

Začínáme se SPARQL

DBpedia

DBpedia je strojově vygenerovaná znalostní báze. Kromě infoboxů z Wikipedie, obsahuje DBpedia i informace z textu (odkazy v relaci wikiPageWikiLink).

SPARQL Server

<https://dbpedia.org/sparql>

Jmenné prostory

Kromě standardních rdf, rdfs, owl, xsd, DBpedia používá svoje vlastní:

dbo - ontology (zde je např. taxonomie Book - Written_Work - Work)

dbr - resource

dbp - property

Navíc používá standardy jako:

dc - Dublin Core

dcterms - DC terms

foaf - Friend of a Friend

geodata - GeoNames

vcard - VCF

skos - Simple Knowledge Organization System

Kompletní seznam jmenných prostorů je zde: <https://dbpedia.org/sparql/?help=nsdecl>

Ontologie

Ontologie říká, na co se můžeme ptát. Jak se jmenují třídy a vlastnosti.

DBpedia T-Box specifikuje třídy (např. Forest, CanadianFootballLeague, FictionalCharacter) a vlastnosti (např. highestPoint má range Place). Některé věci mohou být jak třída (např. Colour), tak vlastnost (colour má range Colour).

<http://archivo.dbpedia.org/download?o=http%3A/dbpedia.org/ontology/&f=owl>

A-Box jsou informace o instancích. Protože je jich hodně, pro stažení jsou v několika

souborech (podle typu tvrzení a jazyka). Např. type statement je

```
<http://cs.dbpedia.org/resource/Loretánské_náměstí>
```

```
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
```

```
<http://dbpedia.org/ontology/Street> .
```

Pro zformulování dotazu je důležitý T-Box.

Dotazy ve SPARQL

Budeme používat DBpedia Virtuoso SPARQL endpoint. Není potřeba do něj psát prefixy. Dokud dotaz nevyładíte, používejte vždy LIMIT.

Jednoduché dotazy

1. Zkusme najít nějaké knihy.
SELECT ?book WHERE {
 ?book a dbo:Book .
}
LIMIT 100
2. Klikněte na libovolný výsledek a prohlédněte si stránku z DBpedie.
3. Prohlédněte si odpovídající stránku z Wikipedie.
4. Vraťte se zpět na stránku v DBpedii, najděte libovolnou vlastnost knihy (nejspíš ve jmenném prostoru dbp, dbo nebo dc) a zkopírujte její název.

Ke složitějším dotazům (graf s více než jednou hranou):

1. Ve SPARQL endpointu rozšířte dotaz o vlastnost, kterou jste předtím našli.
2. Pojmenujte vhodně proměnnou.
SELECT ?book WHERE {
 ?book a dbo:Book .
 ?book <vaše vlastnost> ?vaše_proměnná .
}
LIMIT 100
3. Jaké hodnoty obsahuje ?vaše_proměnná? Jsou to URI nebo literály?

Dotazy s count() a DISTINCT

1. Mějme dotaz:
SELECT ?book
WHERE {
 ?book a dbo:Book .
 ?book dbo:author ?author .
 ?book dbp:releaseDate ?date .
 ?author rdfs:label ?author_name .
 ?book rdfs:label ?title .
}
}
2. Vyzkoušejte dotaz s limitem. Co vidíte?
3. Přidejte DISTINCT (SELECT DISTINCT ?book...)
4. Modifikujte původní dotaz tak, aby spočítal, kolik výsledků vrátí (pozor, nesmí tam být LIMIT)
5. Přidejte do dotazu DISTINCT (SELECT COUNT(DISTINCT(?book)) ...

Dotazy s FILTER a více podmínkami

1. Zkuste dotaz:

```
SELECT ?book ?author ?pages WHERE {  
  ?book a dbo:Book .  
  ?book dbo:author ?author .  
  ?book dbp:releaseDate ?date .  
  ?book dbo:numberOfPages ?pages .  
}  
LIMIT 100
```
2. Přidejte filter na počet stran menší než 100. Je v seznamu nějaká kniha, která má méně než 20 stran?
3. Přidejte filter na počet stran menší než 100, ale větší než 20.

Dotazy s testem na typ proměnné

1. Zkuste dotaz

```
SELECT ?book ?date ?year  
WHERE {  
  ?book a dbo:Book .  
  ?book dbp:releaseDate ?date .  
  BIND(year(?date) as ?year)  
}  
LIMIT 100
```
2. Dotaz skončí s chybou. Co myslíte, že je příčinou? Zkuste zjednodušený dotaz:

```
SELECT ?book ?date  
WHERE {  
  ?book a dbo:Book .  
  ?book dbp:releaseDate ?date .  
}  
LIMIT 100
```
3. Přidejte do původního dotazu filtr na povolený typ hodnoty ?date:

```
FILTER (DATATYPE(?date) = xsd:date)
```

Dotazy s FILTER na datum

1. Zkuste dotaz

```
SELECT ?book ?date  
WHERE {  
  ?book a dbo:Book .  
  ?book dbp:releaseDate ?date .  
  FILTER(?date < "1950-01-01"^^xsd:date)  
}  
LIMIT 100
```

2. Jsou skutečně všechny knihy vydané před rokem 1950?

3. Co může být příčinou?

4. Zkuste dotaz modifikovat:

```
SELECT ?book ?date
WHERE {
    ?book a dbo:Book .
    ?book dbp:releaseDate ?date .
    FILTER(xsd:date(?date) < "1950-01-01"^^xsd:date)
}
LIMIT 100
```

5. Co je za problém? Jakého typu jsou různé hodnoty proměnné ?date?

6. Finální dotaz musí obsahovat jak konverzi proměnné ?date na typ xsd:date, tak specifikaci literálu.

```
SELECT ?book ?date
WHERE {
    ?book a dbo:Book .
    ?book dbp:releaseDate ?date .
    FILTER (DATATYPE(?date) = xsd:date)
    FILTER(xsd:date(?date) < "1950-01-01"^^xsd:date)
}
LIMIT 100
```