

# Web servisai

Įvadas į semantinį žiniatinklį

Osvaldas Grigas

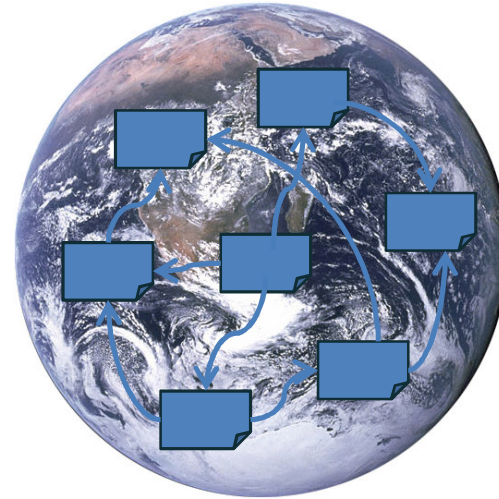
# Apie ką kalbėsime

- Žiniatinklio evoliucija
- Semantinė informacija
- Semantiniai web servisai

# Žiniatinklio evoliucija

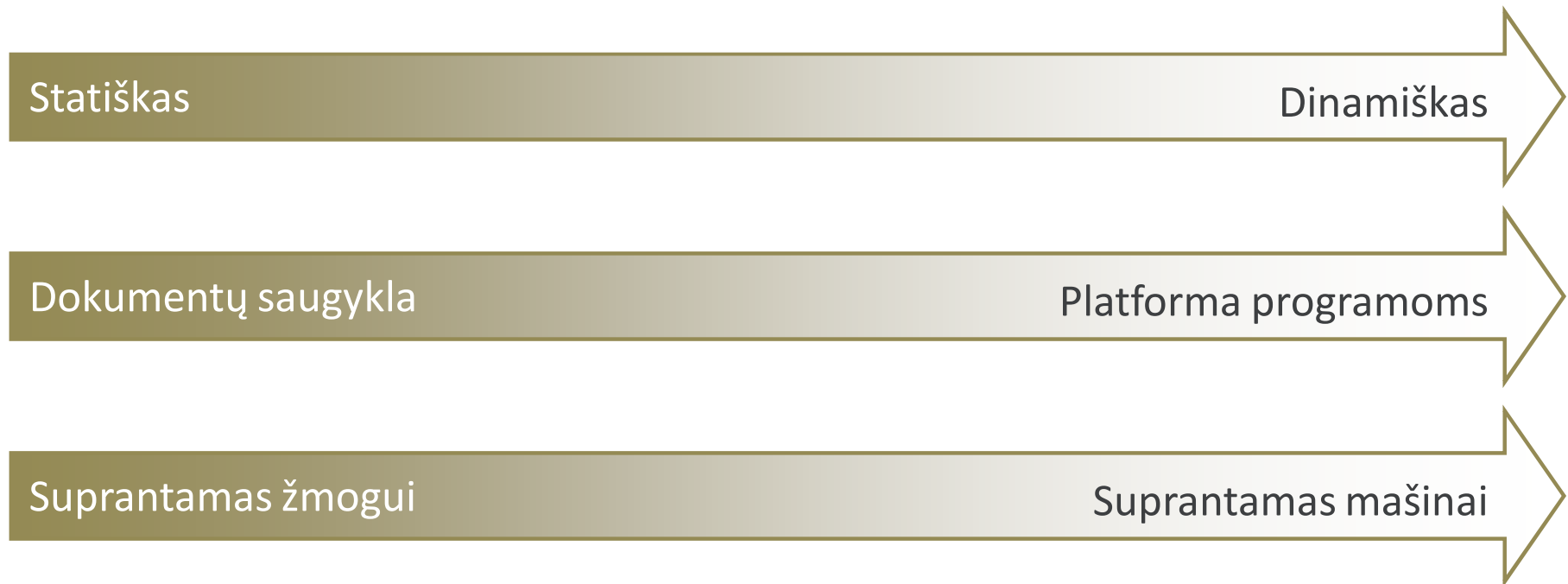
# Žiniatinklio dydis

- Regimasis žiniatinklis (*Surface Web*)
  - Virš 100 milijonų svetainių
  - Mažiausiai 46 milijardai tinklalapių



- Nematomas žiniatinklis (*Deep Web*)
  - Grubiais apskaičiavimais, gali saugoti kelis šimtus kartų daugiau informacijos

# Žiniatinklio evoliucija



## Web 1.0

- Daugiausia statiniai puslapiai
- Informacija skirta žmogui, bet ne mašinai
- Griežta atskirtis tarp turinio kūrėjų ir naudotojų

# Web 2.0

- Turinys dažniausiai generuojamas dinamiškai
- Interneto naudotojai tampa turinio kūrėjais
  - Blogai
  - Wiki
  - Socialiniai tinklai
- Web'as vis dažniau naudojamas kaip platforma programinei įrangai
  - Web servisai
  - Software as a Service (SaaS)
- Dalis turinio aprašoma semantiškai
  - Pateikiama mašinoms suprantama forma
  - Pavyzdžiui, RSS

# Web 3.0

- Pasiiekta kritinė masė semantiškai aprašyto turinio
  - Servisai taip pat aprašomi semantiškai
- Išnyksta riba tarp "desktop" ir "web" programinės įrangos
- Verslo procesai perkeliama į web'ą ir automatizuojami
- Programos remiasi semantine informacija ir savarankiškai priima sprendimus



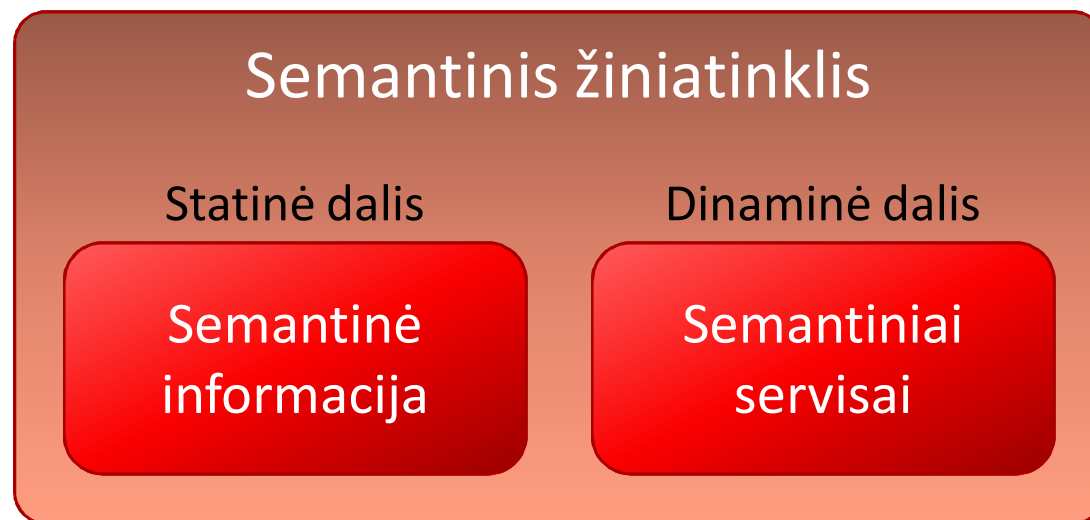
# Web 3.0 agentai



- Agentai - protingos Web programos
  - atstovauja asmenį arba verslą
  - atlieka įvairias užduotis, pvz. ieško prekių/paslaugų/partnerių
  - pačios randa reikalingą informaciją
  - moka deleguoti užduotis kitoms programoms (servisams)

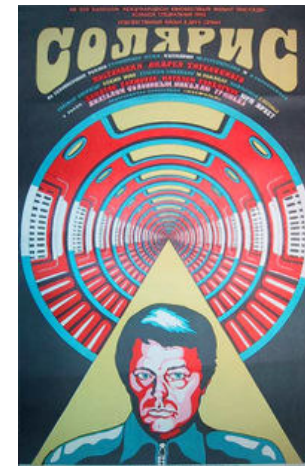
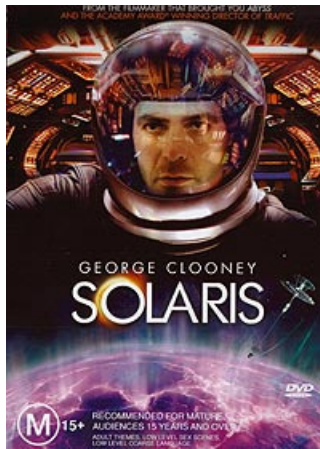
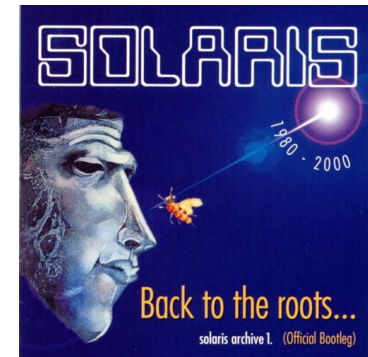
# Web 3.0 vizija - Semantinis žiniatinklis

- Web'as kaip globali duomenų bazė
- Suprantamas ne tik žmogui, bet ir mašinai



# Semantinė informacija

# Paieška pagal užklausą "Solaris"



# Problema

- Dabartinė informacijos paieška duoda labai daug netikslių rezultatų
- Mašina negali prasmingai operuoti šia informacija
  - Kaip automatiškai atskirti "teisingą" rezultatą nuo "neteisingo"?

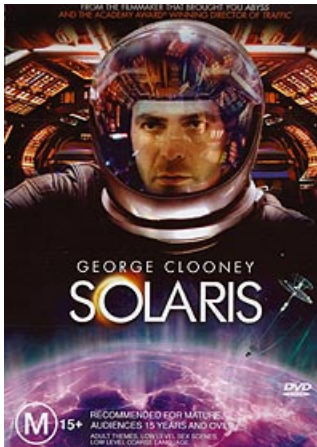
# Problema

- Šiandienos web'e labai trūksta semantikos

**Category: Film**

Year: 2002

Director: Steven Soderbergh



**Category: Film**

Year: 1972

Director: Andrei Tarkovsky



**Category: Operating System**

Developer: Sun Microsystems

Kernel type: monolithic



# Folksonomijos



## Tags from what's rising

What bloggers are saying now

**barack obama** blogs china  
economics food hillary clinton iraq  
john mccain media music **news**  
nfl observer opinion **politics** religion  
technology television united states  
video

- "Socialiniai žodynėliai"
- Pavyzdžiui, blog'ų žymės (*tags*)
- Kategorizuoja turinį
- Gimsta spontaniškai, nekontroliuojamos

# Folksonomijos pavyzdys

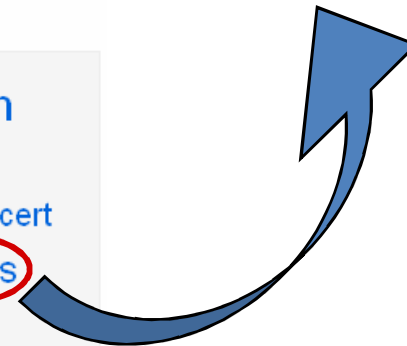


flickr™



## All time most popular tags

07 africa amsterdam animals architecture art asia australia baby band barcelona beach  
berlin birthday black blackandwhite blue boston bw california cameraphone camping  
canada canon car cat chicago china christmas church city clouds color concert  
cute day de dog england europe fall family festival film florida flower **flowers**  
food france friends fun garden geotagged germany girl girls graffiti green  
halloween hawaii hiking holiday home honeymoon house india ireland island italy japan  
july kids la lake landscape light live london macro march me mexico mountain mountains  
museum music nature new newyork newyorkcity newzealand night nikon nyc ocean  
paris park party people photo photos portrait red river rock rome san  
sanfrancisco scotland sea seattle show sky snow spain spring street summer sun  
sunset sydney taiwan texas thailand tokyo toronto tour travel tree trees trip uk urban  
usa vacation vancouver washington water wedding white winter yellow york  
zoo





# Taksonomijos

- Kontroliuojami žodynai - kategorijų sąvadai
- Dažnai tarptautiniai standartai
- Klasifikuoja vieną aiškiai apibrėžtą sritį
- Pakeitimai atliekami nedažnai ir lėtai

# Taksonomijos pavyzdys



International  
Organization for  
Standardization

ISO 3166 - šalių ir regionų kodai

.....	.....
<b>CY</b>	Cyprus
<b>CZ</b>	Czech Republic
<b>DE</b>	Germany
<b>DJ</b>	Djibouti
<b>DK</b>	Denmark
<b>DM</b>	Dominica
<b>DO</b>	Dominican Republic
<b>DZ</b>	Algeria
<b>EC</b>	Ecuador
<b>EE</b>	Estonia
.....	.....



<b>BE</b>	Berlin
<b>BR</b>	Brandenburg
<b>BW</b>	Baden-Württemberg
<b>BY</b>	Bayern
<b>HB</b>	Bremen (Hansestadt)
<b>HE</b>	Hessen
<b>HH</b>	Hamburg (Hansestadt)
<b>MV</b>	Mecklenburg-Vorpommern
<b>NI</b>	Niedersachsen
<b>NW</b>	Nordrhein-Westfalen
<b>RP</b>	Rheinland-Pfalz
<b>SH</b>	Schleswig-Holstein
<b>SL</b>	Saarland
<b>SN</b>	Sachsen
<b>ST</b>	Sachsen-Anhalt
<b>TH</b>	Thüringen

# Taksonomijos pavyzdys



Ekonominės veiklos  
klasifikacija

.....	.....
<b>21</b>	<b>Mining</b>
<b>2111113</b>	Natural gas extraction
<b>212210</b>	Iron core mining
.....	.....
<b>31</b>	<b>Manufacturing</b>
<b>316213</b>	Men's footwear manufacturing
<b>316214</b>	Women's footwear manufacturing
.....	.....
<b>51</b>	<b>Information</b>
<b>51113</b>	Book Publishers
<b>5111312</b>	Children's books
.....	.....
<b>72</b>	<b>Accommodation</b>
<b>7211</b>	Accommodation for travelers
<b>721113</b>	Shared room accommodation for travelers
.....	.....

# Taksonomijos pavyzdys



Produktų ir paslaugų  
klasifikacija

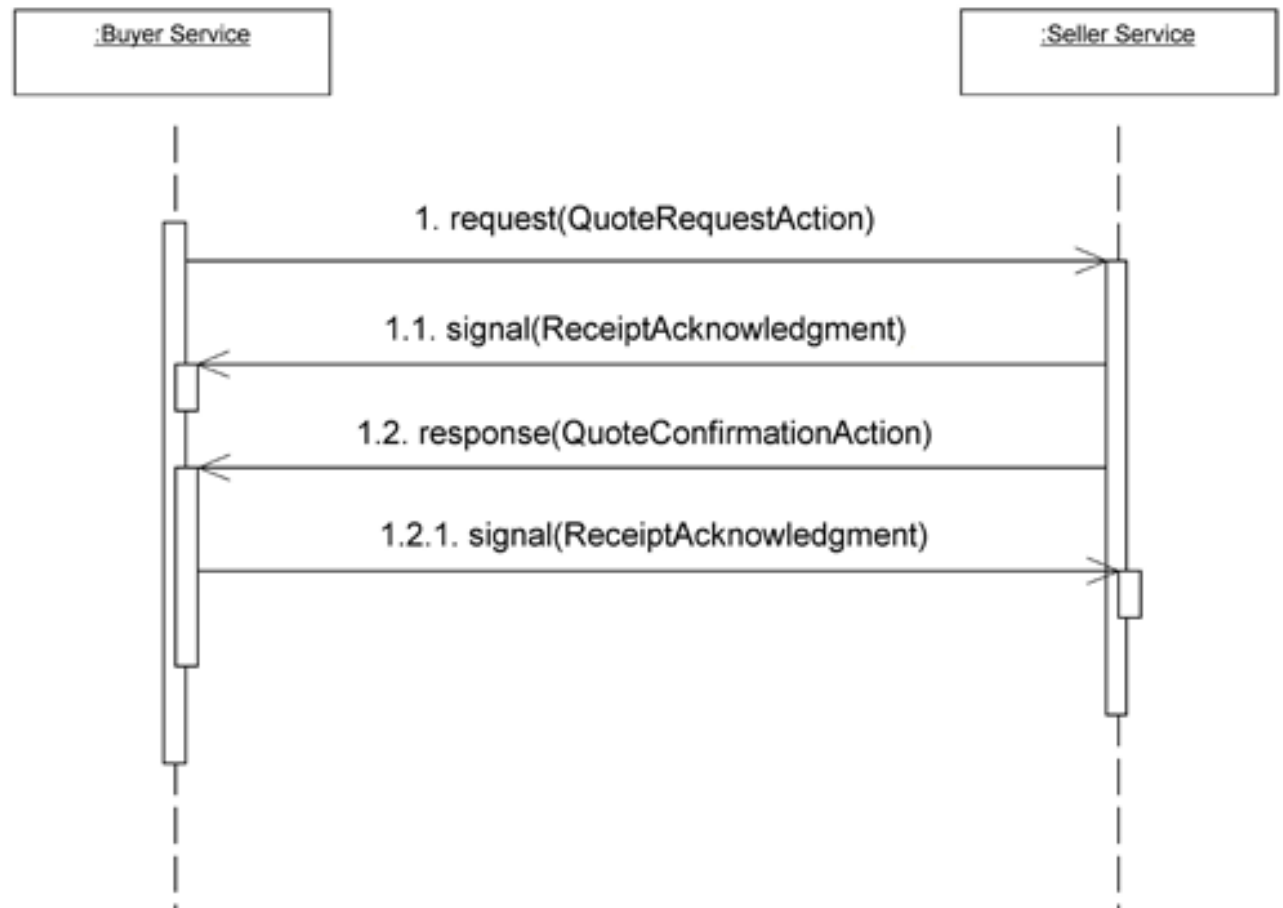
<b>10100000</b>	<b>Live animals</b>
<b>10101501</b>	Cats
<b>10101502</b>	Dogs
.....	.....
<b>44000000</b>	<b>Office Equipment and Accessories and Supplies</b>
<b>44103100</b>	Printer and facsimile and photocopier supplies
<b>44103103</b>	Toner
.....	.....
<b>53000000</b>	<b>Apparel and Luggage and Personal Care Products</b>
<b>53111600</b>	Shoes
<b>53111601</b>	Mens shoes
<b>53111602</b>	Womens shoes
.....	.....

# Taksonomijos pavyzdys

ROSETTANET

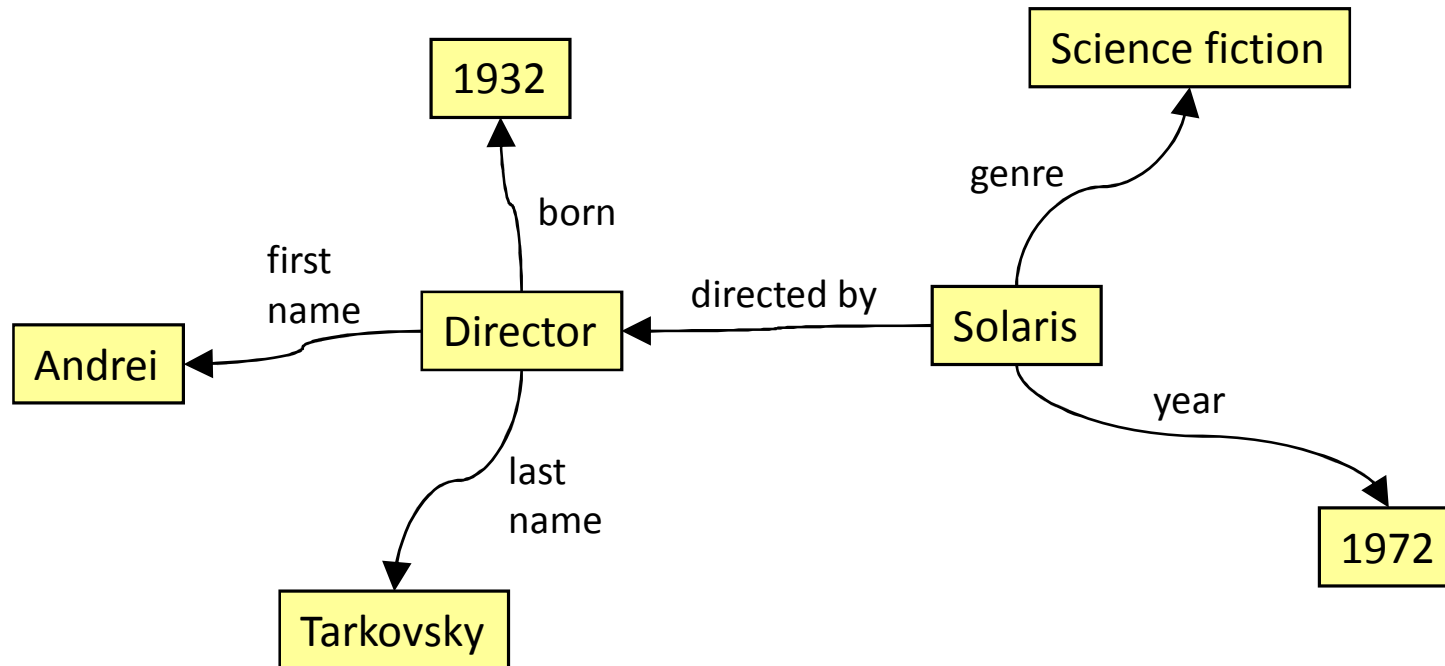
POWERED BY GS1 US

Verslo procesai



# Semantiniai tinklai

- Aprašo realaus pasaulio objektus ir jų ryšius
- Vienas iš būdų pateikti žinias mašinai suprantama forma



# Semantinių tinklų pavyzdžiai

## Socialiniai tinklai



orkut

friendster.

facebook.

Linked in.®

# Semantinių tinklų pavyzdžiai

## WordNet - anglų kalbos žodžių semantinis leksikonas

WordNet Search - 3.0 - [WordNet home page](#) - [Glossary](#) - [Help](#)

Word to search for:

Display Options:

Key: "S." = Show Synset (semantic) relations, "W." = Show Word (lexical) relations

### Noun

- [S: \(n\) movie](#), [film](#), [picture](#), [moving picture](#), [moving-picture show](#), [motion picture](#), [motion-picture show](#), [picture show](#), [pic](#), [flick](#)  
(a form of entertainment that enacts a story by sound and a sequence of images giving the illusion of continuous movement)  
*"they went to a movie every Saturday night"; "the film was shot on location"*
- [S: \(n\) film](#), [cinema](#), [celluloid](#) (a medium that disseminates moving pictures) *"theater pieces transferred to celluloid"; "this story would be good cinema"; "film coverage of sporting events"*
- [S: \(n\) film](#), [photographic film](#) (photographic material consisting of a base of celluloid covered with a photographic emulsion; used to make negatives or transparencies)
- [S: \(n\) film](#) (a thin coating or layer) *"the table was covered with a film of dust"*
- [S: \(n\) film](#), [plastic film](#) (a thin sheet of (usually plastic and usually transparent) material used to wrap or cover things)

### Verb

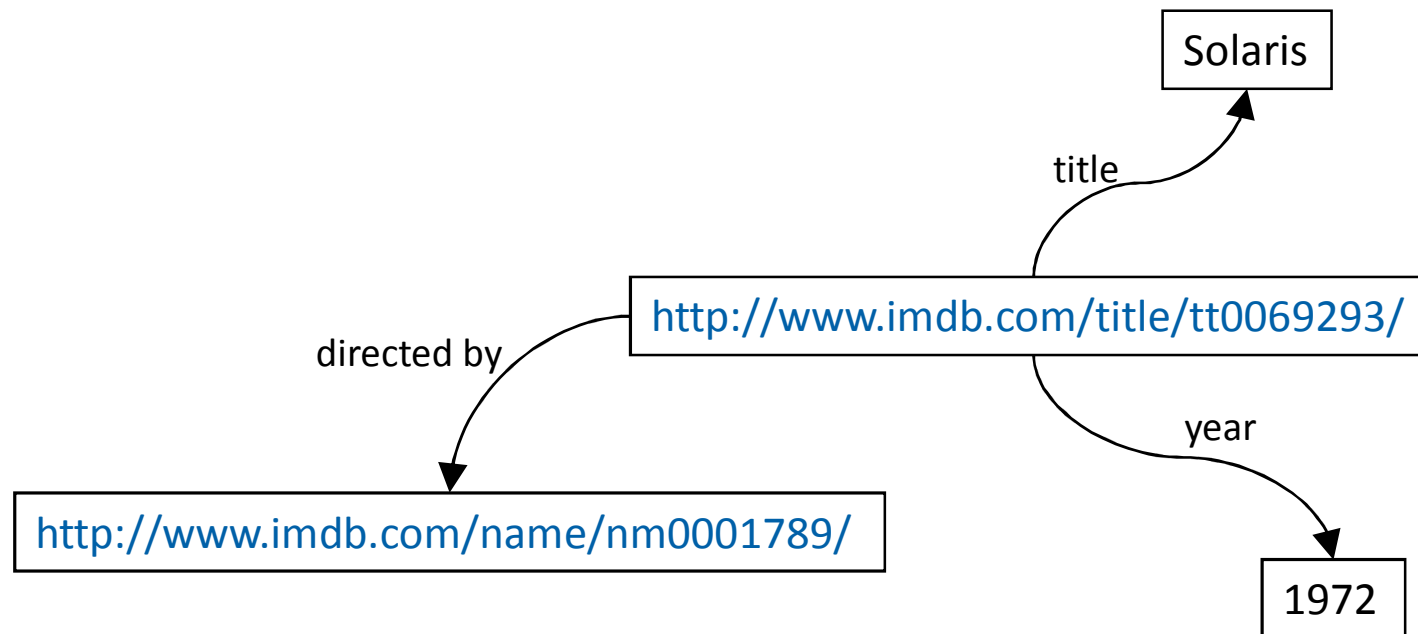
- [S: \(v\) film](#), [shoot](#), [take](#) (make a film or photograph of something) *"take a scene"; "shoot a movie"*
- [S: \(v\) film](#) (record in film) *"The coronation was filmed"*



# Semantinių tinklų kūrimas

## *RDF* - Resource Description Framework

- Būdas semantiškai aprašyti Web resursus ir ryšius tarp jų
- Naudoja XML sintaksę (dažniausiai)
- W3C rekomendacija

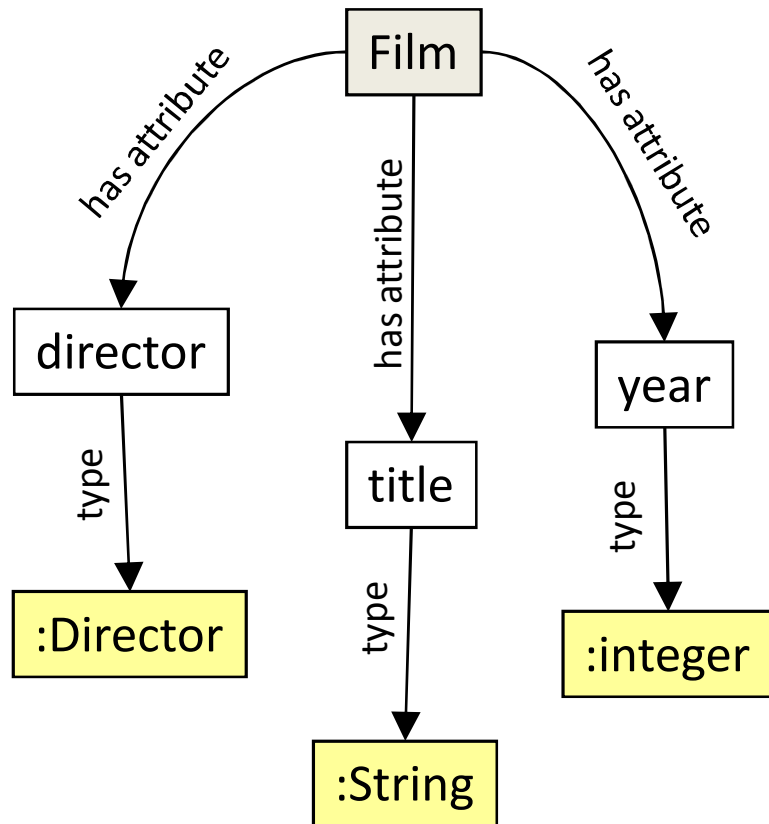


# Ontologijos

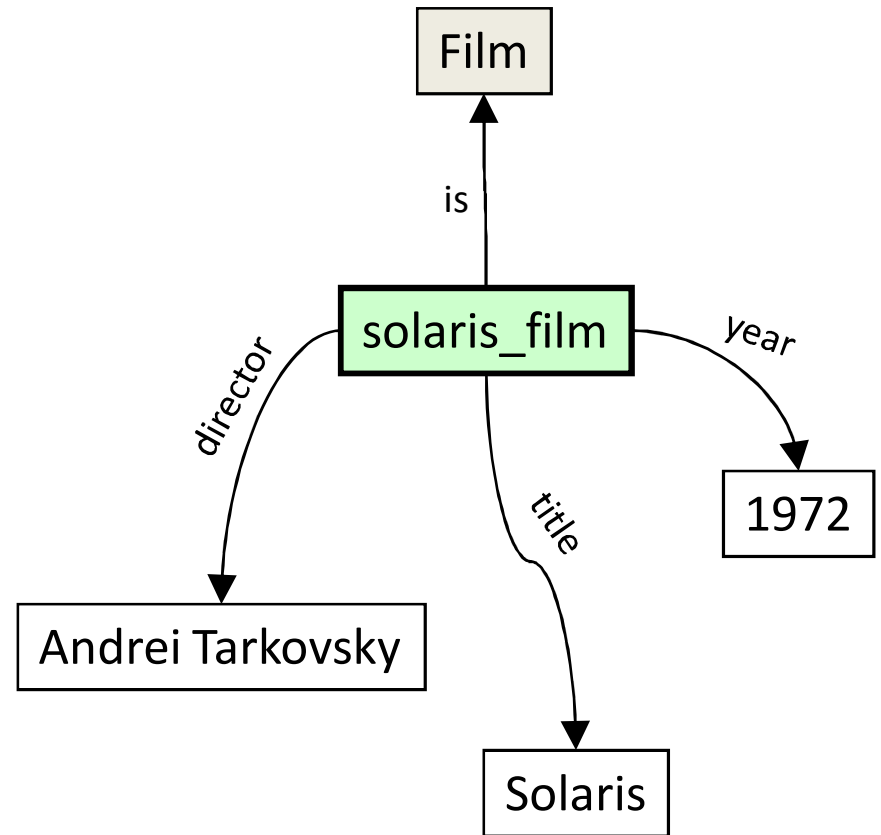
- Kaip semantiniai tinklai, tik ekspresyvesnės
  - objektų klasės
  - apribojimai
  - sąlygos ir taisyklės
- Naudojamos dirbtinio intelekto uždaviniuose
- Semantinio žiniatinklio pagrindas (mūsų vizijoje)
- Užrašomos mašinai suprantama kalba (OWL, SWRL ir kt.)

# Ontologijos pavyzdys

## KLASĖ



## OBJEKTAS



# Šiuo metu kuriamos ontologijos



Teisės ontologija

<http://www.estrellaproject.org/ikif-core/>



The Open Biomedical Ontologies

Biomedicinos ontologija

<http://obofoundry.org/>

• Ir daug daug kitų:

- Kelionių ontologija
- Vyno ontologija
- Giminytės ryšių ontologija
- Bhakti jogos transcendentinės sąmonės būsenų ontologija
- .....

*Semantinė* (kontekstinė) paieška pagal užklausą:

USING ONTOLOGY <http://www.imdb.com/MovieOntology>

FIND Film WHERE Film.title = "Solaris" AND Film.year < 1980



# Semantiniai web servisai

# Verslininkas, vardu Elas



- Gyvena semantinio web'o eroje
- Gamina moteriškus batelius
- Elui priklauso firminė batų parduotuvė
- Jis nori plėsti verslą
- Bet jam labai trūksta laisvalaikio
- Elas norėtų priimti užsakymus *online* ir visiškai automatizuoti batų pardavimą
- Elas norėtų, kad jo *online* servisas būtų pasiekiamas programiškai (klientų agentams)

# Žingsnis #1

- Sukurti viešą *XML* Web servisą
- *XML* formatu atsakyti į užklausas dėl prekių asortimento
- Priimti užsakymus *XML* formatu
- Deleguoti užduotis kitiems servisams:
  - pinigų pervedimas (PayPal)
  - prekių pristatymas (DHL)

Duomenų formatas  
**XML**



# Žingsnis #2

- Aprašyti savo Web serviso sintaksę - mašinai suprantama forma (*WSDL*)
  - Kur yra Web servisas (jo URL adresas)
  - Ką jis daro (operacijos, XML pranešimų struktūros)
  - Kaip jį kviesti (transporto protokolai ir kt.)
- Panaudoti *RosettaNet* verslo procesus
- Paviešinti *WSDL* dokumentą

Sintaksės aprašymas  
**WSDL**

Duomenų formatas  
**XML**

# Žingsnis #3

- Aprašyti serviso semantiką su web servisų ontologija *OWL-S*
- Aprašyti kiekvienos serviso operacijos reikšmę ir poveikį realiam pasauliui

Semantikos aprašymas

**OWL-S**

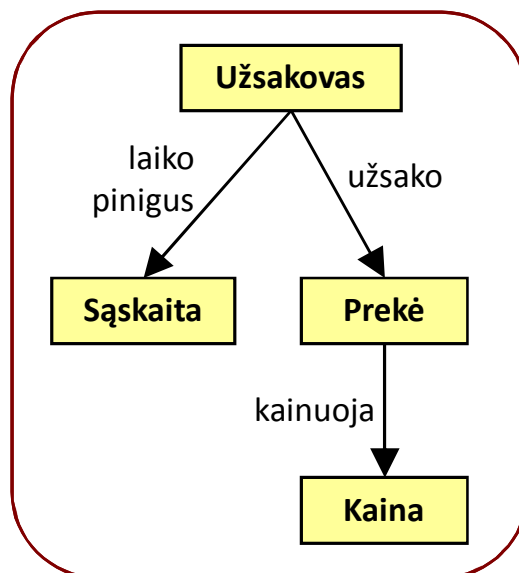
Sintaksės aprašymas

**WSDL**

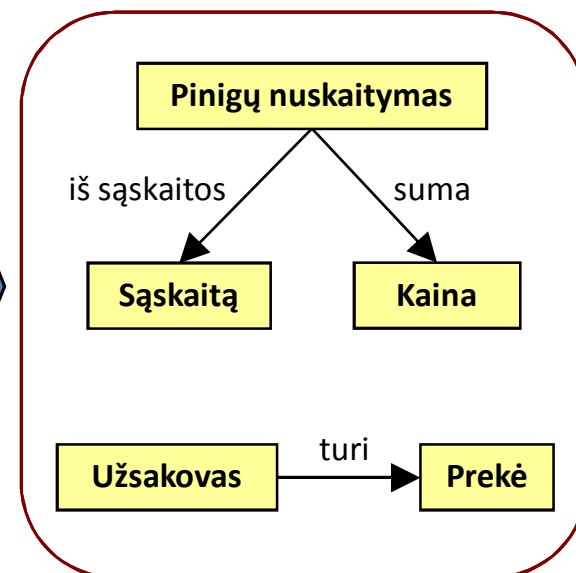
Duomenų formatas

**XML**

## PRECONDITION



## EFFECT



# Žingsnis #4

- Užregistruoti servisą viešame UDDI registre

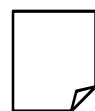
Viešas registravimas  
**UDDI**

Semantikos aprašymas  
**OWL-S**

Sintaksės aprašymas  
**WSDL**

Duomenų formatas  
**XML**

Baltieji puslapiai:



- Įmonės pavadinimas, kontaktai ir veikla
- Panaudoti *ISO 3166* ir *NAICS* taksonomijas

Geltonieji puslapiai



- Servisai, teikiamos paslaugos
- Panaudoti *UNSPSC* taksonomiją

Žalieji puslapiai



- Serviso operacijų techninis modelis
- Nuorodos į OWL-S ir WSDL dokumentus

# Vizija



- Monika ieško sau **raudonų AUKŠTAKULNIŲ** batelių
- Jos *agentas* (programa) pagal pateiktus pageidavimus įvykdo kelias semantines užklausas:
  - Randa viešus *online* parduotuvių registrus
  - Randa moteriškų batų parduotuves
  - Randa Elo servisą
  - Iškviečia Elo servisą, prašydamas **AUKŠTAKULNIŲ** batelių sąrašo
  - Iš rezultatų atrenka **raudonus** batelius ir leidžia Monikai pasirinkti:



# Vizija

- Monika išsirenka brangiausius batelius
- Ir vienu pelės paspaudimu juos nusiperka



**\$100**

