Sissejuhatus semantilisse veebi, 2: Ontoloogiad ja semantilise veebi keeled

21 aastat tagasi - 14.02.2005 Autor: AM

(Arvutimaailm 2/05)

Autor: **Mario Peterson**

Mis asi on ontoloogia ja milleks seda vaja? Millised on semantilise veebi keeled? Sellest alljärgnevalt.

Sõna ontoloogia ehk olemisõpetus on kasutusel filosoofias ja tegeleb eksistentsi probleemidega. Ontoloogia infotehnoloogia kontekstis on teadmiste ja kontseptsioonide spetsifitseerimise ja jagamise mehhanism. Lihtsamalt lahti seletades on ontoloogiad mõistete hierarhiad.

Ontoloogia mõiste

Ontoloogia defineerib terminid, mida kasutatakse teadmiste kirjeldamiseks ja esitlemiseks.

Ontoloogiaid kasutavad inimesed, andmebaasid ja rakendused, mis peavad jagama mingi ala informatsiooni (ala on lihtsalt spetsiifiline teema teadmistes, nagu transport, kinnisvara, finantsjuhtimine jne).

Ontoloogiad sisaldavad antud ala kohta põhialuseid ja nendevahelisi seoseid masinloetaval kujul. Ontoloogiaks kodeeritakse teadmised teatud alalt, samuti selle alaga seotud (ala laiendavad) teadmised. Sel viisil tagatakse nende teadmiste korduvkasutus.

Ontoloogiate abil kirjeldatakse ära:

• klassid (üldisemad asjad);

• suhted, mis saavad asjade vahel eksisteerida;

• omadused (või atribuudid), mis nendel asjadel võivad olla.

Ontoloogiad on semantilise võrgu seisukohalt väga olulised - nende abil saab esitada dokumentide semantikat, mis omakorda võimaldab kasutada semantikat

veebirakendustes ja intelligentsetes agentides.

Ontoloogiate abil saab struktureerida ja defineerida metaandmete termineid, mida parasjagu kogutakse ja standardiseeritakse. Kasutades ontoloogiaid on võimalik luua „intelligentsed” rakendusi, mis suudavad töötada lähemal inimese kontseptuaalsele tasemele.

Ontoloogiad on kriitilised rakendustele, mille abil soovitakse teostada otsingut või võrrelda erinevaid kogumeid.

XML ja XML skeemid on piisavad andmevahetuseks poolte vahel. Seda eeldusel, et definitsioonid on enne kokku lepitud. Probleemiks on aga süsteemi usaldusväärne töö juhul, kui vahetatakse XML -sõnastik, kuna puudub semantika.

Ontoloogiate liigid ja eripärad

Ontoloogiad ei erine mitte ainult sisult, vaid ka struktuurilt ja teostuselt. Neid on võimalik koostada erinevatel tasemetel. Alates lihtsatest sõnaraamatutest lõpetades keerukate klasside hierarhia ja uute mõistete ning seoste loomisega.

Ontoloogiad võivad erineda skoobi valimise ja sisu eesmärgi poolest. Samuti tekivad erinevused reaalse maailma asjade tüübi määramisel.

Raske on üheselt määrata, kas tegemist on eksemplariga või kontseptsiooniga. Näiteks reaalse maailma asja Ford Mustang GT võib kirjeldada kui eksemplari Fordi autost või sõidukit numbrimärgiga „666 XXX”, mis on eksemplar klassist Ford Mustang GT, kusjuures mõlemad väljendused võivad eksisteerida erinevates kehtivates ontoloogiates.

Ontoloogia esitamine ja kasutamine

Ontoloogia peab olema esitatud spetsiaalses eeldefineeritud keeles: ontoloogia keeles. Kõige läbimõeldum ja lihtsamini arusaadav on semantilise veebi mudelis ontoloogiate esitamise osa.

Ontoloogia kirjeldamise keeled baseeruvad kõik XML -keelel. Laiemalt on kasutusel DAML, RDF (Resource Description Framework), RDFS (Resource Description Framework Shema) ja uusim selles valdkonnas OWL (Web Ontology Language). Kõige raskem on olukord loogika kihis. Ei ole ühtset standardit tõestuste jaoks. Samuti on päringute tegemine kohmakas ja standardiseerimata.

Semantilise veebi kataloog Ontaria

Ontaria on otsinguga varustatud semantilise veebi kataloog, mida arendatakse W3C-s ühes semantilise veebiga tegelevas töörühmas. Ontaria kogub ontoloogiaid, mis on kirjutatud kasutades OWL-i või RDF-i. Ontaria kataloogi saab iga info looja ennast registreerida, kasutades veebipõhist liitumisvormi. Lisaks kogub Ontaria informatsiooni ka iseseisvalt veebilehtedel liikudes. Kuna Ontaria on veel arenduses (hetkel on väljas versioon 0.8), siis kõik funktsioonid ei tööta, kuid juba on võimalik ontoloogiaid lisada, otsida ning lehitseda.

Semantilise veebi keelte võrdlus

Ontoloogiate kirjeldamiseks võib kasutada erinevaid keeli, kaasaarvatud üldine loogilise programmeerimise keel Prolog. Allpool võrdleme levinumaid ontoloogia kirjeldamise keeli: XML , RDF(S), DAML+OIL, OIL ja OWL.

W3C semantilise veebi töögrupp on defineerinud disaini kriteeriumid semantilise veebi keeltele:

• ontoloogiate jagamine ja versioonindus;

• koostalitlusvõimelisus;

• tugi järelduste tegemiseks;

• tasakaal väljendusvõime ja skaleeritavuse vahel;

• kasutamise lihtsus ja standarditega ühilduvus.

Siinkohal kasutame võrdluseks vaid olulisemaid omadusi: koostalitlusvõime, järelduste tugi, väljendusvõime ja skaleeritavus ning ühilduvus veebi standarditega.

Faktilised teadmised: andmemudelid

Semantilise keele aluseks olevate andmemudelite juures võime tuua välja järgmised erinevused:

• XML -i põhiliseks andmemudeliks on märgistatud puud. Seega saab infot esitada hierarhilise struktuurina.

• RDF-i andmemudel koosneb kolmest objekti tüübist (ressursid, omadused ja laused). Selle andmemudeli abil RDF-i esitatud avaldised on süntaksi suhtes neutraalsed. See baseerub binaarsetel suhetel, mis on laiendatud mehhanismiga, mille abil saab defineerida suhteid suhete vahel ja lauseid lausete kohta. RDFS kasutab seda andmemudelit defineerimaks RDF-i modelleerimise põhimõtteid.

• OIL, DAML+OIL ja OWL baseeruvad kirjeldusloogikal ja freimipõhisel loogikal. Seepärast on need keeled rikkad klasside, omaduste ja aksioomide poolest, mis muudavad nad suurepäraseks vahendiks maailma kirjeldamisel.

Terminoloogilised teadmised: ontoloogiad

Ontoloogiate abil on võimalik defineerida rikka semantikaga keerukaid objekte, seepärast sobivad nad hästi kirjeldamaks heterogeenseid, hajutatud ja osaliselt struktureeritud infoallikaid veebi kontekstis:

• XML annab mehhanismid dokumendi struktuuri ja sisu defineerimiseks. See lubab kasutada elemendi pärimist, atribuute, andmetüüpide definitsioone ja defineerida kasutajal oma andmetüübid.

• RDF(S)-i saab kasutada, et otse defineerida ontoloogia oma objektide, klasside ja omadustega. RDF-i väljendusrikkus on piiratud: sihiteadlikke definitsioone või keerukamaid seoseid saab defineerida aksioomide kaudu.

• OIL, DAML+OIL ja OWL-i eesmärk on pakkuda täielikku tuge ontoloogiate defineerimiseks. Nad pakuvad rikkalikumaid konstruktsioone keerukamate klassi väljendite ja aksioomide formeerimiseks, et oleks võimalik ontoloogia andmete põhjal järeldusi teha. Oluline omadus nendel keeltel on, et nad on jaotatud kihtideks nii, et rahuldada erinevaid vajadusi ja võimaldada nii lihtsate kui ka keerukate ontoloogiate defineerimine.

Arutlus - teadmiste järeldamine

Arutlus (või argument või tõestus) on lausete hulk, milles üks lause on valitud kui järeldus ja ülejäänud laused on selle lause eelduseks. Teadmiste järeldamise seisukohast on ontoloogia keelte põhilised erinevused:

• XSLT (XSL transformatsiooni keel) võimaldab transformatsioonide väljendamist XML -struktuuris, et anda edasi teatud järelduslikke teadmisi.

• RDF/RDFS - saab kasutada alamklassi seost.

• OIL, DAML+OIL ja OWL lubab defineerida keerukaid reegleid või aksioome.

Kokkuvõtlikult võiks öelda, et kõiki semantilise veebi keeli iseloomustab see, et nad põhinevad XML keelel.

Kõige uuem ja väljendusrikkam neist keeltest on OWL, millega saab avaldada ja jagada ontoloogiaid veebis. Praegusel hetkel on aga ikkagi kõige populaarsem ja enim kasutatav RDF, mis on küllaltki võimas ning selle abil saab edasi anda üpris

detailselt sisu. Üks populaarsuse põhjuseid kindlasti on see, et OWL-i jaoks on veel küllaltki vähe abimaterjale, mida aga ei saa öelda RDF kohta.

Semantilise veebi keeltest, täpsemalt nende keelte vahel valiku tegemisest, räägin aga pikemalt järgmises Arvutimaailma numbris.

- Lahendused