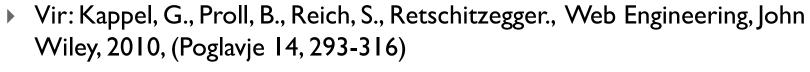
Spletne tehnologije, UL, FRI (VSP) ST 11 – Semantični splet

doc.dr. Mira Trebar

Semantični splet

- Uvod
- Pridobivanje informacij
- Osnove semantičnega spleta
- Tehnološki koncept
- Ontologije
- ▶ RDF Resource Definition Framework
- RDF koda in RDF shema
- Semantične spletne aplikacije
- Ontologije RDF (primeri)



Ucilnica; Semantic web.pdf



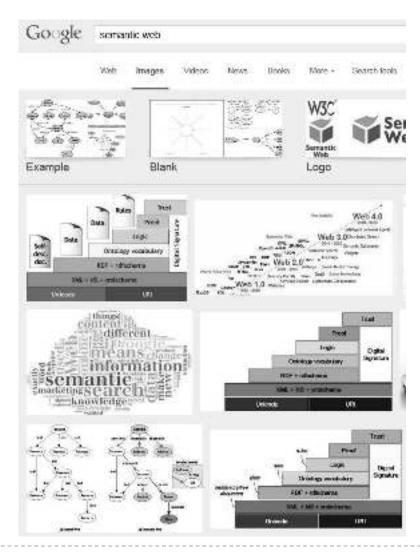
Uvod

- Semantični splet definicija:
 - i-slovar: zamisel Tima Berners-Leeja, avtorja svetovnega spleta, kako splet narediti prijaznejši uporabniku
 - http://www.lavbic.net/delo-in-raziskovanje/semanticni-splet/ (2010)
 - izboljšati trenutni splet, v tej meri, da bodo računalniki lahko procesirali podatke na spletu, jih interpretirali, povezali in na takšen način pomagali ljudem pri iskanju zahtevanega znanja.
 - namen semantičnega spleta je zagotoviti veliko porazdeljeno bazo znanja
 - vzpostavitev skupnega ogrodja, ki bi omogočalo skupno rabo podatkov in ponovno uporabo med aplikacijami, podjetji.
 - skupno sodelovanje, ki ga vodi W3C
- ▶ W3C: http://www.w3.org/standards/semanticweb/
- Semantic web: http://semanticweb.org/wiki/Main_Page

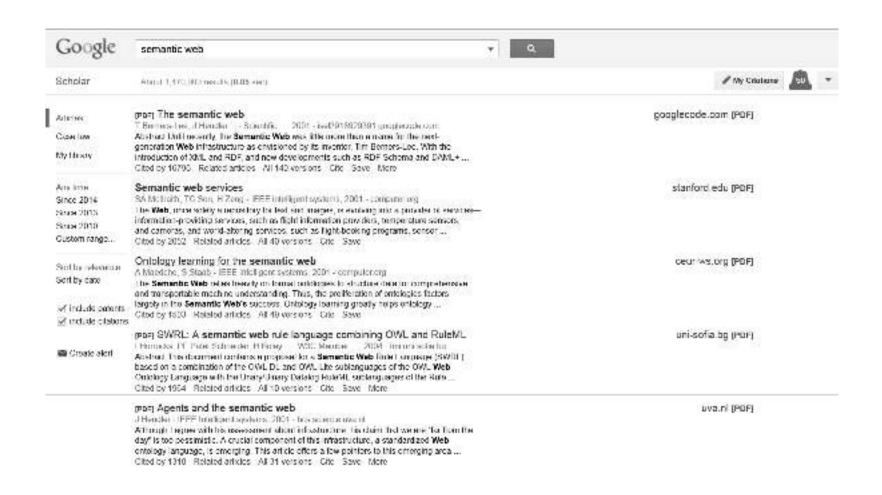
Pridobivanje informacij

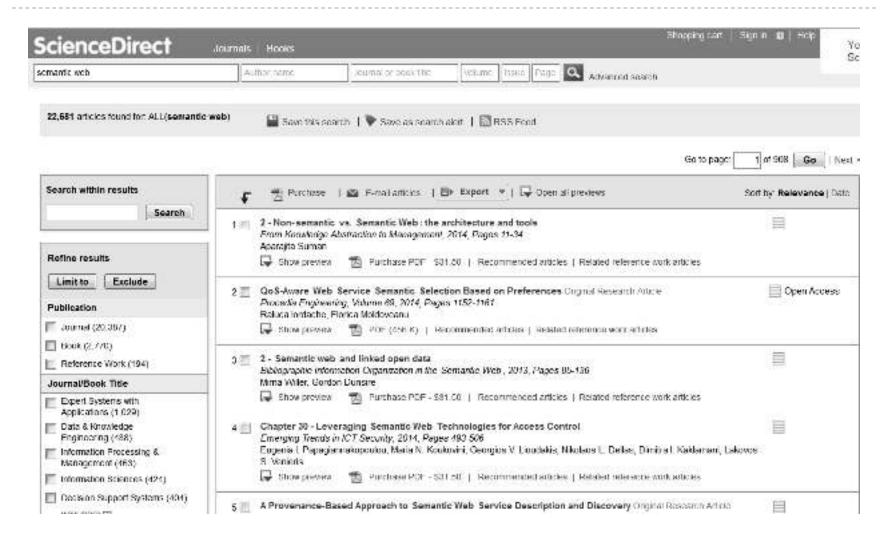
- Informacije:
 - Spletne strani
 - Zbirke podatkov in dokumentov
- Na svetovnem spletu mnogo preveč vsebin, da bi ljudje lahko našli ustrezno/koristno informacijo brez pomoči inteligentnih strojev.
- Spletni iskalniki vrnejo tisoče/sto-tisoče/... odgovorov na eno povpraševanje. Zbirke spletnih iskalnikov omogočajo:
 - avtomatsko odkrivanje in obdelovanje (indeksiranje) dokumentov
 - > shranjevanje samo podatkov o dokumentih, ali pa tudi dokumente same.
- Spletni iskalniki:
 - Google splošen iskalnik, Google Scholar akademsko gradivo
 - Science Direct znansvene objave
 - Zbirke dokumentov: DiKUL, ePrints.FRI (http://eprints.fri.uni-lj.si/),
- ▶ Cilji: Manj rezultatov Več pomena





Google Scholar





Semantični splet

Sestavljajo ga:

- Pomensko označevanje ('semantic mark-up') dobavitelji informacij, ki generirajo spletne vsebine, bodo morali pripraviti semantično označene spletne strani;
- Inteligentni programski agenti ('intelligent software agents') 'programi', ki iščejo in obdelujejo semantično označene spletne strani;
- ▶ Ontologije ('computational ontologies') skupno dogovorjeno razumevanje stvari, ki naredi vsebino razumljivo tudi strojem.

Tehnologije:

- XML (Extensible Markup Language)
- RDF (Resource Description Framework)
- OWL (Web Ontology Language)

Vloga programskih agentov

Agent – je programsko zasnovan računalniški sistem (Michael Wooldridge):

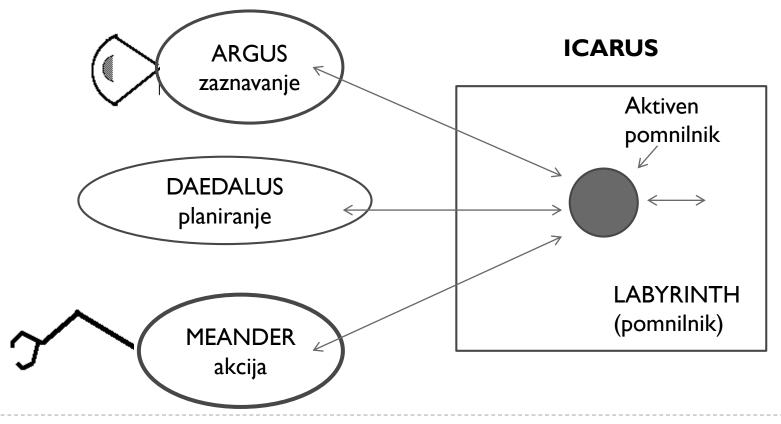
- Avtonomnost delujejo brez posredovanja človeka.
- ▶ **Socialna zmožnost** delujejo vzajemno z drugimi agenti in človekom v neki obliki agentovega komunikacijskega jezika.
- ▶ **Reaktivnost/odzivnost** dojemanje okolja in odzivanje na spremembe v časovni obliki. (spanje, delovanje)
- ▶ Pro-aktivnost sposobni so izražanja ciljno usmerjenega obnašanja s prevzemom iniciative.
- ▶ Časovna kontinuiteta vseskozi delujoči procesi (aktivni v ospredju, ali pasivni v ozadju).
- ▶ Ciljna orientiranost sposobni so obravnavati kompleksne naloge, znajo odločati kdaj je potrebno nalogo razdeliti v podnaloge, v kakšnem vrstnem redu in kako jih izvajati.

Druge zaželene lastnosti so:

- Mobilnost sposobnost agenta, da se premika po omrežju.
- Odkritost predpostavka, da agent ne bo namerno razširjal napačnih informacij.
- ▶ **Naklonjenost** predpostavka, da bo vsak agent vedno poskušal izvesti tisto kar se zahteva od njega.
- ▶ Razumnost predpostavka, da bo vsak agent vedno deloval tako da bi dosegel svoj cilj.
- Prilagodljivost zmožnost prilagoditve navadam uporabnika, delovnim metodam in izbiri.
- Agenti morajo imeti tudi:
 - pomembne funkcionalnosti, da lahko izpolnijo omenjene zahteve.
 - sposobnosti za dojemanje sprememb v okolju.

Primer: zgodnja arhitektura agenta 'ICARUS' (Langely, 1991)

- Štirje moduli: ARGUS, DAEDALUS, MEANDER, LABYRINTH
- Funkcionalnosti, ki v agentu delujejo vzajemno



Vloga semantičnega označevanja

ZAKAJ?

Programski agenti morajo ugotoviti ali je najdeni del informacije uporaben za dani primer.

KAKO?

W3C je določil, da morajo spletne strani imeti **zapis meta-podatkov**, ki je primeren za programsko interpretacijo tako, da bo uporaben za semantični splet.

- Jeziki za semantično označevanje:
 - ▶ SHOE (Simple HTML Ontology Extension), 1997 enostavna struktura
 - DAML+OIL (DARPA Agent Markup Language with Ontology Inference Layer)
 - ▶ OWL Web Ontology Language, 2002

Vloga ontologij

Ontologije so ustvarjalci pojmov

- > za abstraktne izraze, ki
- združijo znanje ekspertov v takšno obliko,
- da jo lahko razumejo in obnovijo eksperti.

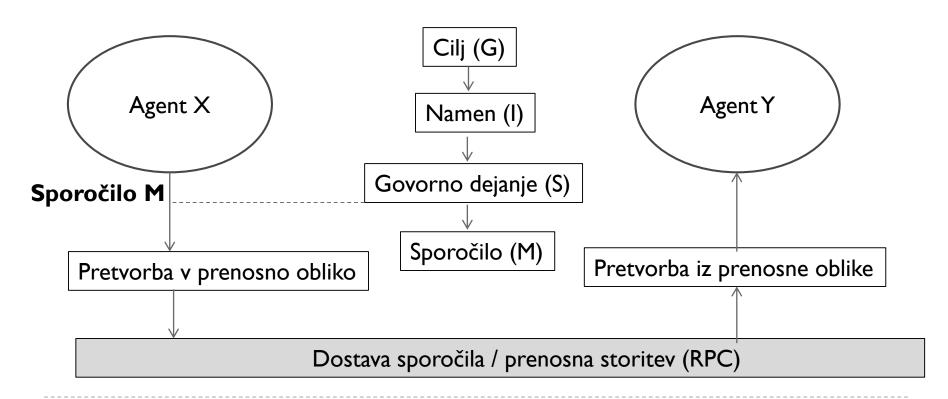
Primer:

Podatkovne baze imajo nedvoumno ontologijo, imenovano shema podatkovne baze. Takšna shema je implementirana s specifičnimi aplikacijami.

- Mnenje v praksi: Mnoge ontologije so samo nadzorovani slovarji.
- Ontologija
 - predstavlja graf pomenskih povezav med pojmi in
 - je zapisana v računalniško razumljivi obliki.
- Ontologije so uporabljene kot:
 - Globalne integracijske sheme
 - Nadzorovan slovar (meta-podatki)

Tehnološki koncepti

- Agenti FIPA standard (1996, http://www.fipa.org/)
 - FIPA 'Foundation for Intelligent Physical Agents'
 - Vir:AgentLink (2005, http://www.agentlink.org/index.php)
- Primer: Model komunikacije med agentoma X in Y



Ontologije - predstavitve opisa znanja

- EER diagrami (Extended Entity Relationship)- relacijski diagrami http://www.lightenna.com/sitefiles/lightenna.com/eer_simple_diagram_big.jpg
- Konceptualni grafi (koncept, konceptualna relacija)
 http://users.bestweb.net/~sowa/ontology/ontoshar.htm

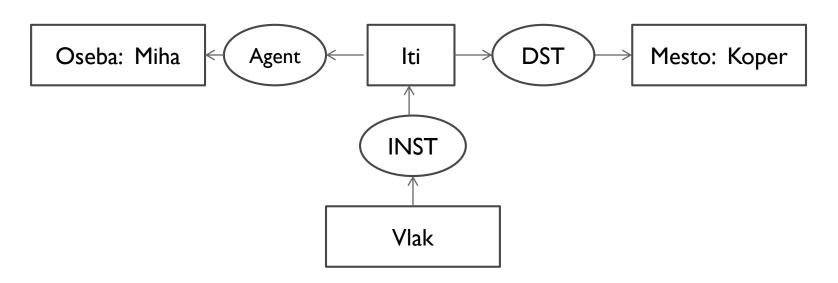
Semantično označevanje na spletu

- Opisni jezik OWL Web Ontology Language hierarhija razredov in podrazredov za objekt (entiteto), npr. računalnik.
 - http://thesemanticway.wordpress.com/2008/11/11/owl-ontology-example/
- ▶ RDF (Resource Description Framework) povezava virov spletnih podatkov za pojavljajoči semantični splet.

Ontologije – formalna predstavitev

- Predstavitev človeškega znanja
- Konceptualni graf: 'Miha gre z vlakom v Koper.'
 - Koncepti:
 - Miha, Iti, Koper, Vlak
 - Koncepti so pripeti na tri konceptualne relacije:

Agent, INST (Instrument), DST (Destination)

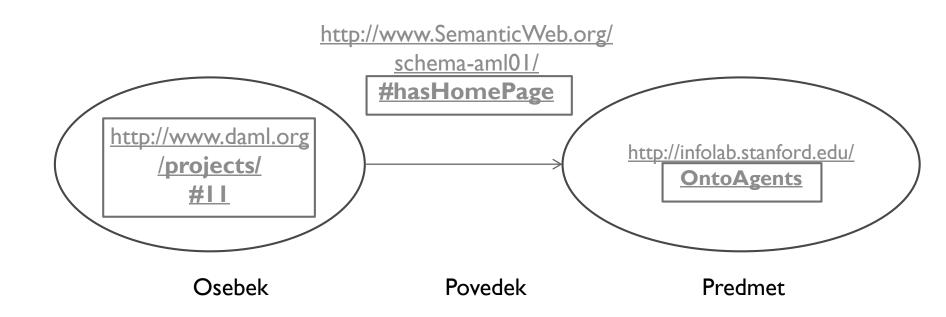


RDF- Resource Description Framework

- Podatkovni model RDF
 - Objektno-zasnovan model
 - Funkcionalen podatkovni model
 - ▶ RDF razlikuje med:
 - entitetami, ki so predstavljene z enoličnimi identifikatorji (UID) in
 - trditvami, ki so veljavne med entitetami.
 - Trditev povezuje osebek (izvorna entiteta) in predmet (ponorna entiteta) s povedkom/lastnostjo (Primer: Janez je študent.)
 - RDF razlikuje med:
 - viri, ki so predstavljene z URI in
 - literali (znakovni niz), ki so veljavni med entitetami.
 - Osebki in povedki so viri
 - Predmet pa je lahko vir ali literal.

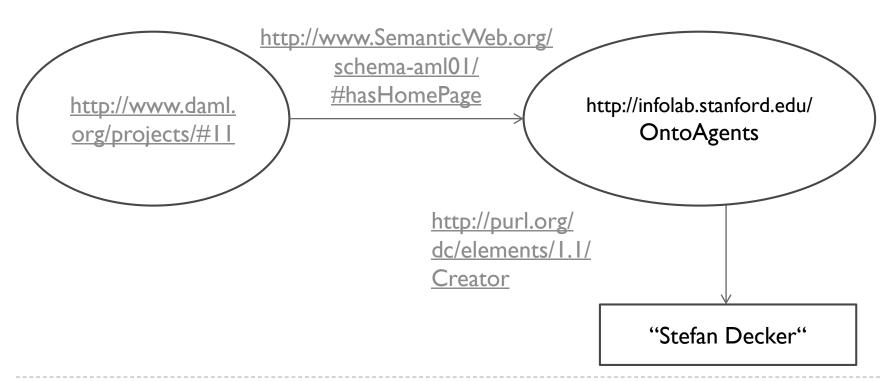
- ▶ RDF graf z osebkom, povedkom in predmetom.
- V shemi je trditev:

"The OntoAgents project has a homepage at (http://infolab.stanford.edu/OntoAgents/)."

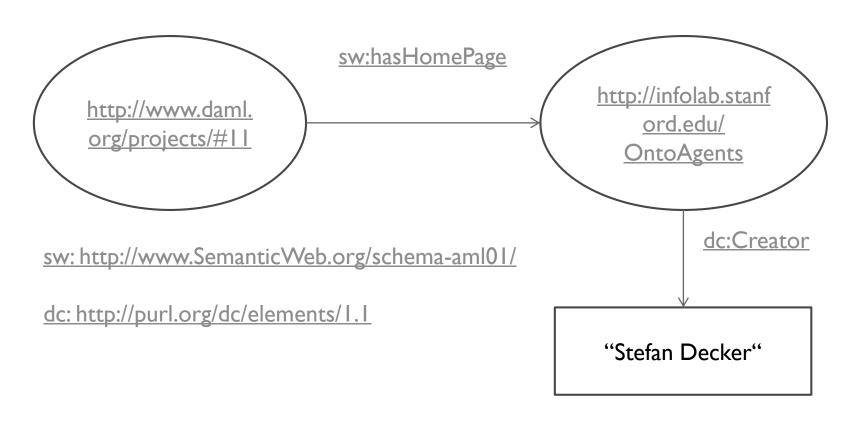


M.Trebar

- ▶ RDF graf z literalom (znakovni niz) dodeljenim kot vrednost.
- "The OntoAgents project has a homepage at (http://infolab.stanford.edu/OntoAgents/)
 and the homepage was created by Stefan Decker."



- ▶ RDF graf z uporabo prostorskih imen.
- Povedki so skrajšani



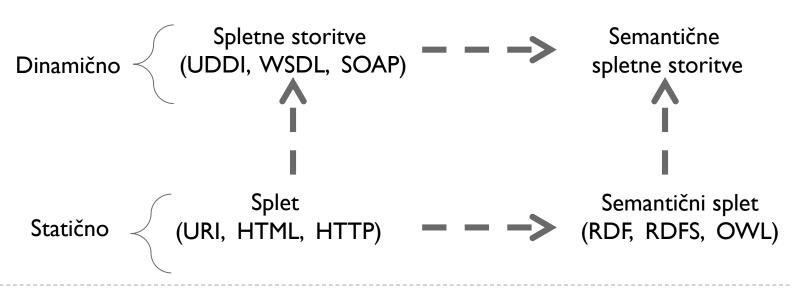
RDF koda (prejšnja prosojnica)

```
<?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-dam101/#"
xmlns:dc = "http://purl.org/dc/element/1.1/">
<rdf:Description
 about="http://www.daml.org/projects/#II">
   <sw: hasHomepage>
     <rdf: Description about=
       "http://www-db.stanford.edu/OntoAgents">
       <dc:Creator>Stefan Decker</dc:Creator>
     </rdf:Description>
   </sw:hasHomepage>
  </rdf:Description>
</rdf:RDF>
```

Semantične spletne aplikacije

- ▶ 2006- še niso dosegle industrijske uporabe
- Obstajajo osnovne komponente in prototipi
- Semantično označevanje:
 - Definirati dober model ontologij,
 - Razlagalci specifičnih ontologij za aplikacije
- Agenti malo pravih sistemov agentov, ki razumejo semantično označevanje
- Ontologije- upoštevati šest principov:
 - Vse je potrebno identificirati z URI
 - 2. Viri in povezave so lahko tipizirani-strojno berljiv način
 - 3. Tolerirati je potrebno nepopolno informacijo

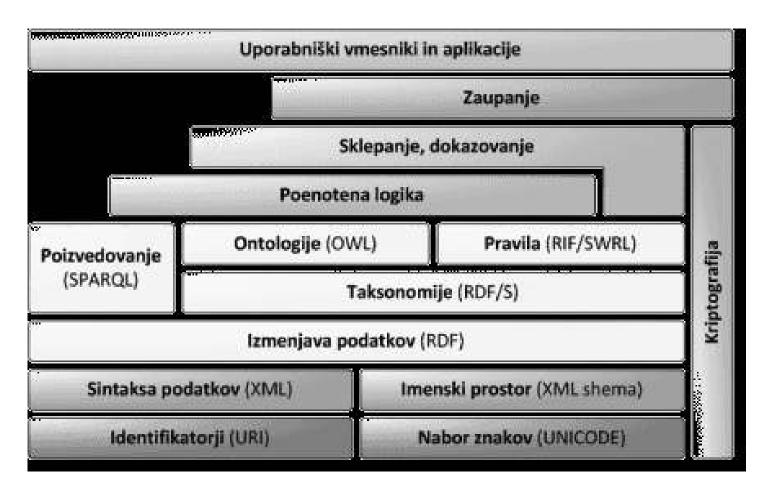
- 4. Nobena trditev ne velja za absolutno resnico
- 5. Možna je evolucija/razvoj informacij in virov
- 6. Enostavnost W3C iniciativa za razvoj enostavnih mehanizmov
- Semantične spletne storitve
 - Definirati semantično označevanje za spletne storitve http://www.w3.org/2004/Talks/0312-rdf-status/



- Vključitev v spletno inženirstvo upoštevati razliko od klasičnih spletnih aplikacij:
 - Agenti Informacijski prostor mora podpirati arhitekturo programskih agentov
 - Predstavitev znanja standardiziran jezik
 - Modeliranje znanja
 - Razlaganje znanja
 - Shranjevanje znanja in povpraševanje po znanju
- Orodja:
 - Ogrodja za razvoj agentov (Agent Frameworks)
 - Urejevalniki ontologij (Ontology Editors)
 - Orodja za zapisovanje (Annotation Tools)

Arhitektura semantičnega spleta

http://www.lavbic.net/delo-in-raziskovanje/semanticni-splet/



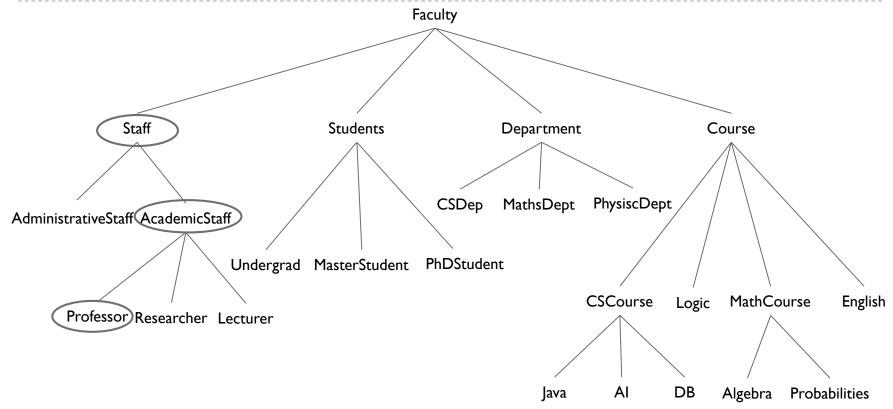
Primer: Ontologije, RDF in OWL

- Vir: Web Data Management http://webdam.inria.fr/Jorge/html/wdmch8.html#x13-1580007.1
- Ontologije zagotavljajo:
 - Organizacijo podatkov: predmeti na univerzi po tematiki, ali po oddelku
 - ▶ Izboljšano iskanje: primer iskanja avtomobila (car:jaguarcountry:USA)
 - Integracijo podatkov: povezuje raznolikost jezikov (študent, student,...)
- Ontologije s primeri Formalni opis področja:
 - niz posameznikov (entitete, objekti),
 - razredi posameznikov
 - obstoječe relacije med temi posamezniki.
- Logične trditve o pripadnosti posameznikov v razredih se nahajajo v bazi dejstev (podatkovna baza). Uporabljajo
 - Uporabljajo se za določanje znanja o razredih in relacijah.
 - Določajo omejitve in tvorijo bazo znanja.

Primer: Ontologija univerze (Faculty)

- Razredi označujejo koncepte, ki so:
 - razumljivi na vseh univerzah
 - po vsem svetu
- Razredi, kjer dvopičje (:) označuje razred
 - :Staff
 - ▶ :Department
 - > :Students
 - :Course
- Podrazredi od razreda :Staff:
 - :AdministrativeStaff
 - :AcademicStaff (:Professor, :Researcher, :Lecturer)
- Primer: Dupond je profesor.
 - ▶ :Dupond je primerek razreda :Professor

Ontologije – hierarhija razredov



 Razred C je podrazred razreda C', če je vsak primerek iz razreda C tudi primerek iz razreda C'

: Dupond je primerek razreda :Professor, potem je primerek razreda :AcademicStaff in razreda :Staff.

28

Ontologije – relacije med razredi

- Predstavitev naravnih relacij med posamezniki v realnem svetu
- Relacije med razredi:
 - TeachesIn(:AcademicStaff, :Course),
 če rečemo "X :TeachesInY", X pripada :AcademicStaff in Y :Course,
 - ▶ Primer: Dupond poučuje CSI0I →TeachesIn(:Dupond,:CSI0I)
- Označevanje vrste udeležencev v različnih povezavah:
 - :TeachesTo(:AcademicStaff,:Student),
 - :Leads(:Staff,:Department).
- Druge omejitve:
 - Razreda :Student in :Staff nimata skupnega posameznika ('disjoint').
 - Vsak oddelek (:Department) ima enoličnega vodjo.
 - Omejitev področja: samo profesorji ali predavatelji lahko poučujejo (:Professor, :Lecturer).

Ontologije - primerki

- Formalna semantika:
 - Dejstvo: Dupond je profesor in vodi oddelek CS.
 - Logično sklepanje, ki sledi:
 - :Dupond je v razredu :Professor
 - :CSDept je v razredu :Department.
- Sklepanje je pomembno pri poizvedbah
- Primer: »Kdo so člani akademskega osebja, ki živijo v Parizu?
 - Predpostavimo, da profesor : Dupond živi v Parizu.
 - ▶ :Dupond se bo nahajal v odgovoru, ker živi v Parizu in je profesor.

Spletni viri, URI, imenski prostor

- Spletni viri (Web resources):
 - spletna stran, del dokumenta, spletna storitev
 - identifikator identitete
 - stvar (predmet), koncept, lastnost
- URI je URL, do katerega dostopa vsak agent (človek, program) ali aplikacija
 - URL ontologije:

http://Webdam.inria.fr/Jorge/OntologiesChapter/Examples#

Primerek Dupond:

http://Webdam.inria.fr/Jorge/OntologiesChapter/Examples#Dupond

- Imenski prostori:
 - ▶ RDF: http://www.w3.org/1999/02/22-rdf-syntax-ns#
 - ▶ RDFS: http://www.w3.org/2000/01/rdf-schema#
 - **...**

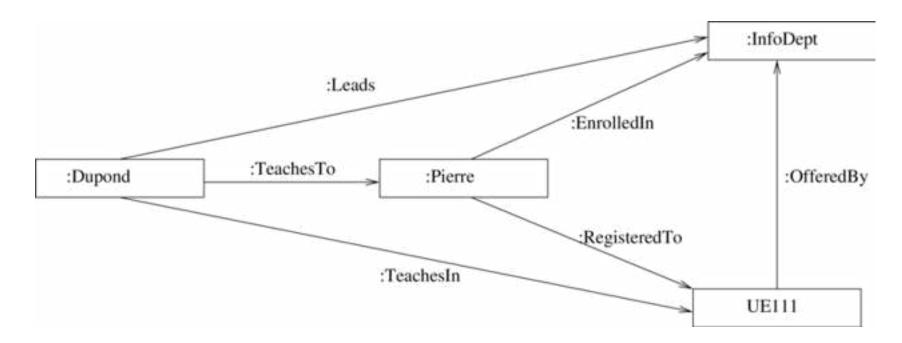
- Omejitve dejstev v posameznih domenah so določene v RDFS ali OWL
- RDF trojčki (relacije določene s predikatom med osebkom in povedkom):
 - Osebek (subject)
 - Povedek (predicate)
 - Predmet (object)
- (a P b) uporabi povedek P in pove, da ima osebek a vrednost b za lastnost P.

Tabela trojčkov RDF	Osebek	Povedek	Predmet
(:Dupond:Leads:InfoDept)	:Dupond	:Leads	:InfoDept
(:Dupond:TeachesIn:UEIII)	:Dupond	:TeachesIn	:UEIII
(:Dupond:TeachesTo:Pierre)	:Dupond	:TeachesTo	:Pierre
〈 :Pierre :EnrolledIn :InfoDept〉	:Piere	:EnrolledIn	:InfoDept
(:Pierre:RegisteredTo:UEIII)	:Piere	:RegisteredTo	:UEIII
〈:UEIII:OfferedBy:InfoDept〉	:UEIII	:OfferedBy	:InfoDept

32

RDF - graf

- Množica dejstev RDF določa:
 - Množico relacij med objekti
 - Graf RDF, kjer so vozlišča osebki in predmeti



Primeri ontologij

- Example Computer Science Department Ontology http://www.cs.umd.edu/projects/plus/SHOE/cs.html
- OWL and Ontologies
 http://courses.ischool.berkeley.edu/i290-14/s05/lecture-27/allslides.html
- Dbpedia infrastruktura za semantični graf znanja
 - http://wiki.dbpedia.org/
 - Ontologija: http://wiki.dbpedia.org/services-resources/ontology
 - Razredi: http://mappings.dbpedia.org/server/ontology/classes/
- SIOC (Semantically-Interlinked Online Communities)
- ► FOAF (Friend of a Friend): http://xmlns.com/foaf/spec/
- ▶ DOAP (Description of open-source projects), 2005

Semantična spletna stran

- Vir: http://www.ibm.com/developerworks/library/x-plansemantic/#c2
- Planiranje semantične spletne strani:
 - Programiranje programskega vmesnika (API) za spletno stran
 - ▶ Pretvorba spletne strani v API (način strukturiranja podatkov XML)
- Analiza podatkov: v povezavi z ontologijami (SIOC, FOAF)
- Informacijski tipi označevanje strukturiranih podatkov
 - Ljudje in organizacije: hCard, RDF vCard
 - ▶ Koledarji in dogodki: hCalendar, RDF Calendar
 - Mnenja, Ocene in pregledi: VoteLinks, hReview
 - Socialna omrežja: XFN, FOAF
 - Licence: rel-licence
 - Značke, ključne besede, Kategorije: rel-tag
 - Seznami in pregledi: XOXO

Nestrukturirana – strukturirana stran

```
Kontaktni podatki
                                       <div class="contactinfo">
                                          Rob Crowther. Web hacker at
                                          <a href="http://example.org"> Example.org </a>.
                                          You can contact me
                                           <a href="mailto:robertc@example.org"> via e-mail </a>
                                           or on my work phone at 0123 456789.
                                       </div>
Kontaktni podatki - vCard
<div xmlns:contact="http://www.w3.org/2001/vcard-rdf/3.0#" class="contactinfo"</pre>
about="http://example.org/staff/robertc">
   <span property="contact:fn">Rob Crowther</span>.
   <span property="contact:title">Web hacker</span> at
   <a rel="contact:org" href="http://example.org"> Example.org </a>.
      You can contact me
    <a rel="contact:email" href="mailto:robertc@example.org"> via e-mail</a>
      or on my
    <span property="contact:tel">
        <span property="contact:type">work</span> phone at
        <span property="contact:value">0123 456789</span>
     </span>.
</div>
```

HTML5 in RDF

http://www.w3.org/TR/rdfa-lite/#the-attributes

```
    My name is Manu Sporny and you can give me a ring via 1-800-555-0199.

    My name is Manu Sporny and you can give me a ring via 1-800-555-0199.
```

37

Semantični splet - Iskalniki



http://swoogle.umbc.edu/



