Sissejuhatus semantilisse veebi 1

21 aastat tagasi - 29.01.2005 Autor: AM

(Arvutimaailm 1/05)

Autor: **Mario Peterson**

Semantilise veebi idee seisneb selles, et veeb oleks kasulik mitte ainult inimene-inimene suhtlemisel, vaid et ka masinad saaksid suhtluses osaleda ja abiks olla - tekib „intelligentsem võrk”.

Käesolev artikkel on sissejuhatus semantilisse veebi, kus annan ülevaate semantilise veebi paradigmast ja selle võimalikust rakendamisest eri valdkondades.

Senise mudeli ammendumine

Viimase 15-20 aasta jooksul on inimkond loonud rohkem informatsiooni, kui kogu eelneva ajaloo jooksul kokku. Sellise infohulga juures on mõeldamatu, et leiduks inimene, kes kogu seda infot valdab, isegi selles orienteerumine on muutumas võimatuks.

90. aastate alguses nähti Internetis tohutut potentsiaali info vahetamises. Siis tundus kõik väga loogiline, kuna traditsiooniline Internet on väga sarnane raamatukoguga - teoste kohta on olemas annotatsioonid ja kartoteegid.

Otsingumootorid nagu näiteks Google, Yahoo , Lycos jt tunduvad olevat suurepärased tööriistad veebimaailmas orienteerumiseks. On loodud tohutu klasterdatud süsteem, mille abil on realiseeritud terabaitidesse ulatuvat infot mahutavad andmebaasid. Selle tulemusena on võimalik sooritada märksõnapõhist otsingut mikrosekundite jooksul, kuid siiski saab otsida vaid märksõnu.

Siit tulenebki HTML-i (Hypertext Markup Language) keele peamine puudus - me saame kirjeldada selle keele abil ainult kujunduslikke elemente (kirjatüüp, suurus jne). HTML ei võimalda aga anda dokumentides kirjeldatud infole tähendust. Seega peab huvitava materjali leidmiseks vaatama ükshaaval üle kõik otsingu tulemused.

Üheks lahenduseks on kitsendada otsingu tingimusi. See lähenemine töötab niikaua kuni infohulk Internetis on väike. Kui aga infohulk on väga suur, siis olemegi olukorras, kus meil on tohutul hulgal päringutulemusi, kuid otsitavat me siiski ei leia.

Näiteks proovisin otsingumootorite abil leida Internetist kõiki artikleid, mille autor on Tim Berners-Lee. Päringu tulemusena sain vastuseks umbes 150 000 veebilehte.

Seega oleks mõistlik see töö anda arvutiprogrammile, mis leiaks vajalikud artiklid, kuid see osutub võimatuks, sest nendel veebilehtedel esitatud info ei ole arvutile arusaadav. Nad lihtsalt ei oma teadmisi ala kohta, mida neilt küsitakse. Semantilises veebis aga sellist olukorda ei tekiks.

Semantiline veeb ja semantiline võrk

Semantilise veebi mõiste on tihedas seoses semantilise võrgu mõistega. Semantilise võrgu mõiste võeti kasutusele oluliselt varem (1968, Ross Quillian), kui hakati rääkima semantilise veebi kontseptsioonist.

Semantiline võrk

Semantilise veebi põhiidee on esitada andmeid semantilise võrguna. Semantilist võrku aga esitatakse tavaliselt suunatud graafina, ehk säärase joonisena, kus objektid on esitatud sõlmedena ning sõlmi ühendavad kaared näitavad ära nendevahelised seosed. Näiteks mingi etteantud ontoloogia graafiline esitlus annab sellest hea ülevaate.

Semantiline veeb

Semantiline veeb on olemasoleva veebi laiendus, kus informatsioonile on lisatud ka selle tähendus. Semantilise veebi kontseptsioon on raamistik, mille kohaselt peaks andmed veebis olema defineeritud ja lingitud selliselt, et neid saaks kasutada efektiivsemalt automatiseerimiseks, integreerimiseks ja erinevate rakenduste vaheliseks taaskasutamiseks. Interneti standardimise ja arendamisega tegelev W3C (World Wide Web Consortium) on võtnud semantilise veebi loomise üheks oma kolmest pikaajalisest põhieesmärgist. See on defineeritud nii: „Luua tarkvara keskkond, mille abil iga kasutaja saaks suurimat kasu nendest allikatest, mis on veebis saadaval”.

Selleks, et kogu semantiline veeb oleks kasutatav, on see vaja kihiliselt üles ehitada. Ehk siis luua infrastruktuur, kus ei ole tegu pelgalt veebilehtedega, vaid

andmebaaside, teenuste, sensorite, isiklike seadmete ja isegi majapidamisseadmetega, mis nii tarbivad kui toodavad veebi informatsiooni.

Algama peaks altpoolt lihtsamatest asjadest ja minema ülespoole keerukamate juurde.

Semantilise veebi kontseptsiooni kasutamine on end juba tõestanud mitmetes eluvaldkondades. Väga palju on uuritud taolise kontseptsiooni kasutamist ja on pakutud ka mitmeid rakendusi:

• Saab teha intelligentsemat otsingut veebist. Populaarseimas otsingusüsteemis Google saab otsingu jaoks sisestada ainult märksõnu. Vastusena kuvatakse vaid märksõnu sisaldavad lehed. Sisulist otsingut aga sooritada ei saa. Näiteks ei saa öelda, et anna kõik artiklid, mille autor on Rein Tõru. Semantiline veeb aga võimaldaks seda.

• Infosüsteemide ühendamise võimalus. Erinevate üksuste andmeid saab kokku viia ilma suuremahulise käsitsitöö tegemist. Vajalik nt suurte üksuste (tanklaketid, pangad jmt) andmete kokkuviimisel.

• Kasutajaõiguste süsteem (kes kuhu ligi pääseb). Luua tavaelust lähtudes paindlik kasutajaõiguste süsteem. Mitte rangelt reglementeeritud süsteem.

• Intelligentsed assistendid. Assistent otsib kasutajale vajalikku infot, nt uudiseid, hindu, pakkumisi jne ning avaldab need eelnevalt kokkulepitud kujul ja tingimustel.

• Meditsiinis haiguste diagnoosimiseks. Ontoloogiate (mõistete hierarhia) abil kirjeldatakse ära haiguste sümptomid. Arst peab sisestama vaid haige teatud parameetrid (temperatuur, vererõhk, kaebused jne) ja süsteem oskab pakkuda võimalikke diagnoose.

• Samuti leiab semantiline veeb kasutamist geenide uurimisel, turvaanalüüsis, elektroonilise info vahetamiseks e-äris jne.

Semantilise veebi visiooni elluviimine eeldab uusi lahendusi ja uusi võrgurakenduste topoloogiaid. Traditsioonilisi klient-server tüüpi lahendusi on mõistlik kasutada seni kuni on võimalik ja mõttekas andmeid ja teenuseid tsentraliseerida. Suurte andmemahtude ja suure arvu allikate puhul muutub see võimatuks.

Intelligentsed agendid

Semantilise veebi peamiseks eesmärgiks on panna arvutid omavahel suhtlema, mitte ainult bitijadana, vaid et nad saaksid aru ka edastatava info sisust. Tekiks n-ö „masinate võrk”, kus arvutid küsiksid ja saaksid üksteiselt infot.

Semantilise veebi eesmärgiks on ka struktureerida veebilehtede sisu, luues keskkonna, kus tarkvaraagendid suudaksid peaaegu iseseisvalt kasutaja huvides sooritada keerukaid ülesandeid.

Siin tulevad appi intelligentsed agendid, ehk väikesed tarkvaraprogrammid, mis oskavad kasutaja poolt tehtud päringutele vastates ammutada informatsiooni Internetist.

Intelligentsed agendid on poolautomaatsed arvutiprogrammid, mis assisteerivad kasutajat mitmesuguste rakenduste kasutamisel. Agendid mitte ainult ei kasuta saadavalolevat semantilist infrastruktuuri, vaid ka täiendavad ja hoiavad seda infrastruktuuri üleval.

Lisaks sellele peaks agent pakkuma kasutajale nii nõu kui ka teenuseid. Hea agent peaks olema võimeline tegutsema ilma, et kasutaja teaks kõiki üksikasju, et kasutajad saaksid delegeerida oma ülesandeid agentidele. Sellisted ülesanded on näiteks informatsiooni otsimine, klassifitseerimine ja hoidmine. Samuti e-kirjade lugemine, kohtumiste kokkuleppimine, päeviku pidamine jne.

Mida vähem on vaja järelvalvet agendi üle, seda parem. Agent peaks olema kohanemisvõimeline - võimeline muutma oma käitumist vastavalt kasutaja tagasisidele ja keskkonnateguritele - st õppimisvõimeline.

Hea Internetiagent peab olema suuteline suhtlema. See on ainult siis võimalik, kui ta „räägib” kasutajaga sama keelt. Agent, kes ei saa aru, mida temalt küsitakse, ei ole abiks. Põhiline probleem otsingumootorite puhul ongi see, et kuigi nad baseeruvad keelel (või vähemalt lingvistilistel sümbolitel), ei ole neil teadmisi ala kohta, mida neilt küsitakse.

Selliseid tarkvaraagente ei eksisteeri veel väljaspool laboratooriumi tingimusi. Nad ei ole veel piisavalt arenenud, et neid Internetis kasutada. Küsimus on ontoloogiates, mille abil kasutaja saaks agendiga suhelda.

Semantiliste ressursside leidmine ja registreerimine

Semantiliste ressursside leidmise ja registreerimise all peetakse silmas otsingumootori süsteemi veebis leiduvate metaandmete ja ontoloogiate tarvis.

Ontoloogiate otsimisel ei saa kasutada inimloetava info otsimiseks kasutatavaid otsingumootoreid (nagu näiteks Google, Yahoo jt), kuna otsijateks on robotid - tarkvaralised agendid.

Semantilise veebi ontoloogiatest rääkides on oluline silmas pidada ka ontoloogiatele viitamist. Enne kui uut ontoloogiat luua, peab otsima juba loodud ontoloogiate hulgast, kas sarnast ontoloogiat ei ole juba loodud. Kui on, siis saab kasutada viitamist sellele ontoloogiale ning ei pea loodavas kõiki mõisteid „lahti seletama”.

Ressursside leidmine ongi intelligentsete agentide puhul üks olulisemaid ja keerukamaid probleeme. Kõige lihtsam oleks kombineerida ressursside register otsingumootoriga. Sel juhul peaks iga allikas oma URI (üldine ressursilokaator, standard Interneti objektide määratlemiseks) registreerima ja kui allikas tehakse muudatusi, peaks otsingumootor informatsiooni uuesti indekseerima.

Üheks lahenduseks on W3C otsingurobot Webbot, mis töötab nagu teisedki robotid, järgides linke lehelt-lehele. W3C Webbot oskab tulemusi salvestada nii logifaili kui ka andmebaasi. Kasutatakse MySQL andmebaasimootorit, mis on kombineeritud failide allalaadimisega.

- Lahendused