
A Szemantikus világháló alapjai

Szeredi Péter Lukácsy Gergely
{szeredi , lukacsy}@cs.bme.hu

Budapesti Műszaki és Gazdaságtudományi Egyetem
Számítástudományi és Információelméleti Tanszék

-
- ① A szemantikus világhálóról általában
 - ② Matematikai alapok: leíró logikák
 - ③ OWL – Web Ontology Language

A SZEMANTIKUS VILÁGHÁLÓ

A SZEMANTIKUS VILÁGHÁLÓRÓL

Szemantikus technológiák

- Jelszó: a számítógépek „értsék” meg amit mondunk nekik
- Legfontosabb alterületek (Google találatok)
 - Szemantikus világháló (52,000,000)
 - Semantic data integration (200,000)
 - Semantic grid, semantic mining, semantic email (< 100,000)

Alappélda: intelligens keresés

- keressünk képet *tengeri emlősökről*
- ehhez tudni kell
 - mi található egy adott képen – **metainformáció**
 - mik a tengeri emlősök – **háttértudás**

METAINFORMÁCIÓK

Metainformációk a weben és azon kívül

- a weblapokon megtalálható a META elem, az oldalak címe, egyéb jellemzői
- a Word dokumentumok tartalmazzák szerző nevét, a dokumentum utolsó módosítási idejét stb.
- egyes képformátumok tartalmazzák azt, hogy milyen eszközzel, mikor, milyen körülmények között készültek

A Szemantikus Web elképzelés célja

- a metainformációkat egységes alakban adhassuk meg
- képesek legyünk következtetni azokon

A W3C ELKÉPZELÉSE

Az RDF (Resource Description Framework) keretrendszer

- az RDF egy nyelv, mely alkalmas arra, hogy *erőforrásokhoz metaadatot* kapcsoljon
 - erőforrás minden, ami URI-val rendelkezik
 - erőforrás többek között egy weblap, a weblap egy része, egy kép, egy tetszőleges állomány, hanganyag, erőforrások egy csoportja stb.
- az *URI-k* alapvető szerepe a Szemantikus Világhálóban
 - egyértelmű állítások fogalmazhatóak meg használatukkal
 - elősegíti a „bárki mondhat bármit” elvet
 - könnyű a metaadatok kombinációja

AZ RDF KERETRENDSZER

RDF példák

- egy RDF leírás *hármások* halmaza (alany, állítmány, tárgy)
- Az állítmány egy olyan URI, amely egy tulajdonságot jelöl.
- Az alany egy URI, a tárgy lehet URI vagy literál.
- példák:

Szeredi Péter levélcíme `mailto:szeredi@cs.bme.hu`

`pelda.jpg` szerzője Béla

Kati szereti Bélat

Kati szereti Bélat

egy adott program írója valaki

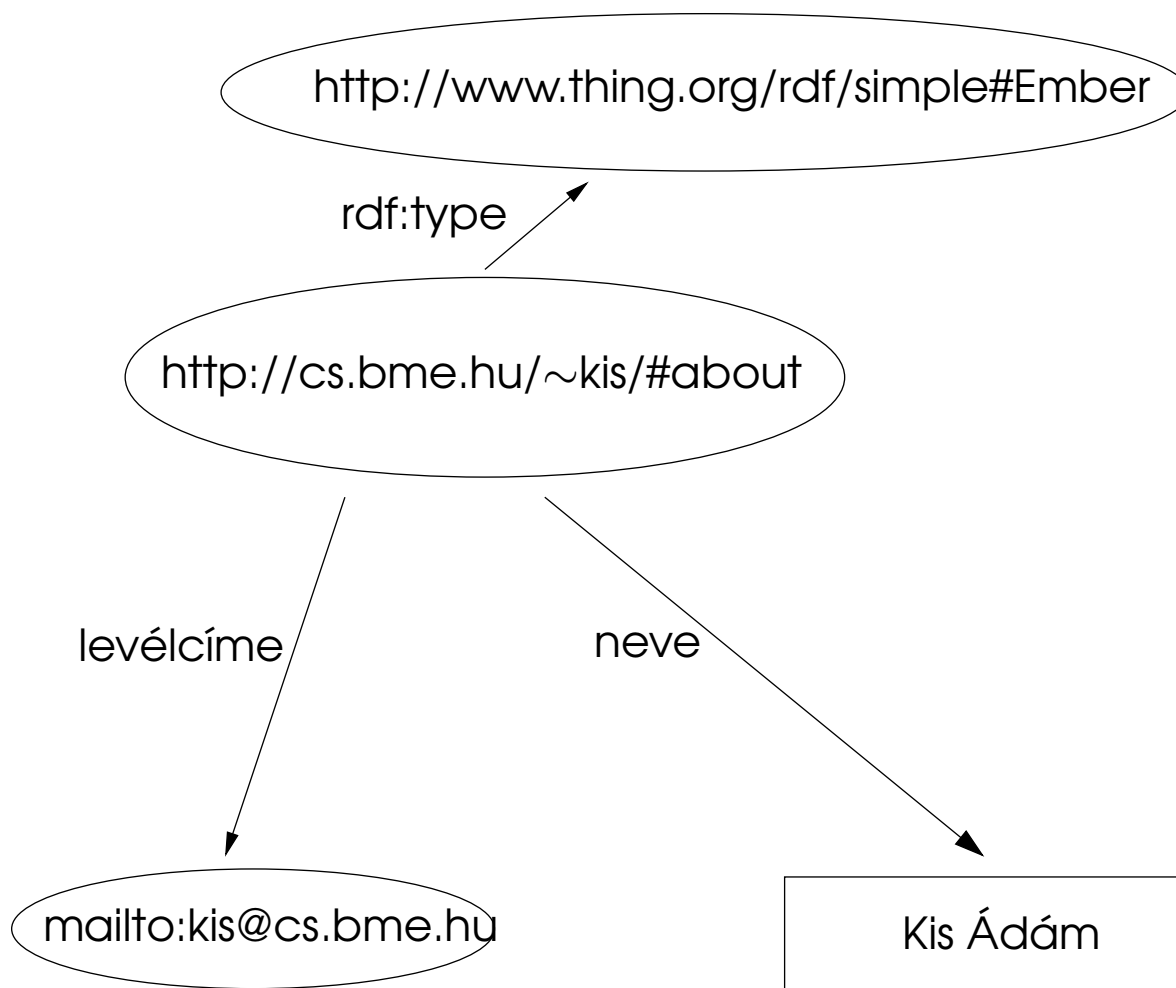
akinek neve Tas

és akinek életkora 23

Kis Ádám típusa Ember (Kis Ádám az Ember osztályba tartozik)

AZ RDF KERETRENDSZER

Egy példa RDF tudásbázis



AZ RDF KERETRENDSZER

Az előbbi RDF gráf XML alakja

```
<?xml version="1.0" encoding="ISO-8859-2"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:s="http://www.utils.org/utils#">

  <rdf:Description rdf:about="http://cs.bme.hu/~kis/#about">
    <s:neve> Kis Ádám</s:neve>
    <s:levélcíme rdf:resource="mailto:kis@cs.bme.hu"/>
    <rdf:type rdf:resource=
      "http://www.thing.org/rdf/schemas/simple#Ember"/>
  </rdf:Description>

</rdf:RDF>
```

AZ RDF KERETRENDSZER

A következtetéshez szükséges **háttértudás**: az RDF séma

→ Osztályhierarchia: osztályok és azok tartalmazási viszonyai

Férfi alosztálya az Embernek

Ember alosztálya az Állatnak

Ikerház alosztálya az Épületnek

→ tulajdonságok és tartalmazási viszonyaik

apja altulajdonsága a szülője tulajdonságnak

(ha egyik embernek egy másik az apja, akkor a másik egyben szülője is az elsőnek)

barátja altulajdonsága az ismerőse tulajdonságnak

→ tulajdonságkorlátozások

az életkora tulajdonság értékkészlete (range) az Egész számok osztálya

AZ RDF KERETRENDSZER

Az RDF sémák - példa

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:base="http://www.ex.org/houses">
  <rdf:Description rdf:ID="Épület">
    <rdf:type rdf:resource="http://.../rdf-schema#Class"/>
  </rdf:Description>
  <rdfs:Class rdf:ID="Ikerház">
    <rdfs:comment>Ikerház házak osztálya</rdfs:comment>
    <rdfs:subClassOf rdf:resource="#Épület"/>
  </rdfs:Class>
  <rdfs:Property rdf:ID="életkora">
    <rdfs:domain rdf:resource="#Épület"/>
    <rdfs:range rdf:resource="http://.../2001/XMLSchema#integer"/>
  </rdfs:Property>
</rdf:RDF>
```

AZ RDF KERETRENDSZER

Az RDF séma hiányosságai

→ Gyenge kifejezőerő

- Osztálműveletek hiánya: nincs metszet, unió stb.
- Osztályok diszjunkttsága nem írható elő
- Lokális tulajdonságkorlátozások hiánya: nem írhatjuk elő, hogy
a cipők **mérete** egész szám, de a pólók **mérete** az S, M, L, XL, XXL jelsorozatok egyike.
- ...

→ Tovább lépés

- Web Ontology Language
- Matematikai alapja: leíró logikák

LEÍRÓ LOGIKÁK

PÉLDA

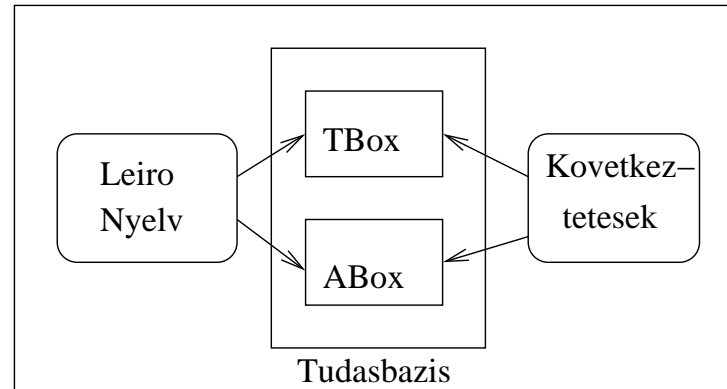
Példa: Anyákat keresünk

- rendelkezésre álló *alapinformációk*: kik az „emberek”, mi az emberek „neme”, kik vannak „szülő-gyerek” kapcsolatban
- rendelkezésre álló *háttértudás*: mit értünk az „anya” fogalmán: olyan nőnemű ember, akinek van gyereke
- elvárás: a gép a fentiek alapján tudjon válaszolni „XY anya?”, ill. „Ki anya?” típusú kérdésekre
- megoldás: *formalizáljuk* az alapinformációkat és a háttértudást, a kérdésre *logikai következtetéssel* kapjuk a választ

Leíró logikai formalizmussal

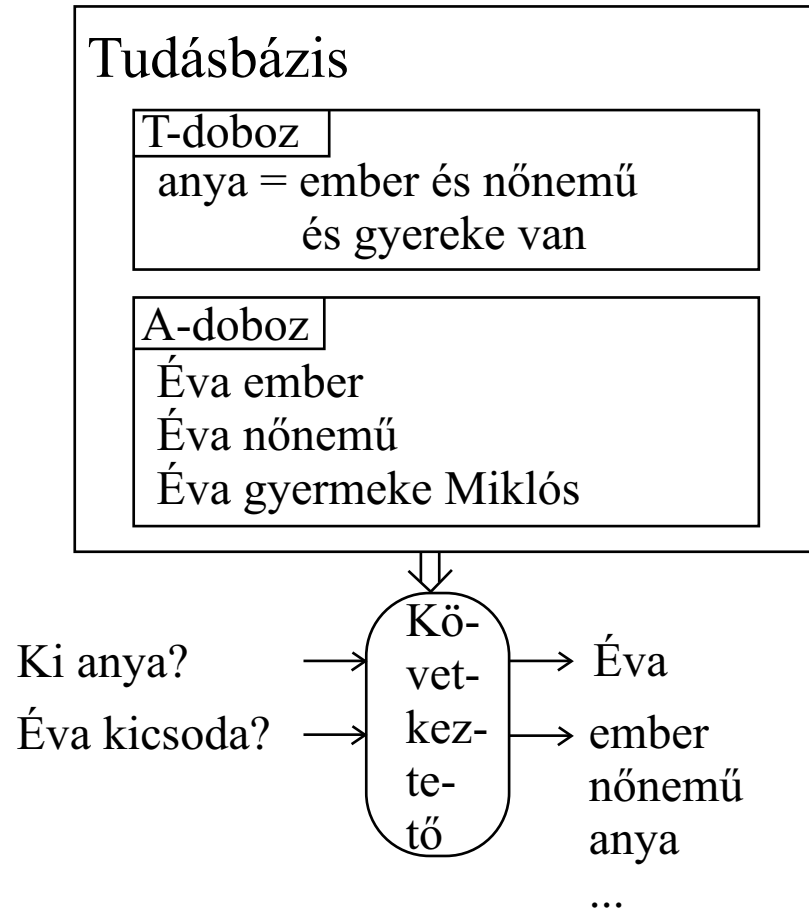
- Ember(ÉVA) Nőnemű(ÉVA) gyereke(ÉVA, MIKLÓS)
- Anya \equiv Ember \sqcap Nőnemű \sqcap \exists gyereke. \top
- Az Anya fogalom *példánykikeresésének* eredménye: ÉVA

LEIRÓ LOGIKÁK MINT A TUDÁSREPREZENTÁCIÓ ESZKÖZEI



- Tudásbázis (KB, knowledge base) = T-doboz (TBox) + A-doboz (ABox):
- T-doboz = terminológiai doboz = terminológiai állítások, **háttértudás**: $\text{Anya} \equiv \text{Ember} \sqcap \text{Nőnemű} \sqcap \exists \text{gyereke}.\top$
- A-doboz = adatdoboz = adatállítások, konkrét **metainformációk**:
 $\text{Ember}(\text{ÉVA}) \text{ Nőnemű}(\text{ÉVA}) \text{ gyereke}(\text{ÉVA}, \text{MIKLÓS})$

PÉLDA LEIRÓ LOGIKAI KÖVETKEZTETÉSRE

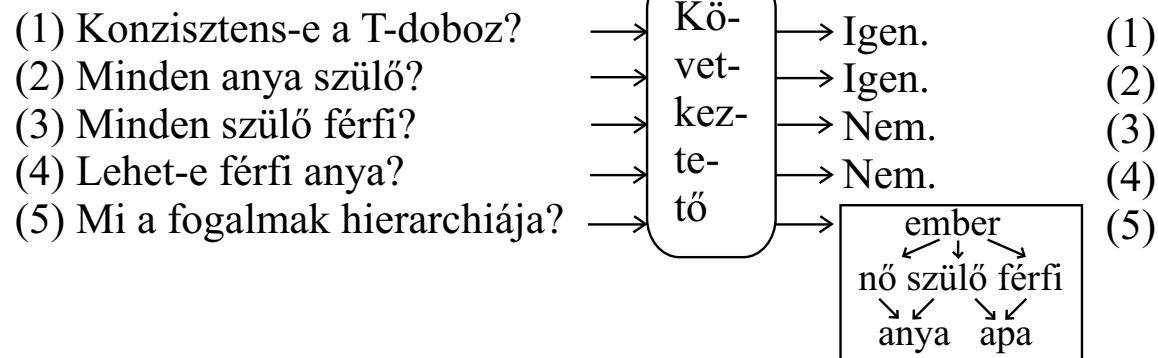


PÉLDA TISZTA T-DOBOZ KÖVETKEZTETÉSRE

Tudásbázis

T-doboz

anya = ember és nőnemű és van gyereke.
nő = ember és nőnemű
férfi = ember és nem nőnemű
szülő = ember és van gyereke
apa = férfi és szülő



A LEÍRÓ LOGIKÁK

Miről szólnak?

- *fogalmak*: a modellezni kívánt világ egyedeiből álló *halmazok* (RDF, OO megfelelője: osztály)
- *szerepek*: a modellezni kívánt világ egyedei között fennálló kétargumentumú *relációk*, azaz egyedpárok halmazai (RDF megfelelője tulajdonság, OO megfelelője: asszociáció)

Hogyan épülnek fel?

- *atomi* fogalmak és szerepek: eleve adottak, pl. Ember, gyereke
- *összetett* fogalmak és szerepek: atomi fogalmakból és szerepekből különféle *konstruktorokkal* felépített kifejezések, pl. $\text{Ember} \sqcap \text{Nőnemű} \sqcap \exists \text{gyereke} . \top$
 gyereke^-

PÉLDÁK FOGALMI AXIÓMÁKRA

Az Anya nem más, mint olyan Ember aki Nőnemű és van gyereke.

$$\text{Anya} \equiv \text{Ember} \cap \text{Nőnemű} \cap \exists \text{gyereke.} \top$$

Minden Tigris Emlős.

$$\text{Tigris} \sqsubseteq \text{Emlős}$$

A boldog emberek gyerekei is boldogak.

$$\text{Boldog} \cap \text{Ember} \sqsubseteq \forall \text{gyereke.} \text{Boldog}$$

A gyermektelen emberek boldogak

$$\forall \text{gyermeke.} \perp \cap \text{Ember} \sqsubseteq \text{Boldog}$$

PÉLDÁK SZEREPAXIÓMÁKRA

A gyereke viszonyban levők leszármazottja viszonyban is vannak.

gyereke \sqsubseteq leszármazottja

A szülője kapcsolat a gyereke kapcsolat megfordítottja (inverze).

szülője \equiv gyereke⁻

A leszármazottja reláció tranzitív

Trans(leszármazottja)

AZ \mathcal{ALC} NYELV SZINTAXISA

- Az \mathcal{ALC} fogalomkifejezések (röviden fogalmak) szintaxisa:

$C \rightarrow A$		(atomi fogalom)	egy halmaz, pl: Ember
\top		(tetőjel, top)	az összes objektum halmaza
\perp		(fenékJel, bottom)	az üres halmaz
$\neg C$		(negálás)	
$C \sqcap D$		(metszet)	
$C \sqcup D$		(unió)	
$\forall R.C$		(értékkorlátozás)	azon egyedek, amelyek minden R -je C -beli
$\exists R.C$		(létezési k.)	azon egyedek, amelyekhez van C -beli R

A atomi fogalom, C, D összetett fogalmak

AZ *ALC* NYELV – CSALÁDI PÉLDÁK

Nő	≡	Ember \sqcap Nőnemű
Férfi	≡	Ember \sqcap \neg Nő
Anya	≡	Nő \sqcap \exists gyereke. \top
Apa	≡	Férfi \sqcap \exists gyereke. \top
Szülő	≡	Anya \sqcup Apa
Nagyanya	≡	Anya \sqcap \exists gyereke.Szülő
Feleség	≡	Nő \sqcap \exists férje. \top
LányosSzülő	≡	Szülő \sqcap \forall gyereke.Nőnemű
Ember	\sqsubseteq	\forall gyereke.Ember
Ember	\sqsubseteq	\forall férje.Férfi
Ember \sqcap \exists férje. \top	\sqsubseteq	Nő

ADATBÁZISOK ÉS ADATDOBOZOK

Adatbázis: zárt világ

- pontosan az szerepel az adatbázisban, ami igaz
- példa-adatbázis (egyetlen tábla egy sorral):
gyereke(ÉVA,KÁIN) – ÉVANak egyetlen gyereke van

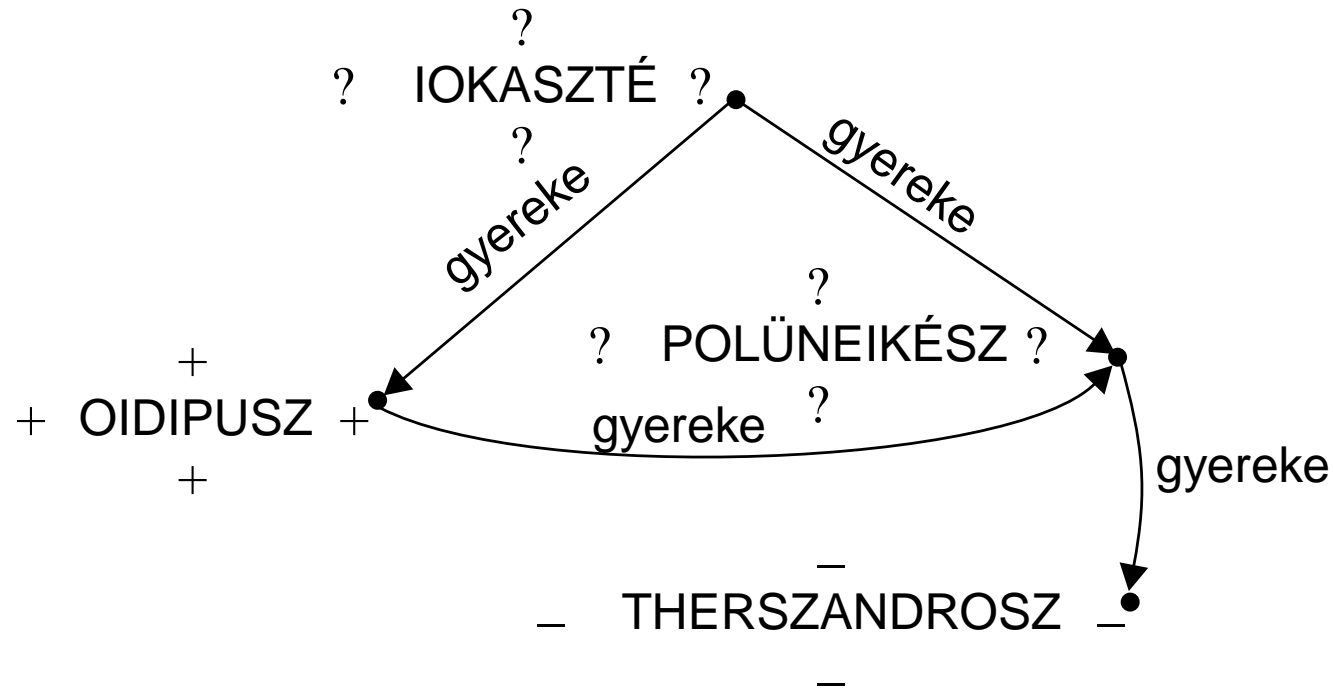
Leíró logikák: nyílt világ

- egyszerre több lehetséges világot kell lefedni
- ami szerepel a tudásbázisunkban az igaz, de más is lehet igaz
- negatív információt is megadhatunk, pl. \neg Ember(ZEUSZ)
- lehet, hogy $\neg C(X)$ és $C(X)$ egyike sem ismert
- példa-tudásbázis (egyetlen adatállítás):
gyereke(ÉVA,KÁIN) – ÉVANak lehet több gyereke is

A szemantikus technológiák körében (pl. a világhálón)

- fontos a nyílt világ szemlélet,
- nem-teljes információ birtokában kell következtetni.

PÉLDA NYÍLT VILÁGBAN VALÓ KÖVETKEZTETÉSRE



→ Az adatdoboz:

gyereke(IOKASZTÉ,OIDIPUSZ)

gyereke(IOKASZTÉ,POLÜNEIKÉSZ)

gyereke(OIDIPUSZ,POLÜNEIKÉSZ)

gyereke(POLÜNEIKÉSZ,THERSZAN.)

Apagyilkos(OIDIPUSZ)

¬ Apagyilkos(THERSZANDROSZ)

→ Egy igaz állítás:

$(\exists \text{gyereke.} (\text{Apagyilkos} \sqcap \exists \text{gyereke.} \neg \text{Apagyilkos}))(\text{IOKASZTÉ})$

OWL – WEB ONTOLOGY LANGUAGE

AZ OWL (WEB ONTOLOGY LANGUAGE) NYELV

Az OWL résznyelvek

- OWL Light
- OWL DL
- OWL Full

Példa: a lányos szülő definíciója OWL-ben

```
<owl:Class rdf:ID="LányosSzülő">
  <owl:intersectionOf rdf:parseType="Collection">
    <owl:Class rdf:about="#Szülő" />
    <owl:Restriction>
      <owl:onProperty rdf:resource="#gyermeke" />
      <owl:allValuesFrom rdf:resource="#Nőnemű" />
    </owl:Restriction>
  </owl:intersectionOf>
</owl:Class>
```

AZ OWL ÉS A LEÍRÓ LOGIKÁK

Az OWL DL alapja a *SHOIN(D)* leíró logikai nyelv

- $\mathcal{S} \equiv \mathcal{ALC}_{\mathcal{R}+}$ (a \mathcal{ALC} nyelv kiegészítve tranzitív szerepekkel), azaz egyes szerepekről (pl. őse) kijelenthetjük, hogy tranzitívak.
- $\mathcal{H} \equiv$ szerephierarchiák. Egy szerephierarchia $R \sqsubseteq S$ alakú állítások halmaza, pl. minden barátja kapcsolat egyben ismerőse kapcsolat is: barátja \sqsubseteq ismerőse.
- $\mathcal{O} \equiv$ egyedfogalmak (objektumok). Olyan fogalmak, amelyeknek csak egy példánya lehet.
- $\mathcal{I} \equiv$ inverz szerepek: egy R szerep mellett annak R^{-} inverzét is használhatjuk, pl. gyereke $^{-} \equiv$ szülője.
- $\mathcal{N} \equiv$ számosság-korlátozások, azaz $\leq nR$ és $\geq nR$ alakú fogalomkifejezések pl. azon emberek akiknek legalább 3 gyereke van: (≥ 3 gyereke)
- $(\mathbf{D}) \equiv$ konkrét adattartományok: egy tulajdonság értékészlete lehet pl. egész, fűzér stb.

ÖSSZEFOGLALÁS

- A szemantikus technológiák: közelebb hozzák a számítógépeket az emberhez
 - A szemantikus világháló az egyik leggyorsabban fejlődő ilyen technológia
- A gépi tudásreprezentációhoz egyszerű logikai nyelvekre van szükség
 - például a leíró logikákra
- A szilárd matematikai alapokra megfelelő mérnöki mérnöki módszerek és eszközök építhetők
 - például RDF, OWL

SZEMANTIKUS TECHNOLOGIÁK

Személyes tapasztalatok

- SINTAGMA (Szemantikai INtegrációs Technológia Alkalmazása Grid-alapú, Modellvezérelt Architektúrákban), NKFP projekt, 2005–2007, IQSYS, BME, SZTAKI, ...
- FUSION (Business process fusion based on Semantically-enabled Service-oriented Business Applications) EU FP6 STREP projekt, 2006-2008, SAP, ... , BME, ...
- „A szemantikus világháló és az ontológiakezelés alapjai” – BME VIK választható tárgy, heti 4 óra, 2004. óta.
- „Szeredi P., Lukácsy G., Benkő T. A szemantikus világháló elmélete és gyakorlata”, tankönyv, Typotex 2005 (angol kiadása várhatóan: Cambridge University Press, 2007 vége)
Jelenleg 50% engedménnyel (2200 Ft) kapható az Olvasók Boltjában, Budapest V.,Pesti Barnabás u. 4.