



Západočeská univerzita v Plzni

Fakulta aplikovaných věd

Semestrální práce z KIV/DB2

Semestrální práce č.3 - RDF (Fuseki)

Jakub Schenk

A21N0069P

schenkj@students.zcu.cz

5. května 2022

Obsah

1	Zadání	2
2	Úvod	3
3	Struktura ttl	3
4	Dotazy nad databází	4
4.1	Dotaz obsahující logickou spojku a podmínku na číselnou hodnotu	4
4.2	Dotaz nad nepovinnou položkou a existenci podřetězce v textové hodnotě	4
4.3	Dotaz s podmínkou na neexistenci trojice/hodnoty;	5
4.4	Dotaz s definicí hledaných n-tic hodnot klíčovým slovem VALUES	5
4.5	Dotaz s agregační funkcí	5
5	Závěr	6
5.1	Nesplněné body zadání	6
5.2	Body k zamyšlení a diskuzi	6
5.3	Shrnutí	7

1 Zadání

Připravte realistický RDF model (ideálně v Turtle serializaci) týkající se tématu. Vložte model do databáze Jena Fuseki. Vytvořte sadu SPARQL dotazů řešící reálné otázky. Vytvořte dokumentaci vysvětlující použitou datovou strukturu a popisující dotazy přirozeným jazykem. Diskutujte svoje řešení. Požadavky:

- Alespoň 10 různých RDF uzlů.
- Model je cyklický graf (tj. obsahuje aspoň alespoň jeden cyklus/kružnici jako svůj podgraf)
- Jsou použity alespoň dvě různé URI namespace schemata pro vytvářené resources.
- Alespoň 4 SPARQL dotazy.
- V sadě dotazů je alespoň jednou použit (lze kombinovat):
 - vzor nad nepovinnou položkou;
 - podmínka na číselnou hodnotu;
 - podmínka na existenci podřetězce v textové hodnotě;
 - podmínka na neexistenci trojice/hodnoty;
 - podmínka využívající logickou spojku A nebo NEBO;
 - vzor využívající Property Path;
 - definice hledaných n-tic (alespoň dvojic) hodnot klíčovým slovem VALUES
 - agregační funkce;

2 Úvod

Za své téma pro tabulky v relaci m:n jsem si vybral Hráč-hraje-Videohry. Toto téma jsem si zvolil, protože je mi relativně blízké, splňuje zadání a přijde mi poměrně názorné.

Jako atributy hráče jsem zvolil id jakožto primární klíč, jeho jméno a příjmení, věk (nepovinný atribut), výpis videoherních konzolí a výpis videoher, které na nich hraje.

3 Struktura ttl

Narozdíl od předchozích prací jsem zvolil jako prvního člena trojice jméno hráče. Hráč má id jakožto celočíselnou hodnotu, věk jakožto nepovinnou celočíselnou hodnotu, seznam vlastněných konzolí jakožto jedna nebo více textových hodnot a seznam hraných videoher jakožto jedna nebo více textových hodnot.

```
<Dobromil_Hladovy>  <id>          1 ;
                      <vek>         21 ;
                      <konzole>     "PC" ,
                                   "PS4" ;
                      <hraje>       "Tetris" ,
                                   "League_of_Legends" .
...

```

4 Dotazy nad databází

Po nahrání ttl souboru do databáze, jsem se začal dotazovat na data v tabulkách následujícími dotazy.

4.1 Dotaz obsahující logickou spojku a podmínku na číselnou hodnotu

```
PREFIX db: <http://localhost:3030/sp3/>
```

```
SELECT ?jmeno ?vek ?id ?konzole
WHERE {
    ?jmeno db:vek ?vek;
           db:id ?id;
           db:konzole ?konzole.
    FILTER(?konzole="PC" && ?vek > 21).
}
LIMIT 20
```

Tímto dotazem lze získat jména hráčů, jejich věk, id a konzole na kterých hrají hry. Získaní hráči musí mít zadaný věk vyšší než 21 let a konzole kterou vlastní musí být "PC".

4.2 Dotaz nad nepovinnou položkou a existenci podřetězce v textové hodnotě

```
PREFIX db: <http://localhost:3030/sp3/>
```

```
SELECT ?id ?jmeno ?vek ?konzole
WHERE {
    ?jmeno db:id ?id.
    OPTIONAL{
        ?jmeno db:vek ?vek.
    }
    ?jmeno db:konzole ?konzole.
    FILTER CONTAINS(?konzole,"P") .
}
ORDER BY DESC(?vek)
```

Narozdíl od prvního dotazu jsou vypsáni i hráči bez zadaného věku. U nich je vypsáno jejich jméno, id, věk (pokud je zadaný) a konzole. Z těchto

záznamů jsou ovšem filtrováni pouze ti hráči, kteří vlastní konzoli obsahující písmeno "P"(tedy "PC"a "PS4").

4.3 Dotaz s podmínkou na neexistenci trojice/hodnoty;

```
PREFIX db: <http://localhost:3030/sp3/>
```

```
SELECT ?id ?jmeno ?vek
WHERE {
    ?jmeno db:id ?id ;
            db:vek ?vek .
    FILTER EXISTS { ?jmeno db:vek ?vek . }
}
ORDER BY(?vek)
```

Výsledkem dotazu by měl být seznam jmen hráčů, jejich id a věk, ale pouze pokud je věk vyplněný. Výsledky jsou pak seřazeny vzestupně podle věku.

4.4 Dotaz s definicí hledaných n-tic hodnot klíčovým slovem VALUES

```
PREFIX db: <http://localhost:3030/sp3/>
```

```
SELECT ?id ?jmeno ?vek
WHERE {
    ?jmeno db:id ?id ;
            db:vek ?vek .
    VALUES ?id {1 2 3}
}
```

Výsledkem dotazu by měli být hráči s id = 1, 2, nebo 3. Jsou vypsaný jejich id, jména a věk. Pokud by hráč neměl zadaný věk, neobjevil by se ve výsledku.

4.5 Dotaz s agregační funkcí

```
PREFIX db: <http://localhost:3030/sp3/>
```

```
SELECT (count(?videohry) as ?pocet)
WHERE {
```

```
?jmeno db:hraje ?videohry .  
}
```

Výsledkem dotazu by měl být celkový počet konzolí vlastněný všemi hráči.

5 Závěr

5.1 Nesplněné body zadání

Bohužel se mi nepodařilo a nenapadá mě, jak nad těmito daty zkonstruovat nějaký smysluplný dotaz, který by využíval "Property PATH". Tento typ dotazu tedy není v sadě dotazů.

5.2 Body k zamyšlení a diskuzi

Členění do souborů; bylo by možné model smysluplně rozčlenit do více souborů? Pro takto malou kolekci dat, nevidím rozumný důvod pro členění do souborů. V případě ukládání většího množství dat, dává smysl ukládat data do více souborů podle nějaké charakteristiky.

Limitace datového formátu RDF pro uložení dat z této domény. Tím že je dotaz tvořen orientovaný graf, nabízí se poukázat na omezení hran grafu na pouze jeden směr. Krom toho je to také poměrně neintuitivní způsob ukládání dat a nepříliš populární způsob dotazování.

Limitace v dotazování – na jaké otázky není možné nebo je těžké odpovědět? Jelikož se při zadávání dat vytváří orientovaný graf, je těžké nebo nemožné pohybovat se po hranách v opačném směru.

Je problematická existence nepovinných údajů? Existenci nepovinných údajů nepovažuji za problematickou. Pokud je v dotazu správně označena klíčovým slovem "OPTIONAL" jsou všechna data vypsána a místa s absencí údajů jsou jednoduše prázdná.

Způsob zapsání relace mezi uzly – uložit pouze hranu orientovanou jedním směrem, nebo uložit i inverzní hranu Záleží na charakteru dat, ale pro většinu případů bych osobně preferoval, kdyby měla data i inverzní hranu, kvůli pohodlnosti v dotazování. Toto řešení by však mohlo být velice paměťově náročné.

5.3 Shrnutí

Prací s touto problematikou jsem se naučil jak pracovat s RDF formátem, soubory typu ttl a databázovým systémem Apache Jena Fuseki. Vypracováním semestrální práce jsem si snad dostatečně připomenul látku z přednášek a cvičení.