

Teoksessa K. Raivio, J. Rydman, A. Sinnemäki (toim.): Rajalla – tiede rajojaan etsimässä. Gaudeamus, 2007.

7.9.2007

Älykäs semanttinen web tietämyksen hallinnan rajoja siirtämässä

– Esimerkkinä suomalainen kulttuuri semanttisessa webissä

Prof. Eero Hyvönen

Teknillinen korkeakoulu, viestintäteknikan laboratorio ja

Helsingin yliopisto, tietojenkäsittelytieteen laitos

Semanttisen laskennan tutkimusryhmä

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

eero.hyvonen[at]tkk.fi

Web on antanut meille maailman tiedon käytettäväksi, mutta samalla ongelmaksi on muodostunut olennaisen tiedon etsiminen, löydettyjen tiedonmurusten yhdistäminen mielekkäiksi laajemmiksi kokonaisuuksiksi ja tietomassan muutosten hallinta. Näitä tietämyksen hallinnan ongelmia ollaan ratkaisemassa webin seuraavalla sukupolvella, semanttisella webillä. Sen tavoitteena on saada tietokone ymmärtämään webin sisältöjä ontologisen tiedon ja päättelyn avulla, jolloin kone voi auttaa ihmistä tiedon haussa, yhdistelyssä ja ylläpidossa. Suomenkieleen perustuvaa semanttista webiä on rakennettu laajassa kansallisessa Suomalaiset semanttisen webin ontologiat -hankkeessa (FinnONTO) 2003–2007, jossa on kehitetty laajoja kansallisia ontologioita, kuten Yleinen suomalainen ontologia (YSO), näiden hyödyntämisessä tarvittavia välineitä sekä useita demonstraatioita käytännön sovelluksista. Artikkelissa esitellään kansainvälisesti ja kotimaassa palkittu kulttuurialan semanttinen portaali ”MuseoSuomi – Suomen museot semanttisessa webissä” sekä ”KulttuuriSampo – suomalainen kulttuuri semanttisessa webissä”, jossa mm. käännetään ensimmäistä kertaa Kalevalaa ja Suomen historiaa tietokoneen ”ymmärtämään” muotoon sekä yritetään ”saada Karjalaa takaisin” virtuaalisesti semanttisessa webissä.

Webin pirstaleisen tietämyksen hallinta

World Wide Web (WWW) on kirjapainotaidon keksimiseen rinnastettava tiedonvälityksen ja tietämyksen hallinnan innovaatio, joka vain noin kymmenen vuoden kuluessa on mullistanut tapamme julkaista, hakea ja jäsentää tietoa. Maailmassa on nykyisin yli miljardi Internetin käyttäjää, jotka hakevat ja selailevat miljardeja toisiinsa linkitettyjä sivuja. Tämän lisäksi web sisältää jättiläisemmän määrän indeksoimatonta tietoa erilaisten tietokantojen muodossa, joita käytetään erillisten hakukoneiden avulla (ns. hidden/deep web). Merkkejä vauhdin hidastumisesta ei ole nähtävissä, päinvastoin. Kirjastojen, arkistojen ja museoiden kokoelmia

digitoidaan kiihtyvällä vauhdilla. Vuonna 2001 perustetun Wikipedian englantilaisessa laitoksessa on tätä kirjoittaessa 2 miljoonaa tietosanakirja-artikkelia, joita toimitetaan vertaisverkossa vapaaehtoisvoimin. Sähköiset tietovarastot ovat nopeasti syrjäyttämässä perinteisiä julkaisumuotoja ja luoneet aivan uudentyypisiä välineitä tiedonvälitykseen. Voimme helposti tuottaa sisältöjä yli miljardin lukijan tavoitettavaksi webersivuilla, blogeilla tai YouTuben kaltaisilla uudenaikaisilla, yhteisöllisillä Web 2.0 -henkisillä julkaisualueilla. Kaikki tämä on yleensä vielä ”ilmaista” loppukäyttäjille.

Maailman tieto on meidän kaikkien tavoitettavissa muutamalla napin painalluksella. Haasteeksi on samalla kuitenkin noussut tiedon hakeminen sekä tietokokonaisuuksien hallinta ja linkittäminen web-tietämyksen heinäsuovassa. Esimerkiksi hakusanalla ”iltatahti” tai ”aamutähti” ei välttämättä löydy Venus-planeettaa käsittelevää sivustoa. ”Nokia”-sanalla löytyvien n. 200 miljoonan kännyköistä ja yritysmaailmasta kertovien sivujen joukkoon hukkuvat niin Nokian kaupungista, soopeli-eläimestä (nokia) kuin F. E. Sillanpään Nokia-nimisestä romaanihenkilöstä kertovat sivut. Hakusanalla ”savu” ei luultavasti löydy tulesta kertovaa tietoa, vaikka kaikki tiedämme, ettei savua ole ilman tulta, ”Helsingillä” haettaessa ”Eiraa” tai ”Suomenlinnaa” käsittelevä tieto saattaa jäädä haun ulkopuolelle jne. Tietokokonaisuuksien hallinnan kannalta webin keskeinen innovaatio on sisältöjen linkittäminen assosiativisiksi kokonaisuuksiksi; web on jättiläismäinen verkosto toisiinsa linkeillä yhdistettyjä sivuja. Tietokokonaisuudet eivät kuitenkaan muodostu automaattisesti, vaan linkittäminen joudutaan pääsääntöisesti tekemään käsityönä. Uusien sisältöjen ilmestyessä webiin ne eivät yleensä automaattisesti yhdisty aiempiin sisältöihin. Vastuu ja vaiva ajantasaisen tiedon etsimisestä ja tietojen sisällöllisestä yhdistelystä jää käyttäjälle. Webissä tarvittaisiin kipeästi älykästä tukea tietämyksen hakuun ja hallintaan.

Luonnollinen ratkaisumalli webin tietämyksen hallinnan ongelmiin on käyttää apuna tietokoneita, joiden avulla sivustot ja sisällöt luodaan, indeksoidaan, välitetään ja luetaan. Ongelmana kuitenkin on, että nykyisen WWW:n tiedot on pääosin kuvattu HTML-kielillä, joka on web-sivuille tarkoitettu ulkoisen tulostuskielen kieli, eikä kerro koneella juuri mitään sivujen merkityssisällöstä, semantiikasta. Tietämyksen hallinnassa on kuitenkin kyse nimenomaan sisällön eikä muodon hallinnasta. Esimerkiksi sivun otsikkomerkkaus <H1>Suomalainen kulttuuri</H1> ohjaa selaimen tulostamaan päätteelle päätason otsikon, muttei kerro mitään otsikon merkityksestä ilman ihmisen tulkintaa. Koneelle otsikko ”Suomalainen kulttuuri” on pelkkä rivi merkkejä, samaan tapaan kuin sama merkkijono on meille käsittämätön kreikkalaisilla kirjaimilla esitetty: Συομαλαινεν κουλττουρι.

Koneellisen tulkinnan ongelmaa on lähestytty yrittämällä tulkita verkon sisältöjä koneella ihmisen tavoin tekoälytekniikoiden avulla ja erilaisin tilastollisin menetelmin. Vaikeutena kuitenkin on, että dokumenteissa käytetty kieli on sisällöllisesti monipuolista luonnollista kieltä, jossa voi olla myös puutteita ja virheitä. Luonnollisen kielen ymmärtäminen tietokoneella ei nykyisin ole mahdollista kuin rajoitetusti. Hakukoneilla esimerkiksi kielitaito rajoittuu yleensä vain erillisten sanojen tunnistamiseen merkkijonoina. Vieläkin hankalampaa ja usein käytännössä lähes mahdotonta on kuvien, puheen, musiikin tai videoiden sisällöllinen tulkinta niiden bittimuotoisista esitysmuodoista. Webin tieto on talletettu koneen kannalta hyvin vaikeasti tulkittavilla tavoilla.

Sisältöjen semantiikasta apu tietämyksen hallintaan

Tietokannoissa, kuten kirjastojärjestelmissä, arkistoissa ja museoiden kokoelmatiedoissa tiedonhallinnan ongelma on perinteisesti ratkaistu metatiedolla, ts. tallettamalla tietokantoihin rakenteellista tietoa tiedosta koneen ymmärtämässä muodossa. Voidaan esimerkiksi kertoa, että tietyn teoksen on kirjoittanut F. E. Sillanpää, että teoksen nimi on ”Ihmiset suviyössä” ja että teos on tyypiltään fiktiota. Tämän perusteella kone esimerkiksi älyää sen eron, että kyseessä oleva romaani ei kerro F. E. Sillanpäästä, vaan on ainoastaan tämän kirjoittama, millä voi olla merkitystä tiedon hakijan kannalta. Semanttisen webin¹ ajatuksena on luoda hieman vastaavaan tapaan nykyisen webin päälle uusi metatietokerros, joka kertoo millaisia sisältöjä missäkin verkko-osoitteessa on käytettävissä ja miten nämä liittyvät sisällöllisesti toisiinsa. Sisältöjen semanttinen kuvailu suoritetaan kuitenkin yleensä huomattavasti syvällisemmin ja täsmällisemmin kuin aiemmin.

Perinteisistä tietokannoista poiketen sisällönkuvailu semanttisessa webissä perustuu täsmällisesti määriteltyihin sanastoihin, joita kutsutaan ontologioiksi. Ontologiat esitetään (yleensä) webin kehitystä koordinoivan W3C-järjestön standardien mukaisilla kielillä RDF² ja Web Ontology Language OWL³. Näiden kielten tulkinnaalle (semantiikalle) on sovittu eräitä yhteisiä, logiikan avulla ilmaistuja yleisiä sovellusriippumattomia sääntöjä, ja näitä voidaan laajentaa tarpeen mukaan. Yksi paljon käytetty sopimus tietämyksen esittämisessä on ominaisuuksien periytyminen käsittehierarkioissa yläluokilta alaluokille. Oletetaan esimerkiksi, että eläimistä kertovan ontologiahierarkian mukaan nisäkkäät imettävät poikasiaan, ja että kissaeläimet ovat nisäkkäitä. Jos lisäksi tiedetään, että leijonat ovat kissaeläimiä, voidaan tästä päätellä, että leijonat imettävät poikasiaan ilman että tätä tietoa on erikseen kerrottu. Vastaavaa yleistä päättelyketjua voidaan soveltaa muihinkin käsittehierarkioihin, kuten taivaankappaleisiin, tuoteluokituksiin, anatomisiin käsitteisiin tms. Ontologisella tietämyksellä ja päättelyllä voidaan näin rikastaa eri alojen sisällönkuvailuja yhteismitallisella tavalla ja tuottaa uutta tietoa, jota kone voi käyttää hyväkseen tiedonhaussa ja sisältöjen linkittämisessä. Hakusanalla ”nisäkäs” voidaan esimerkiksi palauttaa leijonista kertova sivu, vaikkei tällä erikseen olisi mainittu sanaa ”nisäkäs”. Leijonista kertova sivu voidaan myös linkittää nisäkkäiden imettämisestä kertovaan sisältöön automaattisesti ja näin hallita paremmin leijoniin eri tavoin liittyvää tietoa.

Ontologiat: semanttisen webin ydin

Ontologiat ja niihin perustuva päättely ovat semanttisen webin keskeisin elementti. Käytännössä ontologiat ovat eri sovellusalojen terminologisia, formaaleja käsittehierarkioita, joissa määritellään alalla käytettävät termit ja käsitteet ja näiden välisiä suhteita. Kansainvälisiä ontologioita ovat mm. WordNet⁴, joka sisältää n. 155 000 englannin kielen sanaa koottuna 118 000 käsitteeksi (ns. synonyymijoukoksi, synset), ensyklopedinen CYC tietämyskanta⁵,

¹ <http://www.w3.org/2001/SW/>

² <http://www.w3.org/RDF/>

³ <http://www.w3.org/2004/OWL/>

⁴ <http://wordnet.princeton.edu/>

⁵ <http://www.cyc.com/>

jossa on esitetty laajan ontologian avulla miljoonia tosiasioita arkimaailmasta, IEEE:n Standard Upper Ontology⁶ ja J. Paul Gettyn säätiön⁷ kulttuurialan tesauukset Art and Architecture Thesaurus AAT (n. 34 000 käsitettä ja 131 000 termiä), taiteilija- ja toimijarekisteri Universal List of Artist Names (ULAN) (n. 120 000 käsitettä ja 293 000 nimeä) sekä paikkaontologia Thesaurus of Geographical Names (TGN). (n. 912 000 paikkaa). Lääketieteessä on käytössä mm. satoja tuhansia hierarkkisia käsitteitä sisältävä kliininen ontologia Snomed CT⁸, kirjastojen käyttämä Medical Subject Headings (MeSH)⁹ ja monia muita luokittelujärjestelmiä, joissa on ontologisia ominaisuuksia. Mainitsemisen arvoinen on kansainvälinen Open Directory Project -hanke¹⁰, jossa webin sivustoista tuotetaan ontologista metatietoa open source -hengessä yli 75 000 vapaaehtoisen toimittajan voimin. Luokittelevana ontologiana käytetään yhteistä 590 000 kategorian hierarkkia, jonka käsitteiden avulla indeksoituna jo lähes 5 miljoonaa web-sivua. Aineistoa ja luokittelua käyttää mm. Google omassa hakemistopalvelussaan.

Monet laajat tesauukset ja käsitteistöt on kehitetty erityisesti indeksoinnin avuksi. Ongelmana on, että ne eivät ole välttämättä semanttisen tiedonhaun ja yhdistelyn kannalta ontologisesti riittävän tarkkoja ja virheetömiä ja johtavat usein virheellisiin päätelmiin. Monia nykyisin jo käytössä olevia sanastoja ja luokitteluja voidaan kuitenkin hyödyntää käytännössä ns. keveinä ontologioina (light-weight ontology) joko sellaisenaan epätarkkuudet hyväksymällä, tai pienin muutoksin.

Tarkastellaan esimerkkinä sanastojen ja ontologioiden eroista Suomessa laajassa käytössä olevaa Yleistä suomalaista asiasanastoa YSA¹¹, jonka avulla voidaan kuvailla esimerkiksi kirjojen, valokuvien, taiteen yms. sisältöjä. Tesauksissa käytettäviä semanttisia suhteita ovat laajempi, suppeampi, ja rinnakkaistermi. Näiden avulla tiedon indeksoija voi löytää sanaston termikäsitteisiin liittyviä toisia käsitteitä ja suorittaa näitä valitsemalla sisällön kuvailun, ns. asiasanoituksen. YSA-sanaston mukaan esimerkiksi asiasanaa ”Halley'n komeetta” laajempi termi (LT) on ”komeetat” ja tätä laajempi termi on ”aurinkokunnat” kuvan 1 kaavion mukaisesti (vasemmalla). Ongelmana on, että tämän rakenteen tulkinta oikein edellyttää ihmisen piilotietämystä luonnosta, eikä tietokone ei pysty rakennetta täysin hyödyntämään. Sanaston perusteella ei esimerkiksi voi tietää, onko Halley'n komeetta yksittäinen komeetta vai komeettojen aliluokka, vai onko Halley'n komeetta osa komeettaa samoin kuin komeetta näyttäisi olevan osa aurinkokuntaa. Ei ole myöskään selvää, ovatko komeetat aurinkokunnan osa vai aurinkokuntien alityyppi (kuten leijonat ovat eläimiä). Hallittaessa aurinkokuntiin liittyvää tietämystä on välttämätöntä pystyä erottamaan komeetat aurinkokunnista, sillä näillä on hyvin erilaiset ominaisuudet koneellisen päättelyn kannalta. Aurinkokunnilla esimerkiksi on oma planeettajärjestelmä komeettoineen päinvastoin kuin komeetoilla. Jos etsitään

⁶ <http://suo.ieee.org/>

⁷ <http://www.getty.edu/>

⁸ <http://www.ihtsdo.org/>

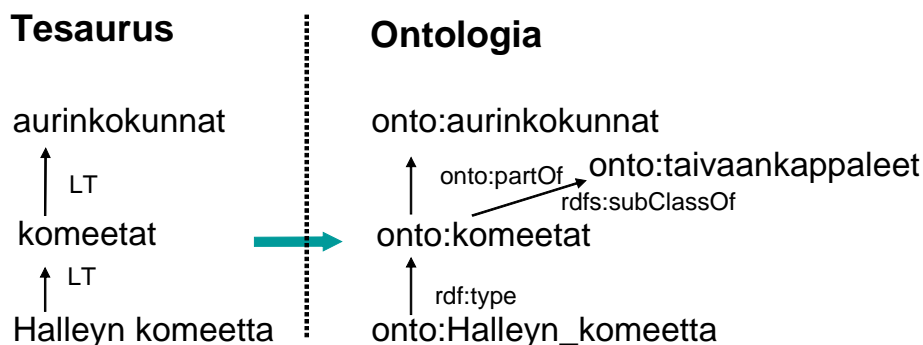
⁹ <http://www.nlm.nih.gov/mesh/>

¹⁰ <http://dmoz.org/>

¹¹ <http://vesa.lib.helsinki.fi/ysa/index.html>

nimenomaisesti aurinkokuntia, ei hakutulokseen saisi tulla komeettoja, koska komeetat eivät ole aurinkokuntia (vaikkakin kuuluvat sellaisiin).

Kuvan 1 oikealla puolella YSA-tesauruksen semanttisia suhteita on täsmennetty yksinkertaiseksi, koneen paremmin ymmärtämäksi ontologiaksi erottamalla yksilöt luokista (type) ja osa-kokonaisuus-suhde (partOf) yläluokka-aliluokkasuhteesta (subClassOf). Syntyneessä ontologiassa kielelliset termit on myös korvattu yksilöivillä URI-tunnisteilla, jotka ovat muotoa *nimiavaruus:tunnus*, kuten web-teknologioissa on tapana. Kuvassa nimiavaruudet *rdf* ja *rdfs* viittaavat semanttisen webin RDF(S) standardeihin, joilla voidaan kuvata yksinkertaisia ontologioita, ja nimiavaruus *onto* viittaa itse määriteltyyn ontologiaan. URI-tunnisteiden nimiavaruuksien avulla voidaan erottaa eri ontologioissa samoille tunnuksille (esimerkiksi komeetat) annetut erilaiset merkitykset toisistaan (esimerkiksi komeetat tähtitieteessä ja komeetat urheiluseurana). Tesauruksen sanastoa on jouduttu myös täydentämään uudella käsitteellä *onto:taivaankappaleet*, jotta käsitteelle *onto:komeetat* voidaan antaa sopiva yläluokka. Sanastoista poiketen ontologiat pyritään rakentamaan systemaattisesti kaikki käsitteet kattaviksi hierarkioiksi ylä-alaluokka-suhteiden avulla.



Kuva 1. Asiasanaston muuttaminen tietokoneen paremmin ymmärtämäksi ontologiaksi.

FinnONTO-hankkeella kohti suomalaista semanttista webiä

Suomalaista semanttista webiä on kehitetty kansallisessa Suomalaiset semanttisen webin ontologiat (FinnONTO) -hankkeessa¹² 2003–2007. FinnONTO on päärahoittajansa Tekesin mukaan sen kaikkien aikojen laaja-alaisin hanke yrityskonsortion koolla mitattuna: mukana on 37 hanketta rahoittavaa yritystä ja julkista organisaatiota. Vastuu hankkeen johtamisesta ja päävastuu tutkimus- ja kehitystyöstä on Teknillisen korkeakoulun viestintätekniikan laboratoriossa ja Helsingin yliopiston tietojenkäsittelytieteenlaitoksella toimivalla Semanttisen laskennan tutkimusryhmällä (SeCo). Mukana yhteistyössä ovat myös Tampereen yliopiston informaatiotutkimuksen laitos, Helsingin yliopiston yleisen kielitieteen ja käännöstieteen laitokset sekä TKK:n geoinformaatio- ja paikannustekniikan laboratorio.

Hankkeen tavoitteena on suomenkieleen ja kulttuuriin perustuvan pohjan luominen kansalliselle semanttiselle webille, joka olisi samalla yhteentoimiva kansainvälisesti. Työn tuloksena on syntynyt useita kansallisia ontologioita, kuten Kansalliskirjaston Yleiseen suomalaiseen asiasanastoon YSA perustuva n. 20 000 käsitteen Yleinen suomalainen ontologia

¹² <http://www.seco.tkk.fi/projects/finnonto/>

YSO¹³. Ontologioiden hyödyntämistä varten on valmistunut työkaluja ja ontologiapalveluita, kuten ONKI-palvelin¹⁴ ontologioiden julkaisemiseksi ja hyödyntämiseksi webissä sekä semanttiseen sisällönkuvailuun tarkoitettu SAHA-editori¹⁵. ONKI-palvelimen ideana on tarjota loppukäyttäjälle, niin ihmiskäyttäjille kuin koneille, sanastojen käyttöön liittyvät palvelut keskitetysti ja kustannustehokkaasti. Esimerkiksi museon luettelointijärjestelmä voidaan kytkeä kulttuurialan MAO-ontologian sisältävään yhteiseen ONKI-palvelimeen tai Maanmittauslaitoksen ja kansainvälisiin paikka-aineistoihin sekä Google Maps -karttoihin perustuvaan ONKI-Paikkaan¹⁶. Liitännään tarvitaan asiakasjärjestelmässä vain pari riviä JavaScript-koodia, joiden avulla ONKI:n käsitteiden hakuun ja tunnisteiden siirtoon toteutettu toiminnallisuus saadaan ”ilmaiseksi” asiakaskäyttöön reaaliajassa ja ajantasaisesti.

Ontologioiden hyödyllisyyttä on demonstroitu useissa semanttisen webin käytännön sovelluksissa. eHealth-alueelle on valmistunut prototyyppi kansalaisten terveyden edistämisen tietoportaalista TerveSuomi.fi¹⁷ osana Kansanterveyslaitoksen koordinoimaa laajempaa kansallista hanketta. eLearning-alueelle tehdään yhteistyössä mm. opetushallituksen ja Yleisradio Oy:n kanssa oppisisältöjen semanttinen julkaisualusta Opintie. Hankkeeseen liittyen on eCommerce-alueella demonstroitu semanttisia keltaisia sivuja Fonecta Oyj:n 020202.fi-palvelun aineistoilla. eGovernment-alueelle valmistui semanttinen kokeiluversio valtiovarainministeriön kansalaisille suuntaamasta suomi.fi-palveluportaalista. Keskeinen sovelluskohde hankkeen alusta alkaen on ollut eCulture, suomalaisten kulttuurisisältöjen julkaiseminen semanttisessa webissä. Hanke julkisti maaliskuussa 2004 Kansallismuseossa pidetyssä tilaisuudessa järjestelmän ”MuseoSuomi – Suomen museot semanttisessa webissä”¹⁸, joka on saanut sittemmin mm. kansainvälisen tutkijayhteisön Semantic Web Challenge -palkinnon Japanissa ja kotimaassa pääministerin kunniamaininnan innovatiivisimmasta web-sovelluksesta hallituksen Tietoyhteiskuntaohjelman Laatia verkkoon -kilpailussa. MuseoSuomi on demonstraatio julkaisukanavasta, jonka kautta maamme eri museot voivat julkaista yhteisöllisesti kokoelmiaan webissä sisällöllisesti yhteismitallisella tavalla. Seuraavassa esitellään lyhyesti MuseoSuomea ja siitä edelleen kehitettävää KulttuuriSampo¹⁹ konkreettisina esimerkkeinä siitä, mitä mahdollisuuksia semanttinen web voi tarjota toisaalta webin loppukäyttäjille toisaalta sisältöjä tuottaville organisaatioille.

MuseoSuomi – Suomen museot semanttisessa webissä

MuseoSuomen ideana on muodostaa kansalaisille webiin virtuaalisesti yksi yhtenäinen kansallinen museokokoelma, joka koostuu saumattomasti toisiinsa yhdistetyistä eri museoiden kokoelmista. MuseoSuomi tarjoaa käyttäjälleen älykkään web-palvelun, jonka kautta voi 1) hakea tietoja samanaikaisesti kaikista järjestelmään kuuluvista kokoelmista ja 2) samoilla

¹³ <http://www.yso.fi/onto/yso/>

¹⁴ <http://www.seco.tkk.fi/services/onki/>

¹⁵ <http://www.seco.tkk.fi/services/saha/>

¹⁶ <http://www.seco.tkk.fi/services/onkipaikka/>

¹⁷ <http://www.seco.tkk.fi/applications/tervesuomi/> ja <http://www.tervesuomi.fi/>

¹⁸ <http://www.museosuomi.fi/>

¹⁹ <http://www.kulttuurisampo.fi/>

kokoelmissa sisältöjä yhdistäviä temaattisia suositteluja seuraten. Sovellus on avoimessa koekäytössä webissä Teknillisen korkeakoulun ylläpitämänä ja sillä on ollut v. 2007 yli 10 000 eri käyttäjää kuukaudessa.

Perinteisistä hakukoneista poiketen MuseoSuomen haku perustuu hakusanojen taustalla oleviin käsitteisiin ja luokitteluihin (ontologioihin), mikä helpottaa merkittävästi hakukyselyiden muodostamista ja johtaa tarkempiin hakutuloksiin. Kysely voidaan tehdä vaivattomasti tekemällä linkkivalintoja yhdeksästä eri luokittelusta (näkyvästä), kuten ”esinetyyppi”, ”materiaali” ja ”käyttöpaikka”. Jokaisen valinnan jälkeen hakutuloksena ovat kohteet, jotka vastaavat samanaikaisesti kaikkia tehtyjä valintoihin. Esimerkiksi esinetyypin valinnalla ”päähine” voidaan löytää erityyppiset päähineet, kuten ”silinterit”, ”ylioppilaslakit” tai ”harakat” ilman että käyttäjän tarvitsee pystyä luettelemaan näitä sanoja erikseen. Samoin käyttöpaikan valinnalla ”Etelä-Suomen lääni” voidaan viitata automaattisesti kaikkiin kyseisen alueen kuntiin, kyliin, kaupunkeihin ja kaupunginosiin kuten luonnollisessakin kielessä. Tällainen sanojen taustalla oleviin käsitteisiin perustuva tiedonhaku on lisäksi mahdollista tehdä samanaikaisesti kohteiden eri ominaisuuksien suhteen. Esimerkiksi valitsemalla kiinnostuksen kohteeksi ”päähineet”, kohteen käyttöpaikaksi ”Etelä-Suomen lääni” ja käyttäjäksi ”nainen” löytyvät kaikki ko. läänin eri kunnissa ja kylissä naisten käyttämät erityyppiset päähineet.

Hakupalveluiden ohella järjestelmä tarjoaa kullekin kohteelle automaattisesti muodostettuja suosittelulinkkejä toisiin kiinnostaviin kohteisiin. Suosittelut perustuvat haun tapaan järjestelmän taustalla olevien kokoelmatietojen muodostamaan assosiatiiviseen semanttiseen verkostoon, joka koostuu n. 10 000:sta käsitteestä seitsemässä eri ontologiassa sekä yli 4000:n kokoelmakohteen metatiedoista. Lisäksi tietokoneelle on ohjelmoitu ”maalaisjärkeä” kulttuurisisältöjen merkityksestä logiikan sääntöinä. Kone esimerkiksi ymmärtää, että ylioppilaslakit liittyvät juhlatapahtumiin, joiden eräs erikoistapaus ovat ylioppilasjuhlat. MuseoSuomen aineistot ovat peräisin pilottihankkeessa mukana olleesta Kansallismuseosta, Espoon kaupunginmuseosta sekä Lahden kaupunginmuseosta.

KulttuuriSampo – suomalainen kulttuuri semanttisessa webissä

MuseoSuomi oli ensimmäinen askel kohti laajempaa järjestelmää ”KulttuuriSampo – suomalainen kulttuuri semanttisessa webissä”. KulttuuriSammon ideana on kehittää MuseoSuomen kaltainen julkaisukanava kaikenlaisten kulttuurisisältöjen, ei ainoastaan museoiden esinekokoelmien julkaisemiseksi ja automaattiseksi linkittämiseksi semanttisessa webissä. Kehitystyön lähtökohtana on ajatus siitä, että webin kautta saatava kulttuurikokemus syntyy kohteiden ohella näiden välisistä kiinnostavista yhteyksistä ja näiden avulla luoduista temaattisista kokonaisuuksista.

Kuva 2 esittää esimerkkinä semanttisen webin mahdollisuuksista KulttuuriSammon ensimmäisen prototyypin käyttöliittymää vuodelta 2005. Siinä käyttäjä tutustuu Ateneumin kokoelmiin kuuluvaan Akseli Gallen-Kallelan maalaukseen ”Kullervon kirous”, joka näkyy vasemmalla. Keskellä näkyvät maalauksen metatiedon ja tämän alla linkkeinä ontologiset käsitteet, joihin maalauksen liittyy. Linkkejä painamalla voidaan löytää muut ao. käsitteisiin liittyvät kohteet. Oikealla näkyvät KulttuuriSammon loogisen päättelykoneen tuottamat suosituslinkit muihin sisältökohteisiin sekä suomenkieliset perustelut linkityksille. Ylin linkki johtaa Kansallismuseon kokoelmissa olevaan tuohikonttiin, koska maalauksen metatiedoista

löytyy tieto siitä, että tämäntyyppinen kontti esiintyy maalauksessa. Keskimäinen linkki johtaa maalauksen luoneen Akseli Gallen-Kallelan biografiaan ja alin linkki Kalevalan runon 31 siihen kohtaan, johon maalaus liittyy.

Temaattisia kulttuurialan virtuaalinäyttelyitä on tehty runsaasti käsityönä sekä CD-levytykseen että webiin. Tämä onnistuu pienille muuttumattomille aineistoille kohtuullisilla kustannuksilla. MuseoSuomessa ja KulttuuriSammossa on kuitenkin kysymys laajoihin ja muuttuviin kokoelmätietokantoihin ja kohdejoukkoihin perustuvista virtuaalinäyttelyistä, jossa voi olla kymmeniä tuhansia erityyppisiä kohteita, kuten esineistöä, valokuvia, maalauksia, videoita, maastokohteita, rakennuksia, runoja, sävellyksiä jne. Lisäksi kokoelmat kasvavat ja sisällöt lisääntyvät vuosi vuodelta. Käsien tehtävä kiinteä teemoitus tai linkitys ei aineistojen monimuotoisuuden, laajuuden, säilytyspaikan ja muuttumisen vuoksi ole käytännössä mahdollista vaan tähän tarvitaan tietokoneen apua.

The screenshot shows the KulttuuriSampo website interface. The main content area displays the painting 'Kullervon kirous' by Akseli Gallen-Kallela. To the right, there is a sidebar with various navigation and search options. Red arrows point from the website to three external images: a small bag (kontti), a portrait of Akseli Gallen-Kallela, and a page from the Kalevala (Runo 31).

Kuva 2. Akseli Gallen-Kallelan teos Kullervon kirous (Ateneumin taidemuseon kokoelmat) ja siihen linkittyneitä sisältöjä KulttuuriSammossa.

Tietokone on hyvä väline myös silloin, kun samasta aineistosta halutaan automaattisesti tuottaa erilaisia näyttelyitä. MuseoSuomessa ja KulttuuriSammossa linkityslogiikkaa voidaan muuttaa ilman että aineistoa tarvitsee muokata. Visiona on, että kertaalleen semanttisesti asianmukaisesti tuotettu aineisto voidaan käyttää sellaisenaan hyödyksi eri toimijoiden eri sovelluksissa, ei ainoastaan kunkin sisällöntuottajan omassa ainutkertaisessa tietojärjestelmässä. Näin säästyy sekä vaivaa, aikaa että rahaa uusien sovelluksia kehitettäessä. KulttuuriSampo on yhteisöllinen sovellus Web 2.0 -hengessä eri sisällöntuottajille. Se on hiili, johon eri sisällöntuottajat voivat yhdessä puhalttaa kansallisen tason KulttuuriSammon takomiseksi.

Tapahtumat sisällönkuvailun ytimenä

KulttuuriSampo laajentaa MuseoSuomea paitsi uusilla sisältötyypeillä myös uudella sisällönkuvailun menetelmällä. Keskeiseen asemaan on nostettu tapahtumat sekä näistä muodostuvat laajemmat kokonaisuudet, prosessit ja tarinat. KulttuuriSammon vision mukaan

kulttuurin sisällöllisen ytimen muodostavat yhteiskunnan ja tarinoiden tapahtumat. Siksi niiden tulee olla keskeisessä asemassa kulttuurisisältöjen kuvailussa, semanttisen haussa ja linkittämisessä. Tapahtumat tarjoavat semanttisen perustan aiempaa tarkempaan kulttuurikohteiden sisällönkuvailuun tietokonetta varten. Niiden avulla voidaan ilmaista sekä sisällönkuvailussa käytettyjä asiasanoja vastaavat ontologiset käsitteet että näiden välisiä keskinäisiä semanttisia suhteita.



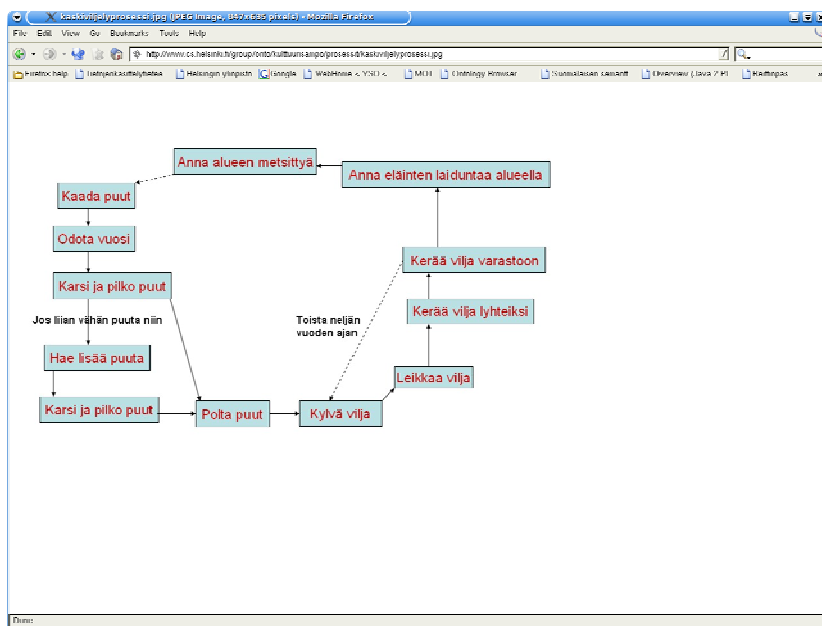
Kuva 3. Akseli Gallen-Kallelan Aino triptyykki (Ateneumin taidemuseo).

Esimerkiksi kuvassa 3 olevassa Akseli Gallen-Kallelan ”Aino triptyykissä” Väinämöinen ajaa takaa Ainoa, joka hukuttautuu järveen. Maalauksen kokoelmätietokannassa olevista metatiedoista löytyvät sisällönkuvailukentässä mm. asiasanat ”Väinämöinen”, ”Aino”, ”hukkuminen” ja ”järvi”. Näiden käsitteiden välisten suhteiden ymmärtäminen edellyttää kuitenkin inhimillistä tietämystä ja tulkintaa maailmasta ja Kalevalasta. Asiasanoista ei esimerkiksi käy ilmi, kuka maalauksessa hukkuu vai liittyykö se vain ylipäänsä hukkumisteemaan. Ongelma ratkaistaan KulttuuriSammossa kuvaamalla Aionon hukkuminen YSO-ontologiasta otetulla ”hukkuminen”-tapahtumakäsitteellä, jonka suhteen ”Aino” on aktiivisena toimijana (agent) ja ”järvi” paikkana (place). Näin tietokonekin voi ymmärtää, että kuvassa nimenomaan Aino hukkuu järveen ja ettei esimerkiksi järvi hukkunut Väinämöiseen.

Tapahtumaperustaista sisällönkuvailua kokeillaan Kalevalan ohella myös kulttuuristen prosessien esittämiseen. Esimerkiksi kuva 4 esittää perinteisen kaskiviljelyn vaiheita prosessina KulttuuriSammon ensimmäisessä prototyyppissä. Tällaisten kuvausten avulla voidaan paitsi tallentaa tietoa katoavasta kansanperinteestä myös yhdistää prosessin eri vaiheisiin liittyvää tietoa, kuten kaskiviljelyssä käytettäviä työkaluja ja kaskiviljelyyn liittyviä kohtia Kalevalassa.

Myös Suomen historiasta on laadittu pieni tapahtumiin perustuva ontologia perustuen Suomen historiaverkko Agricolan²⁰ julkaisemaan aikajanaan. Historialliset tapahtumat ovat tärkeässä asemassa kulttuurisisältöjen linkittämisessä toisiinsa ja ovat toisaalta itse yksi KulttuuriSammon sisältötyyppi. Tieto tapahtumista on keskeistä elämäämme liittyvän tietämyksen hallinnassa.

²⁰ <http://agricola.utu.fi/>



Kuva 4. Kaskiviljelyn semanttinen kuvailu tapahtumina KulttuuriSammossa.

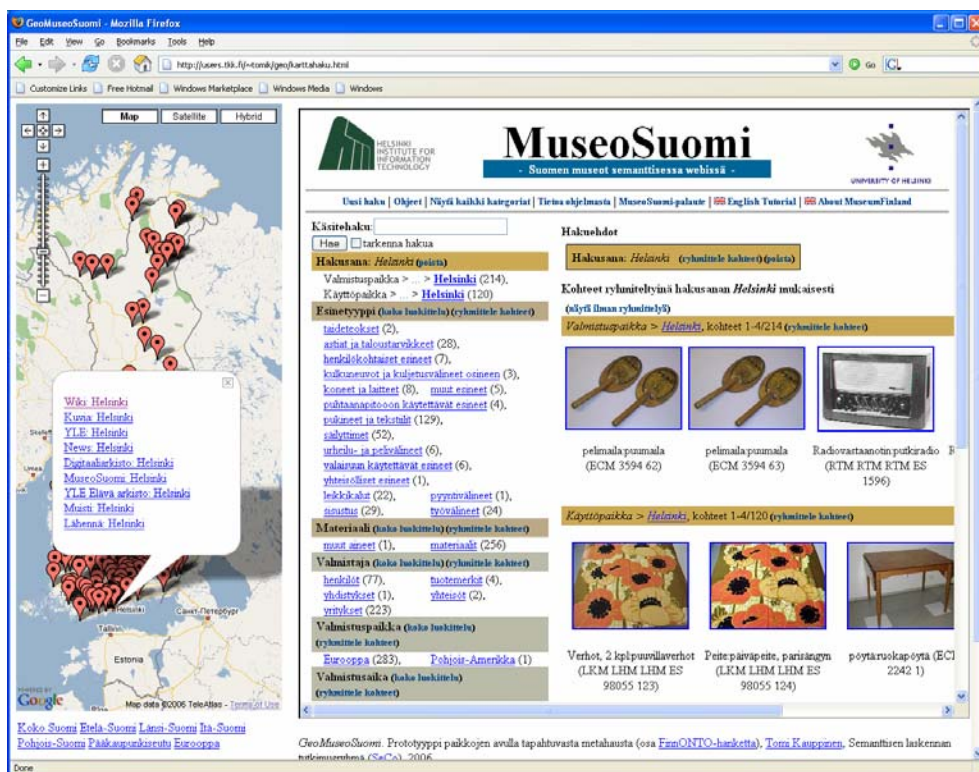
Insinöörien ja tietokoneen Kalevala

Yksittäisiä tapahtumia ketjuttamalla voidaan kuvailla laajempia prosesseja ja tarinoita. Koska Kalevalainen symboliikka ja tarinat muodostavat keskeisen inspiraation lähteen suomalaisessa kulttuurissa, ollaan KulttuuriSammossa kokeilemassa Kalevalan esittämistä tietokoneen ymmärtämässä muodossa YSO-ontologian mukaisina tapahtumina. Näin tarjoutuu mm. mahdollisuus yhdistää maalauksia ja maataloudessa käytettyä esineistöä Kalevalan teksteihin ja tätä kautta toisiinsa. Loppukäyttäjälle tämä ”insinöörien Kalevalaksi” ristitty kokonaisuus näyttäytyy KulttuuriSammossa Kalevalan tekstinä, jonka eri kohtiin muut kulttuurisisällöt samoin kuin eri kohtiin liitetyt selitykset yhdistyvät linkkeinä. Ajatuksena on tarjota KulttuuriSammon sisältö ikään kuin Kalevalan tekstin kautta jäsentyneenä kokonaisuutena.

Kalevala on käännetty n. 60 eri ihmiskielelle. KulttuuriSammossa ollaan ensi kertaa käännetty Kalevalaa tietokoneen ”ymmärtämälle” kielelle. Mukana osahankkeessa on ollut Elias Lönnrotin johdolla v. 1831 perustettu Suomalaisen kirjallisuuden seura. Kansanrunouden ja musiikin saralla ontologisoitityön alla on myös Suomen Kansan Vanhat Runot²¹ (kuitenkin ilman tapahtumaperustaista sisällönkuvailua) sekä näihin liittyvät Suomen Kansan eSävelmät²².

²¹ <http://dbgw.finlit.fi/skvr/>

²² <http://www.jyu.fi/musica/sks/>



Kuva 5. MuseoSuomen ja Google Maps -palvelun yhdistäminen ”mash-up” -sovellukseksi.

Karjala takaisin semanttisessa webissä

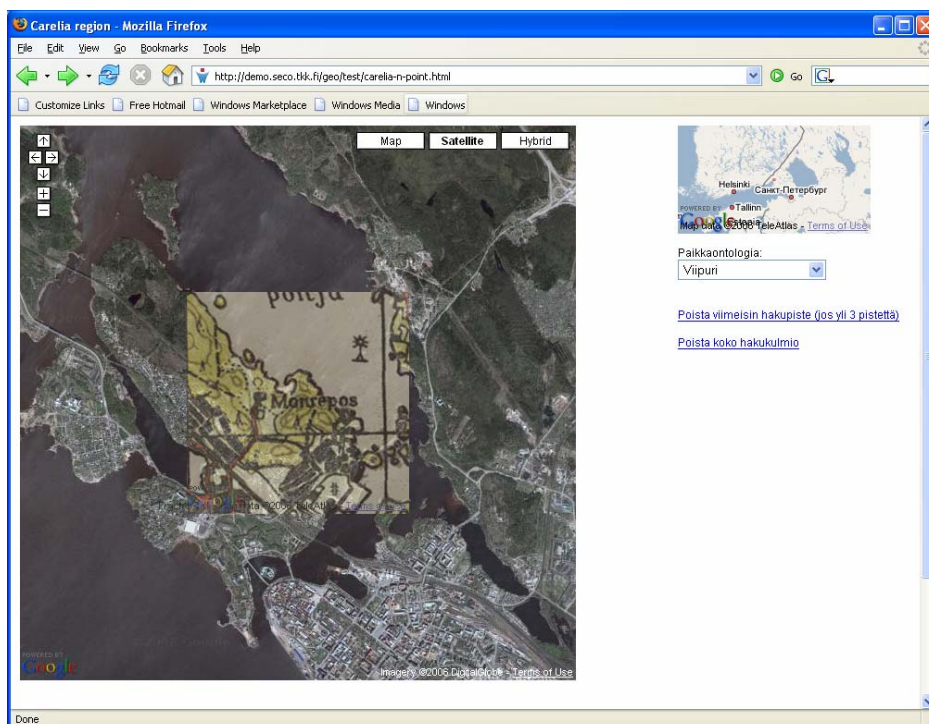
KulttuuriSampoon kerätään alkuvaiheessa aineistoja liittyen erityisesti Karjalaan, Kalevalaan ja 1800-luvun maataloudesta eläneeseen Suomeen. Aineistojen valinnalla pyritään edistämään niiden keskinäistä linkittymistä. Luovutetun Karjalan esineistö kulkeutui ja hajasijoitettiin toisessa maailmansodassa evakuoinnin yhteydessä useaan maamme eri museoon, jotka ovat nyt mukana sisällöntuottajina KulttuuriSampo-hankkeessa opetusministeriön tuella. Web tarjoaa julkaisukanavan, jonka avulla hajasijoitetut Karjalan kulttuurisisällöt voidaan koota verkon avulla virtuaalisesti yhteen. Vastaavanlainen kulttuurisisältöjen hajautuminen eri museoihin ja maihin on hyvin yleistä ja ongelmallista mm. Keski-Euroopassa, puhumattakaan entisistä siirtomaista, joiden kulttuurihistoriallisesti arvokkain esineistö löytyy usein entisten emämaiden suurista museoista.

Tämä osa KulttuuriSampo-portaalia käyttää aineistonaan museoiden ja arkistojen kokoelmia, vanhoja Karjalan karttoja sekä Wikipediassa olevia Karjala-aiheisia tekstisivuja, jotka tarjoavat yhden tekstiperustaisen näkökulman Karjala-teemaan.

Sisältöjen visualisointi ja liittäminen toisiin palveluihin

Sisältöjen hakuun ja visualisointiin käytetään KulttuuriSammossa uusinta web-teknologiaa, mm. Google Maps -palvelun satelliitti- ja karttakuvia yhdistettynä MuseoSuomen ja KulttuuriSammon omiin sisältöihin. Kuvassa 5 on esimerkiksi MuseoSuomen kokoelmien esineistöä peilattu niiden metatiedoissa olevien käyttö- ja valmistuspaikkatietojen perusteella

vasemmalla näkyvälle Google Maps -kartalle. Kartalla olevaa punaista pallukkaa klikkaamalla käyttäjä voi tehdä automaattisesti haun oikealla näkyvään MuseoSuomeen. Kuvassa valintana on ollut ”Helsinki”, jolloin hakutuloksessa näkyvät kaikki Helsingin alueella valmistetut tai käytetyt esineet. Samalla järjestelmä on avannut pallukan kohdalle linkkivalikoiman eräisiin muihin verkossa oleviin hakukoneisiin, kuten Wikipediaan ja Yleisradion uutispalveluun. Linkkiä painamalla voidaan tehdä paikannimeen liittyvä haku linkin osoittamaan palveluun, esimerkiksi hakea Wikipediasta Helsinkiä käsittelevä artikkeli.



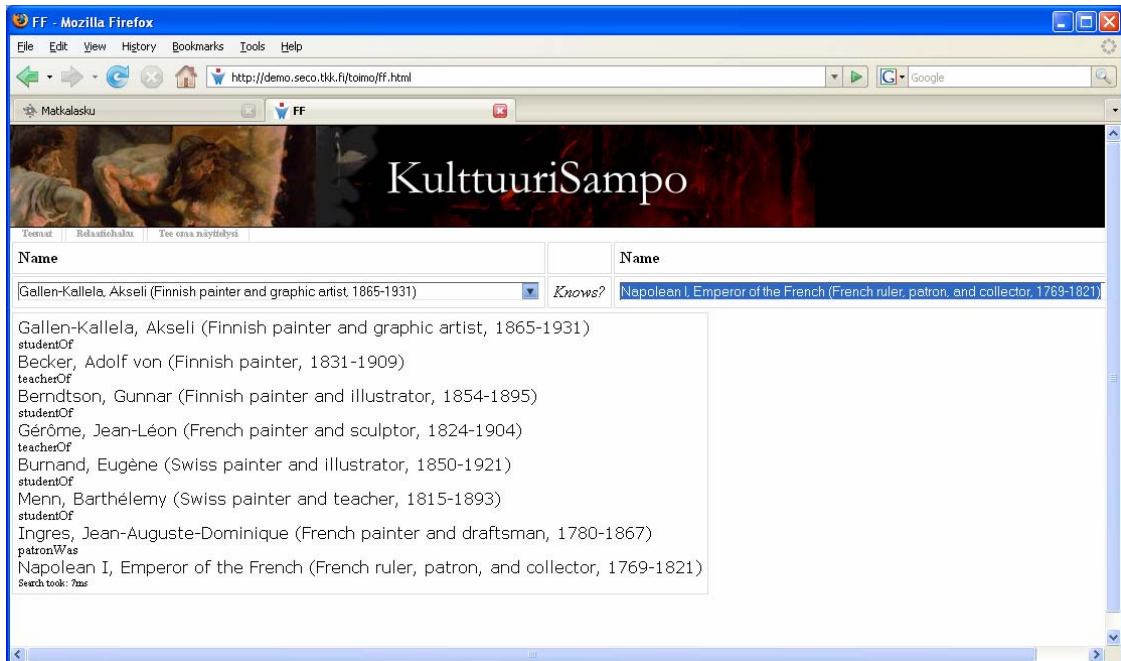
Kuva 6 Luovutetun Karjalan vanhojen karttojen ja Google Maps -palvelun yhdistäminen. Keskiössä Monrepos'n puisto Viipurin alueella.

Kuvassa 6 Google Maps -palvelu on yhdistetty vanhoihin suomalaisiin Maanmittauslaitoksen Karjalan karttoihin toista maailmansotaa edeltäneeltä ajalta. Keskellä oleva neliö on eräänlainen ”historiallinen ikkuna”, joka näyttää samanaikaisesti päällekkäin sekä nykyistä satelliittikuvaa että sotia edeltänyttä vanhaa karttaa Viipurin alueelta. Neliössä näkyvää tarunhoitoista nimeä ”Monrepos” ei nykyisistä kartoista enää välttämättä löydy. Karttapalvelun ideana on auttaa ihmistä hahmottamaan luovutetun Karjalan aluetta ja nimistöä suhteessa nykyiseen Venäjälle kuuluvaan alueeseen.

Relaatiohaku

KulttuuriSammossa kuten semanttisessa webissä yleensä asioiden väliset yhteydet eli relaatiot ovat yhtä tärkeässä asemassa kuin itse asiat. Hankkeessa onkin siksi kehitetty uudenlaista hakutekniikkaa, jossa haun kohteena ovat relaatiot. Kuvassa 7 on esimerkki KulttuuriSammon

relaatiohausta²³. Järjestelmän aineistona on J. Paul Getty -säätiön n. 120 000 taiteilijaa ja kulttuurialan toimijaa sisältävä ULAN-tietokanta²⁴. Käyttäjä on valinnut näistä kaksi: Akseli Gallen-Kallelan ja Ranskan keisari Napoleonin. Relaatiohaun tuloksena on seitsemän ihmissuhteen polku, joka yhdistää Gallen-Kallelan eri aikaan eläneeseen Napoleoniin. Sen mukaan Gallen-Kallela oli Adolf von Beckerin oppilas, joka opetti Gunnar Berndtsonia, joka oli Jean-Leon Geromen oppilas, joka opetti Eugen Burnardia, joka oli Barthelemy Mennin oppilas, joka oli Jean-Auguste-Dominique Ingresin oppilas, jonka suojelija keisari Napoleon oli.



Kuva 7. KulttuuriSammon relaatiohaku on löytänyt seitsemän askeleen yhteyden Akseli Gallen-Kallelan ja Ranskan keisari Napoleonin välillä.

KulttuuriSammosta on kehitetty kaksi tutkimuksellista prototyyppiä vuosina 2005 ja 2006. Sovelluksen lopullinen demonstraatio on tarkoitus ottaa MuseoSuomen tapaan avoimeen koekäyttöön webissä projektin päättyessä vuoden 2007 lopulla. Kokonaisuus sisältää yllä kuvattujen osien lisäksi eräitä muitakin osia, kuten MuseoSuomesta edelleen kehitetyn, uudentyyppisen näkymäperustaisen hakukoneen sekä eräänlaisen automaattisen virtuaalisten näyttelyiden generoijan. Portaalien haku- ja suosituspalvelua tarjotaan käytettäväksi avointen rajapintojen kautta muiden sovellusten sivuilla ”leijukkeina”²⁵, jotka ovat eräänlaisia KulttuuriSammon sisältöjä ja linkkisuosituksia oikeissa yhteyksissä esitteleviä tietoikkunoita.

²³ <http://demo.seco.tkk.fi/toimo/ff.html>

²⁴ http://www.getty.edu/research/conducting_research/vocabularies/ulan/

²⁵ <http://www.seco.tkk.fi/services/floatlets/>

Haasteena semanttinen sisällöntuotanto

Semanttisten web-sovellusten kehittämisessä on runsaasti teknisiä ja käytännöllisiä haasteita. Avainkysymys loppukäyttäjän kannalta on, miten web-sivustoista saadaan aidosti mielenkiintoisia, hyödyllisiä ja helppokäyttöisiä. Suuria käytännön haasteita liittyy tietosisältöjen tuotantoon: miten museoissa, arkistoissa, kirjastoissa ja muissa sisältöä tuottavissa organisaatioissa voitaisiin tuottaa tehokkaasti metatietoa yhteismitallisessa, semanttisesti tulkittavissa olevassa muodossa niin, ettei sisältöjä tarvitsisi myöhemmin korjailta käsin, kun kehitetään MuseoSuomen ja KulttuuriSammon kaltaisia sovelluksia? Nykyinen asiasanoitukseen ja vapaisiin tekstikenttiin perustuva kuvailu ei ole riittävän täsmällistä tietokoneen käytettäväksi, vaan tarvitsee tulkikseen ihmistä. Uusissa Web 2.0 -sovelluksissa käyttöönotetulla vapaalla ”tägäyksellä” on sama ongelma: eri henkilöt käyttävät samoja termejä eri merkityksissä eikä ”tägien” välisiä suhteita huomioida. Esimerkiksi ”lasilla” voidaan viitata joko ikkunaan, materiaaliin, linssiin, juomalasiin tms. Seurauksena on semanttinen sekaannus, joka suuremmilla aineistomäärillä muodostuu yhä häiritsevämmäksi estäen täsmällisen tietämyksen hallinnan.

Käyttämällä sisällönkuvailussa, esimerkiksi kokoelmien luetteloinnissa, semanttisen webin ontologioita ja URI-tunnisteita voidaan sanojen eri merkitykset yksilöidä, määritellä niiden väliset suhteet täsmällisesti ja saada sisällöt semanttisesti yhteentoimiviksi. Tällaisten sisältöjen avulla voidaan toteuttaa aiempaa käyttökelpoisempia älykkäitä haku- ja samoilupalveluita, kuten MuseoSuomi-järjestelmä on osoittanut. Tärkeä ”liiketaloudellinen” näkökohta on, että semanttisen webin avulla sisällöntuottajille voidaan tarjota yhteinen julkaisukanava, jolloin jokaisen toimijan ei tarvitse itse toteuttaa ja ylläpitää kalliita web-palveluita erikseen ja silti asiakkaita voidaan palvella aiempaa paremmin semanttisen yhteisportaalin avulla. Suomessa on laskettu olevan jopa tuhat eri museota – luku edustaa maailman huippua asukasta kohti laskettuna.

Ontologioiden ja semanttisen webin teknologioiden kustannustehokas käyttöönotto edellyttää kuitenkin paitsi ontologioiden kehittämistä myös näiden käyttöön liittyviä palveluita, joihin esimerkiksi eri museoiden luettelointijärjestelmät voitaisiin yhdistää ja jota kautta yhteisesti käytettävät tunnisteet voidaan ottaa helposti käyttöön. FinnONTO-hankkeessa onkin kehitetty teknisiä ratkaisuja myös näihin käytännöllisiin ongelmiin, kuten ONKI-ontologiapalvelin, SAHA-editori sekä puoliautomaattisen sisällönkuvailun apuvälineitä kieliteknologiaan perustuen.

Kiitokset

Yli 30 tutkijaa on osallistunut artikkelissa esiteltyyn tutkimustyöhön sen eri vaiheissa. FinnONTO-hanketta rahoittivat sen loppuvaiheessa Tekes sekä 36 muun yrityksen ja julkisen organisaation muodostama konsortio: AAC Global Oy, Antikvaria-ryhmä ja Museoliitto, CelAmanzi Oy, Connexor Oy, eduskunta, Elisa Oyj, Espoon kaupungin museo, 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 Meteoritti 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.

Kirjallisuutta

Kirjoitus perustuu seuraaviin artikkeleihin, jotka ovat saatavilla sähköisesti Semanttisen laskennan tutkimusryhmän julkaisusivulta <http://www.seco.tkk.fi/publications/>.

Ontologiat

Eero Hyvönen: Miksi asiasanastot eivät riitä vaan tarvitaan ontologioita? *Signum* 5/2005.

Eero Hyvönen, Katri Seppälä, Kim Viljanen ja Matias Frosterus: Yleinen suomalainen ontologia YSO - kohti suomalaista semanttista webiä. *Tietolinja*, no 1, 2007.

MuseoSuomi

Eero Hyvönen: MuseoSuomi - Suomen museot semanttisessa webissä. Järjestelmä museovieraan ja museon näkökulmasta. Helsingin yliopisto ja HIIT, 1.3.2004.

Eero Hyvönen, Eetu Mäkelä, Mirva Salminen, Arttu Valo, Kim Viljanen, Samppa Saarela, Miikka Junnila and Suvi Kettula: MuseumFinland – Finnish Museums on the Semantic Web. *Journal of Web Semantics*, vol. 3, no. 2, 2005.

KulttuuriSampo

Eero Hyvönen, Tuukka Ruotsalo, Thomas Häggström, Mirva Salminen, Miikka Junnila, Mikko Virkkilä, Mikko Haaramo, Eetu Mäkelä, Tomi Kauppinen and Kim Viljanen: CultureSampo – Finnish Culture on the Semantic Web: The Vision and First Results. Teoksessa Klaus Robering (Ed.): *Information Technology for the Virtual Museum*. LIT Verlag, 2007.

Miikka Junnila, Eero Hyvönen and Mirva Salminen: Describing and Linking Cultural Semantic Content by Using Situations and Actions. Teoksessa Klaus Robering (Ed.): *Information Technology for the Virtual Museum*. LIT Verlag, 2007.

Tomi Kauppinen, Riikka Henriksson, Jari Väätäinen, Christine Deichstetter and Eero Hyvönen: Ontology-based Modeling and Visualization of Cultural Spatio-temporal Knowledge. *Developments in Artificial Intelligence and the Semantic Web - Proceedings of the 12th Finnish AI Conference STeP 2006, October 26-27, 2006*.

FinnONTO-hanke

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, IEEE Press, 2007.