl:sameAs ?dbpediaUrl .
FILTER( REGEX( STR(?dbpediaUrl), "dbpedia")) .
SERVICE <http://dbpedia.org/sparql>
{
    ?dbpediaUrl a dbpo:Film .
    ?dbpediaUrl foaf:isPrimaryTopicOf ?wikiPage .
    ?dbpediaUrl dbpo:abstract ?abstract .
    ?dbpediaUrl dbpo:runtime ?runtime .
    ?dbpediaUrl foaf:depiction ?picURL .
    FILTER(lang(?abstract) = "en")
}

```

Poveznice na dodatnu literaturu, dokumentaciju i sl.

Apache Jena i Fuseki server

<http://jena.apache.org/>

Graphite PHP Linked Data Library i sparqllib.php

<http://graphite.ecs.soton.ac.uk/>

Interaktivne upute za pisanje SPARQL upita nad grafom podataka ZPR-a:

<http://zpr.fer.hr/linkeddata.php>

-
- Projekt se radi i prezentira samostalno
 - Izvorni programski kod - rješenje projekta postaviti u vlastiti direktorij LaboratorijPliIS1\P2 na FTP serveru
 - Rok za predaju: **srijeda, 08.02.2017. u 08:00**
 - Prezentacija projekta: **srijeda, 08.02.2017. u 08:00+ u D259**