



25.1.2008

Semanttinen web kansalliseksi voimavaraksi - FinnONTO-hankkeen visio ja tulokset

Eero Hyvönen

Teknillisen korkeakoulun mediatekniikan laitos ja
Helsingin yliopiston tietojenkäsittelytieteen laitos
Semanttisen laskennan tutkimusryhmä

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

Semanttinen web on WWW:n kansainvälistä kehitystä koordinoivan W3C-järjestön ja webin ”isän” prof. Tim Berners-Leen näkymys koko webin seuraavasta kehitysaskelista. Ydinajatuksena on, että tietokoneiden pitäisi kyetä ”ymmärtämään” verkossa olevia sisältöjä niihin liittyvän metatiedon ja yhteisten käsitteistöjen eli ontologioiden avulla. Tämä mahdollistaa verkkosisältöjen yhteentoimivuuden sekä aiempaa olennaisesti ”älykkäämpien” verkkopalveluiden toteuttamisen. Suomessa semanttisen webin suurhanke FinnONTO 2003-2007 loi perustan kansalliselle sisältöinfrastruktuurille verkossa ja demonstroi uusien semanttisen webin ja Web 2.0 teknologioiden mahdollisuuksia laajoissa portaalihankkeissa KulttuuriSampo.fi ja TerveSuomi.fi.

Oletetaan, että etsit ”helsinkiläisiä ravintoloita” webistä. Semanttisessa webissä toimiva palvelu ymmärtää suositella illallista Kalliossa sijaitsevassa pizzeriassa, koska ”Kallio” on osa Helsinkiä, eikä tässä yhteydessä tarkoita luonnonmuodostelmaa tai presidentti Kyösti Kalliota, ja koska ”pizzeria” on ”ravintolan” eräs alatyyppejä. Samalla sinulle voidaan samalla tarjota lisätietoa pizzakulttuurista Wikipedian kautta, suositella lähellä olevan teatterin Italia-aiheista revyytä, jossa on sopivasti pari peruutuspaikkaa, tai tarjota matkapuhelimeen navigointipalvelu.

Kohti kansallista sisältöinfrastruktuuria

Tällainen älykäs kokonaisvaltainen palvelu on mahdollista web-sisältöjen metatietojen ja niihin liittyvien käsitteistöjen eli ontologioiden avulla. Edellytyksenä kuitenkin on, että sisällöt on kuvailtu koneen ymmärtämällä ja yhteentoimivalla tavalla eri organisaatioissa. Tätä varten W3C:ssä on luotu standardit toisaalta metatiedon esittämiseen (RDF), toisaalta ontologisille sanastoille (OWL). Nämä ovat sovellusalue- ja kieliriippumattomia standardeja, eräänlaisia luurankoja ilman lihaa ja elämää. Semanttisen webin syntyminen edellyttää lisäksi yhteistä sopimista metatietojen esittämisestä eri sovellusalueilla, esimerkiksi miten kuvataan yrityksen palvelut tai kokoelmassa oleva museoesine, sekä yhteisten ontologioiden kehittämistä, jotta esimerkiksi sanan ”Kallio” eri merkitykset eivät mene tietokoneessa sekaisin.

Kansallisen FinnONTO-hankkeen 2003-2007 tavoitteena oli luoda maahamme suomenkieleen perustuva ontologiaperustainen ”sisältöinfrastruktuuri”, jonka avulla eri alojen erilaisten sisältöjen yhteentoimivuus olisi mahdollista semanttisessa webissä. Hankkeessa kehitettiin uusinta web-teknologiaa ja sovelluksia. Samalla innovoitiin uudenlainen yhteistoimintaan perustuva ajattelumalli verkkosisältöjen julkaisemiseen webissä, joka samanaikaisesti voi parantaa verkkopalveluiden laatua loppukäyttäjille ja tehostaa eri alojen tietosisältöjen tuotantoprosesseja ja työnjakoa. Esimerkkinä semanttisen webin uusista mahdollisuuksista toteutettiin mm. prototyypit semanttisista portaaleista MuseoSuomi.fi, KulttuuriSampo.fi ja TerveSuomi.fi.

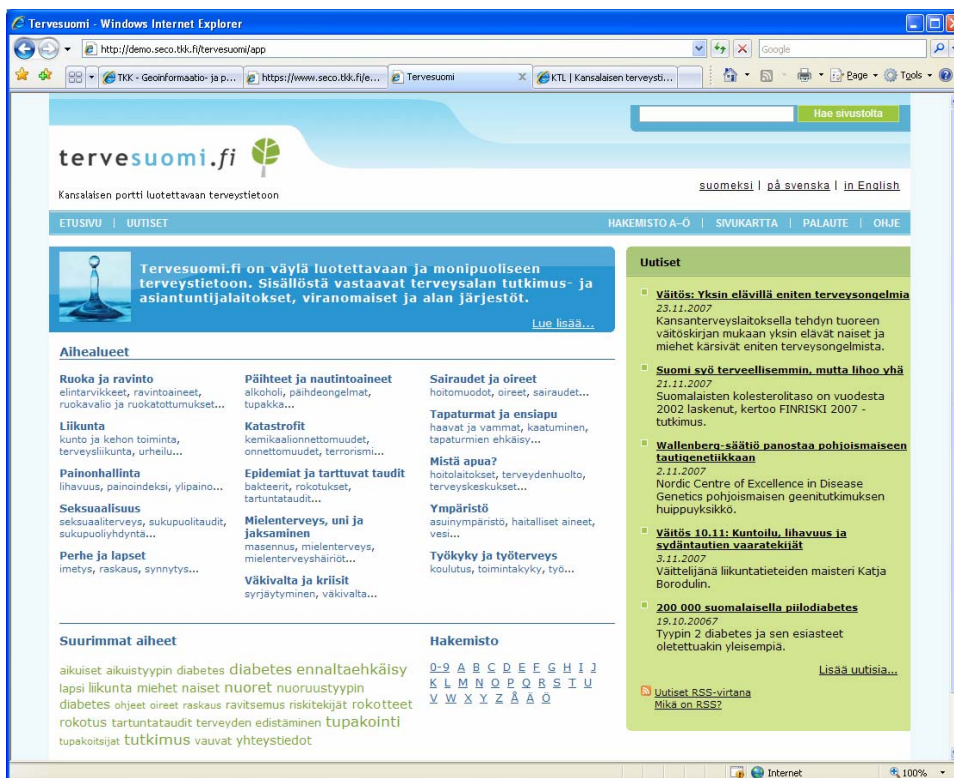
KulttuuriSampo – Suomalainen kulttuuri semanttisessa webissä

Semanttisessa webissä sisällöt ja verkkopalvelut voidaan koostaa eri toimijoiden tuottamista sisällöistä ja palveluista. Tämä on arvokasta sekä palvelun loppukäyttäjille että palvelun tuottajille. KulttuuriSampo-palvelussa loppukäyttäjänä ovat kulttuuritietoa hakevat kansalaiset. Näiden ongelmana on, että kulttuurisisällöt tuotetaan ja julkaistaan Suomessa satojen eri organisaatioiden sivustoilla ja tietokannoissa, jolloin kokonaiskuvan muodostaminen vaikkapa maassamme valmistetuista erilaisista rukiinlavoista, tietyn taiteilijan tuotannosta ja elämästä, tai Pariisin liittymisestä kultakauden taiteeseemme on vaikeaa.



Kansallinen sisällöntuotanto on organisaatiolähtöistä ja tapahtuu muodoissa, jotka ei suoraan vastaa tiedon hakijan tarpeita.

Tilanne on ongelmallinen myös kulttuuritiedon tuottajien kannalta. Sisältöjen julkaiseminen webissä voi olla taloudellisesti tai teknisesti haasteellista ja oma sisältömäärä niin pieni, ettei erillisen palvelun ylläpito ole järkevää. Suomessa on esimerkiksi n. 1000 eri museota ja lukemattomilla keräilijöillä ja muilla organisaatioilla on hallussaan sisältöjä, joilla olisi kysyntää, jos vain sopiva julkaisukanava olisi olemassa. KulttuuriSammon avulla organisaatioissa luotu sisältö, esimerkiksi museon kokoelmatiedot, voidaan (uusio)käyttää yhteisellä alustalla, jolloin kaikki julkaisijat voivat automaattisesti rikastaa ja täydentää omia sisältöjään toistensa avulla. Esimerkiksi KulttuuriSammon ”Semanttisessa Kalevalassa” Kullervosta kertova kohta voidaan yhdistää automaattisesti Valtion taidemuseossa oleviin Aksel Gallen-Kallelan maalauksiin, Kansallisbiografian ja USA:laisen Getty-säätiön tietoihin taiteilijasta ja hänen sosiaalisesta verkostostaan, Wikipedian artikkeleihin, Kansallismuseon esineisiin, Suomalaisen Kirjallisuuden Seuran keräämiin kansansäveliäin, Kalevalan muihin Kullervoa käsitteleviin kohtiin jne. Syntyy win-win tilanne, jossa eri julkaisijat saavat toisiltaan lisäarvoa ”ilmaiseksi” ja samalla loppukäyttäjät parempaa ”yhden luukun” palvelua.



Kuva 1. Semanttisen TerveSuomi.fi-palvelun etusivu.

TerveSuomi – Suomalainen terveystieto semanttisessa webissä

Mahdollisuus eri toimijoiden sisältöjen ja palveluiden yhdistelyyn, rikastamiseen ja uusiokäyttöön haastaa sisällöntuottajia samalla pohtimaan keskinäistä työnjakoa. FinnONTO-hankkeen toisessa pääsovelluksessa, Kansanterveyslaitoksen koordinoimassa kansallisessa TerveSuomi.fi-hankkeessa ideana on koostaa laadukas, eri terveysalan organisaatioiden tuottama terveystieto semanttisen webin avulla yhdeksi palveluksi hieman vastaavaan tapaan kuin KulttuuriSampo.fi on kulttuurialalle suunnattu semanttinen julkaisukanava. TerveSuomi.fi:n hankkeessa kävi pian ilmeiseksi, että maassamme tuotetaan runsaasti päällekkäistä terveystietoa eri organisaatioiden omista lähtökohdista ja velvoitteista johtuen. Esimerkiksi valtiovarainministeriön kansalaisille suunnatulla Suomi.fi-palvelussa on runsaasti linkkejä eri organisaatioiden tuottamaan terveystietoon, Yleisradion sivuilla on terveysivustonsa, samoin lääkäriseura

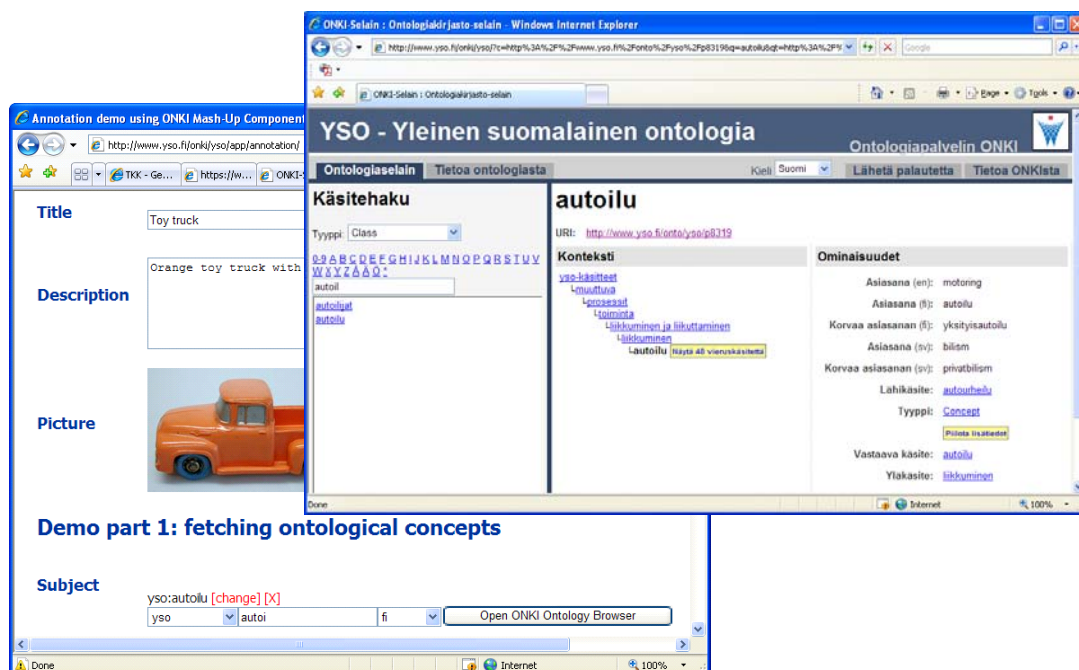


Duodecimilla, Työterveyslaitoksella, UKK Instituutilla, Stakesilla ja Terveystieteiden tutkimuskeskus Tekryn n. 120 jäsenjärjestöllä, kuten Diabetesliitolla, Sydänliitolla jne. FinnONTO:ssa kehitetyn konseptin ja semanttisen portaalin avulla työnjakoa eri organisaatioiden kesken voidaan tehostaa karsimalla päällekkäistä työtä, jolloin eri tahot voivat paremmin keskittyä tuottamaan tietoa omalta erityisalueeltaan. Samanaikaisesti kansalaisille voidaan tarjota laajempia tietosisältöjä aiempaa älykkäämpien haku- ja suosittelumeکانismien kautta, joihin voidaan ohjelmoida terveyden edistämiseen liittyvää asiantuntemustakin. Esimerkiksi ylipainosta kertoviin sisältöihin voidaan linkittää automaattisesti täsmätietoa ravitsemuksesta, kolesterolista, sydäntauteista ja liikunnasta, koska kone ymmärtää ylipainoon liittyviä riskejä ja tietää painonhallinnan keinoista. TerveSuomi.fi-portaali (kuva 1) on kansallinen terveyden edistämisen tietovaranto, jonka sisältöjä on tarkoitus hyödyntää muun muassa terveydenhuollossa asiakastyön tukena.

Kansallinen ontologiakirjasto ONKI

KulttuuriSampo.fi:n ja TerveSuomi.fi:n kaltaisten laaja-alaisten järjestelmien edellytyksenä ovat kansallisella tasolla sovitut, yhteentoimivat metadataskeemat ja ontologiat. FinnONTO-hankkeessa ryhdyttiinkin alusta alkaen kehittämään yhteistä sisältöinfrastruktuuria ja sen kustannustehokkaassa käytössä tarvittavia palveluita. Idea ”kansallisesta sisältöinfrastruktuurista” on analoginen nykyisen sähkö- tai GSM-verkon kanssa, joskin abstraktimpi: ajatuksena on, että kun sisältöä julkaistaan semanttiseen verkkoon, yhdistyisi se automaattisesti tarkoituksenmukaisella tavalla verkon rakenteisiin ja palveluihin.

FinnONTO:ssa valmistui prototyyppinä joukko toisiinsa kytkeytyviä kansallisia ontologioita. Näistä merkittävin on Kansalliskirjaston Yleiseen suomalaiseen asiasanastoon perustuva, n. 20 000 toisiinsa kytkeytyvän käsitteen Yleinen suomalainen ontologia YSO, sekä tätä tarkentavat sovellusala-kohtaiset ontologiat, kuten valokuvausalan VALO, maa- ja metsätalouden AFO, museoalan MAO ja Maanmittauslaitoksen paikannimirekisteriin perustuva, 800 000 paikannimeä sisältävä Suomalainen paikkaontologia SUO. Ontologiat liittyvät toisiinsa muodostaen kansallisten ontologioiden verkoston, mikä mahdollistaa eri alojen sisältöjen yhdistämisen ja hyödyntämisen semanttisessa webissä. Ontologiaverkoston kustannustehokkaaseen käyttöönottoon yrityksissä ja muissa organisaatioissa kehitettiin prototyyppi kansallisesta ONKI-ontologiapalvelusta, jonka ensimmäiset kaupalliset sovellukset ovat valmistumassa. ONKI:n avulla ontologiat saadaan vaivattomasti hyötykäyttöön Web 2.0 -mash-up-tekniikoilla hieman vastaavaan tapaan kuin esimerkiksi Googlen Maps -karttapalvelut (kuva 2).



Kuva 2. Yleinen suomalainen ontologia YSO ONKI-palveluna asiakassovelluksessa.



Historiallisen laaja tutkimushanke

FinnONTO-hanke oli tiettävästi Tekesin historian laajin yrityskonsortion koolla mitattuna, 37 eri rahoittajaa, ja se sai useita kansainvälisiä ja kotimaisia tunnustuksia, kuten alan tutkijayhteisön Semantic Web Challenge -palkinnon Japanissa ja pääministerin kunniamaininnan innovatiivisimmasta verkkosovelluksesta Laatu verkkoon -kilpailussa.

FinnONTO:ssa käynnistynyt semanttisen webin tutkimus- ja kehitystyö jatkuu FinnONTO:akin laajemmassa Semantic Web 2.0 -jatkohankkeessa (Tekes ja 38 muuta rahoittajaa). Sen tavoitteena on Web 2.0:n ja semanttisen webin teknologioiden yhdistäminen ja hyödyntäminen yhteisöllisissä verkkopalveluissa – ajatus joka oli idullaan jo FinnONTO:ssa.

FinnONTO (2003-2007) -konsortion jäsenet

Päävastuun tutkimuksesta kantoi Teknillisen korkeakoulun (TKK) viestintätekniikan laitoksen ja Helsingin yliopiston (HY) tietojenkäsittelytieteen laitoksen Semanttisen laskennan tutkimusryhmä (SeCo). Yhteistyössä olivat mukana Tampereen yliopiston informaatiotutkimuksen laitos, HY:n yleisen kielitieteen laitos ja TKK:n geoinformaatio- ja paikannustekniikan laboratorio. Hanketta rahoitti ja ohjasi sen loppuvaiheessa Tekes (FENIX-ohjelma) sekä 36 muun yrityksen ja julkisen organisaation muodostama konsortio: AAC Global Oy, Antikvaria-ryhmä ja Museoliitto, CelAmanzi Oy, Connexor Oy, eduskunta, Elisa Oyj, Espoon kaupunginmuseo, Geologian tutkimuskeskus GTK, Grip Studios Interactive Oy, Helsingin kaupunginkirjasto, Helsingin yliopiston Viikin tiedekirjasto, Kansalliskirjasto, Kansanterveyslaitos, Kustannus Oy Duodecim, Maa- ja metsätalousministeriö, Maanmittauslaitos, M-Cult ry, Museovirasto, Opetushallitus, Opintoluotsi-projekti/OPM, Osuuspankkikeskus OSK, Sanoma Data Oy, Sanastokeskus TSK ry, Sininen Meteoriiitti Oy, Stakes, Suomalaisen Kirjallisuuden Seura, Suomen maatalousmuseosäätiö, Suomen valokuvataiteen museo, Taideteollisen korkeakoulun kirjasto, Terveyden edistämisen keskus Tekry, TietoEnator Oyj, Valtion taidemuseo, valtiovarainministeriö, Veljekset Karhumäki Oy ja Yleisradio Oy.

Lisätietoja

FinnONTO-hanke <http://www.seco.tkk.fi/projects/finnonto/index.fi.php>
TerveSuomi.fi <http://www.seco.tkk.fi/applications/terveysuomi/> ja <http://www.terveysuomi.fi>
KulttuuriSampo.fi <http://www.kulttuurisampo.fi>
ONKI-ontologiakirjasto <http://www.yso.fi/>
W3C ja semanttinen web <http://www.w3.org/2001/sw/>

Yleisesitys FinnONTO-hankkeesta

Eero Hyvönen, Kim Viljanen, Eetu Mäkelä, Tomi Kauppinen, Tuukka Ruotsalo, Onni Valkeapää, Katri Seppälä, Osma Suominen, Olli Alm, Robin Lindroos, Teppo Käsälä, Riikka Henriksson, Matias Frosterus, Jouni Tuominen, Reetta Sinkkilä and Jussi Kurki: Elements of a National Semantic Web Infrastructure - Case Study Finland on the Semantic Web (Invited paper). Proceedings of the First International Semantic Computing Conference (IEEE ICSC 2007), Irvine, California, September, 2007. IEEE Press.

<http://www.seco.tkk.fi/publications/2007/hyvonen-et-al-elements-2007.pdf>