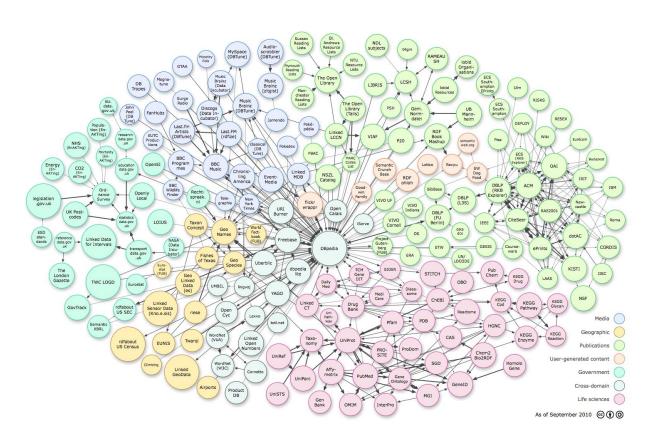
Napredni modeli i baze podataka

Predavanja Siječanj 2016.

11. Semantičke web tehnologije

Semantički web – mreža podataka

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



Što je Web?

- Često se miješaju pojmovi web i Internet, odnosno smatraju se sinonimima
 - Internet je mrežna infrastruktura
 - Web (World Wide Web) je skup dokumenata pohranjen na poslužiteljima koji su povezani Internetom, te koji su dostupni putem HTTP protokola

Web (1)

- World Wide Web često se kraće naziva Web
- Prema W3C-u Web je:
 - Informacijski prostor u kojem su smješteni izvori informacija identificirani URI-em (Uniform Resource Identifier)



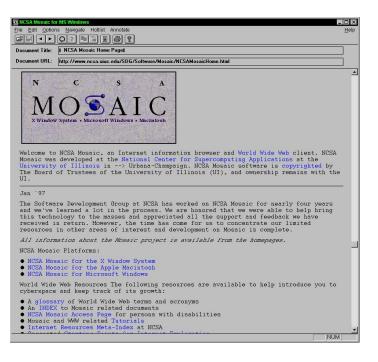
Tim Berners Lee

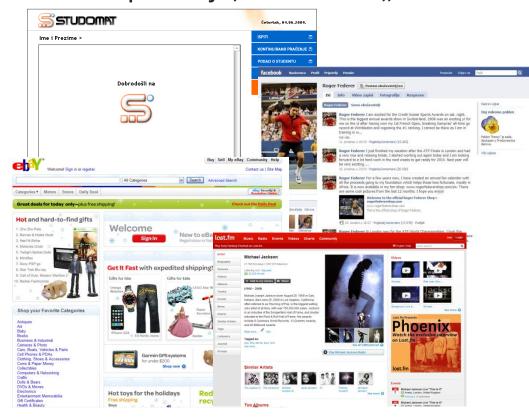
Web (2)

 inicijalno zamišljen kao sustav za distribuiranje statičkog sadržaja, koji se sastojao od jednostavnog teksta i slika

danas je platforma za sofisticirane aplikacije, ali i velika "baza

informacija"





Povijest Web-a

- 1990: Berners-Lee je proizveo radne prototipe Web preglednika i poslužitelja
- 1991: napravljena prva web stranica
- 1993: kreiran je prvi vizualni preglednik: Mosaic
 - Mosaic je integrirao tekst, slike i veze, te je tako surfanje postalo intuitivno
- 1994: osnovan je Netscape, koji je proizveo još jedan popularan preglednik: Netscape Navigator
- 1995: Microsoft je proizveo Internet Explorer → počinje rat preglednika!; osniva se Amazon.com; izlazi Java Applet; izlazi Netscape preglednik 2.0B3 s podrškom za JavaScript
- 1996: osniva se Google; izlazi Macromedia Flash 1.0; proizvodi se IntelliMouse (miš s kotačem); izlazi Internet Explorer 3.0 s podrškom za JavaScript
- 1997: izlazi Hotmail, besplatni e-mail server
- 1999: izlazi Napster
- 1998: pojavljuju se prvi blogovi; osnovan je PayPal; Amazon kupuje Imdb
- 2000: Tim Berners Lee predstavlja viziju semantičkog web-a

Semantički web - definicija

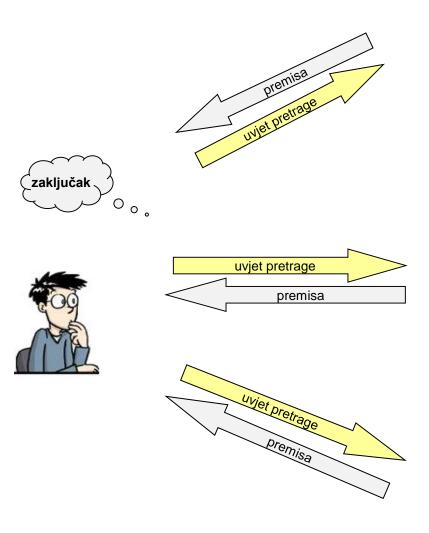
Semantički web je vizija: ideja o mogućnosti da podaci na webu budu definirani i povezani na način na koji bi se mogli koristiti osim za sami prikaz i za automatizaciju, integraciju i ponovnu iskoristivost u različitim aplikacijama. Na taj način obećava se mogućnost pronalaženja, sortiranja i klasificiranja informacija, dakle svih onih zadaća koje troše puno vremena provedenog on i off-line [W3C].

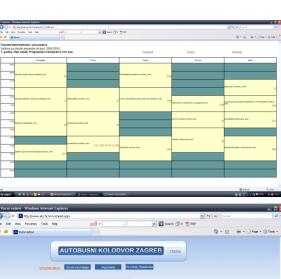


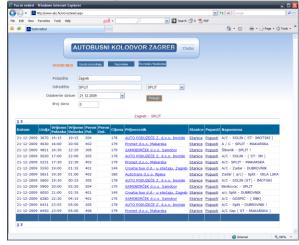
Semantički Web - početna vizija

- semantički web se prema početnoj ideji trebao temeljiti na agentima semantičkog weba koji bi mogli prikupljati podatke iz različitih izvora podataka, procesirati ih i međusobno razmjenjivati s drugim agentima
- semantički web omogućuje
 - razumijevanje računala bez intervencije korisnika
 - donosi zaključke korištenjem umjetne inteligencije

Semantički Web - primjer

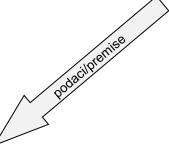








Semantički Web - primjer



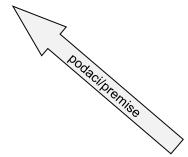
podaci/premise

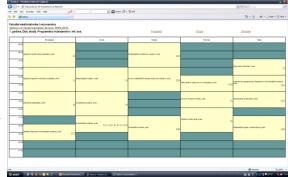


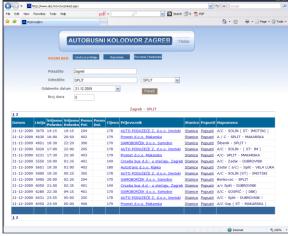














Semantički web - infrastruktura

- ostvarenje ideje semantičkog weba zahtijeva i odgovarajuću tehnološku infrastrukturu
- potrebno je osmisliti i međusobno povezati brojne tehnologije od kojih je svaka namijenjena specifičnom problemu
 - opisivanje podataka
 - logika
 - zaključivanje
 - sigurnost
 - **-**

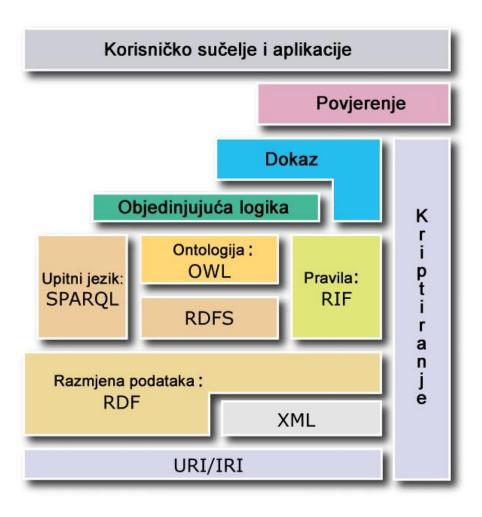
Daljnji razvoj web-a

- 2001: stvorena je Wikipedia
- 2003: osnovan je MySpace, Web 2.0
- 2004: Google najavljuje Gmail; izlazi Firefox 1.0
- 2005: osnovan i predstavljen on-line YouTube; Yahoo kupuje Flickr; pojava Ajax-a
- 2007: Facebook, Twitter (raste globalna popularnost društvenih web aplikacija)
- 2008: Google Chrome
- 2009: raste broj "rich Internet" aplikacija
- 2010: Open Data, Open Government Data, Linked Open Data
- 2011: dominacija Facebooka i mobilnih aplikacija
- 2012: Google Chrome postaje najkorišteniji Web preglednik na svijetu
- 2013: Više korisnika Interneta preko mobilnih uređaja nego preko desktop uređaja
- 2014: RDF 1.1 i HTML5 kao W3C Recommendation, rad na HTML 5.1 i 5.2

W3C - World Wide Web Consortium

- osnovan je 1994.
- W3C je međunarodni konzorcij pri kojem organizacije članice, zaposlenici W3Ca i javnost rade zajedno na razvoju Web standarda
- cilj W3C-a je ostvariti puni potencijal WWW-a razvijajući protokole i smjernice koje bi osigurale dugoročan rast Web-a
- W3C se brine da temeljne tehnologije WWW-a budu međusobno kompatibilne
- objavljivanjem otvorenih web standarda (za web jezike, protokole) nastoji se izbjeći tržišna, odnosno Web fragmentacija
- Primjeri W3C dokumenata:
 - HTML 4.01 Specification 24.12.1999. (W3C recommendation)
 - Media Queries 27.7.2010. (W3C candidate recommendation)
 - SPARQL Protocol for RDF 15.1.2008. (W3C recommendation)
 - HTML 5 28.10.2014. (W3C recommendation)

Semantički web - tehnologije



Od vizije do realizacije

- Tim Berners Lee je očekivao da će razvoj trajati 10 godina, no u tom periodu primarne tehnologije web-a se znatno mijenjaju, pa im se prilagođava i vizija semantičkog web-a
 - društvene web aplikacije postaju sve popularnije
 - gotovo sve web aplikacije su "rich internet" aplikacije (Ajax, Web
 2.0.)

Semantički web danas (1)

- Razlikujemo dvije vizije:
 - mreža podataka Web of data / Linked data
 - otvaranje sadržaja baza podataka za web
 - objava podataka u nekom od otvorenih formata
 - potpunije rješenje
 - mreža oznaka (potiče ga Web 2.0 i društvene aplikacije)
 - označavanje podataka na web-u
 - jednostavna semantika

Semantički web danas (2)

- WWW je bio brzo i široko prihvaćen, prvenstveno zbog sljedećih razloga:
 - jednostavan za korištenje
 - jednostavan za kreiranje
- Semantičkom webu nedostaju alati (pravila) koji bi pridonijeli jednostavnosti korištenja i njegovog kreiranja
 - primarno je osigurati jednostavno označavanje podataka

Semantički web – davanje značenja podacima



https://openhpi.de/course/semanticweb

Semantički web – davanje značenja podacima

- Što je informacija, a što je reklama?
- Što te informacije znače? Koje su bitne?
- Koliko vjerujemo izvoru informacija?
- ljudi imaju:
 - iskustvo
 - kontekstualno znanje
 - znanje o svijetu

```
href="http://raspored.hrt.hr/?raspored=1&mreza=4&datum=2013-12-14&najava=125578" target="blank" rel="nofol
    23:37 <a href="http://www.imdb.com/title/tt1127180/?ref_=fn_al_tt_1" target="_blank" rel="nofollow">Odvuci
    href="http://raspored.hrt.hr/?raspored=16mreza=26datum=2013-12-146najava=117012" target="blank" rel="nofol
   23:45 <a href="http://www.imdb.com/title/tt0112995/?ref =fn al tt_6" target="_blank" rel="nofollow">Pretjer
   href="http://novatv.dnevnik.hr/clanak/filmovi/pretjerana-sila-2-osveta---314342.html" target="_blank" rel="
   01:13 <a href="http://www.imdb.com/title/tt0043408/?ref =fn al tt 1" target="blank" rel="nofollow">Cimarro
   href="http://raspored.hrt.hr/?raspored=1&mreza=2&datum=2013-12-14&najava=125455" target="blank" rel="nofol
   01:35 <a href="http://www.imdb.com/title/tt0443680/?ref =fn_al_tt_1" target="_blank" rel="nofollow">Woojstv
   href="http://novatv.dnevnik.hr/clanak/filmovi/ubojstvo-jessea-jamesa---314343.html" target="_blank" rel="no
   Kompletni programi: <a href="http://raspored.hrt.hr/?raspored=1&mreza=2&datum=2013-12-14" target=" blank" r
  href="http://raspored.hrt.hr/?raspored=1&mreza=3&datum=2013-12-14" target="_blank" rel="nofollow">HTV2</a>,
  raspored=16mreza=46datum=2013-12-14" target="_blank" rel="nofollow">HTV3</a>, <a href="http://raspored.hrt."
  target="blank" rel="nofollog">HTV4</a>, <a href="http://aww.rtl.hr/televizija/tv-raspored/" target=" blank"
  href="http://novatv.dnevnik.hr/raspored/" target="_blank" rel="nofollow">Nova TV</a><span class="actions"><
  programa-subotom/255009/" title="Permalink"><img src="http://resources.monitor.hr/images/design/permalink.g
  title="permalink" /></a></span> <span class="date">(08:12) </span></div><fb:like href="http://www.monitor.hr
 Send="false" layout="button_count" width="130" show_faces="false"></fb:like></div></div></div></div></div></div>
 class="supertitle">Snježne nevolje</div><h2 class="hormal">U Kairu prvi put nakon 112 godina pao snijeg, Je
 SIC="http://resources.monitor.hr/files/images/90x90/s/snileg2 28520.jpg" alt="U Kairu prvi put nakon 112 go
 class="image" /><div class="text">Nakon dugo vremena, na Bliskom istoku sniježi. Jeruzalem je jučer zabilje
 padaline. Gotovo polovica grada uslijed nepogode je ostala bez struje, a gradonačelnik Nir Barkat usporedio
 Stanovnici Egipta padaline doživljavaju kao čudo ili znak da će u svijetu zavladati mir. Ovo je prvi puta n
 eraquo; <a href="http://www.latimes.com/world/worldnow/la-fg-wn-snow-israel-egypt-20131213,0,1691393.story#"
rel="nofollow">LA Times</a>, <a href="http://www.israelnationalnews.com/News/News.aspx/175126#.UqtudNGJISt"
rel="nofollow">israelnationalnews.com</a>, <a href="http://dnevnik.hr/vijesti/svijet/foto-prvi-put-u-112-go
target="blank" rel="nofollou">Dnevnik.hr</a><span class="actions"><a href="http://www.monitor.hr/vijesti/u
jeruzalem-bez-struje/254997/" title="Permalink"><img src="http://resources.monitor.hr/images/design/permali
```

- računala nemaju ništa od toga
 - HTML opisuje kako su informacije prezentirane i povezane
 - HTML ne opisuje značenje informacija

Semantički web – davanje značenja podacima



Tehnologije semantičkog web-a

- W3C standardi
 - URI/IRI
 - XML
 - RDF
 - RDFS
 - OWL
 - SPARQL
 - RDFa
 - RIF

 u semantičkom webu svaki pojam (resurs), apstraktan ili opipljiv, ima jedinstven identifikator – URI

A Uniform Resource Identifier (URI) defines a simple and extensible schema for worldwide unique identification of abstract or physical resources (RFC 3986).

- "ideja URI-ja":
 - WWW -> URL
 - izdavaštvo (knjige, časopisi) -> ISBN, ISSN
 - •

Kolegij Baze podataka http://www.fer.hr/courseInstance/31503/2014-15

Charlie Brown http://www.peanuts.com/charlie.brown

opis pojma ≠ identifikacija



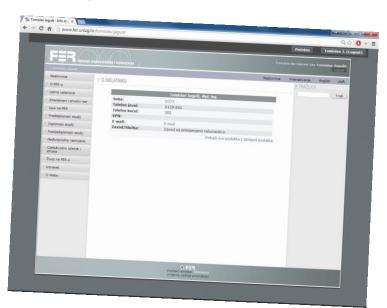
opis pojma ≠ identifikacija

Ako je ...

http://www.fer.unizg.hr/tomislav.jagust

...identifikator osobe koja se zove Tomislav Jagušt,

što je onda identifikator web stranice koja se otvori kada tu adresu upišemo u web preglednik?



- kako razlikovati identifikator osobe od URL-a osobne web stranice?
 - "303 URIs and hash URIs" (R. Cyganiak: Cool URIs for the SemWeb)

http://www.fer.unizg.hr/tomislav.jagust
(HTTP response status code 303 - See Other)
http://www.fer.unizg.hr/about/tomislav.jagust
ili
http://www.fer.unizg.hr/tomislav.jagust

http://www.fer.unizg.hr/tomislav.jagust#about

- što ako za neki pojam ne postoji URI?
 - sami stvorimo novi URI
 - paziti da novi URI već ne postoji (ista pravila kao npr. za Java namespace)
 - omogućiti dohvat opisa pojma u obliku prikladnom za žive korisnike

XML – Extensible Stylesheet Language Transformations

- sve se tehnologije semantičkog weba zasnivaju na XML-u
- XML se koristi samo kao format zapisa semantika je pohranjena u višim slojevima
- W3C odobren način predstavljanja semantičkih zapisa

RDF – Resource Description Framework

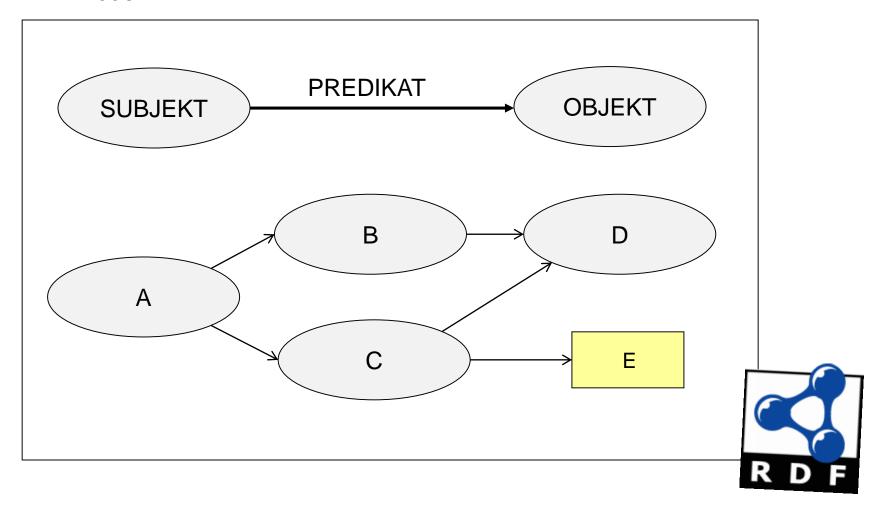
- "okvir" tj. model koji omogućava predstavljanje informacija na Webu.
- namijenjen za spremanje "podataka o podacima"
- prilagođen računalima
- temelji se na konceptu trojki: subjekt predikat objekt
 - subjekt ima određeno svojstvo (predikat) čija je vrijednost objekt
 - subjekt je uvijek resurs, dok objekt može biti resurs ili podatkovna vrijednost
- RDF trojka predstavlja izjavu
- Načini zapisa RDF-a: (RDF notacija)
 - trojna notacija (N3), Turtle
 - XML notacija
 - vizualni prikaz u obliku grafa





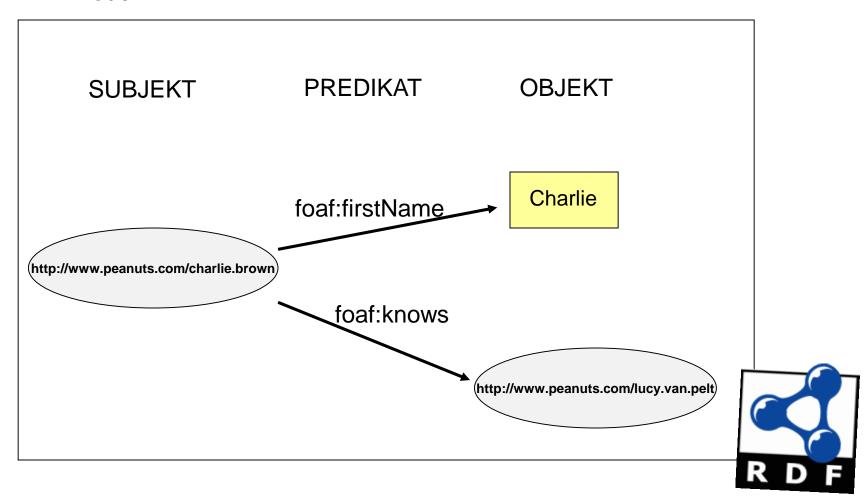
RDF model (1)

RDF model



RDF model (1) - primjer

RDF model



RDF model (2)

XML

```
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:foaf="http://xmlns.com/foaf/0.1/" >
    <rdf:Description rdf:about="http://www.peanuts.com/charlie.brown">
        <foaf:knows rdf:resource="http://www.peanuts.com/lucy.van.pelt"/>
        <foaf:firstName>Charlie</foaf:firstName>
    </rdf:Description>
</rdf:RDF>
```

N3



RDF - objekt

- objekti u RDF izjavama ne moraju biti resursi, mogu imati podatkovnu vrijednost
- vrijednosti objekta može se definirati tip podatka: float, integer, boolean, itd., (tipovi podataka definirani XML Schemom)

```
<http://dbpedia.org/resource/Branko_Gavella>
   :birthdate "1885-07-29"^^xsd:date;
   :dateofbirth "29"^^xsd:integer.
```

 ako je vrijednost objekta tekst, može mu se dodijeliti oznaka jezika

```
<http://dbpedia.org/resource/Branko_Gavella>
          :placeOfBirth "Zagreb, Austria-Hungary"@en;
          :placeOfDeath "Zagreb, SFR Yugoslavia"@en.
```



RDF - predikat

- subjekt pojam (resurs) "URI"
- objekt vrijednost ili pojam
- predikat?
- predikat je svojstvo subjekta, "vrijednost" tog svojstva je objekt
- Mora uvijek biti URI

foaf:firstName?

"Prostor imena" http://xmlns.com/foaf/0.1/

firstName - The first name of a person.

FOAF – rječnik (lista, popis) pojmova koji opisuju osobe, njihove aktivnosti i odnose s drugim osobama ili pojmovima





http://xmlns.com/foaf/spec/

Programski okviri za rad sa podacima u grafu

- JAVA
 - Apache Jena
 - RDF API, SPARQL engine, Triple store ("grafovska baza")...
- PHP
 - ARC / ARC2
 - Graphite PHP Linked Data Library
 - RAP
- C#
 - dotNetRDF
- postoje programski okviri i za druge jezike



PRIMJER: RDF model – kreiranje izjava - Java





PRIMJER: RDF model – kreiranje izjava - PHP

Graphite:

```
<?php
  include once("arc/ARC2.php");
  include once("Graphite.php");
  $personURI
                  = "http://www.peanuts.com/charlie.brown";
  $firstName
                  = "Charlie";
  $friendURI
                 = "http://www.peanuts.com/lucy.van.pelt";
  $graph = new Graphite();
  $graph->ns( "foaf", "http://xmlns.com/foaf/0.1/" );
  $graph->addTriple( $personURI, "foaf:name", $firstName );
  $graph->t( $personURI, "foaf:knows", $friendURI ); // addTriple() == t()
  echo $graph->dump();
?>
                                161.53.18.95/zprweb/inclu ×
                               ← → C ↑ 161.53.18.95/zprweb/include/test.php
                               http://www.peanuts.com/charlie.brown
```

 \rightarrow foaf:name \rightarrow Charlie

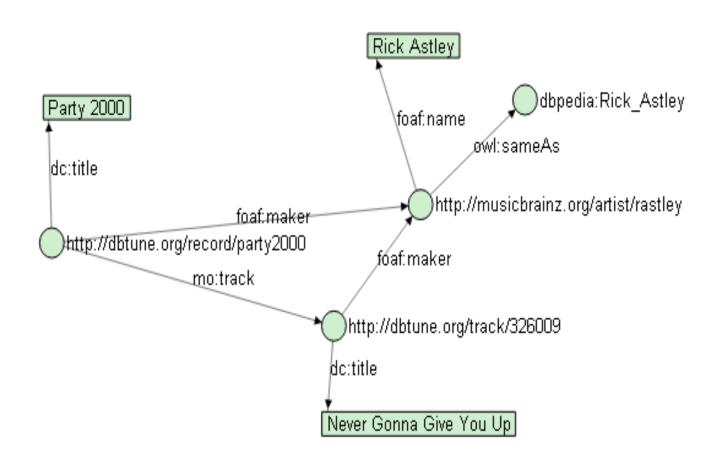


→ foaf:knows → http://www.peanuts.com/snoopy

PRIMJER: RDF model – RDF/XML u graf

```
<?xml version="1.0"?>
<rdf:RDF
      xmlns:owl="http://www.w3.org/2002/07/owl#"
      xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
      xmlns:mo="http://purl.org/ontology/mo/"
      xmlns:foaf="http://xmlns.com/foaf/0.1/"
      xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description rdf:about="http://dbtune.org/album/party2000">
    <dc:title>Party 2000</dc:title>
    <mo:track>
      <rdf:Description rdf:about="http://dbtune.org/track/326009">
        <dc:title>Never Gonna Give You Up</dc:title>
        <foaf:maker rdf:resource="http://musicbrainz.org/artist/rastley"/>
      </rdf:Description>
    </mo:track>
    <foaf:maker rdf:resource="http://musicbrainz.org/artist/rastley"/>
  </rdf:Description>
  <rdf:Description rdf:about="http://musicbrainz.org/artist/rastley">
    <foaf:name>Rick Astley</foaf:name>
      <owl:sameAs rdf:resource="http://dbpedia.org/resource/Rick Astley"/>
  </rdf:Description>
</rdf:RDF>
```

PRIMJER: RDF model – RDF/XML u graf



RDF opisuje resurse s klasama, svojstvima i vrijednostima.

Ali kako "opisati" RDF?

Kako definirati klase i odnose među klasama?

Kako definirati svojstva i odnose među svojstvima?

Kako opisati RDF vokabular? (npr. FOAF)



Odgovor: pomoću RDF-a ©

RDFS - RDF Vocabulary Description Language (1)

RDFS = RDF Schema (stari naziv)

RDFS

- definira značenja, karakteristike i odnose između dozvoljenih termina - omogućuje kreiranje vlastitog RDF rječnika
- prvi korak do "dodatnog znanja"
- omogućava izvođenje zaključaka iz RDFS pravila

RDFS - RDF Vocabulary Description Language (2)

Osnovni elementi:

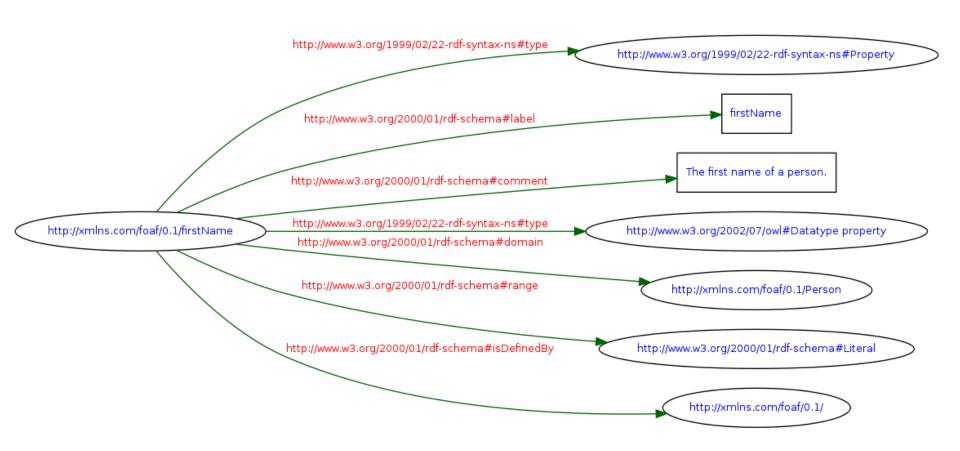
- rdfs:Resource
- rdfs:Class
- rdfs:Literal
- rdfs:Property
- rdfs:domain
- rdfs:range
- rdfs:subClassOf
- rdfs:subPropertyOf

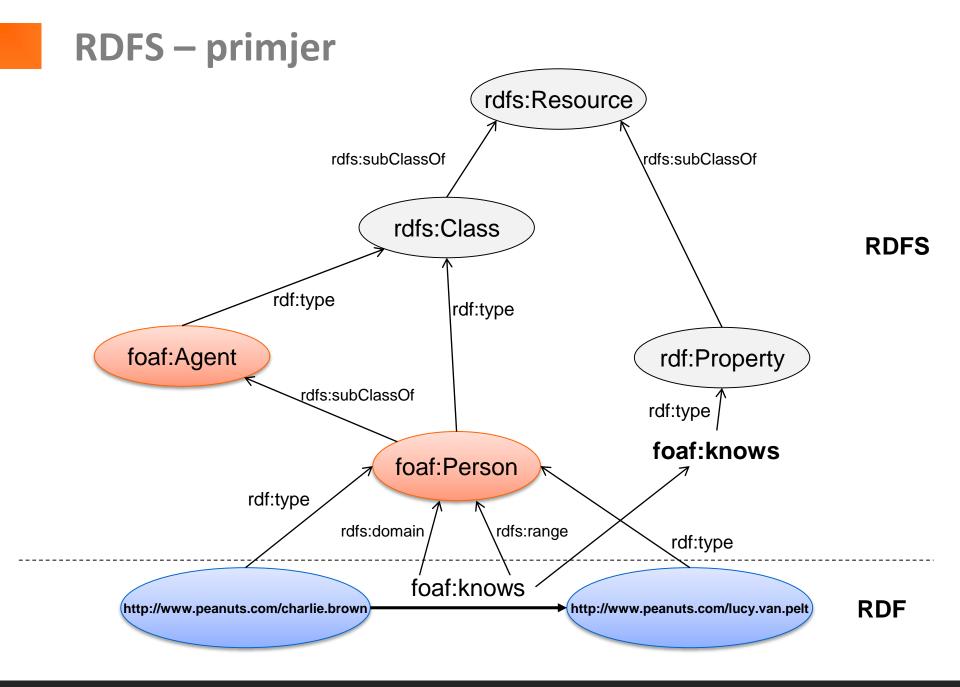
RDFS – primjer

RDFS-om se ne predstavljaju sami podaci, nego znanje o podacima

```
<rdf:Property rdf:about="http://xmlns.com/foaf/0.1/firstName"
    rdfs:label="firstName" rdfs:comment="The first name of a person.">
    <rdf:type rdf:resource = "http://www.w3.org/2002/07/owl#Datatype property"/>
    <rdfs:domain rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
    <rdfs:range rdf:resource="http://www.w3.org/2000/01/rdf-schema#Literal"/>
    <rdfs:isDefinedBy rdf:resource="http://xmlns.com/foaf/0.1/"/>
</rdf:Property>
```

RDFS – primjer





RDFS – primjer zaključivanja

```
<rdfs:Class rdf:ID="vokabular:Student">
        <rdfs:subClassOf rdf:resource="foaf:Person"/>
    </rdfs:Class>
        <rdfs:Class rdf:ID="vokabular:Profesor">
              <rdfs:subClassOf rdf:resource="foaf:Person"/>
        </rdfs:Class>
        <rdf:Property rdf:ID="vokabular:predaje">
              <rdfs:domain rdf:resource="vokabular:Profesor"/>
              <rdfs:range rdf:resource="vokabular:Student"/>
        </rdf:Property>
```

Za vježbu napisati sve trojke. Prije i poslije zaključivanja. <http://www.fer.hr/mirta.baranovic>
<vokabular:predaje>
<http://www.fer.hr/lidia.rovan>

MEHANIZAM ZA ZAKLJUČIVANJE

zaključivan

PRAVILA:

Ako postoji trojka (?x, ?p, ?y) i
(?p, rdfs:range, ?u) onda postoji i
(?y, rdf:type, ?u)

Ako postoji trojka (?x, ?p, ?y) i
(?p, rdfs:domain, ?u) onda postoji i
(?x, rdf:type, ?u)

```
<http://www.fer.hr/mirta.baranovic>
<rdf:type>
<vokabular:Profesor>
<http://www.fer.hr/lidia.rovan>
<rdf:type>
<vokabular:Student>
```