Sissejuhatus semantilisse veebi, 3: Sematilise veebi keelte valik

21 aastat tagasi - 11.03.2005 Autor: AM

(Arvutimaailm 3/05)

Autor: **Mario Peterson**

Eelmises Arvutimaailmas rääkisin ontoloogiast ja sellest millised semantilise veebi keeled on olemas. Käesolevas artiklis aga annan ülevaate keelte erinevustest ning nende plussidest-miinustest valiku tegemisel.

Semantilise veebi keele valik

Põhimõtteliselt võib süsteemi looja semantilist võrku ka paberil kirjeldada ja seejärel „tõlkida" selle mõnda standardsesse keelde, et ka teised seda kasutada saaks.

Keele väljendusrikkus määrab ära, kui täpselt me saame reaalset maailma kirjeldada. Tüüpiliselt ei hoia arendusvahendid infot mõnes semantilise veebi keeles, vaid näiteks relatsioonilises andmebaasis või arendusvahendile spetsiifilises vormingus.

Niisiis on oluline, et arendusvahend võimaldaks koostada vähemalt sama väljendusrikast ontoloogiat kui valitud ontoloogia keel seda võimaldab. Ehk siis arendusvahend peab toetama valitud ontoloogia keelt st peab olema võimalik olemasolevat (antud keelset ontoloogiat) importida. Teisalt peab arendusvahendiga koostatud ontoloogiat olema võimalik teisendada valitud ontoloogia keelde.

RDF ja OWL

Ühte ja sama terminit võib erinevates kontekstides kasutada ka erinevates tähendustes. Ning erinevaid termineid võib kasutada elementide kohta, millel on sama tähendus.

RDF ja RDFS hakkasid lähenema sellele probleemile lubades lihtsa semantika siduda identifikaatoritega.

RDFS-iga saab defineerida klasse, mis võivad omada alamklasse ja superklasse ning defineerida omadusi, millel võib olla alamomadusi, valdkondi ja vahemikke. Selles mõttes on RDFS lihtne ontoloogia keel.

Siiski, et saavutada koostöö paljude autonoomselt arendatud ja juhitud skeemide vahel, on vaja rikkalikumat semantikat. Näiteks ei oska RDFS kirjeldada, et klassid Puuvili ja Lennumasin on seosetud, või et stringi trio taga peitub täpselt kolm muusikut. See ongi üks põhjus, miks on RDF-i asemel mõistlik kasutada uuemat ja väljendusrikkamat keelt OWL.

DAML+OIL ja OWL

Kuna DAML+OIL oli peamine eeskuju OWL-i väljatöötamisel siis on nende erinevus pigem triviaalne. OWL-i abstraktne süntaks on erinev DAML+OIL-i grupeerimise aksioomidest, selle asemel kasutatakse freimi-sarnast struktuuri, mis on omane OIL-ile.

RDF

Eeskätt on RDF mõeldud veebiressursside kohta metaandmete esitamiseks. Veebiressursside identifitseerimiseks kasutab RDF URI-sid. Seega saab RDF-i kasutada kõigi asjadega, mida saab URI-dega identifitseerida isegi, kui neid ei saa veebist otseselt kätte.

RDF on mõeldud kasutamiseks juhtudel kui andmeid kasutavad võrgurakendused. See on raamistik info kirjeldamiseks nii, et seda oleks võimalik vahetada tarkvaraliste rakenduste vahel, ilma et selle tähendus kaduma läheks. Põhiideeks on tagada rakendusest ja platvormist sõltumatu andmete vahetamise raamistik.

Toodud näites (Lihtsustatud RDF dokument XML formaadis) on URI http: // sw. xp. ee/ mario. php veebilehe identifitseerimiseks. Omadus „author” kirjeldab selle veebilehe autorit, mille väärtus on „Mario Peterson”. Omadus „created” kirjeldab, millal see leht loodi ja omadus „modified” näitab, millal seda lehte viimati muudeti. RDF-i terminoloogia kohaselt võib ka öelda, et:

• URI http: // sw. xp. ee/ mario. php on subjekt;

• Element <author> on omadus;

• „Mario Peterson” on väärtus.

Dublin Core

Dublin Core on standard, mida kasutatakse enamasti raamatute kartoteegi koostamisel.

Dublin Core on hulk dokumentide kirjeldamiseks eeldefineeritud omadusi. Esimesed Dublin Core omadused defineeriti Metadata Workshopil 1995. a.USA-s.

Senini tegutseb organisatsioon nimega Dublin Core Metadata Initiative (DCMI), mis tegeleb metaandmete standardite ja raamistike väljatöötamisega.

Dublin Core viimane metaandmete hulga versioon on 1.1. Selles standardis on kirjeldatud kümme välja:

• nimi - silt, mis on antud andmeühikule;

• identifikaator - andmeühiku unikaalne identifikaator;

• definitsioon - lause, mis väljendab selgelt andmeühiku olemust ja sisu;

• kommentaar - märkus andmeühiku rakendamise kohta;

• versioon - andmeelemendi versioon;

• registreerija - kes on volitatud registreerima antud andmeühikut;

• keel - keel, milles andmeühik on kirjeldatud;

• saadavus - kas andmeühik on saadaval pidevalt või mitte;

• andmete tüüp - näitab andmete esitamise tüüpi;

• maksimaalne kogus - palju on antud andmeühikut üldse kogus.

RSS

RSS (Really Simple Syndication) on standard, mida kasutatakse uudiste ja muu uudislaadse info jagamiseks masinloetaval kujul.

RSS-i on kokku seitse erinevat versiooni, põhiliselt kasutatakse versioone 1.0 ja 2.0. Versioon 1.0 on RSS-DEV Working Groupi poolt arendatav ja baseerub RDF-il. Kõik suuremad uudiste portaalid (näiteks: Wired, Slashdot) annavad oma uudiseid välja RSS formaadis.

Nii saab inimene uudised kätte, ilma, et peaks kõiki portaale ükshaaval külastama. Tüüpiliselt on RSS-is kirjeldatud uudise pealkiri, kuupäev, autor, lühikirjeldus ja link täispikale uudisele.

Sellisel kujul uudiseid saab uudisteagent vabalt kasutada. RSS-i lugemiseks on palju meilikliendi sarnaseid rakendusi, nt eestlaste tehtud Feedreader.

Reeglina võimaldavad rakendused registreerida URL-e, millelt saab RSS-i faili allalaadida ja siis näidata infot inimloetavalt. Lisaks pakuvad paljud rakendused lisavõimalusi: otsing, uuest uudisest teavitamine jne.

SHOE

SHOE on HTML-i laiendus, mis võimaldab autoritel luua veebilehtedele masinloetavaid annotatsioone.

SHOE oli suhteliselt hästi arenenud ja teda oli lihtne kasutada. Selle annotatsioonid kirjutatakse vastavalt kasutatavale ontoloogias defineeritud siltide (tagide) vahele. Samuti saab seal ka oma ontoloogia defineerida või laiendada laenatud ontoloogiat.

SHOE on oluline ajaloolistel põhjustel - tema abil loodi esimene semantilise veebi leht - Mindswap. Sama veebileht oli ka esimene (november 2001) veebileht, mis andis andmeid OWL-is.

Tänaseks aga on SHOE keele edasiarendamine lõpetatud, kuna leiti, et vaja on võimsamat keelt.

OWL

OWL on mõeldud andmete esitamiseks nii, et tarkvaralised rakendused saaksid seda kasutada. OWL on semantiline ülestäheldamise keel, mis on loodud, et avaldada ja jagada ontoloogiaid Internetis.

OWL on ühilduv nii OIL-i kui DAML+OIL-iga. OWL on ontoloogia keeltest noorim - lõplik versioon valmis 2004. a veebruaris. Sellest hoolimata on see juba aastaid kasutuses ja peale lõpliku versiooni väljatulekut toetavad seda kõik olulisemad arendusvahendid.

OWL on arendatud RDF-i sõnastiku laiendusena ja tuletatud DAML+OIL veebi ontoloogia keelest.

Alamkeeled - skaleeritavus

OWL-il on kolm alamkeelt, mis tagavad skaleeritavuse:

• OWL Lite sisaldab ainult esmavajalikku: klasside hierarhiat ja lihtsaid piiranguid.

• OWL DL pakub maksimaalset väljendusrikkust, samas on säilinud lõplik arvutatavus (on garanteeritud kõikide järelduste arvutatavus) ja otsustatavus (kõik arvutused lõpetatakse piiratud aja jooksul). OWL DL sisaldab kõiki OWL-i võimalikke konstruktsioone, kuid neid saab kasutada teatud piirangutega (näiteks: klass võib olla paljude klassside alamklass, kuid ta ei saa samal ajal olla mõne teise klassi eksemplar). OWL DL on saanud nime kirjeldusliku loogika järgi (Description Logics).

• OWL Full on mõeldud kasutajatele, kes vajavad maksimaalset väljendusrikkust ja süntaksilist vabadust RDF-ist, samas puuduvad arvutuslikud garantiid. Näiteks saab OWL Fullis käsitleda klassi samaaegselt eksemplaride kogumina kui ka eraldi eksemplarina. On ebatõenäoline, et leidub tarkvara, mis suudab teha järeldusi kogu OWL Fulli võimaluste hulgal.

Kuna iga OWL-i alamkeel on oma eelmise versiooni laiendus siis öeldakse, et:

• iga OWL Lite ontoloogia on samal ajal ka OWL DL ontoloogia;

• iga OWL DL ontoloogia on samal ajal ka OWL Full ontoloogia;

• iga tõene OWL Lite järeldus on samal ajal ka tõene OWL DL järeldus;

• tõene OWL DL järeldus on samal ajal ka tõene OWL Full järeldus.

Sobivama alamkeele valikul tuleks kaaluda, milline alamkeel rahuldab kõige paremini vajadusi. Valik OWL Lite ja OWL DL vahel sõltub kui suures ulatuses on vaja väljendusrikkamaid konstruktsioone OWL DL-ist.

Valik OWL Full-i ja OWL DL-i vahel sõltub kui suures ulatuses kasutatakse metamodelleerimise konstruktsioone RDFS-ist

Järelduste tegemine

Kirjeldusloogika (DL) võimaldab spetsifitseerida terminoloogia hierarhia, kasutades piiratud hulka esimese järgu valemeid. Tänu kirjeldusloogikale on OWL-is võimalik täita olulisi loogilisi nõudeid. Need nõuded sisaldavad kontseptsiooni rahuldatavust, klasside konsistentsust ja instantside kontrollimist.

Kokkuvõttes võib öelda, et OWL on ainuke keel, millesse on versiooninduse funktsionaalsus sisse ehitatud. Järelduste tegemise mehhanism on sama, mis DAML+OIL-il ja see baseerub avatud maailma eeldusel (OWA). OWL on väga väljendusrikas ja kihiline arhitektuur tagab ka skaleeritavuse. Seega oleks OWL teoreetiliselt parim keele valik.

Reaalse süsteemi koostamisel tuleb aga lähtuda ka muudest teguritest nagu keele teostamise keerukus ja olemasolevate arendusvahendite valmisolek.

Seega tuleb loobuda väljendusrikkusest ja järelduste tegemisest ning valida formaat, mida oleks võimalik ka reaalses elus kasutada - RDF.

Järgmises numbris räägin ontoloogiate kasutamist juba reaalse süsteemi loomisel.

- Lahendused