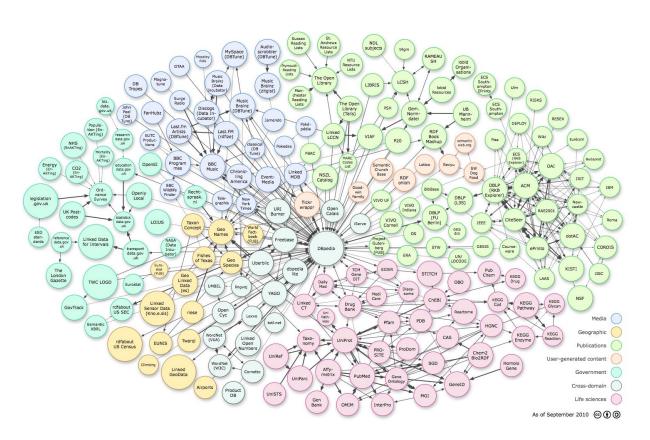
Napredni modeli i baze podataka

Predavanja Siječanj 2016.

13. Semantičke web tehnologije - OWL

Semantički web

- Web
- Semantički web
- Semantičke web tehnologije
 - RDF
 - RDFS
 - SPARQL
 - Ontologije
- Mreža podataka



RDF opisuje resurse s klasama, svojstvima i vrijednostima. RDFS omogućava kreiranje vlastitog RDF rječnika. SPARQL omogućava dohvat i pretraživanje RDF grafova.

Kako definirati veze među klasama ili instancama klasa? npr. "ekvivalentna klasa" ili "isti resurs"

Kako opisati karakteristike i ograničenja svojstava? npr. "inverzno svojstvo"

Ontologije

- Ontologija je model podataka koji predstavlja skup pojmova unutar neke domene i veze između tih pojmova
- Eksplicitna, formalna specifikacija zajedničke konceptualizacije.
 (Thomas R. Gruber, 1993)
 - Eksplicitna: nedvosmislene definicije, svi koncepti moraju biti definirani
 - Formalna specifikacija: "razumljiva računalima"
 - Zajednička: dijeljena, poznatog i općeprihvaćenog značenja
 - Konceptualizacija: apstraktni model (domena, koncepti, odnosi među konceptima)
- Filozofska disciplina grana filozofije koja se bavi prirodom i organizacijom stvarnosti

Ontologije

- Ontologija se sastoji od:
 - rječnika pojmova koji opisuju neku domenu
 - nazivi bitnih koncepata iz domene
 - znanje i ograničenja na pojmove iz domene

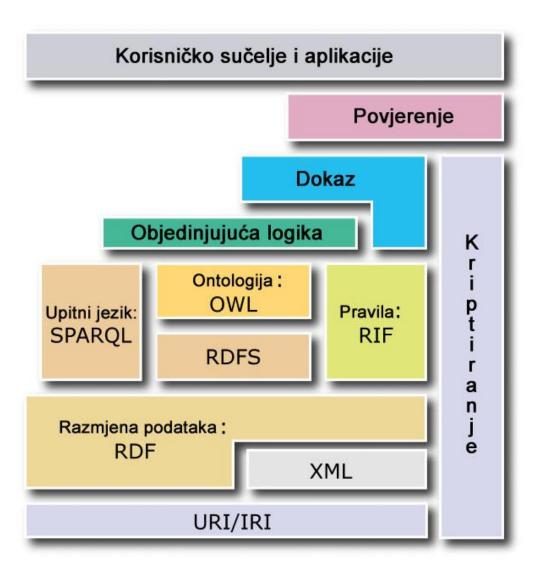
- Jezici ontologija:
 - RDF iskazivanje tvrdnji
 - RDFS opisivanje podataka
 - OWL bogatije opisivanje podataka i veza

OWL – Web Ontology Language

- Jezik Semantičkog Weba stvoren da omogući predstavljanje bogatog i kompleksnog znanja o entitetima (stvarima, pojmovima), grupama entiteta i odnosima među entitetima.
- OWL proširuje mogućnosti RDFS-a
 - karakteristike svojstava (TransitiveProperty, SymmetricProperty, InverseOf, ...)
 - ograničenja svojstava (minCardinality, maxCardinality, ...)
 - preslikavanja (equivalentClass, equivalentProperty, sameAs, ...)
 - složeni razredi (intersectionOf, unionOf, ...)
- Znanje i logika opisani u OWL-u su pogodni za obradu i korištenje pomoću računala



SPARQL – Semantic Web Layer cake





OWL – Web Ontology Language

- Sastoji se od:
 - klasa (i njihove hijerarhije)
 - svojstava koja te klase imaju
 - ograničenja i odnosa među svojstvima (tip, domena, kardinalnost, jednakost)
 - instanca klasa
- "Open World" pretpostavka formalne logike
 - ako nešto ne znamo, ne znači da je "laž"

```
knows(peanuts:charlie.brown, peanuts:lucy.van.pelt);
```

- Znamo da Charlie poznaje Lucy, no ne možemo tvrditi da poznaje SAMO Lucy
- Ne postoji "Unique Name" pretpostavka
 - Moramo eksplicitno izreći vezu između entiteta:

```
owl:sameAs, owl:differentFrom
```

OWL – Web Ontology Language

- Verzije OWL-a:
 - OWL Lite koristi se gdje je dovoljna minimalna izražajnost ontologija
 - OWL DL pruža maksimalnu izražajnost koja garantira izračunavanje
 - OWL Full pruža maksimalnu izražajnost i sintaktičku slobodu RDF-a bez garancije za izračunavanjem
- OWL 2 ima više 'profila' jezika podskupa jezika koji su usmjereni na
 - izračunljivost
 - implementaciju



OWL - Klase

- Postoje dvije predefinirane klase:
 - owl:Thing SVE je Thing
 - owl:Nothing
 - Definicija nove klase:

```
:Automobil a owl:Class <olw:Class rdf:ID="Automobil"/>
```

Podklasa

```
:Automobil rdfs:subClassOf :MotornoVozilo
```

Ekvivalencija dviju klasa

```
:Automobil owl:equivalentClass :Car
```

Disjunktne klase

```
:DizelVozilo owl:disjointWith :BenzinVozilo
```



OWL - Svojstva

definicija svojstva

```
:imaProizvođača a owl:ObjectProperty .
:imaCijenu a owl:DatatypeProperty .
```

domena i doseg svojstva

owl:inverseOf, owl:equivalentPropery, rdfs:subPropertyOf

```
:imaProizvođača owl:inverseOf :jeProizvođač .
```



OWL

veze među instancama

- Spaček je isto što i Citroen 2CV?
- Spaček je Automobil?
- Spaček je motorno vozilo?
- Spaček ima cijenu i proizvođača?
- Spaček je ili benzinac ili dizel vozilo?



OWL - primjer

peanuts:charlie.brown foaf:knows peanuts:lucy.van.pelt .

peanuts:charlie.brown vocab:friendOf peanuts:lucy.van.pelt .



OWL – primjer

```
Koje instance 1-8 su
:MikesBike rdfs:subClassOf :Bicycle .
                                             tipa: Mikes Bike?
:MikesBike owl:equivalentClass [
    rdf:type owl:Restriction ;
                                                 :Bike2
    owl:onProperty :owner ;
                                                 :Bike3
   owl:hasValue :Mike
                                                 :Bike5
                                                 :Car8
:Bike1 a :Bicycle .
:Bike2 a :MikesBike .
:Bike3 a :Bicycle ; :owner :Mike .
:Bike4 a :Bicycle ; :owner :Greg .
:Bike5 a :Bicycle ; :owner [ owl:sameAs :Mike ] .
:Bike6 owl:equivalentClass [ a owl:Restriction ;
          owl:onProperty :owner ; owl:hasValue :Mike
:Mike :owner :Bike7 .
:Car8 :owner :Mike .
```

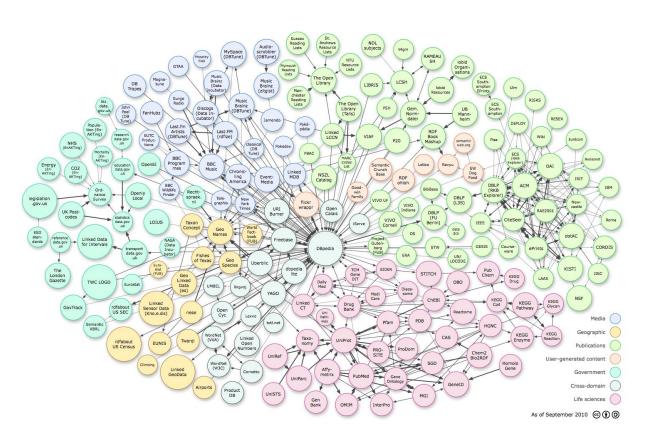
Napredni modeli i baze podataka

Predavanja Siječanj 2016.

14. Semantički izvori podataka – mreža podataka

Semantički web – mreža podataka

- Web
- Semantički web
- Semantičke web tehnologije
 - RDF
 - RDFS
 - SPARQL
 - Ontologije
- Mreža podataka



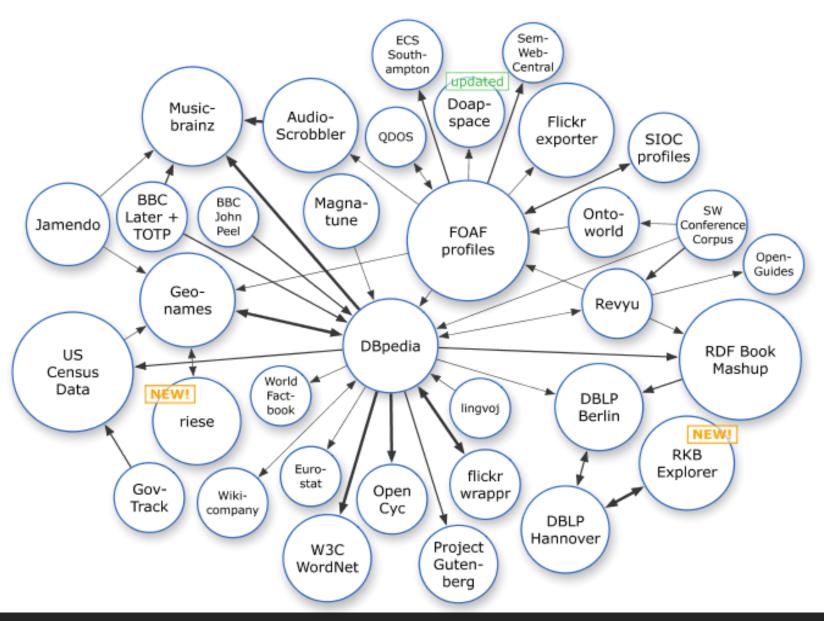
Mreža podataka/povezani podaci – Linked data

Trenutno aktualan pravac istraživanja u semantičkom webu

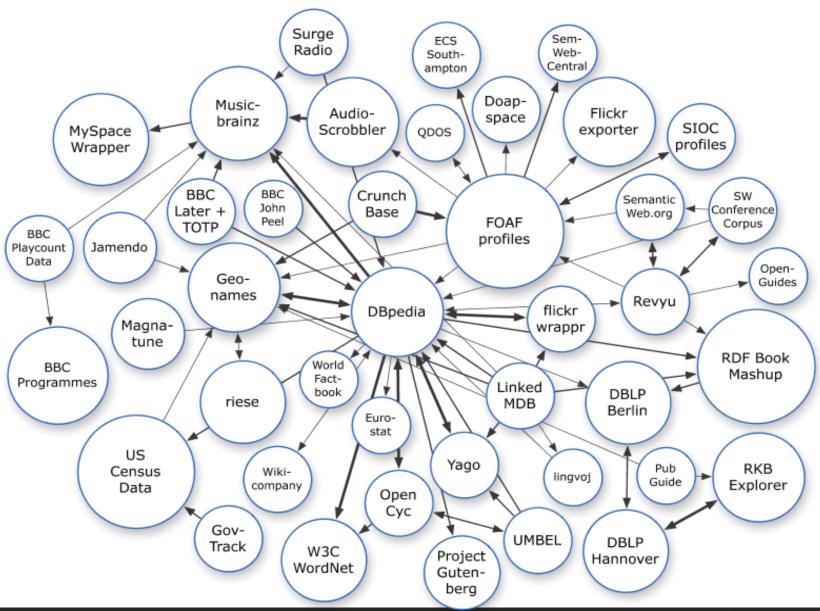
Načela povezanih podataka:

- svi elementi moraju biti identificirani URI-em (nema praznih čvorova)
- svi URI-i moraju se moći razriješiti (eng. dereference) pronaći HTTP
 URL koji daje koristan opis elementa identificiranim zadanim URI-em
- moraju se postaviti veze prema drugim URI-ima kako bi se omogućila daljnja laka potraga za podacima

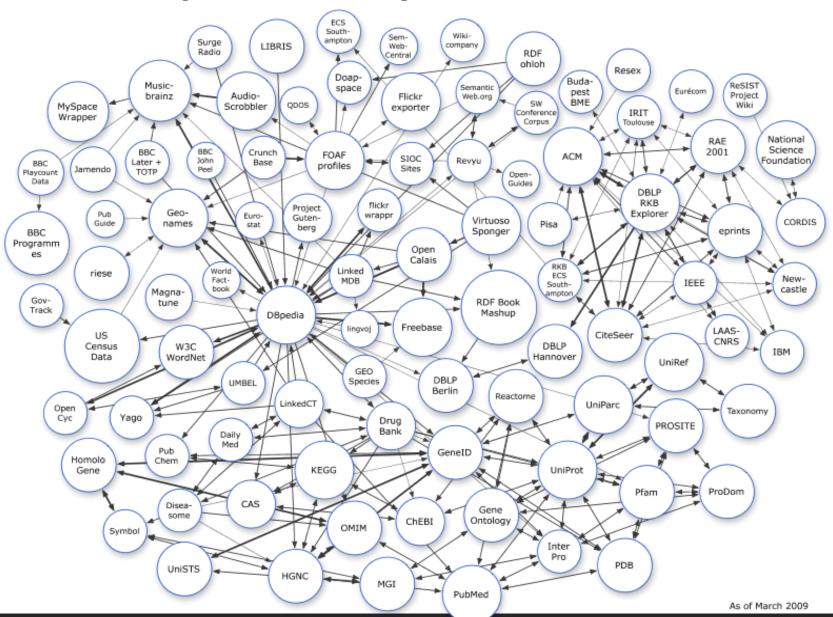
Povezani podaci – ožujak 2008



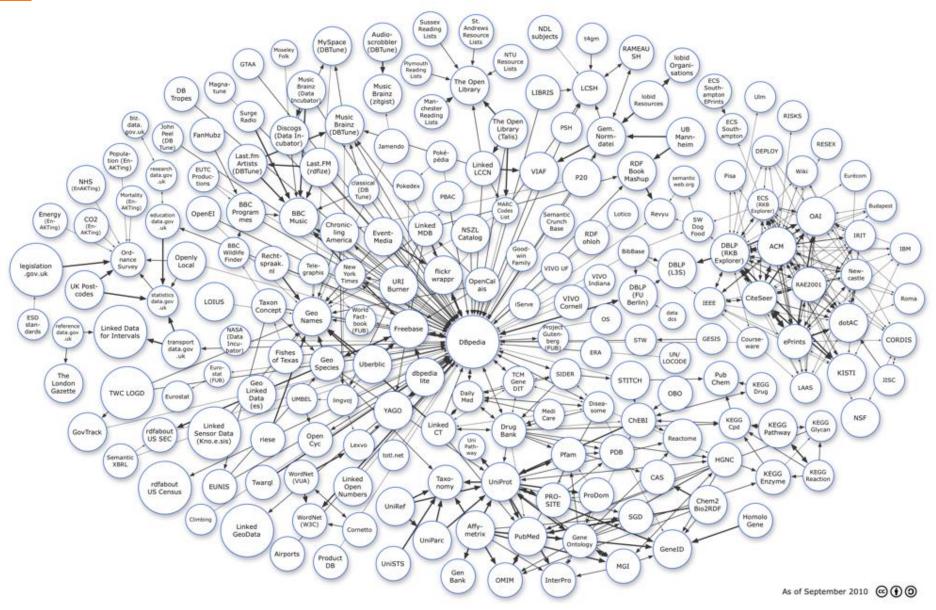
Povezani podaci – rujan 2008



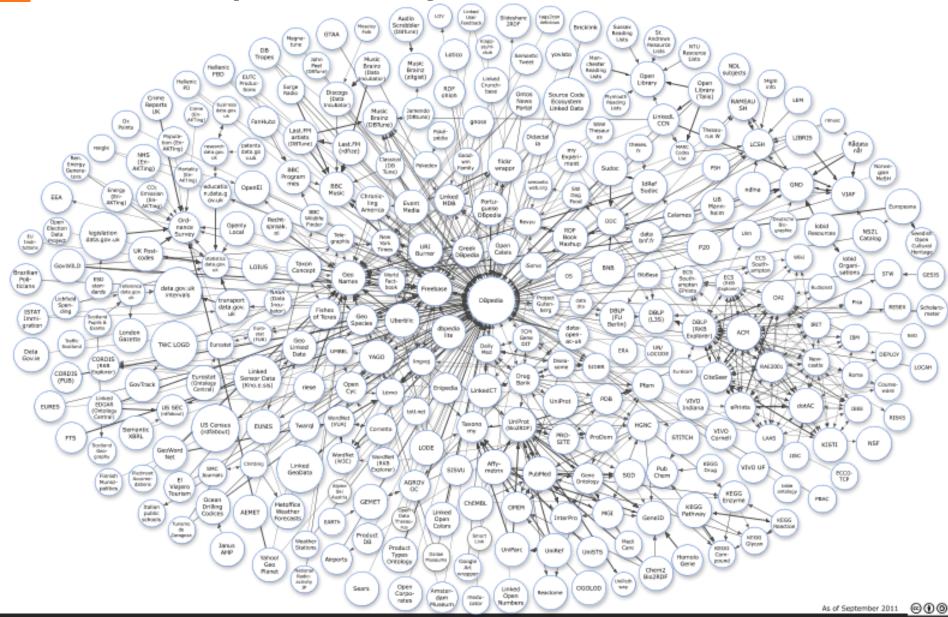
Povezani podaci – ožujak 2009



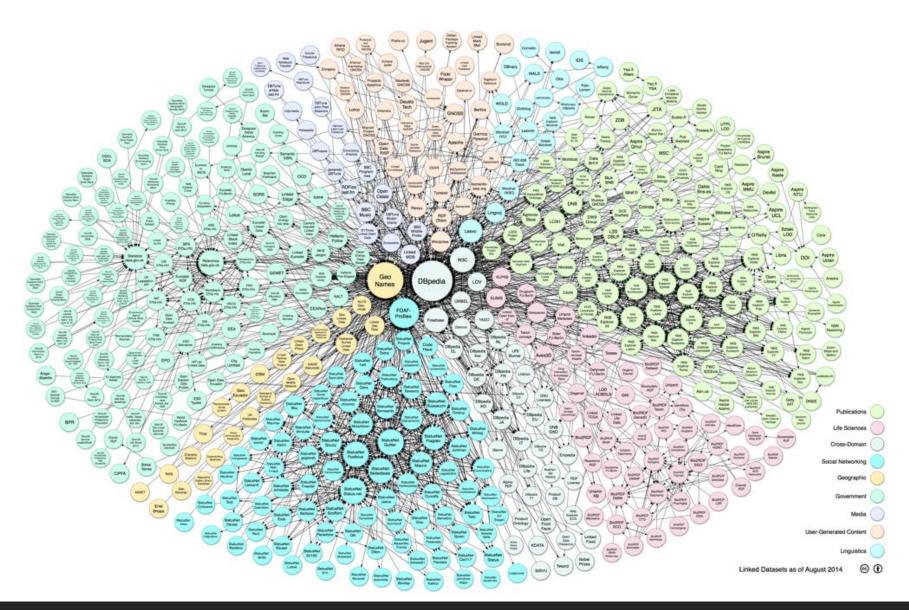
Povezani podaci – rujan 2010



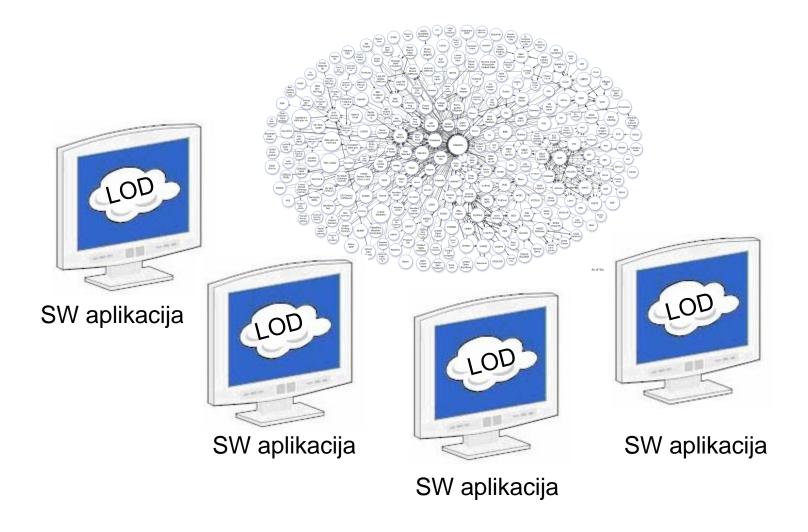
Povezani podaci – rujan 2011



Povezani podaci – kolovoz 2014



Semantički web danas



Početak razvoja mreže podataka – DBpedia (1)

DBpedia

- podaci koji su strukturirano pohranjeni na Wikipediji (infobox) semantički su opisani
- postoji SPARQL servis putem kojeg se mogu postavljati proizvoljni upiti
- budući je DBpedia od početka dio mreže podataka, taj izvor je povezan s velikim brojem drugih izvora

Početak razvoja mreže podataka – DBpedia (2)

```
@prefix dbpedia <http://dbpedia.org/resource/>.
@prefix dbterm <http://dbpedia.org/property/>.

dbpedia:Amsterdam
  dbterm:officialName "Amsterdam";
  dbterm:longd "4";
  dbterm:longm "53";
  dbterm:longs "32";
  dbterm:website <http://www.amsterdam.nl>;
  dbterm:populationUrban "1364422";
  dbterm:areaTotalKm "219";
  ...

dbpedia:ABN_AMRO
  dbterm:location dbpedia:Amsterdam;
  ...
```



DBpedia – način korištenja

- Dozvoljava postavljanje uvjeta pretraživanja kakvi nisu mogući na Wikipediji
 - Npr. nadi sve osobe rođene u Zagrebu prije 1900

```
SELECT ?name ?birth ?death ?person
WHERE {
          ?person dbo:birthPlace :Zagreb.
          ?person dbo:birthDate ?birth .
          ?person foaf:name ?name .
          ?person dbo:deathDate ?death .
          FILTER (?birth < "1900-01-01"^^xsd:date) .
}</pre>
```

DBpedia – princip objave povezanih podataka (1)

- Svaki pojam mora imati URI
 - Npr. http://dbpedia.org/resource/Amsterdam
- Svaki URI mora se moći razriješiti
 - Npr. unese li se link http://dbpedia.org/page/Amsterdam u
 pretraživač, korisnika se preusmjeri na
 http://dbpedia.org/page/Amsterdam

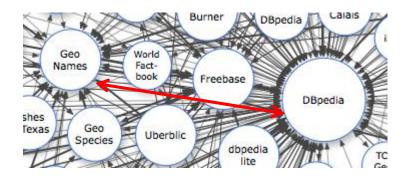
Koji je to način razrješavanja?

DBpedia – princip objave povezanih podataka (2)

Veze između različitih izvora podataka

```
<http://sws.geonames.org/2759793>
  owl:sameAs <http://dbpedia.org/resource/Amsterdam>
  wgs84_pos:lat "52.3666667";
  wgs84_pos:long "4.8833333";
  geo:inCountry <http://www.geonames.org/countries/#NL>;
...

<http://dbpedia.org/resource/Amsterdam>
  owl:sameAs <http://rdf.freebase.com/ns/...>;
  owl:sameAs <http://sws.geonames.org/2759793>;
```



Važnost mreže podataka

- Predstavlja skup podataka nad kojim se mogu stvarati rješenja semantičkog weba
- Web identifikatori za pojmove
 - Npr. http://dbpedia.org/resource/Amsterdam
- Do sad je najveći problem razvoja semantičkog weba predstavljao nedostatak semantički objavljenih podataka – mreža podataka je prvi korak prema ispunjenju tog cilja

Vokabulari

- Nije isto što i ontologija; predstavlja njen podskup
- Za ostvariti integraciju podataka potrebno je međusobno razumijevanje:
 - pojmova
 - "asistent", "viši asistent", "student", "izvanredni profesor", "redovni profesor"
 - kategorija
 - "Osoba", "predmet", "literatura"
 - veza između pojmova i kategorija
 - "student je ujedno i Osoba...", "redovni profesor je uži termin od samog profesora"

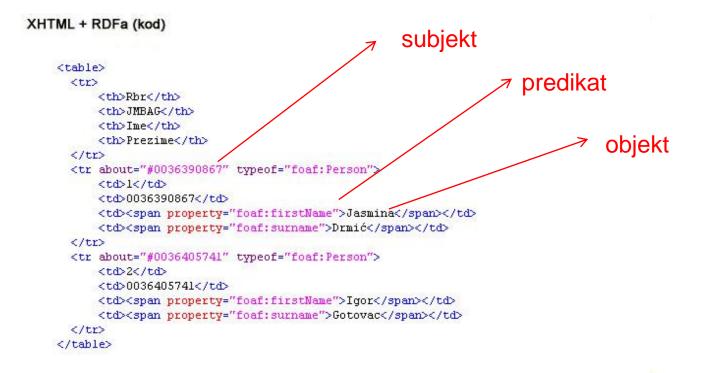
Osnovni vokabulari

- Postoji niz već definiranih vokabulara
 - Dublin Core: sadrži opise pojmova digitalnih knjižnica, autorskih prava, dozvola, i sl.
 - FOAF: opisuje osobe i organizacije
 - DOAP: opis software projekata
 - SIOC: Semantically-Interlinked Online Communities pojmovi društvenih mreža
 - vCard osnovni podaci o osobi (podsjetnica)
 - •
- Vokabulari se moraju dijeliti; pa prije nego se kreira vlastiti vokabular, potrebno je vidjeti koji su već dostupni

RDFa

- RDFa pojednostavljeno: predstavlja RDF trojke unutar XHTML dokumenata
- ugradnja RDF izjava u XHTML omogućena je korištenjem proširenja XHTML-a – sintakse RDFa
- Za izgradnju RDF izjava koriste se atributi XHTML-a (rel, rev, src, href, content, itd...), kao i atributi RDFa sintakse (about, resource, datatype, itd..)
- Trenutno u razvoju RDFa 1.1. W3C working draft
 - RDFa 1.1. Primer
 - RDFa 1.1. Lite podskup RDFa, zamišljen za manje web projekte i početnike u razvoju RDFa omogućenih web stranica

RDFa primjer



XHTML prikaz (preglednik)

Rbr.	JMBAG	Ime	Prezime
1	0036390867	Jasmina	Drmić
2	0036405741	Igor	Gotovac

RDF izjave (nakon parsiranja)

```
#0036390867 foaf:firstName Jasmina.

#0036390867 foaf:surname Drmić.

#0036390867 rdf:type foaf:Person.

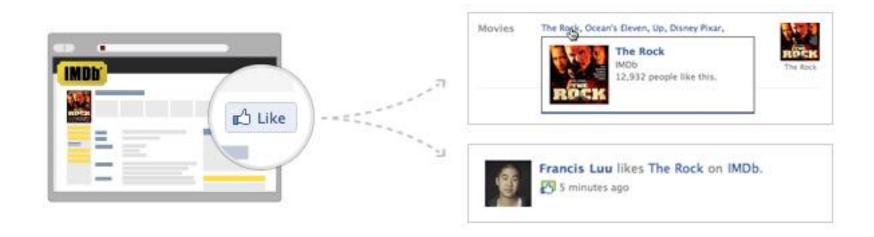
#0036405741 foaf:firstName Igor.

#0036405741 foaf:surname Gotovac.

#0036405741 rdf:type foaf:Person.
```

Primjena Open Graph protokola (1)

Semantički Web na Facebooku



- Strukturirano opisani podaci na izvorištu
- Facebook 'razumije' sadržaj i može integrirati sadržaj

Primjena Open Graph protokola (2)

```
→ -Vokabular Open-Graph
<html xmlns="http://www.w3.org/1999/xhtml"</pre>
      xmlns:og="http://ogp.me/ns#"-
      xmlns:fb="http://www.facebook.com/2008/fbml">
  <head>
                                                       property vokabulara
    <title>The Rock (1996)</title>
    <meta property="og:title" content="The Rock"/>
    <meta property="og:type" content="movie"/>
    <meta property="og:url" content="http://www.imdb.com/title/tt0117500/"/>
    <meta property="og:image" content="http://ia.media-imdb.com/rock.jpg"/>
    <meta property="og:site_name" content="IMDb"/>
    <meta property="fb:admins" content="USER ID"/>
    <meta property="og:description"</pre>
          content="A group of U.S. Marines, under command of
                   a renegade general, take over Alcatraz and
                   threaten San Francisco Bay with biological
                   weapons."/>
  </head>
</html>
```

Integracija heterogenih izvora podataka

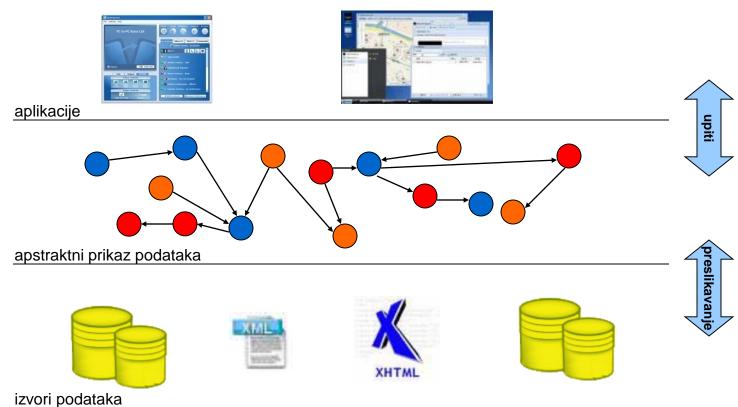
- podaci se nalaze u različitim oblicima
 - polustrukturirani (web dokumenti)
 - strukturirani (XML dokumenti, baze podataka, ...)
- tehnologije semantičkog web-a mogu pomoći pri integraciji heterogenih izvora podataka

Integracija baza podataka

- problem integracije je u tom što su baze podataka obično heterogene po strukturi i sadržaju
- Zašto (kada) je potrebna integracija?
 - velik broj servisa zahtjeva rukovanje podacima iz više baza podataka
 - nakon udruživanja tvrtki
 - povezivanje raznih administrativnih podataka (npr. za porez)
 - biološka, genetska, farmaceutska i medicinska istraživanja
 - itd...

Integracija podataka

- Osnovni koraci integracije podataka:
 - 1. Dati značenje podacima preslikati ih u uvriježeno značenje
 - 2. Udružiti preslikane podatke
 - 3. Obavljanje upita nad cijelim skupom podataka



Relacijske baze i RDF (1)

- n-torke relacija su slične RDF trojkama
 - sama n-torka se može gledati kao subjekt (čvor)
 - atribut kao *predikat*
 - vrijednost atributa pojedine n-torke kao objekt
- Ipak preslikavanje nije tako jednostavno
 - kako odrediti URI, odgovara li on primarnom ključu?
 - u bazi podataka iste trojke se mogu javiti više puta
 - prilagođavanje tipova podataka odabranoj ontologiji
 - šifrirani podaci (npr. spol 'M' i 'Ž')
 - ponekad je potrebno spajati više vrijednosti atributa (npr. ime i prezime čine jednu vrijednost u ontologiji)
 - NULL vrijednosti
 - **-** ...

Relacijske baze i RDF (2)

GRAD

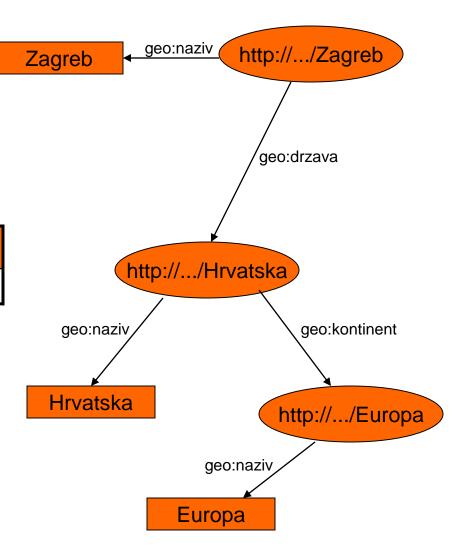
ID	naziv	oznDrzava
1	Zagreb	HR

DRŽAVA

oznDrzava	naziv	kontinentID
HR	Hrvatska	1

KONTINENT

ID	naziv
1	Europa

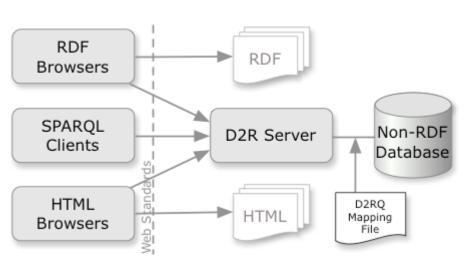


Relacijske baze i RDF (3)

- Postojeće tehnologije preslikavanja:
 - D2RQ / D2R server
 - Virtuoso
 - DartGrid
 - SPASQL

Relacijske baze i RDF - D2R Server

- alat za objavljivanje sadržaja relacijskih baza podataka na semantičkom webu
- koristi D2RQ za mapiranje podataka iz relacijske baze u RDF
- upiti koji stižu na D2R Server se prepisuju u SQL upite, koji se izvršavaju na bazi u pozadini, rezultati se formatiraju u RDF i prikazuju korisniku u odabranom formatu
- omogućava tri različita "sučelja" za pristupa podacima:
 - "Linked data sučelje"
 - "SPARQL" sučelje
 - klasični HTML pristup



Relacijske baze i RDF - D2R Server

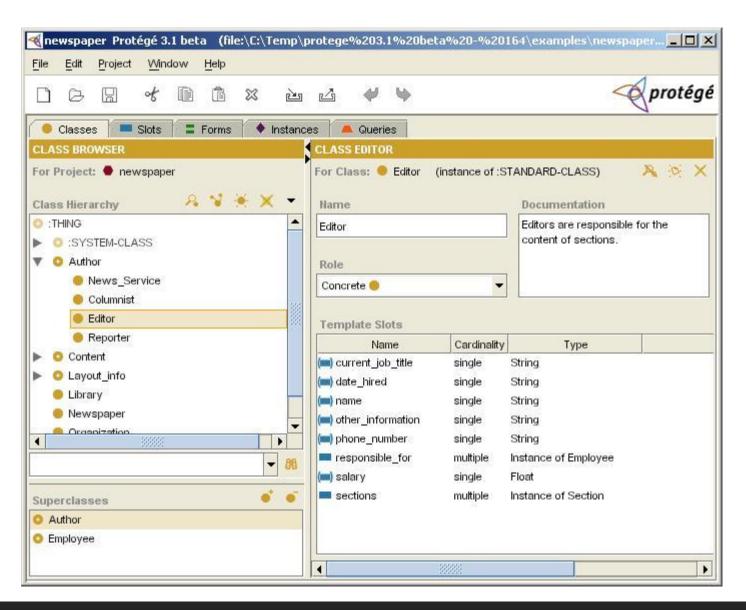
Primjer datoteke mapiranja

```
<> a d2r:Server;
    rdfs:label "Semantic ISVU - Educational Program - D2R Server";
    d2r:baseURI <http://kent.zpr.fer.hr:8080/educationalProgram/>;
    d2r:port 8080;
    d2r:autoReloadMapping false;
map:database a d2rq:Database;
   d2rq:jdbcDriver "com.informix.jdbc.IfxDriver";
   d2rq:jdbcDSN "jdbc:informix-
   sqli://127.0.0.1:9089/educationalProgram:INFORMIXSERVER=projekt;DB LOCALE=HR
   HR.8859-2; DELIMIDENT=Y";
   d2rq:username "username";
   d2rq:password "password";
```

Odabir ontologije

- za obavljanje preslikavanja treba odabrati ciljnu ontologiju
- budući da se u semantičkom web-u potiče dijeljenje ontologija, potrebno je pregledati dostupnost odgovarajuće ontologije (ili barem pojma)
 - npr. http://swoogle.umbc.edu/, www.sindice.com
- ukoliko primjenjiva ontologija nije dostupna, treba ju kreirati
- ukoliko postoji djelomično primjenjiva ontologija, treba ju nadograditi

Izrada ontologije



D2RQ / D2R (1)

- definira se datoteka koja sadrži preslikavanja strukture relacijske baze u ontologiju
- preslikavanje se definira upotrebom posebnog jezika

preslikavanje relacija u razrede

preslikavanje atributa u predikate

D2RQ / D2R (2)

```
map:organization a d2rq:ClassMap;
    d2rg:dataStorage map:database;
   d2rg:uriPattern "organization/@@orgjed.siforgjed@@";
   d2rq:class foaf:Organization;
   d2rg:classDefinitionLabel "Organization";
map:organization name hr a d2rq:PropertyBridge;
   d2rq:belongsToClassMap map:organization;
   d2rq:property foaf:name;
    d2rq:propertyDefinitionLabel "name";
    d2rg:condition "nazorgjed.padez = 'no' AND nazorgjed.jezik = 'hr' AND sifvrorgjed < 33";
    d2rg:column "nazorgjed.nazorgjed";
    d2rq:lang "hr";
   d2rq:join "orgjed.siforgjed => nazorgjed.siforgjed";
map:organization name en a d2rq:PropertyBridge;
   d2rq:belongsToClassMap map:organization;
   d2rq:property foaf:name;
    d2rg:propertyDefinitionLabel "name";
    d2rq:condition "nazorgjed.padez = 'no' AND nazorgjed.jezik = 'en' AND sifvrorgjed < 33";
    d2rg:column "nazorgjed.nazorgjed";
    d2rg:lang "en";
   d2rq:join "orgjed.siforgjed => nazorgjed.siforgjed";
```

D2RQ / D2R (3)

```
map:place a d2rq:ClassMap;
   d2rg:dataStorage map:database;
   d2rq:uriPattern "place/@@nazmjesto.ozndrzava@@/@@nazmjesto.postozn@@";
   d2rg:class sisvu:Place;
    d2rg:condition "nazmjesto.ozndrzava = 'HR'";
   d2rg:classDefinitionLabel "Place";
map:place postozn a d2rq:PropertyBridge;
    d2rg:belongsToClassMap map:place;
    d2rq:property sisvu:zip;
    d2rg:join "mjesto.postozn => nazmjesto.postozn";
    d2rq:join "mjesto.ozndrzava => nazmjesto.ozndrzava";
    d2rg:condition "nazmjesto.jezik = 'hr' AND nazmjesto.padez = 'no'";
    d2rq:column "nazmjesto.postozn";
map:place geonamesid a d2rq:PropertyBridge;
    d2rg:belongsToClassMap map:place;
    d2rq:property owl:sameAs;
    d2rq:join "mjesto.postozn => nazmjesto.postozn";
    d2rq:join "mjesto.ozndrzava => nazmjesto.ozndrzava";
    d2rq:condition "nazmjesto.jezik = 'hr' AND nazmjesto.padez = 'no'";
    d2rq:uriPattern "http://sws.geonames.org/@@mjesto.geoname id@@/";
```

D2RQ / D2R (4)

- preslikanim podacima se može pristupati preko SPARQL endpointa
- SPARQL upit prevodi se u SQL upit koji se izvodi nad relacijskom bazom te se dobiveni rezultati transformiraju u rezultate SPARQL upita
- sve promjene u bazi su odmah vidljive i na RDF modelu s obzirom da RDF model nije učitan u memoriju nego se upiti uvijek izvode nad relacijskim modelom

Integracija podataka – primjer (6)

- udruživanje dobivenih modela
- treba izjednačiti URI-e istih pojmova
 - Primjer:

```
<http://dbpedia.org/resource/Berlin> owl:sameAs
<http://sws.geonames.org/2950159>
```

- povezati značenja različitih ontologija
 - Primjer:

Integracija podataka – trajna spremišta

- podatke dobivene u RDF formatu (RDF trojke) treba negdje pohraniti
- RDF spremišta nude i metode za dohvaćanje podataka
- sadržana u memoriji
 - Jena Joseki / Fuseki, Sesame
- prirodna spremišta
 - Virtuoso, Sesame Native, Oracle 11g, AllegroGraph
- temeljen na relacijskim bazama podataka
 - Jena + (MySQL, PostgreSQL, Oracle, DB2)
- najbolje performanse pokazuju prirodna RDF spremišta

Zaključivanje

- Mehanizam za zaključivanje
 - Zaključivanje nad pravilima definiranim u OWL-u
 - Rezultat zaključivanja su dodatne trojke
 - Neki od mehanizama za zaključivanje: Jena Resoner, Pellet, Racer,...
- Kod integracije podataka zaključivanje se obavlja u svrhu izjednačavanja pojmova

Integracija podataka – zaključivanje

