

14.11.2005



Kohti suomalaista semanttista webiä

**Suomalaiset semanttisen webin ontologiat
(FinnONTO)-hankkeen esittely**

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

Eero Hyvönen

1 Tausta: webin seuraava semanttinen sukupolvi

Internet-kupla särkyi vuosituhaten vaihteessa siihen, ettei tietoverkkoihin pystytty tuottamaan sellaisia sisältöjä ja palveluita, jotka olisivat olleet aidosti hyödyllisiä käyttäjilleen kustannuksiinsa nähden. Yksi selitys web-palveluiden toteuttamisen vaikeudelle on se, että webin sisällöt on esitetty muodossa, joka on tarkoitettu vain ihmisen eikä koneen tulkittavaksi. Kun kone ei ymmärrä niitä sisältöjä, joita se käsittelee, on hyödyllisiä älykkäitä palveluita vaikeaa ja kallista toteuttaa.

Ongelma tiedostettiin WWW:n kansainvälistä kehitystä koordinoivassa W3C-järjestössä jo 1990-luvun lopulla. Ongelman ratkaisumalliksi on muodostunut ns. semanttinen web (Berners-Lee, 2000), jonka aktiivinen kehitystyö käynnistyi keväällä 2001 W3C:n toimesta erillishjelmana (activity). Suomessa alan lähtölaukaus ammuttiin puoli vuotta myöhemmin syksyllä 2001 Semantic Web Kick-off in Finland -seminaarina (Hyvönen, 2002), jossa kuultiin mm. W3C-konsortion edustajan esitys.

Semanttisen webin voi mieltää nykyistä webiä täydentävänä sisällöllisenä metatietokerroksena, jonka avulla voidaan kertoa Internetin koneille webin sisältöjen merkitys. Näin voidaan toisaalta ohjelmoida entistä älykkämpiä web-palveluita, toisaalta saada erilaiset tietosisällöt ja -järjestelmät keskenään yhteentoimiviksi. Semanttinen web perustuu täsmällisesti määriteltyihin sanastoihin ja käsitteistöihin, ns. ontologioihin, joiden avulla metatieto esitetään. Ilman yhteisesti sovittuja sanastoja eivät ihmiset ja koneet voi ymmärtää toisiaan. Semanttisen webin tunnetuin puolestapuhuja on W3C:n johtaja ja webin ”isä” Tim Berners-Lee, jolle myönnettiin kesällä 2004 suomalainen miljoonan euron Millennium-palkinto.

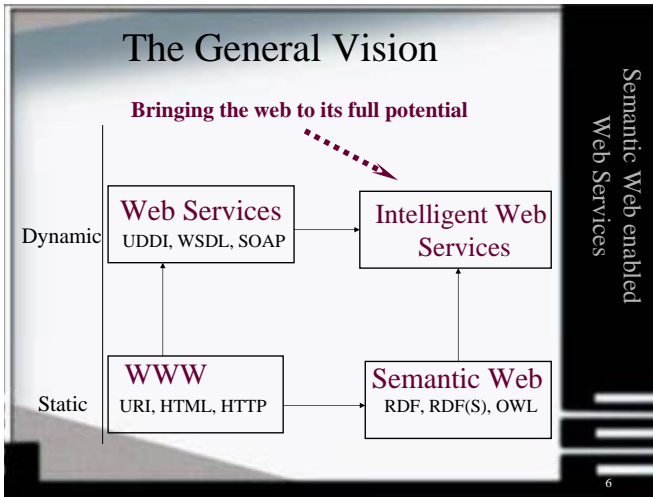
Semanttinen web (Davies et al., 2003) ei voi toteutua ilman voimakasta panostusta tutkimukseen¹. Esimerkiksi sisältöjen semanttiseen esittämiseen, tietojen ontologiseen yhdistämiseen, web-palveluiden hajauttamiseen, semanttiseen tiedonhakuun ja linkitykseen, kollaboratiivisiin semanttisiin järjestelmiin, automaattiseen annotointiin ja luokitteluun ja semanttisen webin työvälineiden ja julkaisualustojen tutkimukseen liittyy suuria tutkimuksellisia ja teknisiä haasteita. Suomalaisen semanttisen webin kehityksessä erityisen ongelman muodostavat vielä suomen kieleen liittyvät erityiskysymykset, sillä semantiikka on vahvasti kieleen sidottua. Suomalaisen semanttisen webin sisällöllisen infrastruktuurin kehittäminen tarvitsee tuekseen tutkimusta ja julkista tukea vastaavaan tapaan kuin rautatie- tai tietoliikenneverkoston rakentaminen aiemmin. Hyödyllisen infrastruktuurin varaan voidaan jatkossa toteuttaa kustannustehokkaasti sekä julkisia web-palveluita että synnyttää uutta liiketoimintaa.

Semanttisen webin ohella toinen webiä mullistava kehitysalue on konetulkittavat web-palvelut, ”Web Services”. Siinä on kysymys toiminnallisten palveluiden tuottamisesta ja julkaisemisesta webissä muodossa, jota toiset palvelut voivat käyttää. Webistä on tämän vision ja megatrendin mukaisesti muodostumassa ikään kuin valtava hajautettu tietokone, jossa palveluita koostetaan kustannustehokkaasti toisaalla julkaistujen modulaaristen palveluiden avulla. Web-palveluiden idea on suurten IT-alan yritysten (IBM, Microsoft, SUN ym.) käynnistämä kehityssuunta, joka on jo monelta osin alan yritysten arkipäivää (esim. Microsoftin .NET) tai standardisoituvaa teknologiaa (esim. web-palveluiden standardit SOAP, WSDL, UDDI ym.). Web-palveluiden taso on kuitenkin semanttiselta kannalta tarkasteltuja jäänyt vaatimattomaksi. Web-alan visiona onkin viime aikoina ollut yhdistää semanttisen webin sisältökuvaukset web-palveluihin. Tällöin syntyy tutkimus- ja kehitysalue, josta käytetään nimitystä ”Intelligent Web Services” tai ”Semantic Web Services” (kuva 1.1).

Semanttisen webin ja semanttisten web-palveluiden tutkimus muodostaa uuden tutkimusalueen, jolla on jo omat kansainväliset konferenssisarjansa, kuten International Semantic Web Conference ISWC, European Semantic Web Conference ESWC ja vuodesta 2006 alkaen myös Asian Semantic Web Conference (ASWC).

¹ Alan tutkimus- ja kehitystyötä esitellään esimerkiksi portaalissa <http://semanticweb.org> ja W3C:n Semantic Web Activity -ohjelman sivustolla <http://www.w3.org/2001/sw/>.

Perinteisessä W3C:n World Wide Web Conference -sarjassa (WWW) semanttinen web on nykyisin erittäin suosittu tutkimusaihe. Semanttisen webin alalle on syntynyt myös uusia tieteellisiä joulnealeja, kuten Journal of Web Semantics, International Journal of Metadata, Semantics and Ontologies, Applied Ontology sekä International Journal on Semantic Web and Information Systems.



Kuva 1.1. Webin megatendit prof. Dieter Fenselin kaavion mukaan vuodelta 2002: Semantic Web ja Web Services yhdistyvät Intelligent Web Services -alueeksi.

Webin syntyminen lienee merkittävön yhteiskuntakehitystä kohdannut tekniikkavetoinen mullistus sitten tietokoneen keksimisen. Kehitys jatkuu nopeana ja toimii samalla monen sovellusalueen kehityksen veturina. Alan tutkimus ja kehitys Suomessa on jäänyt aiheen merkityksen suhteen hämmästyttävän vähäiseksi.

2 Visio ja kehitystavoitteet

Kansallinen Suomalaiset semanttisen webin ontologiat (FinnONTO) -hanke² käynnistyi syksyllä 2003 ja jatkuu suunnitelman mukaan v. 2007 puoliväliin saakka osana Tekesin Fenix-ohjelmaa. **Hankkeen tavoitteena on käynnistää Suomessa kansallisen semanttisen webin julkisen sisältöinfrastruktuurin rakentaminen.** Visiona on, että suomalainen semanttinen web tarvitsee sisältöinfrastruktuurin hieman vastaavaan tapaan kuin liikenne tarvitsee julkisen tieverkoston, kuin energiahuolto sähköverkon ja standardit tai kuin tietoliikenne tarvitsee GSM- tai laajakaistaverkon.

Kehittävän suomalaisen semanttisen webin infrastruktuurin keskeisiä komponentteja ovat

- yleiset kansalliset *ontologiat*,
- ontologioiden kehittämiseen ja hyödyntämiseen liittyvät *ontologiapalvelut* sekä
- tiedonesitys- ja muut web-standardit sekä työvälineet.

Hankkeessa syntyviä ontologioita ja palveluita testataan ja pilotoidaan käytännön web-sovelluksissa, joista useimmat käyttävät hyväksi ja yhdistelevät olemassa olevien web-portaalien tietosisältöjä. Tällä hetkellä työn alla olevia pilottisovelluksia ja niihin liittyviä portaalieja ovat

² <http://www.seco.tkk.fi/projects/finnonto/>

- kulttuurialan MuseoSuomi-järjestelmä³ ja siitä kehitettävä KulttuuriSampo-portaali,
- sosiaali- ja terveysministeriön rahoittama kansalaisen terveydenedistämisen portaalihanke,
- valtiovarainministeriön suomi.fi,
- kirjastoalan tiedonportti.fi,
- Stakesin sosiaaliportti.fi ja
- opetusalan opintoluotsi.fi.

Semanttinen web ja siksi myös sen sisältöinfrastruktuuri perustuu ontologioihin (Fensel, 2003; Staab, Studer, 2003). Ilman yhteisiä ontologioita aitoa semanttisuutta ei webiin saada, semanttinen web pirstaloituu yhteentoimimattomiksi saarekkeiksi, sisältötyötä tehdään turhaan moneen kertaan ja web-sovellusten kehittäminen tulee vaikeammaksi ja kalliimmaksi. Kansallisena ongelmana on, että suomalaisia ”koneymmärrettäviä” ontologioita ei juurikaan ole käytettävissä, ainoastaan ihmiskäyttäjälle tarkoitettuja asiasanatesaureja (Foskett, 1980; Aitchison et al., 2000). Nykyisin käytössä olevat tesaurukset ovat arvokas semanttinen voimavara ja käytettävissä on paljon niiden avulla jo indeksoituja sisältöjä. Tesaurusten käyttö semanttisen webin sovelluksissa sellaisenaan ei valitettavasti yleensä kuitenkaan riitä, vaan ne on ontologisoitava ensin koneymmärrettävämpään muotoon. FinnONTO-hankkeen yksi perusteeksi onkin, että *maassamme pitäisi siirtyä kehittämään ontologioita nykyisten tesaurusten sijasta* (Hyvönen, 2005).

Ontologiatyöhön pitäisi ryhtyä laajalla kansallisella rintamalla, sillä semanttinen web mahdollistaa eri alojen ja organisaatioiden tietovarantojen sisällöllisen yhdistämisen. Tämä antaisi usein lisäarvoa loppukäyttäjille ja säästäisi sisällön tuottajien resursseja. Esimerkiksi valtionhallinnon yhdessä portaalissa tuotettu sisältö voitaisiin käyttää uudelleen muissa portaaaleissa ja sovelluksissa, jolloin tarpeeton päällekkäinen sisältöyminen minimoituisi. XML:n käytön keskeisenä sovellusideana on ollut saman sisällön monikanavainen vaivaton julkaiseminen erilaisina syntaktisina ilmiöinä kuten painatteina, web-sivuna jne. Semanttinen web yleistää ja syventää tämän ajatuksen eräänlaiseen ”monisovellusjulkaisemiseen”: kerran tuotettu semanttinen sisältö voidaan käyttää sellaisenaan hyväksi erilaisten älykkäiden sovellusten perustana.

FinnONTO-vision mukaan suomalaisia ydinontologioita pitäisi ylläpitää yleensä julkisin varoin ja open source -hengessä, jotta yhteisten ontologioiden ja standardien käytölle saataisiin mahdollisimman laaja käyttöaste ja sitoutuminen. Hyvä ja laajasti käytetty infrastruktuuri tarjoaa parhaat liiketoimintamahdollisuudet tuottaa lisäarvoa esimerkiksi sovelluskohtaisina ontologioina ja semanttisen webin sovelluksina.

FinnONTO-projektissa ratkotaan ontologioiden avulla informaatiotieteissä ja tietotekniikassa käytettyihin sisällönkuvailun menetelmiin liittyviä useita ongelmia:

- **Yhteentoimivuuden ongelma.** Eri organisaatioiden ja henkilöiden käyttämät terminologiat eivät ole yhteensopivia. Esimerkiksi eri tieteenalat käyttävät toisistaan poikkeavia termejä samastakin asiasta, yksittäisillä organisaatioilla on omia sisäisiä sanastoja, eri kielten käännöstermien välillä on semanttisia eroja jne. Semanttinen yhteensopimattomuus vie pohjan täsmälliseltä tiedonhaulta ja estää järjestelmien yhteiskäytön tietoverkoissa.
- **Semanttinen köyhyys.** Nykyiset asiasanatesaurukset ja sisällönkuvailun formaatit sisältävät vain hyvin primitiivisiä semanttisia suhteita, kuten ”laajempi termi”, ”suppeampi termi” ja ”rinnakkaistermi”. Tämä estää käsitteistöjen käytön syvällisempää analyysiä edellyttävissä, merkityspersustaisissa sovelluksissa.

³ Portaali on käytettävissä osoitteessa <http://www.museosuomi.fi>. Järjestelmä sai syksyllä 2004 kansainvälisen Semantic Web Challenge Award –teknologiapalkinnon ja pääministerin kunniamaininnan innovatiivisimmasta sovelluksesta Tietoyhteiskuntaohjelman Laatusuunnitelman –kilpailussa.

- **Kollaboratiivinen sisällönhallinta.** Mikään taho ei yksin pysty eikä ole halukaskaan ylläpitämään kaikkien eri alojen sanastoja tai ontologioita. Esimerkiksi geologian tai taiteen käsitteistön hallinta ei voi kuulua yleiskäsitteistöä tällä hetkellä ylläpitävälle Kansalliskirjastolle, eikä Geologian tutkimuskeskuskella tai Valtion taidemuseolla ole mahdollisuuksia laajojen yleiskielen sanastojen ylläpitoon, mikä taas kuuluu luontevasti kirjastojen toimenkuvaan. Sanastotyö on kyettävä hajauttamaan eri intressiryhmille ja yhdistämään tulokset WWW:n kautta sovelluksissa, joissa usein tarvitaan samanaikaisesti eri ontologioita.
- **Muutosten hallinta.** Käsitteistö ja sen kuvaamisessa käytetyt ontologia muuttuvat ajan kuluessa. Esimerkiksi Tšekkoslovakiaa tai sotaa edeltävän Suomen Karjalan kuntia ei enää ole olemassa nykyisessä valtio-ontologiassa, mutta paljon aineistoa on indeksoitu Tšekkoslovakian ja luovutetun Karjalan eri paikkojen nimillä. Jotta eri aikoina erilaisilla ontologioilla kuvattuja aineistoja voitaisiin myöhemmin hakea, tarvitaan menetelmiä muutosten ja sanastojen versioiden hallintaan.
- **Kuvausten tuottamisen ongelmat.** Asiasanoitusten tuottaminen sisällöille eli *annotointi* on vaikeaa, aiheuttaa kustannuksia ja sen laatu vaihtelee suuresti indeksoijasta riippuen. Kun sisältöjä tuottaa ja niitä indeksoi jatkossa yhä laajempi ja siksi ei-ammattimaisempi webin käyttäjäkunta, tarvitaan yhä parempia välineitä annotoinnin tukemiseksi ja automatisoimiseksi. Sisältöjen vaivaton kuvailu on semanttisen webin menestyksen avainkysymyksiä.
- **Monikielisyys.** Suomalaiset WWW-sisällöt ovat paljolti suomenkielisiä, mutta WWW on leimallisesti monikielinen ja kansainvälinen julkaisu- ja palveluntarjonnan mekanismi. Käsitteet ovat lähtökohtaisesti kieliriippumattomia abstraktioita. Ne mahdollistavat erikielisten termien väliset kuvaukset ja tätä kautta monikieliset WWW-palvelut, laajan saavutettavuuden ja käyttömukavuuden.
- **Käytettävyys.** Semanttiset kuvaukset antavat uusia sovellusmahdollisuuksia webin käyttöliittymien kehittämiseen, esimerkiksi semanttiseen hakuun, samoiluun, sisältöjen semanttiseen visualisointiin ja multimodaaliliittymiin, jossa yhdistyy puheen/äänän ja näyttöruudun tekstin/kuvan/videon sekä näppäinten/hiiren käyttö.
- **Vapaa avoin sisältö.** Nykyiset sanastot eivät ole yleensä vapaasti saatavilla tutkimus- tai sovelluskäyttöön. Keskeisimmät kansalliset käsitteistöt, kuten Yleinen suomalainen asiasanasto YSA, pitäisi olla esitetty avointen standardien avulla ja olla vapaasti kaikkien saatavilla. Koneen ymmärtämässä avoimessa muodossa saatavaa aineistoa voitaisiin helposti hyödyntää erityyppisissä sovelluksissa.

Hankkeen teknologiapohjan muodostavat semantic web -standardit, suositukset ja työkalut kuten RDF(S), OWL, Jena, SWI-Prolog jne. Ontologioiden osalta pyritään hyödyntämään mahdollisimman paljon jo olemassa olevia sanastoja ja ontologioita.

3 Osatehtävät

Projekti jakautuu moniin osatehtäviin, joista keskeisimpiä esitellään lyhyesti alla.

3.1 Ontologioiden kehitystyö

Hankkeessa kehitetään menetelmiä ja malleja ontologisen tietämyksen esittämiseen kohdealueilla sekä erikielisten terminologioiden yhdistämiseen. Samalla seurataan kansainvälisen standardointityön etenemistä ja hyödynnetään sitä.

Työn tuloksena syntyy

- **Yleinen suomalainen ontologia YSO.** YSO perustuu Kansalliskirjaston ylläpitämän Yleisen suomalaisen asiasanaston (YSA) käsitteisiin semanttisesti analysoituna ja rikastettuna. Huomioon otetaan myös yle-

sisä kirjastoissa käytettävät luokitusjärjestelmät: Yleisten kirjastojen luokitusjärjestelmä (YKL) ja Helsingin kaupunginkirjaston luokitusjärjestelmä (HKLJ). Ontologiat toteutetaan avoimilla tietokoneen ymmärtämällä semantic web -teknologioilla.

- **Joukko YSO:aan liittyviä erityisontologioita.** Koska sovelluksissa joudutaan yleensä yhdistelemään yleis- ja erityisontologioita, hankkeessa kehitetään myös joukko YSO:aan liittyviä, sitä tarkentavia vertikaalisia erityisontologioita.

Tärkeimpiä hankkeessa kehitteillä olevia ontologioita on esitetty taulukossa 3.1. Keskeisin ja laajin osahanke on n. 20 000 käsitteen YSO. Sen sanastoa käytetään laajasti Suomessa eri sovelluksissa.

Ontologia	Kuvaus
YSO	Yleinen suomalainen ontologia. Kansalliskirjaston Yleisen suomalaisen asiasanaston (YSA) ontologisoitu versio.
MAO	Museoviraston Museoalan asiasanastosta kehitetty Museoalan ontologia. Ensimmäinen versio käytössä MuseoSuomessa.
PAIKKO	Suomalainen paikkaontologia.
SAPO	Suomen ajallinen paikkaontologia eri aikakausina. Ensimmäisen version tavoitteena on kuvata Suomen kunnat ja näiden muutokset 1900-alusta lähtien.
ICONCLASS	Kansainvälisesti käytetty, Hollannissa kehitetty taiteen kuvailusysteemi, jota Valtion taidemuseo on suomentanut taidemuseoiden käyttöön. Järjestelmä muunnetaan ontologiaksi ja yhdistetään YSO:aan.
VALO	Valokuvataiteen museon kolmesta asiasanastosta muodostettu ontologinen kokonaisuus.
TOIMO	Henkilöiden, yritysten, organisaatioiden yms. toimijoiden ontologia. Ensimmäinen versio käytössä MuseoSuomessa.
TAO	Tapahtumien ja prosessien ontologia. Perustuu YSO:aan.
MATO	Maatalousalan ontologia. Perustuu osittain Viikin tiedekirjaston Argiforest sanastoon.

Taulukko 3.1. Hankkeessa kehitettäviä sisällön kuvailussa käytettäviä ontologioita.

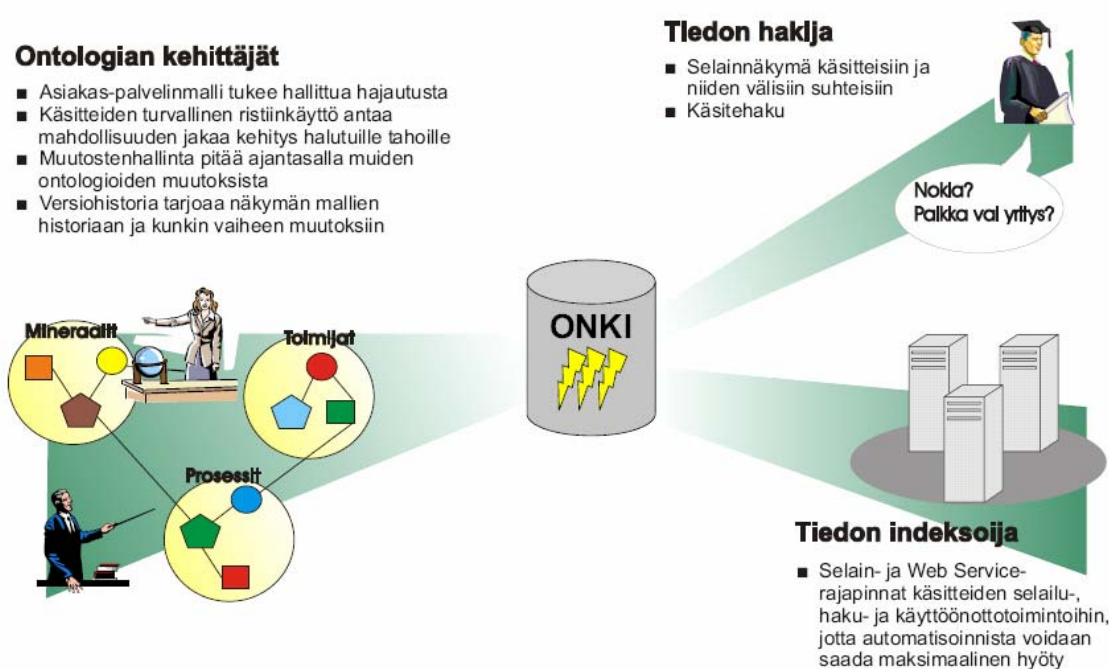
3.2 Kansallinen ontologiakirjastopalvelin ONKI

Ontologioiden julkaisemista, hyödyntämistä, ylläpitoa ja jatkokehitystä varten (Ding, Fensel, 2002) kehitetään ontologiakirjastopalvelin nimeltä ONKI (Komulainen et al., 2005). ONKI tarjoaa palveluita kolmelle käyttäjäryhmälle (kuva 3.1):

1. **Ontologioiden kehittäjät.** Ontologioiden kehittäjäyhteisöille ONKI tarjoaa ympäristön, joka tukee eri kehittäjäryhmien yhteistyötä, ontologioiden versiointia, päivittämistä ja julkaisemista (Valo et al., 2005).
2. **Tietosisältöjen tuottajat.** ONKI:in liitettävä web-palvelurajapinta mahdollistaa aineistojen kuvailuissa tarvittavien URI-viittausten siirtämisen toisiin sovelluksiin. Sisällön tuottaja, esimerkiksi kirjaston tai museon luetteluoja, löytää kuvailuun tarvittavat käsitteet nopeasti ONKI-selaimen avulla ONKI-palvelimelta ja voi nappia painamalla siirtää tarvittavat tiedot omaan sovellukseensa. ONKI-järjestelmän

palvelurajapintaan kytkeytymällä ei luettelointi- yms. ohjelmistoissa tarvitse erikseen toteuttaa ontologioiden selailuun ja käsitteiden hakuun liittyviä palveluita, vaan voidaan hyödyntää webissä jo olevaa julkista palvelua. Käytettävissä on aina ajantasainen, standardi käsitteistö. Jaettujen ontologioiden vaivaton saatavuus ja käyttömahdollisuus on ensiarvoisen tärkeää tiedon indeksoijille. Palvelimeen lisätään myös tietämystekninen sisällöntuottajan apuri, joka ontologisen tietämyksensä perustella osaa auttaa käyttäjäänsä sopivien ja kattavien kuvailujen luomiseksi. Tavoitteena on sisällönkuvailujen laadun parantaminen ja yhdenmukaistaminen eri sovelluksissa, mikä on koettu tärkeäksi käytännölliseksi kehityskohteeksi.

3. **Hakupalveluiden käyttäjät.** ONKI-selaimen kolmas käyttäjäryhmä on tiedonhakupalveluiden loppukäyttäjät. ONKI-selaimen hakupalveluiden avulla voidaan helposti löytää tietokantahauissa tarvittavien avainsanojen eri merkitykset ja näitä vastaavat käsitteet. Myös tätä palvelua voidaan käyttää avoimen rajapinnan kautta toisista sovelluksista.



Kuva 3.1. ONKI-ontologiakirjaston palveluita kolmelle käyttäjäryhmälle.

Projektin tuloksena ei ole vain kertaluonteinen ONKI-järjestelmän toteutus ja joukko ontologiaprototyyppiä, vaan tavoitteena on kehittää maamme eri asiantuntijaorganisaatioille hajautettu, jatkuva WWW-perustainen prosessi tukemaan ontologioiden kehitystyötä. Tällaisen mekanismin synnyttäminen on tarpeen, jotta suomalaisen semanttisen webin keskeisten ontologioiden kehitystyötä voidaan projektin päätyttyäkin jatkaa eteenpäin. Ajatuksena on, että hankkeeseen osallistuvat tahot, kuten Kansalliskirjasto, Museo- ja muut toimijat voisivat projektin kehittämän välineistön avulla ryhtyä ylläpitämään erillisten asiantuntijajensa sijasta paremmin hyödynnettävissä olevia yhteisiä ontologioita.

3.3 Sisältöjen semanttinen yhdistäminen: KulttuuriSampo

Yksi keskeinen tutkimuskohde hankkeessa on selvittää, miten yhdistää webin merkitykseltään erityyppisiä sisältöjä mielekkäiksi kokonaisuuksiksi ja sovelluksiksi. Työ on tältä osin loogista jatkoa MuseoSuomi-hankkeelle (Hyvönen et al., 2005). Siinä tietokantojen yhteentoimivuusongelmalle (interoperability) kehitettiin ratkaisu tilanteessa, jossa sisällöt ovat heterogeenisten tietokantojen tietueita, mutta sisällöltään saman-

tyyppisiä museoesineitä. Nyt tutkimus- ja kehityskohteenä on tätä haastavampia sovellukset, joissa 1) yhdistettävät aineistot ovat sisällöltään erityyppisiä ja 2) joiden kuvailussa on käytetty erilaisia ontologioita.

Tutkimustyön kohteena olevia sisältötyyppejä, sisältötuottajia ja sisällön kuvailussa tarvittavia ontologioita ja aineistoja on lueteltu taulukossa 3.2.

Sisältötyyppi	Sisällöntuottaja	Ontologioita	Aineistoja
Museoesine	Kansallismuseo, Antikvaria-ryhmä, Suomen maatalousmuseo Sarka	MAO, SAPO, TOIMO	MuseoSuomi- ja Suomen museot online, Museoiden tietokannat
Luontokohde	Geologian tutkimuskeskus GTK, Museovirasto	MAO, SAPO	GTK:n kuvatietokanta, Museoviraston muinaismuistorekisteri
Rakennettu kohde	Museovirasto, Suomen maatalousmuseo Sarka ym.	MAO, SAPO, TOIMO	Museoviraston muinaismuistorekisteri, kirkkorakennukset, maatilat
Taideteos	Valtion taidemuseo	YSO, ICONCLASS	Valtion taidemuseon kokoelmat
Valokuva	Valokuvataiteen museo, GTK	YSO, VALO	Valokuvataiteen museon kokoelmat, Suomen museot online, GTK:n kuvatietokannat
Elokuva	YLE	YSO	Yleisradion opetuohjelmat ja elokuva-arkistot
Arkistodokumentti	SKS, Kansalliskirjasto	YSO	Suomalaisen kirjallisuuden seuran kokoelmat
Web-sivu/palvelu	Käyttäjäjyhteisöt	YSO	Web
Musiikki, audionäyte	YLE, SKS	YSO	YLE:n arkistot, SKS:n äänitteet
Prosessi	Suomen maatalousmuseo Sarka, HY	TAO	Kirjallisuus ja museoiden kokoelmat
Taito	Espoon kaupunginmuseo	TAO	Mm. vanhat elokuvat
Tarina	TKK, HY	TAO	Kalevala ym.
Henkilö	SKS, TKK; HY	TOIMO	Kansallisbiografia

Taulukko 3.2. KulttuuriSuomi-hankkeessa tutkittavia aineistotyyppisiä, sisältötuottajia, tarvittavia ontologioita ja demo-järjestelmässä käytettäviä aineistoja. (Alustava suunnitelma, muutokset mahdollisia.)

Erityyppisten tietosisältöjen ja ontologisten kuvausten yhdistämistä tutkitaan ja pilotoidaan KulttuuriSampo-järjestelmässä, joka on MuseoSuomi-portaalin seuraava kehitysaste. Sovellusalueena on kulttuuri- ja opetusala.. Kehittävät ratkaisut ovat kuitenkin yleisiä ja soveltuvat myös muille sovellusalueille kuten yritysten tietämyksenhallinnan sovelluksiin (knowledge management), media-alan sisällönhallintaan jne.

3.4 Ontologinen tiedonhaku

Hankkeessa kehitetään semanttista, suomenkielen morfologian ja ontologisia suhteita huomioon ottavaa indeksointi- ja hakuteknologiaa. Uutena tutkimuskohteena 2005-2007 on kehittää menetelmä, jonka avulla haku voidaan kohdistaa sisältöjen välisiin implisiittisiin suhteisiin eikä vain tieto-objekteihin. Tällaisessa uudentyyppisessä ”koostehaussa” voidaan esimerkiksi kysyä, miten kaksi historian henkilöä liittyvät toisiinsa ja vastauksena on erilaiset ko. henkilöitä yhdistävät semanttiset yhteydet.

3.5 Puoliautomaattinen sisällönkuvailu ja luokittelu

Tärkeä uusi pilotointikohde hankkeessa on (puoli)automaattinen sisällönkuvailu (annotointi). Automaattisessa annoitoinnissa järjestelmä saa syötteenään tekstimuodossa olevan tiedon ja päättelee sen perusteella annetun ontologian mukaisen sisällönkuvailun. Automaattisen sisällönkuvailijan keskeinen komponentti on semanttinen disambiguoija, joka kykenee erottamaan sanojen merkityksiä niiden kontekstin perusteella.

3.6 Ontologiaperustainen kysymys-vastauspalvelu

Osahankkeessa tutkitaan, miten ontologisesti annoitettua kysymys-vastaus-tietoa voitaisiin käyttää hyväksi ongelmanratkonnassa help-desk-tyyppisissä sovelluksissa. Tutkimusaineistona on kirjastot.fi-portaalin n. 20 000 kysymys-vastausparia, jotka on jo kerätty ja asiasanoitettu ”Kysy kirjastonhoitajalta” -palvelun kautta.

3.7 Ontologiaperustainen metahakukone

Osahankkeessa tutkitaan ja kehitetään ontologiaperustaista metahakuteknologiaa. Ontologinen metahakukone on hakupalvelu, jolla loppukäyttäjän kysely voidaan muuttaa automaattisesti taustalla olevien tietokantahakujen tai toisten hakukoneiden kyselyiksi. Osahankkeen uutena ideana on käyttää tässä hyväksi päättelyä, joka perustuu toisaalta YSO/ONKI-tyyppiseen laajaan ontologiseen tietämuskantaan ja web-palveluun, toisaalta tietokannoista ja niiden indeksoinnista saatavaan metadataan (esim. tietoon siitä, millä ontologialla ja millä kielellä tietokannan aineisto on annotoitu).

3.8 Semanttisten portaalien yhdistäminen

Suomeen on syntynyt ja on parhaillaan kehitteillä lukuisia portaaleja, joiden aineistot ovat osittain päällekkäisiä. Esimerkiksi Sosiaali- ja terveysministeriöllä on vireillä hanke Kansalaisen terveydenedistämisen portaaliksi, joka on osin päällekkäinen valtiovarainministeriön kehittämän suomi.fi-portaalin kanssa terveyden edistämiseen liittyvän aineiston osalta, ja toisaalta kirjastojen tietokannoissa on paljon terveysalan tietoa ja artikkeleita. Osahankkeessa tutkitaan, miten eri portaaleissa tehtävä tarpeeton päällekkäinen tiedontuottamistyö voitaisiin estää. Lähtökohtana on, että kertaalleen tuotettu semanttinen aineisto voitaisiin käyttää uudelleen toisissa portaaleissa ja hakupalveluissa, mikä johtaisi merkittäviin taloudellisiin säästöihin. Ratkaisuideana on kehittää muunnosmenetelmiä ontologioiden välille niin, että tietyllä ontologialla kuvailtu aineisto voidaan projisoida toiselle ontologialle, jolloin haku voidaan ulottaa yli ontologiarajojen

3.9 Monikielinen vuorovaikutteinen ontologinen tiedonhaku

SeCo-ryhmän kehittämän OntoViews-työkalun (Mäkelä et al., 2004) monikielisyystukea kehitetään edelleen. Ontologioiden osalta monikielisyyttä pilotoidaan mm. YSO:lla, jonka termeille on saatavilla käännökset ruotsin kielelle.

3.10 Sosiaali- ja terveysalan ontologiat ja palvelut

Osahankkeen tavoitteena on tutkia ontologiapohjaisten tiedonesitystekniikoiden soveltuvuutta Stakesin edustamalle sosiaali- ja terveysalueelle ja kehittää uusia tapoja hallita älykkäämmin ja joustavammin alan palvelutietoja, potilastietoja yms. tietosisältöjä. Tavoitteena on erityisesti kehittää Stakesin tarjoamia, valtakunnallisesti käytössä olevia perinteisiä tiedonhallinnan infrastruktuureja (esim. koodistopalvelin) kohti yhä joustavampia ja monipuolisempia palveluita tarjoavia järjestelmiä. Tällainen infrastruktuuri mahdollistaa osaltaan maamme ja tulevaisuudessa koko EU:n sosiaali- ja terveysalan tietojen yhteiskäyttöä.

3.11 Kansalaisen terveyden edistämisen portaali

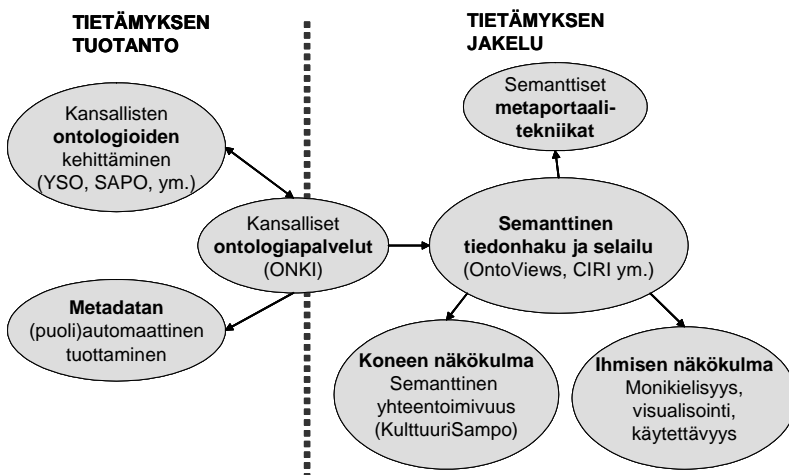
Kehittäviä teknologioita pilotoidaan Sosiaali- ja terveysministeriön rahoittamassa, Kansanterveyslaitoksen vetämässä Kansalaisen terveyden edistämisen portaali -hankkeessa. FinnONTO-projektin osuus liittyy portaalin suunnitteluvaiheeseen, jossa selvitetään ja demonstroidaan semanttisen webin teknologioiden lisäarvo Kansalaisen terveyden edistämisen portaalin kaltaisessa hankkeessa, jolle on luonteenomaista seuraavat seikat:

- Eri organisaatiot tuottavat hajautetusti tietosisältöjä, jotka on tallennettu keskenään yhteensopimattomiin tietokantoihin, mutta joiden tietoja pitäisi voida yhdistellä asiakkaan tietotarpeen tyydyttämiseksi.
- Julkaistavat tietosisällöt, tässä terveyden ylläpitoon eri tavoin liittyvät tiedot ja osaaminen, ovat monimuotoista ja liittyvät toisiinsa monin eri tavoin. Eri tietojen väliset yhteydet ovat keskeinen osa portaalin antia asiakkaalle.

Työhypoteesina on, että semanttisen webin teknologioiden avulla eri organisaatioiden erityyppiset tietosisällöt voitaisiin yhdistää sekä teknisesti että sisällöltään paremmin kuin perinteisillä tekniikoilla ja tarjota loppukäyttäjillä aiempaa olennaisesti parempia verkkopalveluita.

3.12 Oppimateriaalien semanttinen haku ja yhdistäminen

FinnONTO-hankkeen tämän osion taustalla on opetushallituksen ja Yleisradion yhteistyöryhmästä lähtenyt ajatus verkko-oppimateriaalin hakupalvelun kehittämiseksi semanttisen webin teknologioiden avulla. Taustalla on myös Digitaaliset tietovarannot -työryhmän työ minimimetatiedon määrittämiseksi ja verkkomateriaalien saatavuuden ja käytön edistämiseksi. Työryhmässä oli vuosina 2002-2003 opetusalan, Yleisradion, kirjasto- ja museotoimen, virtuaaliyliopisto- ja virtuaaliammattikorkeakouluhankkeiden ja kaupallisten kustantajien edustus.



Kuva 3.1. Projektin tutkimukselliset teemat web-sovellusten kannalta.

3.13 Tutkimusteemat

Kuva 3.1 hahmottaa yleisemmin projektin keskeisiä tutkimusteemoja ja näiden välisiä synergioita käytännön web-sovellusten kannalta. Teemat jakautuvat toisaalta tietämyksen tuotannon eli ontologioiden ja metadatan tuottamisen kysymyksiin, toisaalta tietämyksen jakeluun liittyviin teemoihin. Yhdistävänä tekijänä toimivat ontologiapalvelut, joita voidaan käyttää sisällön tuotantopuolella ontologioiden luomiseen ja tietojen annotointiin sekä sisällön jakelussa käsiteperustaiseen hakuun ja selailuun. Haku- ja selailupalvelut voidaan toteuttaa toisaalta yksittäisinä semanttisen webin sovelluksina toisaalta useiden sovellusten sisältöjä yhdistävinä metaportaaleina.

4 Hyödyntäminen

Hankkeessa kehitettävä teknologia tarjoaa uusia tapoja ratkaista tiedonhaun, sisällöllisen yhteentoimivuuden ja palveluiden käytön ongelmia kolmelta kannalta:

- **Tiedon loppukäyttäjälle** teknologia tarjoaa sanastoja ja välineitä, jolla tietoa voidaan hakea tietokannoista ja webistä aiempaa joustavammin, monipuolisemmin ja täsmällisemmin.
- **Tiedon tuottajalle** teknologia tarjoaa kehyksen, jonka avulla sisältöjä voidaan kuvata aiempaa täsmällisemmin, sisällöllisesti rikkaalla tavalla ja tasalaatuisemmin.
- **Järjestelmien kehittäjille** avoimet semantic web -esitysmuodot sekä standardit ontologiat tarjoavat teknologisen perustan tiedonhakupöytäkirjojen kehittämiseen webiin. Lisäksi ne tarjoavat perustan hajautettujen järjestelmien yhteiskäytölle verkossa, esimerkiksi museo-, terveydenhuollon- tai kirjastojärjestelmien yhdistämiselle. Web-palvelurajapintojen (web service) kautta voidaan kerran toteutettu palvelu ottaa toisissa sovelluksissa käyttöön vaivattomasti ja pienin kustannuksin.

FinnONTO-konsortio muodostaa poikkeuksellisen laajan kansallisen yhteistyö- ja tutkimusverkoston, johon kuuluu toimijoita web-palveluiden ja julkaisemisen pitkän ja monimuotoisen arvoketjun eri osista:

- Mukana on runsaasti sisällöntuotannon organisaatioita, mm. edustajia kaikista muistiorganisaatioista (kirjastot, museot, arkistot) ja media-alan yrityksistä.
- Mukana on suuri systeemi-integraattori ja lukuista pienempiä yrityksiä, jotka ovat toteuttaneet laajoja web-portaaleja ja jotka harjoittavat myös ohjelmistotuotealaa projektin tutkimusalueella.
- Mukana on suuria organisaatioita ja yrityksiä, joiden liiketoimintaan kuuluvat tiedon tuottamisen ohella (web)jakeluun liittyvät kysymykset.
- Myös webin teknisen infrastruktuurin edustajia on mukana konsortiossa.
- Mukana on useita julkishallinnon ministeriöitä ja organisaatioita, joiden kautta tuloksia voidaan jatkossa hyödyntää julkisella sektorilla yhdessä järjestelmiä toteuttavien yritysten kanssa.

Hankkeessa edistetään laaja-alaista teknologista ja sisällöllistä yhteentoimivuutta ja avoimen semanttisen webin infrastruktuurin luomista maahamme. Luontevaa onkin, että FinnONTO-hankeen toimesta on johdettu Fenix-ohjelman Tietämyksen hallinta ja semanttinen web -aihehrymän toimintaa. Konkreettisenä toimenpiteenä verkottumisen lisäämiseen järjestettiin STeP-2004 päivien yhteydessä Web Intelligence -symposio (Hyvönen et al., 2004) Tiedekeskus Heurekaassa yhteistyössä Suomen Tekoälyseuran, XML Finland ry:n ja W3C:n Suomen toimiston kanssa. 16.11.2005 FinnONTO-projekti järjesti Teknisessä korkeakoulussa yksipäiväisen symposion Kohti suomalaista semanttista webiä, johon ilmoittautui yli 300 osallistujaa.

5 Tuloksia

FinnONTO-hanke jakautuu kahteen n. 2-vuotiseen jaksoon 2003-2005 ja 2005-2007. Ensimmäisen jakson aikana saatuja keskeisimpiä tuloksia on lueteltu alla.

5.1 Ontologiat

Ontologiakehityksen osalta valmistuivat MuseoSuomi-järjestelmän yhteydessä kehitetyt 7 ontologiaa, jotka ovat koekäytössä nykyisessä MuseoSuomi-portaalissa (Hyvönen et al., 2005). Ontologioista laajin on n. 7000 käsitteen Museoalan ontologia MAO. Valokuvausalan ontologia VALO on parhaimmillaan kommentointi-

kierroksella alan asiantuntijoilla. Yleisen suomalaisen ontologian YSO ensimmäinen versio valmistunee vuoden 2006 alussa. Suomen ajallisesta paikkaontologiasta SAPO on valmistunut demonstraatio ja päättelykone koeaineistona alue Karjalan kannaksella (Kauppinen, Hyvönen, 2005). FinnONTO: on v. 2005 liittyneessä Intelligent Web Services –hankkeessa on kehitetty useita ontologioita keltaiset sivut –tyyppisten aiheistojen sisällöllistä esittämistä varten. Tampereen yliopistossa on kehitetty ontologioita mm. elintarvikelaan liittyen.

5.2 Ontologiapalvelut

ONKI-ontologiakirjastosta on valmistunut demonstraatio, joka näyttää, miten ONKI:a voidaan käyttää tiedon indeksoijan apuvälineenä, esimerkiksi osana museon luettelointijärjestelmää. Systemiä on alustavasti testattu Suomen maatalousmuseo Sarkan luettelointijärjestelmässä.

ONKI-demonstraattoria voi kokeilla osoitteessa <http://www.seco.tkk.fi/applications/onki/>.

5.3 Työvälineet

FinnONTO-hankkeen laajin kehityspanos työkalujen osalta on kohdistunut semanttisten portaalien kehittämiseen tarkoitetun OntoViews-ohjelmiston toteuttamiseen. OntoViews-työkalusta valmistui toteutus, joka on saatavilla webistä MIT open source -lisenssillä⁴. Välinettä on testattu ja käytetty alla esiteltävissä pilottidemonstraatioissa.

Hajautetun semanttisen sisällöntuotannon tarpeisiin on valmistunut demonstraatio SAHA-nimisestä editorista⁵. Sen avulla voidaan tuottaa metatietoa esimerkiksi ulkoisista web-sivustoista linkkikirjastoon. Järjestelmä on yhdistetty ONKI-palveluun ja sitä ollaan ottamassa koekäyttöön SeCo-ryhmän portaalihankkeissa. SAHA:an ollaan liittämässä kehitteillä olevaa puoliautomaattista tekstitiedon annotointityökalua POKA⁶.

Helsingin yliopiston yleisen kielitieteen laitoksella on valmistunut demonstraattoreita sanojen merkitysten automaattiseen erottamiseen liittyen.

Tampereen yliopistossa on selvitetty ontologiaperustaisen tiedonhaun etuja ja haittoja käyttäjätutkimuksilla kehitetty mm. ShOE4CIRI- ja CIRI2-nimiset työkalut ontologioiden kehittämiseen ja ontologiaperustaiseen tiedonhakuun annotoimattomasta tekstiaineistosta.

5.4 Pilottisovelluksia

FinnONTO-hankkeen tunnetuin tulos⁷ on kansainvälisesti ja kotimaassa palkittu portaalit MuseoSuomi – Suomen museot semanttisessa webissä (kuva 5.1).

⁴ <http://www.cs.helsinki.fi/group/seco/museums/dist/>

⁵ <http://www.seco.tkk.fi/applications/saha/>

⁶ <http://www.seco.tkk.fi/applications/poka/>

⁷ MuseoSuomi kehitettiin suurimmaksi osaksi FinnONTO:a edeltävässä tutkimushankkeessa v. 2002–2003, mutta järjestelmä vii-meisteltiin ja otettiin pilottikäyttöön vasta v. 2004 FinnONTO-hankkeessa.



The screenshot shows the MuseoSuomi website interface in Microsoft Internet Explorer. The search results are for 'pukineet ja tekstiilit'. On the left, there are filters for 'Esimettyppi', 'Materiaali', 'Valmistaja', 'Valmistuspaikka', 'Valmistusaika', 'Käyttäjä', 'Käyttöpaikka', and 'Käyttötilanne'. The main content area displays a grid of search results, each with a small image and a caption. The results include items like 'harjalampaannahka', 'gasellinahka', 'peski, turkki', 'korkolappukantalappu', 'Kaularöyhelo', 'Tukilivit', and 'Sukkanauhahelit'. The website header includes the MuseoSuomi logo and navigation links.

Kuva 5.1. MuseoSuomi-portaalin hakunäkymä valinnan ”pukineet ja tekstiilit” jälkeen. Vasemmalla on yhdeksän eri näkökulmaa (kahdeksan näkyvässä) aineistoihin ja lomake semanttista käsitteitä varten. Oikealla näkyvät hakutulokset luokiteltuna erityyppisiin pukineisiin ja tekstiileihin.

MuseoSuomesta ollaan nyt takomassa KulttuuriSampo⁸, josta valmistui v. 2005 ensimmäinen pieni demonstraatio (kuva 5.1). KulttuuriSampo yhdistää suomalaisen kulttuuriin eri tavoin liittyviä sisältöjä laajaksi semanttiseksi kansalliseksi kulttuuriportaaliiksi. Temaattisena kohdealueena on perinteinen suomalainen elämänmuoto ja Kalevalan maailma. Yhtenä KulttuuriSammon osatavoitteena on maamme eri museoihin hajautuneen karjalaisen aineiston kokoaminen yhteen portaaliin semanttisen webin tekniikoiden avulla. Osa-hankkeen työnimenä on ”Karjala takaisin mutta semanttisessa webissä”.

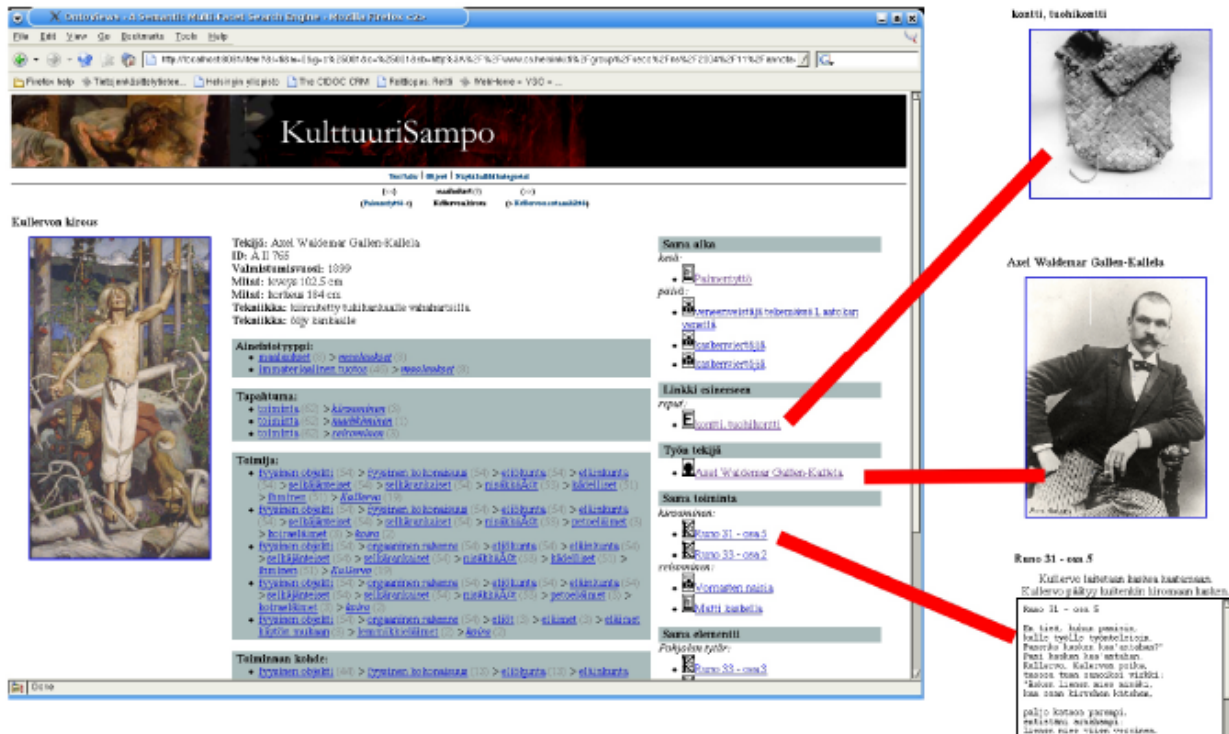
Yleisradion Klaffi-portaalin⁹ yli 2000 videoklipin ja opetusaineiston pohjalta on valmistunut Klaffia vastaava semanttinen portaali Orava¹⁰, joka tukee MuseoSuomen tavoin semanttista moninäkökulmaista ja samoilua. Uusina ominaisuuksina järjestelmään kehitettiin ja toteutettiin semanttinen sanahaun täydennysominaisuus (”semantic autocompletion”) (Hyvönen, Mäkelä, 2005). Siinä käyttäjän kirjoitusta täydennetään ontologisiin hakukategorioiden ja hakutuloksiin ”lennossa” samalla aikaa, kun käyttäjä kirjoittaa tekstiä. Oravassa kokeiltiin myös ensi kerran eri semanttisten portaalien sisältöjen yhdistämistä, tässä tapauksessa Klaffin ja MuseoSuomen. Katsottaessa esimerkiksi Klaffin videota, jossa Paasikivi pitää puheen Helsingin olympialaisten

⁸ <http://www.seco.tkk.fi/applications/kulttuurisampo/>

⁹ <http://www.yle.fi/klaffi/>

¹⁰ <http://www.museosuomi.fi/orava/>

osallistujille 1952, järjestelmä osaa linkittää videon Kansallismuseossa oleviin urheilupalkintoihin ja -välineisiin, joita löytyy MuseoSuomesta (kuva 5.2).



Kuva 5.1. KulttuuriSampo linkittää automaattisesti kulttuurisisältöjä. Kuvassa Ateneumin kokoelmiin kuuluva maalaus Kullervon kirous on yhdistetty Kansallismuseossa olevaan tuohikonttiin (jollainen esiintyy maalauksessakin), Akseli Gallen-Kallelan biografisiin tietoihin ja ontologisesti kuvattuun, Kalevalassa olevaan runoon Kullervon kirouksesta.

eGovernment-alueelle valmistui demonstraatio semanttisesta suomi.fi –portaalista (Sidoroff, Hyvönen, 2005). Demonstraatio osoittaa, miten valtion kansalaiselle tuottamat web-sisällöt voidaan löytää aiempaa joustavammin monien eri näkökulmien, esimerkiksi loppukäyttäjän elämäntilanteiden kautta. Semanttisen tietojen yhdistelyn avulla valtionhallinnon eri tahoilla tuotettu tieto voidaan koostaa automaattisesti kansalaisen kannalta mielekkäiksi kokonaisuuksia, eikä hänen tarvitse etsiskellä tiedon murusia eri tahojen erillisiltä web-sivustoilta. Esimerkiksi elämäntilanteeseen ”Lapsi on syntynyt” voidaan linkittää terveydenhoitoon liittyviä sivuja, tietoa eri kirkkokuntien palveluista, lainopillisia sivuja, tietoa lapsiperheiden sosiaalieduista jne.

eCommerce alueeseen liittyen on toteutettu prototyyppi Veturi (Mäkelä et al., 2005) semanttisista keltaiset sivut –hakemistosta perustuen Fonecta Oy:n 020202.fi-portaalin yli 200 000 ilmoitukseen. Portaaliin ideana on kuvata yritysten ja julkishallinnon palveluiden sisältö ontologioiden avulla, jolloin tiettyyn ongelmaan tai tarpeeseen liittyvät palvelut voidaan löytää loppukäyttäjän näkökulmasta helpommin ja linkittää toisiinsa liittyviä palveluita palvelukokonaisuuksiksi.



Kuva 5.2. Yleisradion Klaffi-portaalin sisällöstä tuotettu semanttinen Orava-portaali (<http://www.museosuomi.fi/orava/>).

5.5 Julkaisuja ja lisätietoa

Semanttisen laskennan tutkimusryhmä SeCo on tuottanut v. 2003-2005 edellä esiteltyyn tutkimukseen liittyen yli 40 tutkimusartikkelia ja 7 opinnäytetöitä, jotka on saatavilla elektronisessa muodossa ryhmän julkaisusivulta¹¹. Tampereen yliopiston informaatiotutkimuksen laitoksen¹² ja Helsingin yliopiston kielitieteen laitoksella¹³ tehtyyn tutkimukseen liittyvää tietoa ja artikkeleita löytyy ao. laitosten kotisivuilta.

6 Projektin organisaatio

6.1.1 Tutkimusosapuolet

Tutkimusosapuolen muodostama verkosta on monitieteinen ja edustaa maamme suurimpia yliopistoja pääkaupunkiseudulla ja Tampereella:

¹¹ <http://www.seco.tkk.fi/publications/>

¹² <http://www.info.uta.fi/>

¹³ <http://www.ling.helsinki.fi/laitos/index.shtml>

FinnONTO -- Suomalaiset semanttisen webin ontologiat

Johtoryhmän kokoonpano (1.10.2005-)

Yrityskonsortio	Jäsen	Varajäsen
AAC Global Oy	Janne Nevasuo	Arto Sinkkonen
Antikvaria-ryhmä	Merja Heiskanen	Hannu Mustonen
Connexor Oy	Pasi Tapanainen	Sirkku Paajanen
Elisa Oyj	Aimo Maanavilja	
Espoon kaupunginmuseo	Liisa Ropponen	Suvi Kettula
Geologian tutkimuskeskus GTK	Olli Rantala	Maija Pennanen
Grip Studios Interactive Oy	Aki Kivelä	
Helsingin kaupunginkirjasto	Erkki Lounasvuori	Matti Sarmela
Kansalliskirjasto	Juha Hakala	Jani Stenvall
Kansanterveyslaitos	Eija Hukka	Jaason Haapakoski
Maa- ja metsätalousministeriö	Antti Vertanen	
Maanmittauslaitos	Reino Ruotsalainen	
M-Cult ry	Minna Tarkka	
Museovirasto	Sirkka Valanto	Vesa Hongisto
Opetushallitus	Liisa Lind	
Opintoluotsi-projekti/OPM	Minna Pesiö-Kokko	Marko Kuparinen
Sanastokeskus TSK ry	Lena Jolkkonen	Katri Seppälä
Sininen Meteoriitti Oy	Aki Antman	
Sosiaaliportti-projekti/Stakes	Ani Kajander	
Stakes (luokituskeskus)	Matti Ojala	Pirkko Kortekangas
Suomalaisen Kirjallisuuden Seura	Lauri Harvilahti	Heli Kautonen
Suomen maatalousmuseosäätiö	Reijo Bäckman	Juha Kuisma
Suomen valokuvataiteen museo	Asko Mäkelä	Anne Isomursu
Tekes	Marko Heikkinen	Matti Sihto
TietoEnator Oyj	Riitta Alkula	Lasse Akselin
Valtion taidemuseo	Elina Heikka	Eeva Melvasalo
Valtiovarainministeriö	Riku Moisio	Pirkko Romakkaniemi
Yleisradio Oy	Tomi Nummi	Marjaana Kuusinen
Tutkimusorganisaatio	Jäsen	Varajäsen
TKK (viestintäteknikka)	Eero Hyvönen	Eetu Mäkelä
Helsingin yliopisto (TKTL)	Eero Hyvönen	Tomi Kauppinen
TKK (kartografia & geoinformatiikka)	Kirsi Virrantaus	
Helsingin yliopisto (YKL)	Lauri Carlson	Kimmo Koskenniemi
Tampereen yliopisto	Kalervo Järvelin	Jaana Kekäläinen

Taulukko 3.3. FinnONTO-hankeeseen osallistuvat yritykset ja julkishallinnon organisaatiot sekä johtoryhmän jäsenet 1.10.2005.

- **Teknillinen korkeakoulu, viestintäteknikan laboratorio** (Prof. Eero Hyvönen). Projektin vetovastuu on Teknillisen korkeakoulun viestintäteknikan laboratoriossa ja Helsingin yliopiston tietojenkäsittelytieteen laitoksella toimivalla Semantic Computing Research Group (SeCo) -tutkimusryhmällä. Ryhmän tutkimus- ja kehitystyö on kohdistunut semanttiseen webiin ja verkkopalveluihin (web services).
- **Helsingin yliopisto, tietojenkäsittelytieteen laitos** (Prof. Eero Hyvönen). Hankkeen hallinnollinen koordinaatio tapahtuu Helsingin yliopiston tietojenkäsittelytieteen laitoksen toimesta. Laitoksen tutkimustyö tehdään SeCo-ryhmän toimesta.
- **Teknillinen korkeakoulu, kartografian ja geoinformatiikan laboratorio** (Prof. Kirsi Virrantaus). Uutena tutkijaosapuolena hankkeeseen liittyy v. 2005 TKK:n kartografian ja geoinformatiikan laboratorio.

- **Helsingin yliopisto, yleisen kielitieteen laitos** (Prof. Lauri Carlson ja Prof. Kimmo Koskenniemi). HY:n yleisen kielitieteen laitos on Suomen johtava kieliteknologian tutkimusyksikkö ja kansainvälisesti arvostettu. Laitoksen tutkimustyö on synnyttänyt maahamme useita kieliteknologian yrityksiä. Tässä hankkeessa laitoksen erityisintressissä ovat mm. semanttinen disambiguointi, ontologioihin liittyvät monikielisyyskysymykset ja sanastojen semanttiset esitysmuodot.
- **Tampereen yliopisto, informaatiotutkimuksen laitos** (Akatemiaprof. Kalervo Järvelin). Tampereen yliopiston informaatiotutkimuksen laitos on kansainvälisesti tunnustettu, Pohjoismaiden suurin alan tutkimusyksikkö. Laitoksen tutkimuksen tärkeitä painopistealue ovat olleet käsiteperustainen ja monikielinen tiedonhaku.

7 Hanketta rahoittavat yritykset ja julkiset organisaatiot

FinnONTO-hankeeseen osallistuvat ja sitä rahoittavat yritykset ja julkishallinnon organisaatiot sekä näiden edustajat projektin johtoryhmässä on esitetty taulukossa 3.3.

Projektin on budjetiltaan 2005-2006 n. 800 000€vuodessa. Hieman yli 80% tutkimustyöstä tehdään Semanttisen laskennan tutkimusryhmässä TKK:n Viestintäteknikan laboratorion tiloissa.

8 Hankkeeseen osallistuminen

FinnONTO-projektin konsortioon liittyminen on vielä mahdollista hankkeen viimeisen vuoden 2006-2007 osalta keväällä 2006.

Tekesin Fenix-ohjelman Tietämyksen hallinta ja semanttinen web –aiiheryhmän toiminnasta kiinnostuneet voivat ilmoittautua ryhmän postituslistalle lähettämällä siitä pyynnön osoitteeseen eero.hyvonen@tkk.fi.

Viitteet

Aitchison J., Gilchrist A., Bawden D.: *Thesaurus Construction and Use: A Practical Manual*. Europa Publications, London, 2000.

Berners-Lee T.: *Weaving the Web*. HapersCollins, New York, 2000.

Davies J., Fensel D., Harmelen F. (eds): *Towards the Semantic Web*. Wiley, 2003.

Ding Y., Fensel D.: *Ontology Library Systems: The Key to Successful Ontology Re-use*. Free University of Amsterdam, 2002. URL: citeseer.nj.nec.com/455227.html

Fellbaum C. (ed.): *WordNet: An Electronic Lexical Database*. MIT Press, 1998.

Fensel D.: *Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce*. Springer-Verlag, 2001.

Foskett D.: *Thesaurus*. In: *Encyclopaedia of Library and Information Science*, Vol. 30, Marcel Dekker, 1980.

Hyvönen E. (ed.): *Semantic Web Kick-Off in Finland*. HIIT Publications 2002-01, Helsinki, 2002.

Hyvönen E., Klemettinen M. (eds): *Towards the Semantic Web and Web Services*. Proceedings of XML Finland 2002. HIIT Publications 2002-03, Helsinki, 2002.

Hyvönen E., Kauppinen T., Salminen M., Viljanen K., Ala-Siuru P. (eds): *Proceedings of the 11th Finnish AI Conference, vol 2., Web Intelligence*. Finnish AI Society, Finland, 2004.

- Hyvönen E.: Miksi asiansanastot eivät riitä vaan tarvitaan ontologioita? Artikkel, TKK Viestintätekniikka, Otaniemi, 2005. URL: <http://www.cs.helsinki.fi/group/seco/ontologies/AsiasanastoistaOntologioihin.pdf>
- Hyvönen E., Mäkelä E., Salminen M., Valo A., Viljanen K., Saarela S., Junnila M., Kettula S.: MuseumFinland – Finnish Museums on the Semantic Web. *Journal of Web Semantics*, Vol 3., No. 2, 2005.
- Hyvönen E., Mäkelä E.: Semantic autocompletion. TKK Media Technology, Espoo, paper, 2005. URL: <http://www.cs.helsinki.fi/group/seco/publications/2005/autocompletion.pdf>
- Kauppinen T., Hyvönen E.: Modeling and Reasoning about Changes in Ontology Time Series. In: Rajiv Kishore, Ram Ramesh and Raj Sharman (eds.): *Ontologies in the Context of Information Systems*. Springer-Verlag, 2005. In press.
- Komulainen V., Valo A., Hyvönen E.: A Collaborative Ontology Development and Service Framework ONKI. *Proceeding of ESWC 2005, poster papers*, Heraklion, Crete, 2005.
- Mäkelä E., Hyvönen E., Saarela S., Viljanen K.: OntoViews – A Tool for Creating Semantic Web Portals. *Proceedings of the 3rd International Semantic Web Conference (ISWC 2004)*, Hiroshima, Japan, 2004, Springer-Verlag, 2004.
- Mäkelä E., Viljanen K., Lindgren P., Laukkanen M., Hyvönen E.: Semantic Yellow Page Service Discovery: The Veturi Portal. *Poster paper, 4th International Semantic Web Conference*, Galway, Ireland, Nov. 6-10, 2005.
- Sidoroff T., Hyvönen E.: Semantic E-government Portals – A Case Study. *Proceedings of the ISWC-2005 Workshop Semantic Web Case Studies and Best Practices for eBusiness SWCASE05*, Galway, Ireland, Nov. 7, 2005.
- Staab S., Studer R. (eds): *Handbook on Ontologies*. Springer-Verlag, 2003.
- Valo A., Hyvönen E., Komulainen V.: A Tool for Collaborative Ontology Development for the Semantic Web. *Proceedings of International Conference on Dublin Core and Metadata Applications (DC 2005)*, Madrid, Spain, 2005.