



Univerza v Ljubljani  
Fakulteta za računalništvo  
in informatiko



# Semantični splet



# Uvod

- ▶ semantični splet je razširitev svetovnega spleta skozi standarde, ki jih je postavil W3C
- ▶ definicija W3C pravi, da:
  - »semantični splet zagotavlja skupno ogrodje ki omogoča deljenje in ponovno uporabo podatkov preko meja aplikacij, podjetij in skupnosti«
- ▶ izraz je sproduciral Tim Berners Lee in označuje splet podatkov, katere lahko procesirajo računalniki
  - računalniki morajo znati prebrati pomen podatkov
  - vizija spleta, kjer bi lahko računalniki sprocesirali VSE podatke na spletu: vsebino, povezave in transakcije med računalniki in ljudmi
  - tak splet bo omogoča inteligentne agente

# Uvod

- ▶ semantični splet je (definicija W3C):
  - splet dokumentov (standarden splet)
  - splet podatkov (nadgradnja, ki omogoča semantični splet)
  - omogoča splet povezanih podatkov
    - podatkovna skladišča: podatkovne baze, datoteke, e-pošta, ...
    - besednjak
    - pravila za upravljanje s podatki





# Zakaj semantični splet

- ▶ splet je idealen medij za širjenje podatkov
  - spletne strani
  - zbirke podatkov
- ▶ na spletu obstaja več kot 1.8 milijarde spletnih strani (miljarda je bila presežena septembra leta 2014)
- ▶ težko je najti relevantno informacijo
- ▶ za večino iskanj iskalniki vrnejo nekaj tisoč rezultatov
- ▶ potrebna je pomoč pametnih algoritmov oziroma aplikacij



# Zakaj semantični splet

- ▶ spletni iskalniki
  - splošni: Google, Baidu, Bing, Yahoo
  - specializirani: ScienceDirect
  - iskalniki po zbirkah: RUL, ePrints, ...
- ▶ iskalniki vrnejo rezultate glede na iskani niz
- ▶ ponavadi rezultati vsebujejo tudi relevantno informacijo
- ▶ informacija je lahko skrita med prevelikim številom rezultatov
- ▶ zmanjšati je potrebno število rezultatov in dvigniti relevantnost rezultatov



# Semantični splet

- ▶ koncept semantičnega modela mreže že v 60ih letih, kot način predstavitve semantično strukturiranega znanja
- ▶ ideja o semantičnem spletu že kmalu po »nastanku« svetovnega spleta, v zgodnjih 90ih letih
- ▶ uporaba koncepta na spletu pomeni razširitev hiperpovezanih strani, ki so berljive za človeka, z dodajanjem metapodatkov o straneh in njihovih medsebojnih relacijah, ki so berljivi za stroje,
- ▶ to omogoča inteligentnim agentom inteligentnejši dostop do spleta in opravljanje večjega števila nalog za uporabnike
- ▶ uporabljene so različne tehnologije, ki omogočajo strojem, da razumejo podatke



# Elementi semantičnega spleta

- ▶ metapodatki – strukturirane informacije, opisujejo in omogočajo lažje pridobivanje ali upravljanje virov
  - opisovalni metpodatki: opisujejo vire
  - strukturni metapodatki: opisujejo kako so deli objektov sestavljeni
  - administrativni podatki: dajejo informacijo, ki pomaga pri upravljanju
- ▶ ontologije – dogovorjeno poimenovanje stvari, ki naredi neko stvar razumljivo tudi strojem
- ▶ inteligentni agenti – programi, ki analizirajo metapodatke in ontologije na spletu, na njih lahko preložimo nekatere stvari, sposobni so učenja

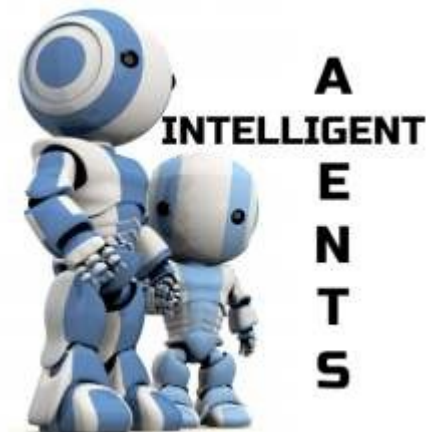
# Programski agenti

- ▶ so programi, ki izvajajo nek nabor operacij v imenu uporabnika ali drugega programa z določeno stopnjo neodvisnosti ali avtonomije in pri tem uporabljajo znanje ali reprezentacijo uporabnikovih ciljev ali želja (IBM)
- ▶ agent je računalniški sistem (Wooldrige) z naslednjimi lastnostmi
  - avtonomnost → deluje brez posredovanja
  - socialna zmožnost → delovanje z drugimi agenti in človekom
  - odzivnost → dojemanje okolja, odzivanje na časovne spremembe (spanje, ...)
  - ciljna orientiranost → lahko obravnavajo kompleksne naloge, zmožni razbitja nalog v podnaloge in pametnega sekvenčnega izvajanja
  - pro-aktivnost → zmožni so ciljno usmerjenega obnašanja in prevzema iniciative
  - časovna kontinuiteta → vseskozi delujoči



# Programski agenti

- ▶ zaželenne lastnosti
  - mobilnost → premikanje po omrežju
  - odkritost → ne razširjajo namerno napačnih informacij
  - naklonjenost → agenti poskušajo izvesti zahtevano nalogo
  - razumnost → delovanje v smeri dosege zastavljenega cilja
  - prilagodljivost → fleksibilnost glede navad uporabnika, delovnim metodam, ...
- ▶ agenti morajo imeti funkcionalnosti, ki omogočajo izpolnitev nalog
- ▶ agenti morajo imeti sposobnost dojemanja sprememb v okolju



# Zakaj semantično označevanje

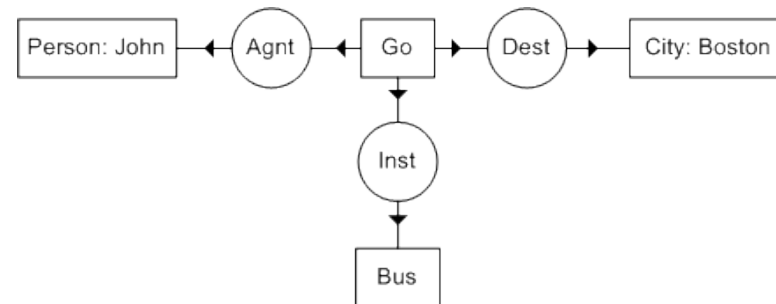
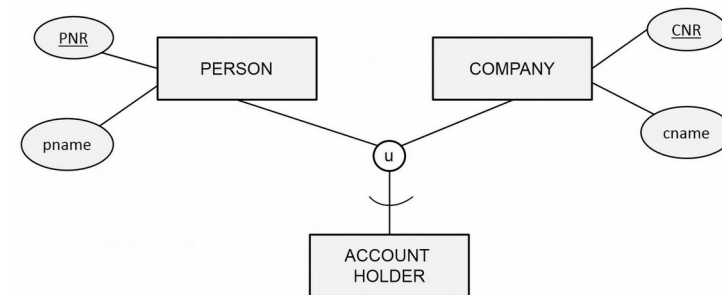
- ▶ agenti morajo ugotoviti če je najdena informacija ustrezna glede na zastavljene cilje
- ▶ v ta namen morajo imeti spletne strani zapis meta-podatkov, ki je primeren za strojno interpretacijo
- ▶ tehnologije, ki omogočajo semantično označevanje
  - SHOE (Simple HTML Ontology Extension) – razširitev HTML
  - DAML + OIL (DARPA Agent Markup Language with Ontology Inference Layer) – sintaksa za opis množice dejstev
  - OWL (Web Ontology Language) – družina jezikov za pisanje ontologij

# Zakaj ontologije

- ▶ ontologije so formalno imenovanje in definicije tipov, lastnosti in povezav entitet, ki obstajajo v določeni domeni
  - združijo znanje ekspertov
  - omogočajo razumljivost in obnovljivost s strani ekspertov
- ▶ primer ontologije za podatkovno bazo je entitetna shema, ki vsebuje tipe, lastnosti in povezave med posameznimi entitetami
- ▶ ontologija razdeli potrebne spremenljivke in vzpostavi razmerja med njimi
- ▶ splošne komponente ontologij vsebujejo
  - razrede → množice, zbirke, koncepte, tipe objektov ali podobno
  - lastnosti → aspekte, značilnosti, karakteristike, ki jih imajo objekti
  - povezave → kako so razredi in posamezniki povezani med sabo

# Ontologije

- ▶ predstavitev opisa znanja
  - relacijski diagram  
(Extended Entity Relationship – EER diagram)
  - konceptualni grafi je grafična notacija za logiko temelječo na grafih in dopolnjeno s komponentami iz lingvistike; omogoča lažje preslikave v in iz naravnih jezikov (John gre v Boston z avtobusom)
    - koncepti: John, iti, Boston, Avtobus
    - konceptualne relacije na katere so pripeti koncepti Agnt (agent), Inst (sredstvo), Dest (cilj)



# Resource Description Framework

- ▶ je podatkovni model, ki se uporablja za predstavitev podatkov na semantičnem spletu
  - je objektno zasnovan
  - omogoča predstavitev podatkov v obliki grafov
  - temelji na trojčku: osebek – predikat – predmet
  - razlikuje med entitetami, ki so predstavljene z identifikatorji, in trditvami, ki veljajo med entitetami → **Danko** (izvorna entiteta/osebek) **je** (trditev/predikat) **študent** (ponorna entiteta/predmet)
  - izvirne entitete in trditve so viri, ki so predstavljeni z URI, ponorne entitete pa so lahko viri ali pa znakovni nizi, ki so veljavni med entitetami



# Primer

- ▶ RDF graf  
Podana trditev

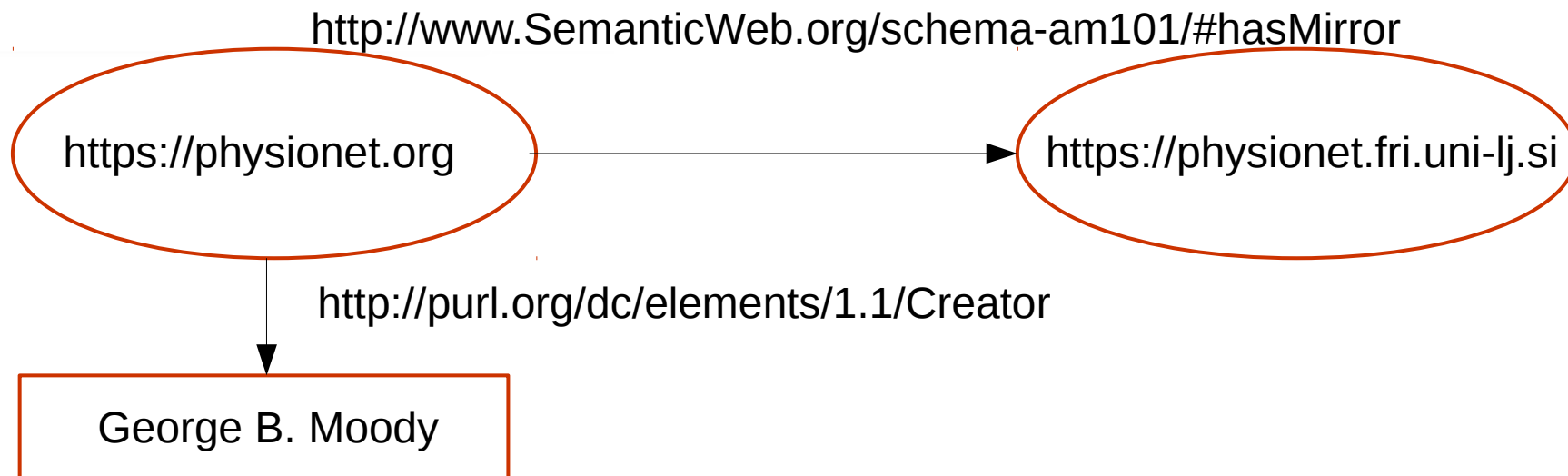
»Physionet ima zrcalni strežnik na [physionet.fri.uni-lj.si](https://physionet.fri.uni-lj.si)«

<http://www.SemanticWeb.org/schema-am101/#hasMirror>



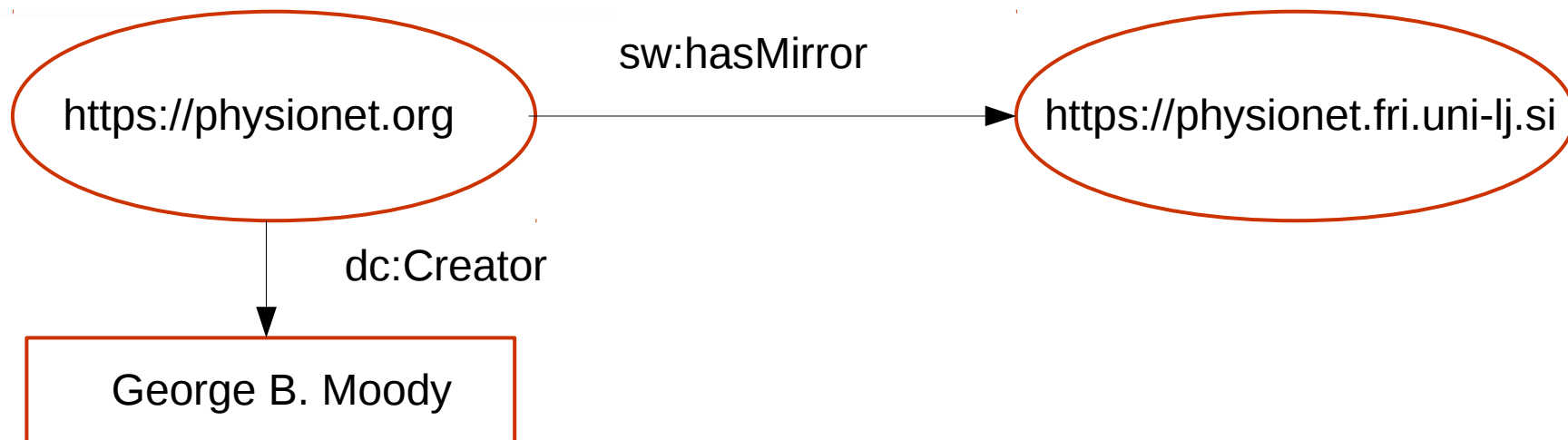
# Primer

- ▶ RDF graf z znakovnim nizom, dodeljenim kot vrednost  
Podana trditev  
»Physionet ima zrcalni strežnik na [physionet.fri.uni-lj.si](https://physionet.fri.uni-lj.si); avtor  
Physioneta je George B. Moody«



# Primer

- ▶ RDF graf z uporabo prostorskih imen (namespace)  
Trditvi (povedka) sta skrajšana in določena z zapisom:  
sw: <http://www.SemanticWeb.org/schema-am101/>  
dc: <http://purl.org/dc/elements/1.1>





# Koda primera

```
<?xml version='1.0'?>
<rdf:RDF
  xmlns:rdf='http://www.w3.org/1999/02/22/rdf-syntax-ns#'
  xmlns:sw='http://www.SemanticWeb.org/schema-dam[0]/#'
  xmlns:dc='http://purl.org/dc/element/1.1/'>
  <rdf:Description about='https://physionet.org'>
    <sw:hasMirror>
      <rdf:Description about='http://physionet.fri.uni-lj.si'>
        </rdf:Description>
      </sw:hasMirror>
      <dc:Creator>George B. Moody</dc:Creator>
    </rdf:Description>
  </rdf:RDF>
```

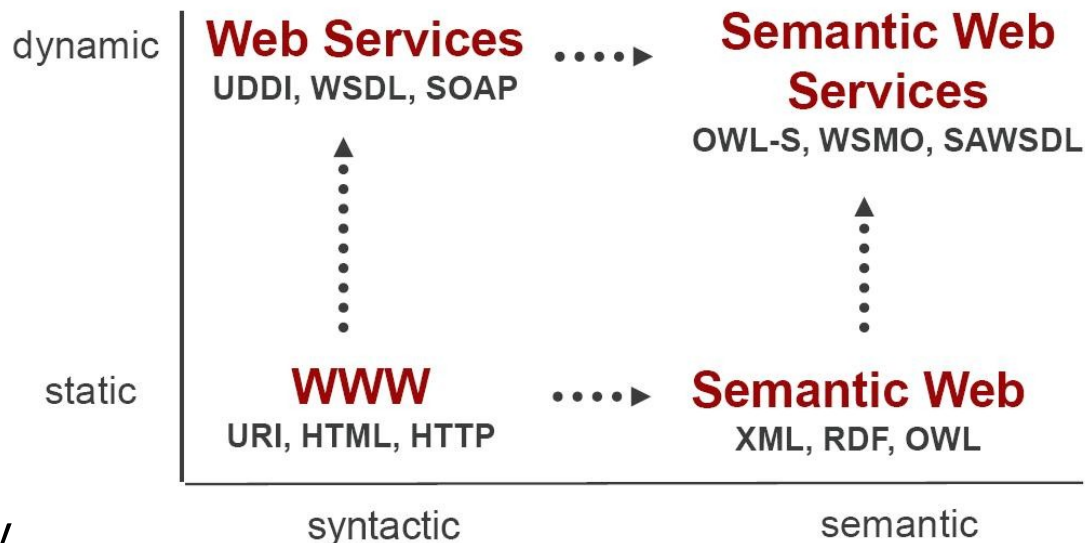


# Semantične spletne aplikacije

- ▶ počasi dosegajo industrijsko uporabo
- ▶ osnovna orodja že obstajajo: komponente in prototipi
- ▶ za delovanje je potrebno definirati:
  - model ontologije
  - sheme specifičnih ontologij za aplikacije
- ▶ malo rešitev z agenti, ki upoštevajo in razumejo semantično označevanje
- ▶ pri ontologijah je potrebno upoštevati naslednje principe
  - 1) vse indentificirati z URI
  - 2) viri in povezave so lahko tipizirani
  - 3) dovoljevati je potrebno nepopolno informacijo
  - 4) nobena trditev ni absolutna resnica
  - 5) možen je razvoj informacij in virov, ki se lahko nadomestijo
  - 6) enostavnost

# Semantične spletne storitve

- ▶ definirati je potrebno ustrezno semantično označevanje
  - objava – omogočanje storitve za uporabo → opis parametrov
  - iskanje – iskanje storitve za določeno nalogo
  - izbira – več podobnih storitev je lahko na voljo
  - sestavljanje – sestavljanje storitev za izvedbo naloge
  - mediacija – posredovanje pri neujemanju podatkov/protokolov
  - izvedba



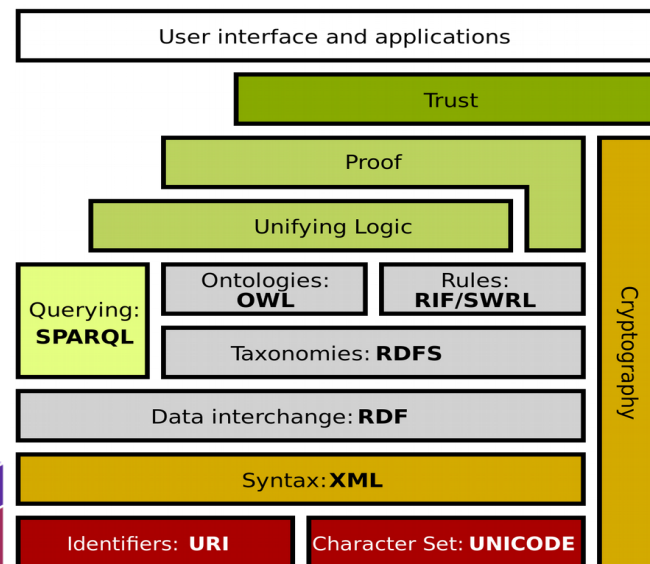
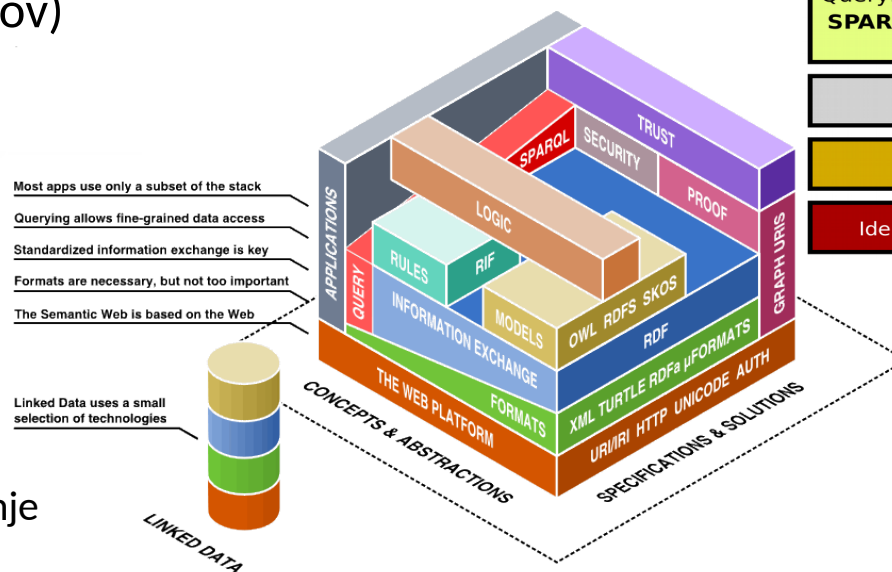


# Semantične spletne storitve

- ▶ potrebno je upoštevati razlike glede na klasične spletne aplikacije
  - prisotnost agentov → informacijski prostor mora podpirati arhitekturo agentov
  - predstavitev znanja → uporabljati je potrebno standardiziran jezik
  - modeliranje znanja → potrebno je uporabljati ontologije za predstavitev znanja
  - razlaga znanja → informacije so zasnovane na ontologiji
  - shranjevanje znanja
  - povpraševanje po znanju
- ▶ uporaba orodij za razvoj agentov, urejanje ontologij in za označevanje/razlago

# Sklad semantičnega spleta

- ▶ Unicode/URI (nabor znakov/identifikator)
- ▶ XML/XML Schema (sintaksa in imenski prostor)
- ▶ RDF (izmenjava podatkov)
- ▶ RDFS (taksonomije)
- ▶ OWL (ontologije)
- ▶ SWRL/RIF (pravila)
- ▶ SPARQL (poizvedbe)
- ▶ višje plasti
  - poenotena logika
  - sklepanje in dokazovanje
  - zaupanje
  - kriptografija





# Ontologije , OWL in RDF

- ▶ ontologije omogočajo
  - organizacijo podatkov
  - izboljšanje iskanja → iskanje avtomobila (car:focus country:Germany)
  - integracijo podatkov → povezovanje raznolikosti jezikov (študent, student, ...)
- ▶ ontologije s primeri omogočajo formalni opis področja
  - niz posameznikov (entitete, objekti)
  - razredi posameznikov
  - relacije med posamezniki
- ▶ trditve o pripadnosti posameznikov v razredih so v bazi dejstev (podatkovni bazi)
  - za določanje znanja o razredih in relacijah
  - določajo omejitve in tvorijo bazo znanja

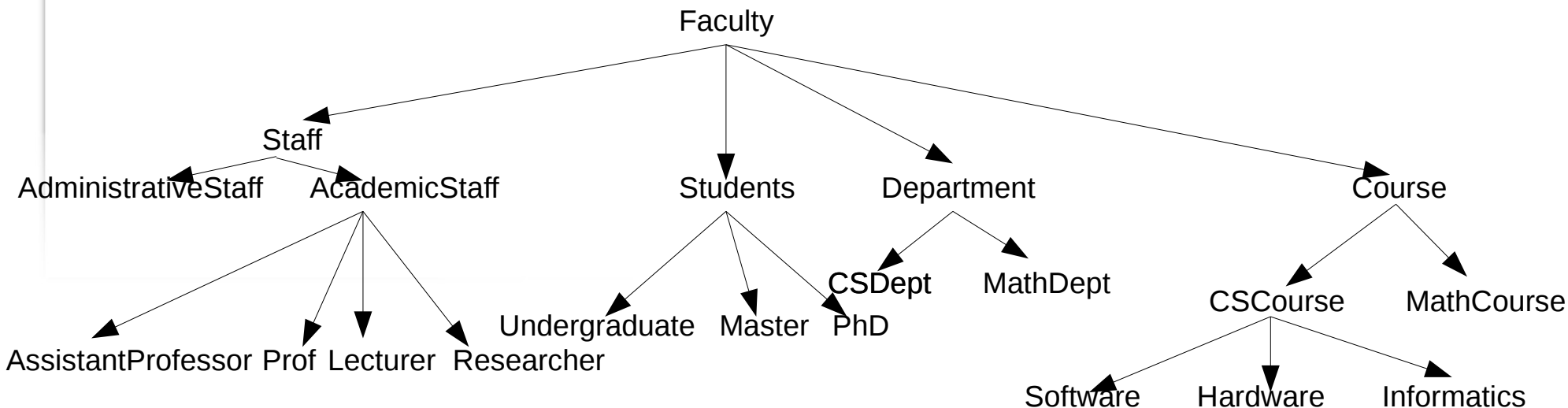


# Primer → ontologija univerze

- ▶ izrazi ontologije imajo predpono »:« za posameznike (:Smerdel) in za razrede (:Students)
- ▶ Razredi so:  
:Staff, :Department, :Students, :Course
- ▶ razredi imajo lahko podrazrede, podrazredi razreda :Staff so:  
:AdministrativeStaff in :AcademicStaff (:Professor, :AssistantProfessor, :Researcher, :Lecturer, ...)
- ▶ Tako imamo na primer »Smerdel je docent«
  - :Smerdel je primerek razreda :AssistantProfessor



# Primer → hierarhija razredov



- ▶ razred C je podrazred razreda C' če je vsaka instanca razreda C tudi instanca iz razreda C'
- ▶ Primer  
:Smerdel je instanca razreda :AssistantProfessor, je instanca :AcademicStaff in instanca :Staff



# Primer → razmerja med razredi

- ▶ prikazujejo naravna razmerja med osebki v resničnem svetu
- ▶ ontologija univerze vključuje razmerje :TeachesIn
- ▶ Razmerja združujejo primerke, na primer  
:TeachesIn(:Smrdel, :WT) → pomeni Smrdel predava predmet WT
- ▶ določanje področja razmerij
  - TeachesIn(:AcademicStaff, :Course) za :TeachesIn(:X, :Y) velja, da pripada X :AcademicStaff in Y :Course
  - Leads(:Staff, :Department)
- ▶ omejitve
  - razreda :Student in :Staff nista povezana → nimata skupnega posameznika
  - vsak oddelek ima enoličnega vodjo
  - samo predavatelji, docenti in profesorji lahko poučujejo



# Primer → primerki

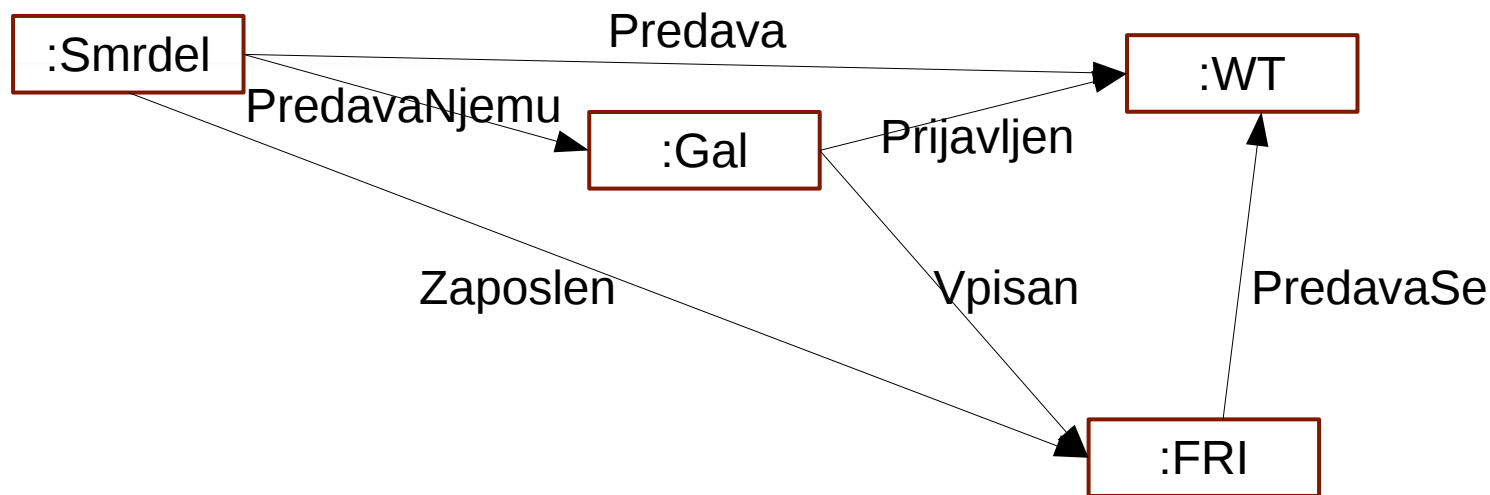
- ▶ formalna semantika
  - dejstvo: Smrdel predava predmet WT
  - Logično sklepanje iz ontologije je
    - :Smrdel je v razredu :AssistantProfessor
    - podobno izpeljemo tudi druge trditve → :CSDept je v razredu :Department
- ▶ sklepanje je pomembno pri odgovorih na poizvedbe
- ▶ recimo, da nas zanima, kdo je član akademskega osebja, ki živi v Kranju
  - predpostavimo, da Smrdel živi v Kranju
  - zakaj bo njegovo ime v odgovoru?
  - ker živi v Kranju in je docent

# Resource Description Framework → RDF

- ▶ RDF je družina W3C specifikacij
- ▶ uporablja se kot splošna metoda za konceptualni opis ali modeliranje informacij, ki je implementirana v spletnem viru
- ▶ omejitve dejstev v posamezni domeni določene v RDFS ali OWL
- ▶ RDF temelji na trojčku <osebek, povedek, predmet> → relacije določene s predikatom (povedkom) med osebkom in predmetom
  - <:Smrdel :Predava :WT>      Os → :Smrdel, Pov → :Predava,      Pred → :WT
  - <:Smrdel :PredavaNjemu :Gal>      Os → :Smrdel, Pov → :PredavaNjemu,      Pred → :Gal
  - <:Gal :Vpisan :FRI>      Os → :Gal,      Pov → :Vpisan,      Pred → :FRI
  - <:Gal :Prijavljen :WT>      Os → :Gal,      Pov → :Prijavljen,      Pred → :WT
  - <:WT :PredavaSe :FRI> →      Os → :WT,      Pov → :PredavaSe,      Pred → :FRI

# RDF graf

- ▶ množica dejstev določa razmerja med objekti
- ▶ dejstva se lahko predstavijo z grafom, kjer so vozlišča osebki in predmeti, predikati pa povezave med vozlišči





# RDF graf

- ▶ primer ontologije za študente:  
[www.cs.umd.edu/projects/plus/SHOE/cs.html](http://www.cs.umd.edu/projects/plus/SHOE/cs.html)

Person

Worker

Faculty

Professor

AssistantProfessor

AssociateProfessor

FullProfessor

VisitingProfessor

Lecturer

PostDoc

Assistant

ResearchAssistant

TeachingAssistant

AdministrativeStaff

Director

Chair {Professor}

Dean {Professor}

ClericalStaff

SystemsStaff

Student

UndergraduateStudent

GraduateStudent

Organization

Department

School

University

Program

ResearchGroup

Institute

Publication

Article

TechnicalReport

JournalArticle

ConferencePaper

UnofficialPublication

Book

Software

Manual

Specification

Work

Course

Research

Schedule



# Semantična spletna stran

- ▶ planiranje
  - implementacija API-ja za spletno stran
  - pretvorba strani v API
- ▶ analiza podatkov → ontologije: Semantically Interlinked Online Comm. ter Friend Of A Friend
- ▶ označevanje strukturiranih podatkov:
  - ljudje, organizacije → hCard, RDF vCard
  - dogodki, koledarji → hCalendar, RDF Calendar
  - menja, ocene → Votelinks, hReview
  - socialna omrežja → FOAF
  - značke, ključne besede, kategorije → rel-tag
  - sezname, pregledi → XOXO

# Prehod na semantično spletno stran

- ▶ spodaj je nesemantična stran
- ▶ uporabimo hCard mikroformat
- ▶ lahko uporabimo vtičnike

```
<!doctype html>
<html>
<head>
  <title></title>
  <meta charset="utf-8">
</head>
<body>
  <div class="kontakt">
    Aleš Smrdel. Docent na
    <a href="https://www.fri.uni-lj.si">
      FRI</a>.
    Kontaktirate me lahko preko
    <a href="mailto:ales.smrdel@fri.uni-lj.si">
      emajla</a>
    ali na telefon 01 1111 111.
  </div>
</html>
</body>
```

```
<!doctype html>
<html>
<head>
  <title></title>
  <meta charset="utf-8">
  <link rel="profile"
    href="http://microformats.org/profile/hcard">
</head>
<body>
  <div id="hcard-AS" class="vcard">
    <a class="url fn" href="https://www.fri.uni-lj.si">
      Aleš Smrdel</a>
    <a class="email"
      href="mailto:ales.smrdel@fri.uni-lj.si">emajl</a>
    <div class="adr">
      <div class="street-address">Večna pot 113</div>
      <span class="locality">Ljubljana</span>
      <span class="region">Osrednje slovenska</span>
      <span class="postal-code">1310</span>
      <span class="countr">Slovenija</span>
    </div>
    <div class="tel">01 1111 111</span>
  </div>
</body>
</html>
```

# RDFa

- ▶ RDFa je W3C priporočilo, ki dodaja množico atributov dokumentnim tipom HTML, XHTML in nekaj XML tipom
- ▶ omogoča vgradnjo metapodatkov znotraj spletnega dokumenta
- ▶ RDFa Lite kot standardizirana alternativa mikropodatkom (specifikacija za gnezdenje metapodatkov na spletnih straneh)
- ▶ izbira slovarja:

```
<p vocab="http://schema.org"> Moje ime je Danko</p>
```

- ▶ določitev tipa

```
<p vocab="http://schema.org" typeof="Person"> Moje ime je Danko</p>
```

- ▶ določitev lastnosti

```
<p vocab="http://schema.org" typeof="Person"> Moje ime je  
<span property="name">Danko</span></p>
```





# Semantični splet

- ▶ veliko o semantičnem spletu lahko zveste tudi na:  
<https://www.slideshare.net/marinasantini1/lecture-ontologies-and-the-semantic-web>
- ▶ dober vir je tudi:  
<http://www.cambridgesemantics.com/semantic-university/example-semantic-web-applications>
  - kaj je pomembno za dobro semantično aplikacijo
  - primeri uporabe semantičnih tehnologij
    - Microsoft, Google in Yahoo uporabljajo Schema.org, ki ima RDFa predstavitev
    - Ecommerce strani uporabljajo GoodRelations, ki tudi uporablja RDFa