

Tietämyksen mallintaminen semanttisessa webissä¹

Tomi Kauppinen, Tuukka Ruotsalo ja Mirva Salminen
TKK Viestintäteknikka ja
Tietojenkäsittelytieteen laitos, Helsingin Yliopisto.

Semanttisessa webissä mallinnetaan semantiikkaa koneiden käyttöön toisin kuin perinteisessä webissä, jonka dokumentit ovat ihmisille tarkoitettuja. Näin koneet saadaan valjastettua sekä tietämyksen tarkkaan tallennukseen että valtavien tietomäärien käsittelyyn. Tässä käytetään apuna ontologioita eli koneymmärrettävästi ilmaistuja käsitteitä ja niiden välisiä suhteita. Sovelluskohteina ovat esimerkiksi tietämyksen ja aineistojen hallinta sekä elektroninen kaupankäynti.

Semanttisen webin teknologiat ja ontologiat

Semanttisessa webissä [1] mallinnetaan semantiikkaa koneiden käyttöön toisin kuin perinteisessä webissä, jonka dokumentit ovat ihmisille tarkoitettuja. Semantiikka mallinnetaan ontologioilla, eli koneymmärrettävästi ilmaistuilla käsitteillä ja niiden välisillä suhteilla.

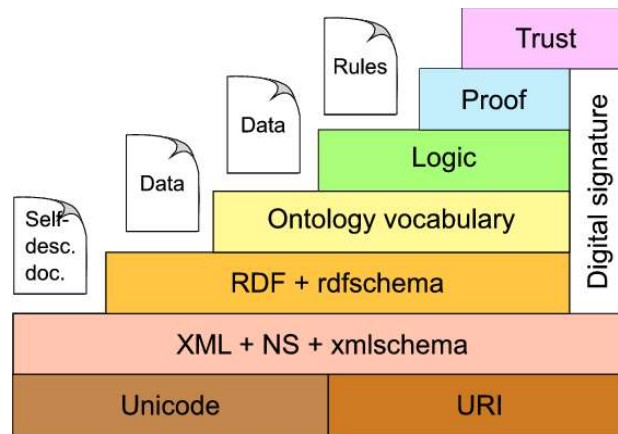
W3C:n johtajan Ivan Hermanin mukaan [2] semanttiselle webille on tapahtumassa sama kuin webille tapahtui useita vuosia sitten: kunnollisten työkalujen, järjestelmien ja sovellusten myötä liiketoiminta vakiintuu. Tämä on mahdollista, koska semanttisen webin teknologiat ovat jo tulossa yritysten perusteknologioiksi.

Semanttisen webin teknologiat [1] helpottavat ontologiatyön tekoa käytännön tasolla, toisin sanoen ne tarjoavat välineet, joiden avulla on mahdollista määritellä semanttisen tiedon eli käsitteet sekä niiden väliset suhteet täsmällisesti ja koneymmärrettävästi. Käsitteet identifioidaan yksiselitteisesti URIen (Universal Resource Identifier) avulla, jolloin niihin voidaan viitata globaalisti. Ontologioiden kuvaamiseen on kehitetty esimerkiksi OWL (Web Ontology Language).

Ontologian käsitteitä voidaan käyttää esimerkiksi metadatana sisällönkuvaailussa aina julkisten museoiden ja kirjastojen kokoelmista yksityisten ihmisten musiikkikokoelmiin ja yritysten verkossa tarjoamiin palveluihin asti. Kun mahdollisimman moni jakaa saman käsitteistön, jossa käsitteillä on identifioivat tunnisteet, voidaan järjestelmien integrointi suorittaa tehokkaammin. Lisäksi voidaan saada selville aineistojen välisiä mielenkiintoisia ja yllättäviäkin suhteita päättelykoneiden avulla.

Kuvassa 1 on esitetty Tim Berners-Leen esittelemä Semanttisen webin "teknologiakakku" (layer cake) [3]. Kakku kuvaa kuinka perusteknologioiden Unicoden ja URIn varaan on rakennettu syntaksikerros XML:n nimiavaruuksien (NS) avulla ja tietomallikerros RDF:n ja RDF skeeman avulla. Tällä hetkellä rakennetaan erityisesti sekä seuraavina tulevaa ontologiakerrosta (ontology vocabulary) että logiikkakerrosta (logic). Seuraavaksi rakennetaan todistuserrosta (proof) ja lopulta ideana on rakentaa luottamuskerros (trust).

¹Julkaistu Systeemyö-lehden teemanumerossa "Tiedon ja tietämyksen hallinta" 4/2005.



Kuva 1 Semanttisen webin teknologiakakku (layer cake) Tim Berners-Leen mukaan [3].

Ontologioilla mallinnetaan siis käsitteitä ja niiden välisiä suhteita. Esimerkiksi *purjehduksessa käytetään purjevenettä*, jonka yksi osa on *purje*. *Purjevene* puolestaan on eräänlainen *vene*. Samoin moottoriveneen osa on moottori ja se myös on *vene*. Suhteista ja käsitteistä muodostuu mallinnuksen tuloksena verkko, johon perustuen voidaan tehdä päätelmiä käyttäen logiikkaa ja heuristisia sääntöjä. Päätelykone voi tehdä ontologian avulla eron esimerkiksi purjeveneen ja moottoriveneen välillä sekä toisaalta nähdä missä mielessä niillä on yhteisiä piirteitä. Ontologioiden käyttö mahdollistaa myös vaikeiden ja pitkien päättelyketjujen teon: esimerkiksi kuvaa laajemman ontologian perusteella voitaisiin löytää yhteys navigointivälineiden ja purjeveneen välillä.

Yläontologiat yhdistävät kohdealue tietämystä

Yksittäisten sovellusten käyttämä tieto esitetään usein kohdealuetta kuvaavalla ontologialla tai tietomallilla. Semanttisen webin keskeisenä ajatuksena on yhdistää metatietoja verkossa, eli kohdealueiden tietomalleja semanttiseksi verkoksi, jossa päättely ja automaattinen tietojen yhteiskäyttö on mahdollista.

Eräänä ratkaisukeinona kohdealueiden yhteentoimivuuden saavuttamiseksi on tarjottu yläontologioita. Ne tarjoavat yhteisen, laajennettavan peruskäsitteistön, joiden avulla mahdollistetaan erilaisten sovellusten ja niiden hyödyntämien tietomallien yhteentoimivuus [4]. Ideana on siis, että yläontologiassa esitellään joukko käsitteitä ja suhteita, joita tarkennetaan kohdealuekohtaisissa ontologioissa. On epärealistista ajatella, että kaikki jonkin kohdealueen tietomallit voidaan mallintaa yhdessä sovitulla tavalla. Tämän vuoksi tarvitaan keinoja kytkeä uusia käsitteitä ja kohdealueontologioita yhteen.

Esimerkiksi yläontologian käsitteen *vene*, suhteen voitaisiin toisessa kohdealueessa kuvata tarkemmin käsitteet *purjevene* ja *moottorivene*. Käyttämällä sopivaa sääntökieltä voidaan luoda myös monimutkaisempia käsittemalleja. Esimerkiksi käsite *purjehtija* voidaan kuvata käsitteen *purjehtiminen* suhteen säännöllä, jossa määrätään *purjehtija* käsitteen ilmentymiksi kaikki ne *ihminen* -käsitteen ilmentymät, jotka on kytketty tapahtumaan *purjehtiminen* millä tahansa kaarella. Näin mahdollistetaan toisistaan poikkeavien kohdealueontologioiden käyttäminen osana laajempaa tietämuskantaa. Menetelmien kehittäminen voidaan lisäksi kohdistaa ontologioihin kohdealuekohtaisten tietomallien sijaan.

Esimerkkinä kohdealueontologioista voi mainita, että tiedonhallinnan alueella on DAMA Finlandissa käynnistetty tiedonhallinnan keskeisen käsitteistön kokoaminen ontologiaksi.

Muuttuvan maailman mallintaminen ontologioilla

Ontologioilla pyritään kuvaamaan maailmaa --- mutta ongelmaksi muodostuu helposti se, että mallinnettava maailma muuttuu. Esimerkissämme *vene*-käsitteen tarkentavina käsitteinä voi nykyisin olla käsitteet *moottorivene* ja *purjevene*, mutta aikaa ennen moottoriveneen keksimistä kuvaavassa ontologia-aikasiivussa käsite *moottorivene* puuttuu.

Myös maantieteellinen käsitteistö muuttuu. Jos esimerkiksi lähitulevaisuudessa tehdään päätös joidenkin kuntien yhdistämisestä ja siten kuntien kokonaismäärän vähentämisestä, niin myös paikkojen ontologiaa on muokattava. Tämänkaltaisten muutosten mallinnusta voidaan tehdä ontologia-aikasarjan ja muutossiltojen [5] avulla ja hyödyntää tätä tietoa hakukoneissa. Niiden avulla on esimerkiksi saatu selville, että nykyinen Lappeenranta peittää 1900-luvun vaihteen Viipurista n. 12.5%. Sen ajan Viipurilla indeksoitu museoesineistö on siis todennäköisesti osittain peräisin nykyisen Suomen alueelta --- tieto, jota voidaan hyödyntää aineistojen hallinnassa. Päättelykoneita käyttämällä voidaan semanttisen webin ontologioista siis kaivaa yllättävää tietämystä ja tarkentaa siten hakutuloksia.

Kirjoittajat

Tutkijat Tomi Kauppinen, Tuukka Ruotsalo ja Mirva Salminen työskentelevät Teknillisen korkeakoulun viestintätekniikan laboratoriossa ja Helsingin yliopiston tietojenkäsittelytieteen laitoksella toimivassa semanttisen laskennan tutkimusryhmässä (SeCo). Tomi Kauppinen on toiminut DAMA Finlandin hallituksessa sen perustamisesta lokakuusta 2004 lähtien ja XML Finland ry:n hallituksessa vuosina 2000-2005.

Lisätietoja:

W3C:

<http://www.w3c.org/2001/sw/>

Semanttisen laskennan tutkimusryhmä (SeCo):

<http://www.seco.tkk.fi/>

FinnONTO-hanke:

<http://www.cs.helsinki.fi/group/seco/ontologies/>

Dama Finland:

<http://www.damafinland.org/>

Lähteet:

- [1] Tim Berners-Lee, Jim Hendler, and Ora Lassila. The Semantic Web. *Scientific American*, 284(5):34–43, May, 2001.
- [2] Ivan Herman. Towards the Semantic Web. Invited talk and tutorial at Web Intelligence symposium. The Finnish AI Conference, 20 year Anniversary (STeP 2004), Heureka - the Finnish Science Centre, Vantaa, Finland, September 2-3, 2004.
- [3] Tim Berners-Lee. Semantic Web Wedding Cake (or "layer cake") in a conference talk, XML 2000 conference, Washington DC, 2000.
- [4] Eero Hyvönen, Arttu Valo, Ville Komulainen, Katri Seppälä, Tomi Kauppinen, Tuukka Ruotsalo, Mirva Salminen and Anu Ylisalmi. Finnish National Ontologies for the Semantic Web - Towards a Content and Service Infrastructure. *Proceedings of International Conference on Dublin Core and Metadata Applications (DC 2005)*, Madrid, Spain, Sep, 2005.
- [5] Tomi Kauppinen and Eero Hyvönen. Modeling and Reasoning about Changes in Ontology Time Series. Chapter in book. *Ontologies in the Context of Information Systems* (Rajiv Kishore, Ram Ramesh and Raj Sharman (eds.)), Springer-Verlag, Berlin, to appear, 2006. In press.