

FinnONTO-malli kansallisen semanttisen webin sisältöinfrastruktuurin perustaksi - visio ja sen toteutus

Eero Hyvönen

Professori, tutkimusjohtaja

TKK viestintäteknikan laitos ja Helsingin yliopisto, tietojenkäsittelytieteen laitos

PL 5500, 02015 TKK

eero.hyvonen [ät] tkk.fi

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

Artikkelissa luodaan katsaus FinnONTO-hankkeen (2003-2008 ja 2008-2010) visioon kansallisesta semanttisen webin sisältöinfrastruktuurista, jonka ensimmäinen prototyyppi otetaan avoimeen Living Laboratory -käyttöön 12.9.2008 Teknillisellä korkeakoululla pidettävässä julkistustilaisuudessa. Prototyypissä toteutettuja ydinosa-ia ovat Yleisestä suomalaisesta ontologiasta YSO ja sitä täydentävistä alaontologioista koostuva Kansallinen ontologiajärjestelmä KOKO sekä Kansallinen ontologiakirjastopalvelu ONKI, joka tarjoaa alustan ontologioiden yhteisölliseen kehittämiseen, julkaisemiseen verkkopalveluina sekä hyödyntämiseen asiakasjärjestelmissä kustannustehokkaasti Web 2.0 -teknologialla. Järjestelmä on ensimmäinen laatuaan maailmassa.

1 Miksi tarvitsemme kansallisen semanttisen webin sisältöinfrastruktuurin?

1.1 Semanttinen web on tietojen verkko

World Wide Web on joukko toisiinsa linkeillä yhdistettyjä verkkosivuja (web of pages), joiden hakemiseen ja käyttämiseen on tarjolla erilaisia hakupalveluita ja muita verkkosovelluksia. Semanttinen web¹ on webin kehitystä koordinoivan W3C-järjestön v. 2001 käynnistämä kehityssuunta², jonka pyrkimyksenä on täydentää perinteistä, sivujen verkosta muodostuvaa webiä siinä esiintyvien käsitteiden ja (meta)tietojen muodostamalla semanttisella verkolla (web of data). Semanttisen webin käsitteitä ja tietoja ymmärtävät paitsi ihmiset myös verkkosovellukset, ts. tietokoneet, mikä mahdollistaa aiempaa älykkäämpien ja keskenään yhteentoimivampien verkkosovellusten kehittämisen.

¹ <http://www.w3.org/2001/sw/>

² <http://www.w3.org/>

Oletetaan, että etsit ”helsinkiläisiä ravintoloita” webistä. Semanttisessa webissä toimiva palvelu ymmärtää suositella sinulle illallista ”Kalliossa” sijaitsevassa pizzariassa, koska webin taustalla oleva semanttinen verkko tietää, että ”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 (koska pizza on italialainen ruokalaji), jossa on sopivasti pari peruutuspaikkaa, tai tarjota matkapuhelimeen navigointipalvelu.

Tällaisten älykkäiden palveluiden kehittämisen edellytyksenä on, että verkossa toimivien järjestelmien käytettävänä on yhteisesti sovittuja, tietokoneen ”ymmärtämässä” muodossa määriteltäviä käsitteitä ja näiden välisiä suhteita, kuten käsiteluoikka ”kaupunginosa” ja siihen kuuluva yksilö ”Kallio”, joka on osa kaupunkiluokan yksilöä nimeltä ”Helsinki”. Käsitteitä ja niiden välisten suhteiden määrittelyverkkoja kutsutaan web-alalla ontologioiksi (Fensel, 2003; Staab, Studer, 2008).

Semanttisessa webissä sisällönkuvailu tapahtuu määrämuotoisten metadataformaattien avulla kuten perinteisissäkin informaatiojärjestelmissä. Esimerkiksi valokuvasta voidaan kertoa kuvan ottohetki ja -paikka, kuvaaja, kuvan aihe, kameratyyppi jne. Verkkosisältöjen kuvailuun käytetään kuitenkin perinteisten asiasanojen, ”tägien” ja tekstikuvailujen sijasta ontologioista otettuja käsitteitä, joiden merkitys yksilöidään URI-tunnisteiden avulla. Tämä mahdollistaa sisältöjen aiempaa täsmällisemmän ja monipuolisemman haun, yhdistelemisen ja visualisoinnin verkkopalveluissa (Hyvönen, 2005). Kuvailussa voidaan käyttää myös käsitteistä muodostuvia rikkaampia rakenteellisia kuvauksia. Jos esimerkiksi Akseli Gallen-Kallelan Aino-triptyykin maalauksen luettelointitietoihin kuuluvat asiasanoina merkkijonot ”Väinämöinen”, ”Aino”, ”hukkuminen” ja ”Kullervo”, voidaan nämä tiedot semanttisessa webissä täsmentää niin, että nimenomaan Kalevalan henkilöhenkilö Aino hukkuu, eikä Väinämöinen (tai muut Aino-nimiset henkilöt). Tämä voi olla tarpeen, kun esimerkiksi etsitään hukkuneita fiktiivisiä henkilöitä. Sisältöjen ontologinen kuvailu, annotointi, voidaan tehdä käsityönä tai olla automaattista, mikäli tämä on teknisesti mahdollista.

Semanttisen webin perustana on metatietojen ja niihin kytkeytyvien ontologioiden muodostama semanttinen verkko. Verkon koko ja laatu riippuvat sovelluksesta. Useimmat käytännön sovellukset liittyvät rajattuun sisältöalueeseen. Esimerkiksi MuseoSuomessa³ (Hyvönen et al., 2005b) ontologisia käsitteitä on käytössä n. 10 000 ja hakukohteita hieman yli 4000 yhdellä metadataformaattilla esitettynä. Kulttuurisammossa⁴ taas on käytössä kymmeniä tuhansia käsitteitä ja sisältökohteita indeksoituna lukuisten eri metadataformaattien avulla. Koko verkon mittakaavassa tapahtuvasta semanttisesta sisältötyöstä mainittakoon Open Directory Project⁵, jossa nykyisin yli 80 000 vapaaehtoista toimittajaa indeksoi verkkosisältöjä yli 590 000

³ <http://www.museosuomi.fi/>

⁴ <http://www.kulttuurisampo.fi/>

⁵ <http://www.dmoz.org/>

kategoriaa sisältävän yksinkertaisen RDFS-ontologian avulla, sekä Linked Data Initiative⁶, jossa ajatuksena on julkaista avoimesti ontologioita ja metatietoja ja tarjota niihin perustuvia verkkopalveluja. Hankkeen RDF-muotoisessa tietovarastossa oli maaliskuussa 2008 jo yli 2 miljardia semanttista kaarta, aineistoina mm. Wikipediasta tuotettu semanttinen versio DBpedia.

1.2 Kansallinen haaste: ontologiat

Semanttisen webin ytimenä ovat yhteisesti käytettävät, tietokoneen ymmärtämät ontologiat, jotka 1) mahdollistavat sisältöjen koneellisen ”ymmärtämisen” ja 2) joiden avulla verkon monimuotoiset, hajautetusti tuotetut eri alojen sisällöt ja palvelut voidaan saada keskenään yhteentoimiviksi (interoperability) sisällöllisellä (semanttisella) tasolla. Webin muuttuminen vähitellen yhä semanttisemmaksi asettaa kansainvälisen haasteen standardintyölle, jota on tehty W3C-järjestössä määrätietoisesti v. 2001 alkaen. Työn keskeisimpinä tuloksina on otettu käyttöön joukko XML-perustaisia suosituksia tietomalleista ja rakenteista liittyen metadatan, sanastojen ja ontologioiden esittämiseen ja hakuun, kuten Resource Description Framework RDF(S)⁷, Simple Knowledge Organisation System SKOS⁸, Web Ontology Language OWL⁹ ja SPARQL¹⁰. Nämä standardit määrittelevät yleisiä, yhteentoimivuuden ja päättelyn kannalta keskeisiä semanttisia (loogisia) periaatteita, joita noudattamalla yhteentoimivaa semanttista webiä nykyisin rakennetaan. Esimerkiksi RDFS-suosituksessa määritellyllä subclassOf-suhteella voidaan ilmaista käsitettä yleisempi käsite, esimerkiksi että ”leijonat” ovat ”kissaeläimiä”.

W3C:n standardit eivät ota kantaa ontologioiden tai metadata formaattien varsinaiseen sisältöön, vaan ovat luonteeltaan sovellusriippumattomia ja perustuvat logiikkaan. Esimerkiksi subclassOf suhde mahdollistaa leijonaa yleisemmän käsitteen kissaeläin ilmaisemisen, mutta varsinaiseen eläinlajien taksonomian muodostamiseen ei oteta kantaa. Sisältö- ja ontologiatyö on jätetty kunkin alan asiantuntijatahojen tehtäväksi.

Suomessa ontologioita alettiin kehittää laajamittaisemmin semanttisen MuseoSuomi-järjestelmän yhteydessä v. 2002 alussa. Työssä tarvittiin käsitteistöjä mm. museoesineille, materiaaleille, henkilöille, paikoille ja tapahtumille. Hankkeessa kävi heti ilmi, ettei maassamme ei ollut käytettävissä ontologioita ja ettei kansainvälisiä järjestelmiä voitu sellaisinaan juuri hyödyntää suomenkielestä ja kansallisista sisällöistä ja käytännöistä johtuen. Kansainvälisistä rekistereistä ei esimerkiksi löytynyt paljoa suomalaisia paikkoja, henkilöitä tai organisaatiota ja kulttuurimme on muutenkin esimerkiksi esineistön osalta paljolti omaleimasta. Tästä havainnosta kypsyi ajatus siitä, että maamme tarvitsee semanttisen webin ontologisen sisältöinfrastruktuurin, jonka varaan esimerkiksi MuseoSuomen kaltaiset julkiset ja yksityiset palvelut voitaisiin rakentaa kustannustehokkaasti. Hanketta toteuttamaan koottiin aluksi 14 yrityksen ja julkisen organisaation rahoittajakonsortio, jonka työ

⁶ <http://linkeddata.org/>

⁷ <http://www.w3.org/RDF/>

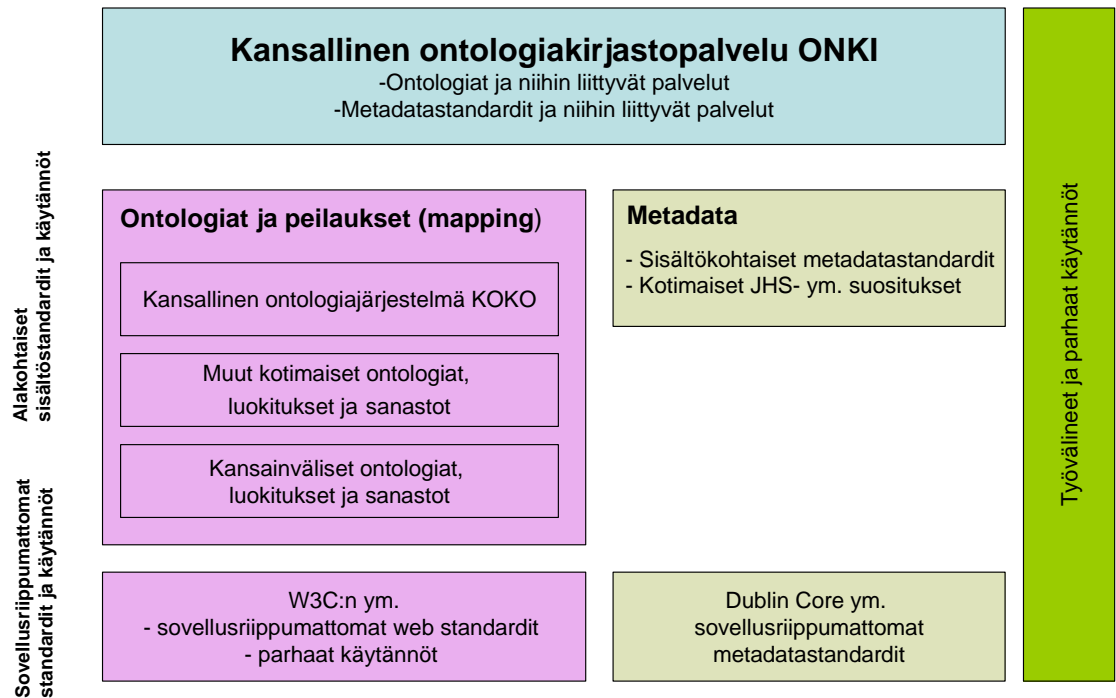
⁸ <http://www.w3.org/2004/02/skos/>

⁹ <http://www.w3.org/2004/OWL/>

¹⁰ www.w3.org/TR/rdf-sparql-query/

käynnistyi Tekesin tuella v. 2003 FinnONTO-hankkeena¹¹. Vuosien kuluessa FinnONTO-hankkeen koko on kasvanut 39 rahoittajaan tehden projektista tiittävästi Tekesin historian laajimman hankkeen konsortion koolla mitattuna. FinnONTO-hanketta on vetänyt Teknillisen korkeakoulun ja Helsingin yliopiston Semanttisen laskennan tutkimusryhmä SeCo¹², jonka suojissa myös pääosa tutkimustyöstä on tehty.

Kansallinen semanttisen webin sisältöinfrastruktuuri FinnONTO



Kuva 1.1. Kansallisen semanttisen webin infrastruktuurin komponentit.

1.3 Ehdotus sisältöinfrastruktuurin malliksi

Ajatusta semanttisen webin sisältöinfrastruktuurista voi verrata tie-, sähkö- ja puhelinverkkojen muodostamiin perinteisiin infrastruktuureihin. Semanttisten käsitteiden verkosto on luonnollisesti abstrakti ja näkymätön, mutta mahdollistaa vastaavaan tapaan yhteyksiä, liikkumista ja tiedonsiirtoa kuin vaikkapa tie- tai puhelinverkko. Visiona on, että jos FinnONTO-infrastruktuurin pelisäännöillä tuotettu sisältö julkaistaan verkossa, voidaan se kytkeä ja hyödyntää automaattisesti muiden toimijoiden semanttisissa verkostoissa hieman vastaavaan tapaan, kuin uusi maantienpätkä tieverkoston osana. Kantava ajatus on, että uuden sisällön arvo rikastuu ”ilmaiseksi” infrastruktuurin avulla verkon muusta sisällöstä (vastaavasti kuin uuden tieosuuden arvo syntyy sen kytkeytymisestä muihin teihin), ja toisaalta muiden verkon julkaisijoiden sisältöjen arvo rikastuu uuden tiedonsirpaleen avulla (vastaavasti kuin uusi tie parantaa muiden teiden keskinäistä saavutettavuutta). Sekä

¹¹ <http://www.seco.tkk.fi/projects/finnonto/>

¹² <http://www.seco.tkk.fi/>

itse infrastruktuuri että siinä jo oleva tieto voidaan hyödyntää toisissa sovelluksissa sellaisenaan, mikä säästää merkittävästi järjestelmien kehityskustannuksia.

FinnONTO-hankkeessa kehityksen ratkaisumallin (Hyvönen et al., 2008ab) mukaan sisältöinfrastruktuurin keskeisimmät komponentit ovat kuvan 1.1 mukaisesti:

1. Ontologiat ja niiden väliset peilaukset (ontology matching/mapping) (Euzenat, Shvaiko, 2007), joista muodostuu yhtenäinen kokonaisontologia KOKO
2. Metadastandardit
3. Ontologioiden ja metadastandardien kehittämiseen, julkaisemiseen ja hyödyntämiseen tarkoitettu Kansallinen ontologiakirjastopalvelu ONKI
4. Joukko työvälineitä ja käytäntöjä

Ontologioiden osalta infrastruktuurissa pyritään noudattamaan erityisesti W3C:n suosituksia. Samoin metadastandardien määrittelyissä yhteentoimivuus kansainvälisiin standardeihin ja hankkeisiin, kuten Dublin Core Metadata Initiative¹³, on tärkeää. Käytetyistä ja itse kehitetyistä työkaluista mainittakoon ontologiaeditori Protege¹⁴, ONKI:a hyödyntävä metadastandardien editoria SAHA (Valkeapää, 2006; Valkeapää et al., 2008), (puoli)automaattisen annotoinnin väline POKA¹⁵ (Alm, 2007), erilaiset muuntimet formaattien välillä, ontologioiden peilausten muodostajat ja metadatan semanttisen validoinnin työkalut.

Tässä artikkelissa esitellään FinnONTO-hankkeen sisältöinfrastruktuurin keskeisimmät osat ontologioiden osalta:

1. **Kansallinen ontologiajärjestelmä KOKO.** KOKO on kokoelma suomalaisia ydinontologioita, jotka on liitetty peilauksilla toisiinsa. Näitä ovat ensivaiheessa eri alojen yleiskäsitteitä sisältävä Yleinen suomalainen ontologia YSO ja joukko sitä laajentavia ja tarkentavia erityisalojen ontologioita. Tavoitteena on, että ontologioita voitaisiin jatkossa kehittää ja ylläpitää yhteistoiminnallisesti (collaborative development) ja koordinoitusti eri alojen ammattilaisryhmien toimesta, joita ovat mm. nykyiset asiantuntijajoukot ja luokitusten kehittäjät.
2. **Kansallinen ontologiakirjastopalvelu ONKI.** ONKI mahdollistaa ontologioiden, luokitusjärjestelmien ja sanastojen ja kustannustehokkaan, keskitetyn julkaisemisen verkossa ja hyödyntämisen vaivattomasti asiakassovelluksissa. Palvelut voidaan ottaa käyttöön joko valmiina käyttöliittymäkomponentteina ja toiminnallisuuksina Google Maps –palvelun tapaan (AJAX-teknologia) tai perinteisempinä Web Service –palveluina (SOAP- ja WSDL-rajapinnat).

¹³ <http://dublincore.org/>

¹⁴ <http://protege.stanford.edu/>

¹⁵ <http://www.seco.tkk.fi/tools/poka/>

Ontologioita ja ONKI-palvelua on kehitetty ja hyödynnetty erityisesti FinnONTO-hankkeen kahdessa laajassa pääsovelluksessa:

1. **Kulttuurisampo** on järjestelmädemostratio kansallisesta, museoiden, kirjastojen, arkistojen ja muiden kulttuurisisältöjen tuottajien semanttisesta julkaisukanavasta ja älykkästä portaalista¹⁶ (Hyvönen et al., 2007b). Prototyypin sisältölajeja ovat mm. museoesineet, valokuvat, kulttuuripaikat, kansanrunot, musiikki, kartat, biografiat, Kalevala, kaunokirjallisuus, maalaukset, videot ja Wikipedia-sivut. Kulttuurisammon aineisto on peräisin yli kahdestakymmenestä eri organisaatiosta ja sen tietämuskanta perustuu KOKO-ontologiaan ja metadataskeemoihin erilaisille aineistotyypeille. Kulttuurisampo-portaalin prototyyppi julkaistaan Kullervon päivänä 25.9.2008 Kansallismuseossa pidettävässä tilaisuudessa¹⁷.
2. **TerveSuomi** (Hyvönen et al., 2007; Suominen et al., 2007) on vastaavantyyppinen järjestelmä terveystiedon hajautettuun sisällöntuotantoon ja julkaisemiseen semanttisessa webissä. Portaalin prototyyppi julkaistaan 12.9. Teknillisellä korkeakoululla. TerveSuomi-hanketta johtaa Kansanterveyslaitos Sosiaali- ja terveysministeriön rahoituksella ja siihen on osallistunut n. 30 terveystietoa tuottavaa organisaatiota. Semanttiseen webiin perustuva julkaisukonsepti ja teknisen ratkaisun prototyyppi on kehitetty osana FinnONTO-hanketta.

Seuraavassa esitellään lyhyesti FinnONTO-infrastruktuurin ydinkomponentit: KOKO-ontologiajärjestelmä siihen liittyvine kehitysprosesseineen sekä ontologiakirjastopalvelu ONKI.

2 Kansallinen ontologiajärjestelmä KOKO

Semanttisen webin tuleminen on käynnistänyt ja vilkastuttanut sanasto- ja ontologiatyötä monilla eri aloilla. Esimerkiksi terveysalalla on kehitetty jättimäinen, satojen tuhansien käsitteiden Snomed Clinical Terms¹⁸ ja kulttuurialalla käytetään Getty-säätiön kulttuurikäsitteitä, taiteilijoita ja historiallisia paikkoja mallittavia tesauuksia AAT, ULAN ja TGN¹⁹. Geonames-palvelu²⁰ sisältää miljoonia nykyisiä paikkoja. WordNet (Fellbaum, 1998) on laajasti käytetty englannin kielen ontologinen rakenne, josta on kehitetty versioita muillekin kielille. Kansallisten semanttisen webin ontologioiden tulisi olla yhteentoimivia kansainvälisten käsitteistöjen kanssa. Toisaalta kansalliset käytännöt, sisällöt ja kielet joudutaan ottamaan huomioon.

Semanttisen webin sisällöllinen ontologiatyö on ollut paljolti alakohtaista ja perustunut olemassa olevien luokitusten käyttöön ja muokkaamiseen. Eri alojen

¹⁶ <http://www.kulttuurisampo.fi/>

¹⁷ <http://www.seco.tkk.fi/events/2008/2008-09-25-kulttuurisampo/>

¹⁸ <http://snomed.org/>

¹⁹ http://www.getty.edu/research/conducting_research/vocabularies/

²⁰ <http://www.geonames.org/>

välisen työn koordinaatiota ei ole juurikaan ollut. Eri sovelluksissa tarvittava ontologinen tieto on yleensä ainakin osittain tapauskohtaista, mistä johtuen aiemmin kehitettyjen ontologioiden uusiokäyttö toisissa sovelluksissa on haasteellisempaa kuin helposti kuvitellaan. Työn tuloksena semanttisen webin kehittäjien käytettävissä on kuitenkin jatkuvasti kasvava joukko ontologioita. Esimerkiksi ontologioiden hakuun kehitetty Swoogle hakukone²¹ on indeksoinut yli 10 000 verkossa saatavilla olevaa ontologiaa. De facto -standardeja on muodostumassa suurten ontologiahankkeiden kuten WordNetin ja Getty säätöön sanastojen ympärille.

2.1 FinnONTO-ontologiatyön periaatteita

FinnONTO-hankkeessa ontologiatyö käynnistyi kansallisten käytäntöjen pohjalta ja perustuen olemassa oleviin asiasanastoihin. Asiasanastot (Aitchison et al., 2000) esitetään yleensä ISO 2788, BB 5723 (Iso-Britannia) ja ANSI/NISO Z39.19 (USA) standardien mukaisessa muodossa. Sanasto koostuu termeistä, joiden merkitys on kuvattu joukolla suhteita toisiin termeihin. Tärkeimmät termien väliset suhteet, joita käytetään mm. suomalaisissa asiasanastoissa, on esitetty taulukossa 2.1. Sanaston semanttiset suhteet jakaantuvat kolmeen pääryhmään: ekvivalenssisuhteisiin (KÄYTÄ, KORVAA), hierarkkisiin suhteisiin (LT, ST) ja assosiativisiin suhteisiin (RT).

Suhde englanniksi	Suhde suomeksi	Merkitys
BT Broader term	LT	Laajempi termi, hierarkkinen suhde, käänteissuhde ST
NT Narrower term	ST	Suppeampi termi, hierarkkinen suhde, käänteissuhde LT
RT Related term	RT	Rinnakkaistermi, assosiativinen suhde
USE Use	KÄYTÄ	Ilmaisee suositeltavan synonyymien, käänteissuhde KORVAA
UF Used for	KORVAA	Ilmaisee ei-suositeltavan synonyymien, käänteissuhde KÄYTÄ
SN Scope note	HUOMAUTUS	Luonnehtii termin merkitystä ja käyttöä

Taulukko 2.1. Tärkeimpiä termien välisiä semanttisia suhteita asiasanastoissa.

FinnONTO-hankkeessa on pyritty hyödyntämään maksimaalisesti maassamme jo tehty sanastotyö. Näin tietokantoihin jo indeksoitu aineisto saadaan helpommin semanttisen käsittelyn piiriin ja sanastojen kehittäjien työ hyödynnettyä. Nykyisiä sanastoja ja luokituksia voidaan yrittää hyödyntää joko sellaisenaan semanttisessa webissä tekemällä mekaanisia muunnoksia esimerkiksi SKOS-formaattiin. FinnONTO:ssa monessa tapauksessa näin on meneteltykin. Yleisenä tavoitteena käytössä olevien asiasanatesaurusten muuntamisessa on kuitenkin täsmentää niiden

²¹ <http://swoogle.umbc.edu/>

semantiikkaa ja rakennetta siten, että niiden käyttö semanttisen webin sovelluksissa kuten päättelyssä ja visualisoinnissa tulee paremmin mahdolliseksi. Asiasanastojen kevyeen ontologisointiin kehitettiin yksinkertainen menetelmä (Hyvönen et al., 2008a), jonka avulla joukko sanastoja on muutettu semanttisen webin kannalta käyttökelpoisempaan muotoon. FinnONTO-hankkeen ehdotus asiasanastojen kehittäjille on astua askel eteenpäin ja kehittää asiasanastojen sijasta jatkossa yksinkertaisia ontologioita (light weight ontology) (Hyvönen, 2005).

Sanastojen muuttaminen ontologioiksi on prosessi, johon liittyy paitsi teknisiä myös yhteisöllisiä ja rahoituksellisia kysymyksiä ratkaistavaksi. Alla on lueteltu FinnONTO-hankkeen lähtökohtia, teesejä ja ehdotuksia siitä, miten ontologisointi prosessissa kannattaisi kokemustemme mukaan edetä:

Perustan hyödyntäminen uudelleen eri yhteyksissä. Semanttisen webin sovellusten edellytyksenä on usein laajojen ontologioiden käyttö. Näiden kehittäminen on kuitenkin usein liian kallista yksittäistä sovellusta varten. FinnONTO:n ontologiatyön lähtökohtana on tehdä tuota perustyötä yhdessä julkisin varoin, jakaa tulokset yleisesti käytettäväksi ja saada näin aikaan merkittäviä säästöjä yksittäisiä sovelluksia laajemmalla kansallisella tasolla. Kehitystyö maksaa itsensä uusina palveluina ja uuden liiketoiminnan kautta.

Päällekkäisen työn eliminointi ja koordinointi. Eri alojen sanastoissa on paljon yhteisiä termejä. Esimerkiksi Museoalan ontologiassa MAO on 2330 yhteistä käsitettä Kansalliskirjaston Yleisestä suomalaisesta asiasanastosta YSA muodostettuun YSO-ontologiaan. Lisäksi nämä molemmat ovat osin päällekkäisiä mm. Taideteollisuusalan ontologian TAO kanssa. Sanasto- ja ontologiatyö leikkaavilla alueilla johtaa helposti pyörän keksimiseen moneen kertaan eri tahoilla ja ei-yhteentoimiviin ratkaisuihin. Tämän estämiseksi tarvitaan yhteistyötä ja koordinaatiota.

Avoin lähdekoodi kansallisille ydinsisällöille. FinnONTO-ontologiat julkaistaan lähtökohtaisesti avoimena lähdekoodina liiketoimintaystävällisellä Creative Commons -lisenssiversiolla, jossa ontologioiden käyttöä, rakenteiden muokkausta tai koodin sulkemista myöhemmin ei rajoiteta millään tavoin. Lisenssi edellyttää ainoastaan järjestelmän kehittäjien mainintaa. Tavoitteena on tehdä ontologioiden käyttö mahdollisimman esteettömäksi ja näin edistää laajasti käytettyjen ontologioiden ja käytäntöjen syntymistä ja alan standardoitumista, mikä on sisällöllisen yhteentoimivuuden edellytys.

Käytön edistäminen maksuttomuudella. FinnONTO-vision mukaan Kansallisen ontologiapalvelun ONKI ylläpitäminen pitäisi tapahtua julkisin varoin samaan tapaan kuin esimerkiksi tieverkosto on Suomessa maksuton ja kaikkien käytettävissä. Hyvin toimivan ja yleisesti käyttöön otetun infrastruktuurin varaan voidaan rakentaa kustannustehokkaammin palveluita ja liiketoimintaa. Tästä on Suomessakin hyviä kokemuksia mm. GSM-standardien mahdollistaessa osaltaan Nokian maailmanvalloitusta 1990-luvulla.

Yhteistyöverkoston rakentaminen ja tukeminen. FinnONTO-infrastruktuurin on perustuttava eri alojen yhteistoimintaan, sillä minkään tahon asiantuntemus ja resurssit eivät voi kattaa kaikkea ontologiatyötä. ONKI-palvelun tavoitteena on tarjota keskitetty alusta ja palvelu, jonka kautta eri kehittäjäryhmille tarjotaan tukea yhteistyön edistämiseksi. Jos esimerkiksi ontologian A käsite K on otettu käyttöön ontologiassa B, ja ontologian A kehittäjät muuttavat K:ta, voitaisiin tästä automaattisesti varoittaa ontologian B kehittäjiä.

Moniarvoisuus ja virheiden sietäminen. FinnONTO:n ontologiatyön yhtenä lähtökohtana on tunnustaa se tosiasia, ettei yhteen, eri tahoja ja tarpeita tyydyttävään ontologiseen maailmanmalliin ole mahdollista päästä tai edes tarkoituksenmukaista pyrkiä. Samaten on selvää, etteivät ontologiat ole koskaan ”valmiita”, vaan niistä löytyy aina parannettavaa - vielä enemmän kuin asiasanastoista. Tämä johtuu ontologisen mallintamisen vaikeudesta ja menetelmien kehittymättömyydestä. FinnONTO:n yhteistyömallissa eri kehittäjäryhmät voivat siksi ottaa käyttöön itsenäisesti omia ratkaisujaan käsitteiden organisoinnissa hierarkioiksi. Esimerkiksi Museoalan ontologiassa MAO tai Taideteollisuusalan ontologiassa TAO voidaan ottaa käyttöön erilainen käsitepuu samoille käsitteille kuin Yleisessä suomalaisessa ontologiassa, jos se katsotaan tarpeelliseksi MAO-laisten ja TAO-laisten piirissä.

Minimaalinen ontologinen sitoutuminen. KOKO-järjestelmään liittyviltä ontologioilta edellytetään vain hyvin minimaalisten ontologisten periaatteiden kunnioittamista: RDFS-semantiikan mukainen yksinkertainen luokkahierarkia, sama tulkintaa asiasanastoista periytyville yksinkertaisille ominaisuuksille (kuten RT) ja eräitä yksinkertaisia harmonisoituja nimeämiskäytäntöjä. Esimerkiksi ekvivalenteiksi merkityillä käsitteillä pitää olla sama nimike (label) eri ontologioissa. Näiden yhteisesti käytössä olevien minimivaatimusten avulla ontologian soveltaminen eri käyttötarkoituksissa tulee joustavammaksi. KOKO-järjestelmässä kuitenkin pyritään edistämään ja tukemaan yhteentoimivien ja jaettujen käsitteiden ja rakenteiden käyttöä. Jos erilaisiin ratkaisuihin ollaan päätyvässä, pyritään se tekemään vähintään tietoisena toisista ja kertomalla eksplisiittisesti, miten omat ratkaisut kuvautuvat naapuriontologioille ala/yläluokka- ja ekvivalenssisuhteilla.

Ontologioiden yksityiskohtaisemmat semanttiset ratkaisut ja sitoumukset jätetään eri sovellusten harkittaviksi. Ontologioissa voi yhteisten periaatteiden ja ominaisuuksien ulkopuolella olla vapaasti omia lisäyksiä, kuhan ne eivät ole ristiriidassa yhteisten pelisääntöjen kanssa. Esimerkiksi kysymys siitä, mitkä käsitteet pitäisi mallintaa luokkina ja mitkä yksilöinä, ei ole aina välttämättä selvää. Pitäisikö esimerkiksi museoesineen materiaalina mainittu ”villa” olla viittaus yleisesti luokkaan vain yksilöön luokasta villa, koska esine on kuitenkin valmistettu juuri tietystä villasta? Ongelma on ratkaistu KOKO-ontologiassa sitoutumatta kumpaakaan tulkintaa: kaikki käsitteet ovat luokkia ja sovelluksen asiaksi jää päättää se, milloin luokista on tarpeen luoda yksilöitä.

2.2 Ontologiat

FinnONTO-hankkeessa kehitettyjä tietämyksen organisoinnin järjestelmiä (knowledge organization system, KOS²²) voidaan jäsentää seuraaviin kategorioihin: ontologiat, luokitukset, tesauukset (hierarkkiset sanastot), asiasanastot ja sana/käsitelistat. Tarkka rajanveto on monasti epäselvää ja vaikeaa eikä tarkoituksenmukaistakaan.

Ontologioita voidaan ryhmitellä edelleen luokkaontologioihin ja yksilöontologioihin. Luokkaontologiat sisältävät asiasanastojen tapaan yleiskäsitteitä luokkina (class), kuten ”kissat”, ”kansantuote” tai ”kylät”. Luokka edustaa siihen kuuluvia yksilöitä (individual, instance) ja määrittelee näille mahdollisesti yhteisiä ominaisuuksia. Esimerkiksi kaupungit-luokan yksilöitä voivat olla Helsinki tai New York, ja kaupungin sijainti voidaan ilmaista koordinaateilla ja sitä laajempi hallinnollinen alue valtio-yksilöön viittaavalla ominaisuudella.

Seuraavassa käsitellään FinnONTO:n ontologioiksi luokiteltavia tietämusrakenteita, erityisesti kansallisen semanttisen webin ytimeksi ehdotettavaa KOKO-ontologiajärjestelmää.

2.2.1 KOKO-luokkaontologiat

FinnONTO:n luokkaontologiat on pääsääntöisesti muodostettu maassamme käytössä olevista asiasanastoista. Nämä on ensin muunnettu semanttisen webin OWL-muotoon ja siirretty Protege-editoriin. Tämän jälkeen sanaston hierarkiat on täydennetty ja semanttiset suhteet (LT laajempi termi, ST suppeampi termi, RT rinnakkaistermi) täsmennetty ja korjattu käsityönä. Lopputulos on semantiikaltaan RDF(S)-suosituksen mukainen yksinkertainen luokkahierarkia, jonka luokista luotavien yksilöiden pitää olla kaikkien yläluokkiensa yksilöitä (transitive closure).

Kansallisista luokkaontologioista keskeisin on kansalliskirjaston Yleisen suomalaisen asiasanaston YSA pohjalta muodostettu Yleinen suomalainen ontologia YSO. YSA on nykyisin yli 25 000 termin sanasto, joka on laajassa käytössä maassamme eri aloilla sisällönkuvailussa. YSO muodostaa FinnONTO-infrastruktuurin yhteisen yläontologian (upper ontology). Ajatuksena on, että muut alakohtaiset ontologiat täsmenävät YSO:n käsitteistöä omilla alueillaan. Esimerkiksi maa- ja metsätalouden ontologia AFO luettelee YSO:n käsitteen ”madot” alaluokkia, kuten ”sukkulamadot” ja ”ankeroiset”.

2.2.2 KOKO-ontologioiden skeema ja URI-käytäntö

YSO-ontologialle määritelty skeema on otettu yleiseksi malliksi kaikissa KOKO-järjestelmän ontologioissa. Siinä luokkahierarkia koostuu alla luetelluista käsitetyypeistä:

- **Indeksointikäsitteet** ovat nimensä mukaisesti sisällönkuvailussa usein asiasanojen tapaan käytettäviä käsitteitä. Niistä jokaisella on pysyvä URI-tunniste, joka yksilöi merkityksen.

²² http://www.db.dk/bh/lifeboat_ko/concepts/knowledge_organization_systems.htm

- **Ryhmittelevät käsitteet** (group concept) jäsentävät ontologian alaluokkia ihmisen kannalta mielekkäisiin, usein vaihtoehtoisiin ryhmiin. Esimerkiksi ryhmittelevän käsitteen ”aine olomuodon mukaan” alta löytyvät indeksointikäsitteet ”kaasut”, ”kiinteät aineet” ja ”nesteet”. Ryhmitteleviä käsitteitä ei ole tarkoitettu käytettäväksi indeksointitunnisteina konetta varten vaan käsitteiden ryhmittelyyn ihmistä varten.
- **Koostekäsitteet** ovat asiasanastosta periytyviä monimerkityksiä käsitteitä, joita voidaan käyttää indeksoinnissa. Esimerkiksi ”lapset (kooste)” vastaa YSA-sanaston termiä ”lapset”, joka on ontologisesti monimerkityksinen merkityksillä ”lapset (perheenjäsenet)”, ”lapset (ikäryhmänä)”, ”lapset (sosiaalinen ryhmä)”. Ontologisoinnin yhteydessä koostekäsitteet on jaettu osamerkityksiinsä, jotka on sitten sijoitettu omille paikoille ontologian eri haaroihin. Alkuperäistä käsitettä ei kuitenkaan ole mahdollista sijoittaa sellaisenaan mihinkään paikkaan hierarkiassa. Jos monimerkityksinen käsite on kuitenkin hyödyllinen indeksoinnissa, voidaan se sijoittaa ontologiaan hierarkian ulkopuolella koostekäsitteenä, jonka merkitys määritellään sen osamerkitysten joukkona. Esimerkiksi ”lapset (kooste)” tarkoittaa samanaikaisesti kaikkia kolmea merkitystään. Sitä voi olla joskus helpompi käyttää ilman tarvetta eritellä ja luetella eri merkityksiä tarkemmin. Koostekäsitteet toimivat samalla peilauksina monimerkityksisten asiasanojen ja tarkennettujen ontologisten käsitteiden välillä.

Eri käsitetyypeille voidaan asettaa vain niille ominaisia ominaisuuksia. Indeksointikäsitteen keskeisiä ominaisuuksia ovat erikieliset nimikkeet, viittaukset yläluokkiin ja ekvivalentteihin luokkiin omassa tai toisissa ontologioissa sekä viittaukset lähikäsitteisiin (RT) omassa ontologiassa. Ryhmittelevän käsitteen ominaisuuksia ovat erikieliset nimikkeet sekä yläkäsitteet. Koostekäsitteen skeemassa taas tarvitaan mm. viittaus alkuperäiseen asiasanaan sekä viittaukset indeksointikäsitteisiin, joiden merkitysten yhdistelmää koostekäsite tarkoittaa.

Jokaiselle käsitteelle on annettu nimikkeen ohella yksilöivä pysyvä URI-tunniste HTTP-skeeman mukaan, mikä on vakiintunut semanttisen webin alalla käytännöksi. Nimet ovat muotoa:

http://www.yso.fi/onto/ontologian_nimi/käsitteen_tunniste

Esimekiksi YSO-ontologian käsite ”nimet” on saanut tunnisteiden:

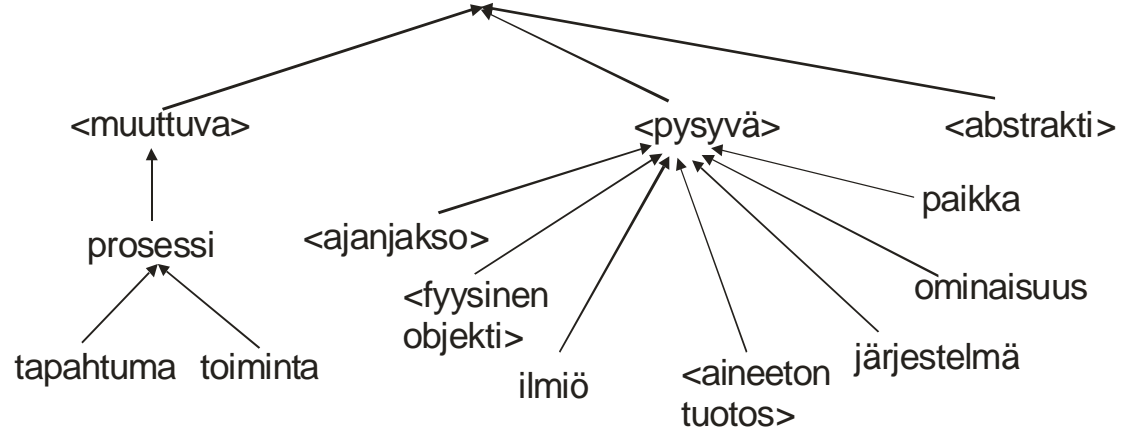
<http://www.yso.fi/onto/yso/p1638>

Käsitteiden tunnisteet ovat kirjain-numeroyhdistelmiä ilman mitään semantiikka, mikä edesauttaa niiden käyttöä kielirajojen yli ja pysyvyyttä. Sama paikallinen käsitteen tunniste kuten p1638 voi olla käytössä useissa eri ontologioissa.

Keskeisiä syitä HTTP-skeeman käyttöön URI-tunnisteissa ovat: 1) nimipalvelu (Domain Name System DNS) on jo maailmanlaajuisessa käytössä webissä ja mahdollistaa 1) domain-nimen perusteella tapahtuvan yksilöinnin sekä 2) URI:in

liittyvän tietoverkon (RDF-määritelmä) hakemisen webistä, mikä on mm. yksi Tim Berners-Leen ja W3C:n kiteyttämistä Linked Data ajattelumallin periaatteista²³. Muut eri yhteyksissä esitetyt nimeämisjärjestelmät ja URI-skeemat, kuten URN ja OID, ovat näissä suhteissa ongelmallisia.

2.2.3 KOKO-ontologian rakenne ja sisältö



Kuva 2.1. YSO-ontologian ylimmät luokat. Ryhmittelevät käsitteet on merkitty kulmasuluilla.

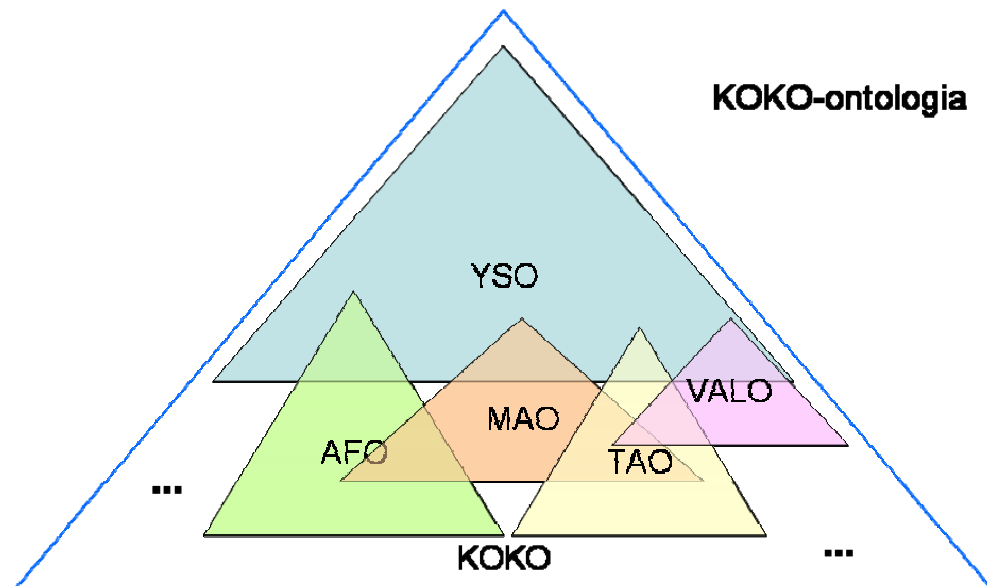
YSO:n käsitteet on jaettu DOLCE-järjestelmän (Gangemi et al., 2002) perusteella kolmeen pääluokkaan (kuva 2.1): ”muuttuva” (perdurant), joka sisältää tapahtumiin ja toimintoihin liittyvät käsitteet, ”pysyvä” (endurant), joka sisältää mm. fyysisiin objekteihin, ajanjaksoihin ja paikkoihin liittyviä luokkia sekä ”abstrakti” (abstract), joka sisältää lähinnä suureisiin ja yksiköihin liittyviä käsitteitä. Prosessien erottaminen muusta käsitteistöstä on yksi keskeinen periaate tietämyksen esittämisessä tekoälyssä (Sowa, 2000) ja mahdollistaa mm. Kulttuurisammon tapahtumakeskeisen sisältöjen mallintamisen ja skeemaintegraatiota (Ruotsalo, Hyvönen, 2008).

Kuva 2.2 havainnollistaa YSO:n suhdetta sitä tarkentaviin alaontologioihin. Alaontologiat kehitetään lähtökohtaisesti YSO-ontologian laajenuksina peilaamalla niiden leikkaavat käsitteet YSO:n vastaaviin. Sääntönä on, että jokaisella alaontologian käsitteellä pitää olla vähintään yksi ekvivalentti tai yläkäsite YSO:ssa. Näin rakentuu yhtenäinen hierarkia ja semanttinen yhteys alaontologioiden ja YSO:n välille. Samalla syntyy silta YSO:n kautta myös eri alaontologioiden välille. Alaontologiat voivat leikata toisiaan myös suoraan, ei vain YSO:n kautta. Tällöin voidaan muodostaa myös suoria peilauksia eri alaontologioiden välille.

Käytännössä ontologioiden peilaus on FinnONTO:ssa tehty ensin alustavasti tietokoneella termien nimien perusteella. Tämän jälkeen rakenne on tarkistettu, korjattu ja muodostettu käsityönä Protege-editorin avulla. Kokemustemme mukaan alaontologian muodostaminen peilaamalla se YSO:oon helpottaa ontologiatyötä

²³ <http://www.w3.org/DesignIssues/LinkedData.html>

merkittävästi, kun YSO:n yhteydessä jo tehtyä hierarkisointityötä voidaan yleensä hyödyntää sellaisenaan.



Kuva 2.2. Yhteisöllinen kokonaisontologia KOKO koostuu yläontologiasta YSO ja sitä tarkentavista alaontologioista, joilla on yhteisiä käsitteitä YSO:n ja toisten alaontologioiden kanssa. Ontologioiden keskinäiset leikkaukset kuvassa ovat vain viitteellisiä ja tarkoitettu yleiskuvan havainnollistukseksi.

Eri alaontologioiden kehitysversiot YSO:n (ja mahdollisesti muiden ontologioiden) suhteen peilattuna löytyvät ONKI-palvelusta omasta osiostaan. Tätä kirjoittaessa KOKO-järjestelmään on integroitu ensimmäiset versiot taulukossa 2.1 esitellyistä ontologioista, yhteensä yli 37 000 luokkakäsitteen ontologiat. Taulukossa 2.2. on lueteltu ontologioiden peilauksen tuloksena löydettyjen ja muodostettujen ekvivalenssisuhteiden määriä. Lisäksi ontologioiden käsitteiden välille on lisätty yläluokkasuhteita (subclassOf).

Nähtäväksi jää millaiseksi rajanveto ja työnjako eri ontologioiden välille lopulta muotoutuu, kannattaako esimerkiksi pienempiä ontologioita missä määrin yhdistää. FinnONTO:n yleisenä ratkaisumallina kehitystyölle on käynnistää Web 2.0 -henkinen yhteisöllinen kehitysprosessi eri ontologioiden kehittäjäryhmien välillä, jota tuetaan ONKI-palveluiden kautta.

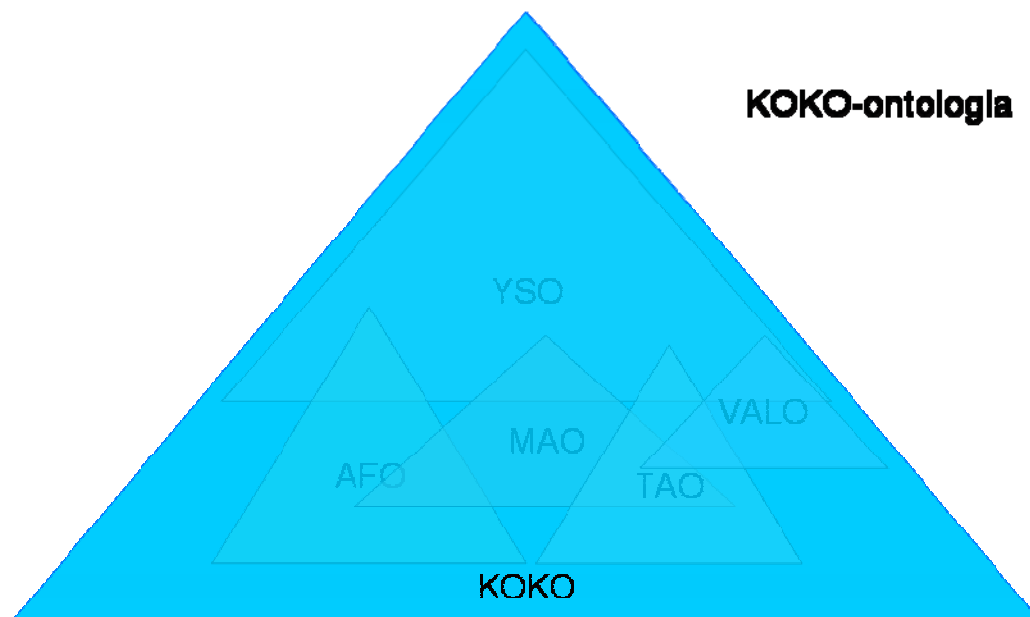
Alakohtaiset yhdistelmäontologiat on tarkoitettu vain ontologioiden kehittäjille. Sisällön kuvaajille ja halupalveluiden käyttäjille KOKO-ontologia näyttäytyy yhtenäisenä ontologiana kuvan 2.3 havainnollistamalla tavalla. Annotoinneissa suositellaan käytettäväksi KOKO-ontologiassa olevia URI-tunnisteita ja rakenteita. Indeksointi hyvinkin tarkoilla käsitteillä on mahdollista ilman että yhteys yleisempiin yläkäsitteisiin katkeaisi. Tämä edesauttaa eri tarkkuustasoilla kuvailtujen aineistojen yhteentoimivuutta. Kokonaisuudesta voidaan myös tarpeen mukaan poistaa tietyssä sovelluksessa tarpeettomia alaontologioita.

Nimi	Ontologian ala	Pohjana oleva asiasanasto	Koko
YSO	Yleinen suomalainen ontologia Yleiskäsitteet, monialainen	Yleinen suomalainen asiasanasto YSA. Kansalliskirjasto. http://vesa.lib.helsinki.fi/	20 600
MAO	Museoalan ontologia	Museoalan asiasanasto MASA. Museovirasto (Leskinen, 1997).	6800
AFO	Maa- ja metsätalouden ontologia. Maa ja metsätalous	Agriforest tesaurus. Viikin tiedekirjasto.	5500
TAO	Taideteollisuusalan ontologia	Käsi- ja taideteollisuuden asiasanasto (Krogerus et al., 1997)	2600
VALO	Valokuvausalan ontologia	Valokuvakirjallisuuden asiasanasto (Ulla Karttunen). Valokuvaustekniikan asiasanasto. Suomen valokuvataiteen museo.	1900

Taulukko 2.1. KOKO-ontologian osaontologioita. Uusia alaontologioita liitetään järjestelmään muuntamalla asiasanastoja ontologioiksi.

Leikkaavat ontologiat	Yhteisten ekvivalenttien käsitteiden määrä
YSO + TAO	1990
YSO + MAO	2330
YSO + VALO	950
MAO + TAO	1190

Taulukko 2.2. KOKO ontologian osaontologiat jakavat runsaasti yhteisiä ekvivalentteja käsitteitä YSO:n kanssa ja keskenään.



Kuva 2.3. KOKO-ontologia näkyy indeksoijalle ja tiedon hakijalle yhtenäisenä kokonaisuutena (vrt. kuvan 2.2 kehittäjänäkymä).

2.2.4 KOKO-yksilöontologiat

FinnONTO:ssa ensimmäisenä kehitettyjä yksilöontologioita ovat paikkaontologiat ja toimijaontologiat (henkilöt ja organisaatiot), joita on tarvittu kulttuurialan portaaleissa. Paikkojen osalta keskeinen ontologia on Suomalainen paikkaontologia SUO (Henriksson et al., 2008). Se koostuu n. 700 paikkaluokasta, joista on sitten muodostettu suuri joukko yksilöitä Maanmittauslaitoksen paikannimirekisterin (PNR) pohjalta (n. 800 000 kotimaista paikkaa) ja kansainvälisistä paikka-aineistoista (GEONet Names Server GNS, miljoonia paikkoja). Suomen historiallisista kunnista ja niiden välisistä muutoksista on laadittu erillinen Suomen historiallinen paikkaontologia SAPO (Kauppinen, Hyvönen, 2007; Kauppinen et al., 2008), joka on yhdistetty SUO:hon.

Henkilöiden ja organisaatioiden osalta FinnONTOssa käytetään Getty-säätiön Union List of Artis Names (ULAN) käsitelistää, jota on tarkoitus rikastaa kotimaisilla aineistoilla, kuten Suomalaisen kansallisbiografian tiedoilla, kotimaisilla auktoriteettitietokannoilla ja Kulttuurisampo-portaalin kokoelmien metatiedoissa esiintyvillä toimijoilla.

2.2.5 Muita ontologioita

KOKO-ontologioiden ohella hankkeessa on kehitetty ja tuotettu lukuisia muitakin ontologioita, joita on sitten peilattu YSO:on (ja sitä kautta muihin alaontologioihin) eri tavoilla. Näistä mainittakoon seuraavat:

- **Iconclass**²⁴ on Hollannissa kehitetty laaja käsitteistö taiteen ikonograafiseen kuvailuun, joka on osittain suomennettu Valtion taidemuseon toimesta. Systeemi linkitettiin tätä kautta YSO-ontologiaan (Haaramo, 2006).
- **European multilingual thesaurus on health promotion (HPMULTI)**²⁵ on 12-kielinen eurooppalainen terveyden edistämisen tesaurus, jossa on n. 1200 käsitettä. Ontologiaa käytetään TerveSuomi-portaalissa.
- **Medical Subject Headings MeSH**²⁶ on kansainvälinen monikielinen lääketieteen sanasto, jota käytetään indeksointiin mm. lääketieteellisissä kirjastoissa. Sen hierarkiassa on n. 24 400 käsitettä. Sanaston suomennos FinMeSH ja saatu käyttöön lääkäreiden Duodecim-järjestöltä; ruotsinnoos taas on saatu Ruotsista. Systeemi on peilattu soveltuvin osin YSO:oon ja edellä mainittuun HPMULTI-ontologiaan (Suominen, 2008). Kokonaisuus on käytössä TerveSuomi-portaalissa.
- **Helsingin kaupunginkirjaston (HKK) luokitusjärjestelmä HKLJ**²⁷ on HKK:ssa käytetty laaja järjestelmä, jonka luokat on kuvailtu joukolla YSA:n termejä. Systemistä on muodostettu oma, YSO:on linkittyvä ontologinen rakenne. HKLJ:ssä käytetyt YSA-termit on käännetty HKK:n toimesta

²⁴ <http://www.iconclass.nl/>

²⁵ <http://www.hpmulti.net/>

²⁶ <http://www.nlm.nih.gov/mesh/>

²⁷ <http://hklj.kirjastot.fi/>

englanniksi ja saatu tätä kautta liitettyä osaksi YSO:a – alkuperäinen Kansalliskirjasto YSA-sanasto on vain kaksikielinen. Kokonaisuutta on käytetty Opas-järjestelmässä (Vehviläinen 2006, Vehviläinen et al., 2008), josta on tuoteistettu versio Kirjasto.fi-portaaliin.²⁸

2.2.6 Asiasanastoja SKOS-muodossa

Lisäksi FinnONTO:ssa on muutettu asiasanastoja SKOS-muotoon ilman erillistä ontologisointia, ja julkaistu niitä ONKI palveluna taulukon 2.2 mukaisesti.

Nimi	Sanasto ja alka	Lähde	Koko
YSA	Yleinen suomalainen asiasanasto	Kansalliskirjasto, http://vesa.lib.helsinki.fi/	25600
NVAS	Valtioneuvoston asiasanasto	Valtioneuvoston kanslia	6300
Kaunokki	Kaunokirjallisuuden asiasanasto	http://www.kirjastot.fi/	4400
MESA	Merenkulkualan asiasanasto	Merenkululaitos	1500
MUSA-CILLA	Musiikin asiasanasto	Kansalliskirjasto, http://vesa.lib.helsinki.fi/musa	900

Taulukko 2. ONKI-palveluna julkaistuja SKOS-muotoisia asiasanastoja.

3 Kansallinen ontologiakirjastopalvelu ONKI

FinnONTO-vision mukaan pelkkä ontologioiden olemassaolo ei riitä kansalliseksi sisältöinfrastruktuuriksi, vaan lisäksi tarvitaan pysyviä palveluita, jotka mahdollistavat ontologioiden hyödyntämisen kustannustehokkaasti asiakasjärjestelmissä ja tukevat järjestelmän ylläpitoa ja kehittämistä.

3.1 Koordinoitu ontologiakehitysprosessi

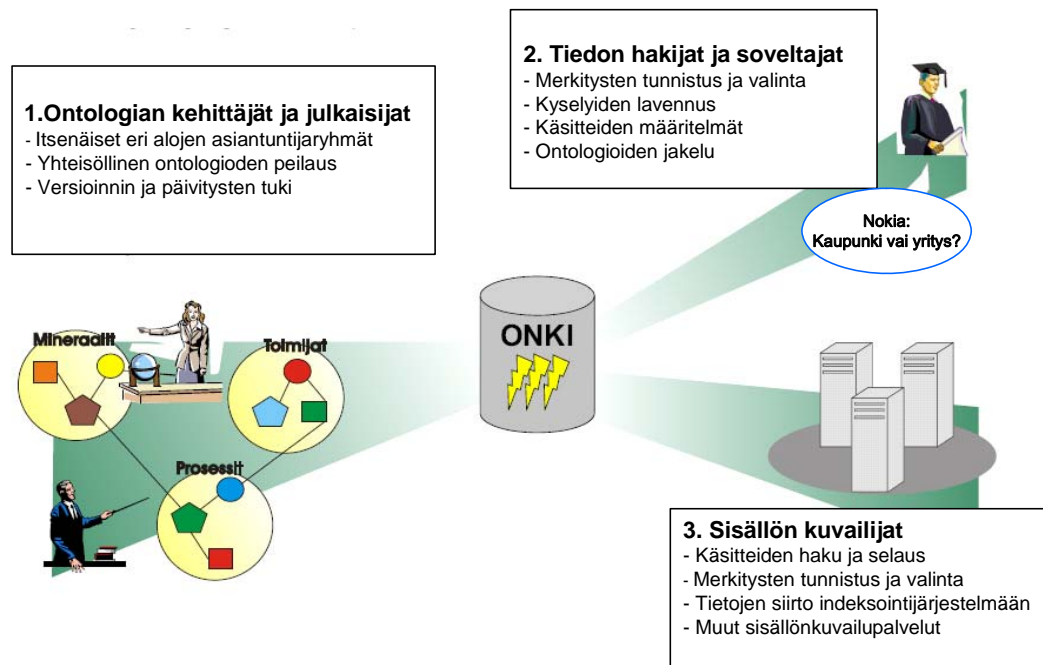
Ontologiat kehittyvät, lisääntyvät, yhdistyvät ja elävät ajan kuluessa, ja prosessit muuttuvat niitä kehittävien tahojen muuttuessa. FinnONTO-projektin ehdotus on käynnistää koordinoitu kansallinen ontologiakehitystyö nykyisten sanastojen kehittämishankkeiden luomalle perustalle.

Tällaisen prosessin käynnistäminen ja ylläpito edellyttää:

1. Yhteistä **tahtotilaa** semanttisen webin sisältöinfrastruktuurin luomiseksi maahamme
2. Sanastoja kehittävien tahojen **organisointumista** yhteiseen tavoitteeseen pääsemiseksi ja työn koordinoitua
3. Riittävää ja jatkuvaa työn **resurssointia**

²⁸ <http://www.kirjastot.fi/fi-FI/tietopalvelu/>

4. Teknisiä kustannustehokkaita ratkaisuja, joilla prosessi saadaan toimimaan ja sen tuottamat ontologiat hyödynnettäviksi asiakassovelluksissa



Kuva 3.1. Kansallisen ontologiakirjastopalvelu ONKI:n palvelut eri käyttäjäryhmille.

FinnONTO:n tavoitteena on synnyttää tahtotila ja vahvistaa sitä käytännön demonstraatioiden avulla, kuten MuseoSuomi, Orava, Opas, Kulttuurisampo ja TerveSuomi²⁹. Ajatuksena on, että tahtotilan seurauksena syntyisi organisoitumista ja tarve resurssoinnille. Aluksi ontologiatyötä voidaan tehdä FinnONTO-hankkeen toimesta ja rahoituksella pilotti-hankkeina. Tavoitteena on siirtää työn koordinointi hankkeen päättyessä tutkimushankkeelta siihen sitoutuneille organisaatiolle, näillä näkymin vuoden 2009 tai 2010 lopussa. Luontevana koordinaattorina voisi jatkossa olla mm. YSA:n kehityksestä nykyisin vastaava Kansalliskirjasto, josta on vähitellen tullut FinnONTO:n suurin yksittäinen rahoittaja Tekesin jälkeen. Uudet velvoitteet luonnollisesti edellyttäisivät työn jatkuvaa rahoitusta. Sanastotyö on maassamme perinteisesti ollut niukasti resurssoitua, mutta tilanne parantunee semanttisen webin myötä selkeän hyödyntämispolun syntyessä sanasto- ja sisältötyöstä käytännön verkkosovelluksiin.

Keskeinen uusi elementti kansallisesti koordinoitussa ontologiaprosessissa ovat uudet tekniset mahdollisuudet, joilla prosessia voidaan tukea ja saada sen tulokset kustannustehokkaasti hyödynnettäviksi eri sovelluksissa. FinnONTO-hankkeen tekninen ratkaisumalli ja alusta tähän on Kansallinen ontologiapalvelu ONKI. ONKI tarjoaa palveluita kolmelle käyttäjäryhmälle kuvan 3.1. mukaisesti:

1. Ontologian kehittäjille ja julkaisijoille on nyt tarjolla keskitetty palvelu

²⁹ <http://www.seco.tkk.fi/applications/>

<http://www.yso.fi/>

jonka kautta standardimuotoiset ontologiat, luokituksen ja sanastot voidaan julkaista välittömästi ja käytännössä ”ilmaiseksi” ilman lisätyötä: tyypillisesti tarvitaan vain ONKI SKOS -palvelimen pieni konfigurointi ja käynnistäminen. Tämän lisäksi palvelu tulee tarjoamaan yhteisöllistä Web 2.0 –henkistä tukea oman ontologian peilaamiseksi YSO:on tai sen alaontologioihin kansallista KOKO-ontologiaa varten sekä tuen ontologioiden versiointia ja päivityksiä varten.

2. Tiedon hakijoille ja soveltajille tarjotaan palveluja mm. hakusanojen merkitysten erotteluun. Ajatuksena on, että jos esimerkiksi Google-tyyppiseen hakukenttään kirjoittaa sanan ”johtaminen”, niin kenttään kytketty ONKI-palvelu voi tarjota semanttisen täydennyksen (Hyvönen, Mäkelä, 2006) avulla vaihtoehtoja sanan eri merkityksille johtaminen (sähkötekniikka), johtaminen (musiikin), johtaminen (matematiikassa), johtaminen (liiketaloudessa), jolloin eri valinnat voivat johtaa aivan erilaisiin tulosjoukkoihin. Toinen peruspalvelutyyppi, johon ontologioita voidaan käyttää, ovat erilaiset kyselyiden ontologiset laajentamiset (term expansion). Jos esimerkiksi hakusovelluksessa kysytään EU-maiden asioista, voidaan kysely laventaa paikkaontologian avulla TAI-kyselyksi Suomesta, Sloveniasta, Ranskasta jne. ilman että kysyjän tarvitsee tietää tai luetella kaikki EU-maat nimeltä. Tällaisesta palvelusta on tekeillä ensimmäinen demonstraatio Metsämuseo Luston Kantapuu.fi palveluun opetusministeriön rahoituksella.

ONKI tarjoaa soveltajille valmiita, mahdollisimman korkean jalostusasteen toiminnallisia palveluita käyttöliittymätasolla Web 2.0 leijukkeina (widget) (AJAX-tekniologia) ja perinteisinä Web Service –palveluina. Asiakaspäässä tarvittavan koodin luomista varten on ONKI-palvelussa tarjolla helppokäyttöisiä interaktiivisia verkkopalveluja. Mikäli asiakas tarvitsee koko ontologian omaa sovellustaan varten, on sen ajantasaisen version lähdekoodi ladattavissa ONKI:sta vastaavaan tapaan kuin perinteisissä koodistopalvelimissa. ONKI-vision mukaan suurin etu kuitenkin saavutetaan valmiiden palveluiden avulla. ONKI-palvelun kautta on esimerkiksi mahdollista liittää kymmeniä ontologioita tai sanastoja omaan sovellukseen valmiina toiminnallisuutena leijuke-valitsemella, joka syntyy yhdellä rivillä Javascript-koodia.

3. Sisällön kuvailijoille ONKI tarjoaa palvelut indeksointikäsitteiden vaivattomaan hakuun semanttisen täydentäjän avulla, joka pyrkii arvaamaan jokaisen kirjoitetun merkin jälkeen, mitä käsitettä ollaan hakemassa. Monipuolisen selaimen avulla voidaan tarkastella käsitteen semanttista ympäristöä parhaan käsitteen merkityksen selvittämiseksi ja siihen liittyvien muiden käsitteiden tarkastelemiseksi. Semanttista selailua varten on tarjolla IRMA-väline (Sinkkilä, 2008; Sinkkilä et al., 2008), joka yhdistelee toisiinsa liittyviä käsitteitä ontologian eri puolilta - myös useamman eri ontologian välillä, Tavoitteena on helpottaa indeksoijan työtä parhaiden käsitevalintojen tekemissä.

3.2 ONKI Living Laboratory: miten julkaisen oman ontologiani?

ONKI ontologiakirjastopalvelu on Living Laboratory –ympäristö, joka takaa ONKI-palveluille pysyvän ylläpidon ja tuen ainakin FinnONTO-hankkeen jo varmistuneen

rahoituskauden loppuun saakka 31.12.2009. Vakaana aikeena on toiminnan vakinaistaminen pilottivaiheen jälkeen. Tavoitteena on, että loppukäyttäjät voisivat näin luottavaisin mielin testata ONKI-palveluita omissa järjestelmissään. ONKI-palvelut tarjotaan toistaiseksi Teknillisen korkeakoulun ja tulevan Aalto-yliopiston palvelinten kautta.

ONKI Living Laboratory tarjoaa kaikille halukkaille mahdollisuuden julkaista omia sanastojaan ja ontologioitaan ONKI-palvelussa lähtökohtaisesti maksutta. Proseduuri oman ontologian tai sanaston julkaisemiseksi on seuraava:

1. Julkaisija tekee vapaamuotoisen anomuksen ONKI-kehitystiimille aloitteesta. Siinä perustellaan, miksi ehdotettu käsiterakenne olisi syytä julkaista ONKI-palveluna, ketkä sitä käyttäisivät ja mihin, ja onko sanasto julkaistavissa palvelun suosittamalla Creative Commons -lisenssillä (mitä lähtökohtaisesti toivotaan).
2. ONKI tiimi arvioi hakemuksen ja tekee julkaisupäätöksen.
3. Julkaisija toimittaa tietämysrakenteen jossakin ONKI:n tukemassa standardiformaatissa ONKI-tiimille. Jos sanasto on erityisen arvokas kansallisesta näkökulmasta, FinnONTO voi kohdistaa omiakin resursseja sen muokkaamiseen.
4. Tietämysrakenne julkaistaan ja siihen voi kytkeytyä ONKI-palvelinten kautta.

3.3 ONKI-palvelimet

ONKI-palvelimista on kehitetty kolme eri versiota erilaisia aineistoja varten:

1. ONKI SKOS (Tuominen, 2008; Viljanen et al., 2008b) on tarkoitettu SKOS-muotoisten sanastojen ja yksinkertaisten OWL-ontologioiden julkaisemiseen.
2. ONKI Geo (Lindroos, 2008) on paikkatiedon ontologiapalvelin.
3. ONKI People on kehitetty henkilö- ja organisaatio -ontologioita varten.

Palvelimet tarjoavat:

- **ihmiskäyttäjälle** haku- ja selainliittymän (browser) ja valitsinkomponentin
- **sovelluksille** AJAX ja Web Service –rajapinnat

Seuraavassa esitellään näitä palveluita tarkemmin käytännön esimerkkien avulla.

3.3.1 ONKI valitsin ja selaimet: käyttöesimerkki

Ensin näytetään, miten ONKI valitsin (selector widget) voidaan ottaa käyttöön kuvitteellisessa asiakasovelluksessa (Viljanen et al., 2008a). Tavoitteena on kehittää verkkosovelluksissa yleisesti käytetty HTML-lomake metatietojen syöttämistä varten valokuvista. Oletetaan esimerkiksi, että halutaan annotoida Helsingin sanomissa julkaistu valokuva nuorten autoilijoiden kolarista Pyhäjärvellä 5.9.2008. Valokuvien

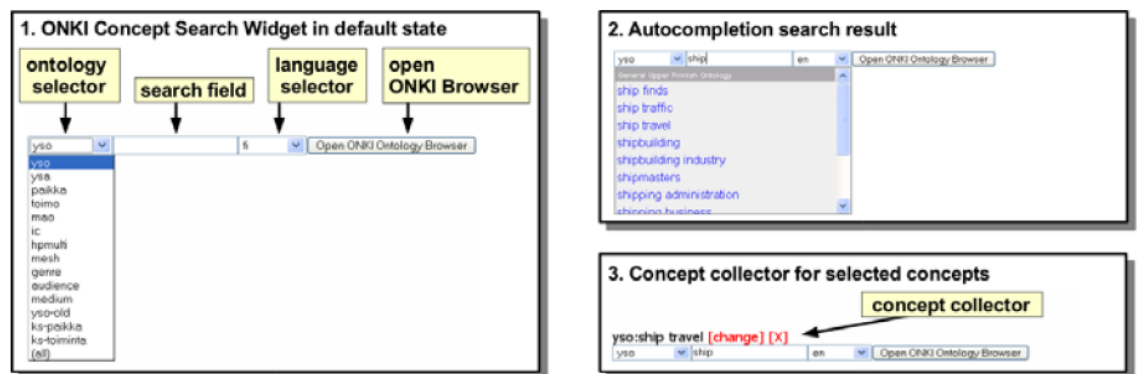
metadataskeema olkoon Dublin Core standardin mukainen ja käytössä on kenttä dc:subject, joissa kerrotaan kuvan aihe ("nuoret" merkityksessä nuoret aikuiset; "kolari" merkityksessä liikenneonnettomuus eikä Kolarin kunta), ja kenttä dc:coverage, jossa kerrotaan, missä kuva on otettu (yksi tietty Pyhäjärvi Suomen lukuisista Pyhäjärivistä paikkana eikä henkilönä).

ONKI voidaan kytkeä luettelointijärjestelmään luomalla metadatakentille - tässä kuvan aiheelle ja paikalle - syötekentät ja kytkemällä ne sopiviin ONKI-palvelimiin. Aiheen (asiasanojen) osalta syötekenttä voidaan kytkeä useampaakin ontologiaan ja paikkatiedon osalta paikkaontologiaan SUO. Tämä onnistuu lisäämällä ensin HTML-lomakkeen HEAD osioon viittaus ONKI-palvelussa olevaan Javascript kirjastoon vakiorivillä:

```
<script type="text/javascript" src="http://www.yso.fi/onki.js"></script>
```

Tämän jälkeen jokaiselle syötekentällä (input field) annetaan kytkentä valittuun ontologiaan. Esimerkiksi lomakkeen asiasanakentälle dc:subject voidaan muodostaa input-kenttä ja liittää se ONKI-ontologioihin seuraavalla rivillä:

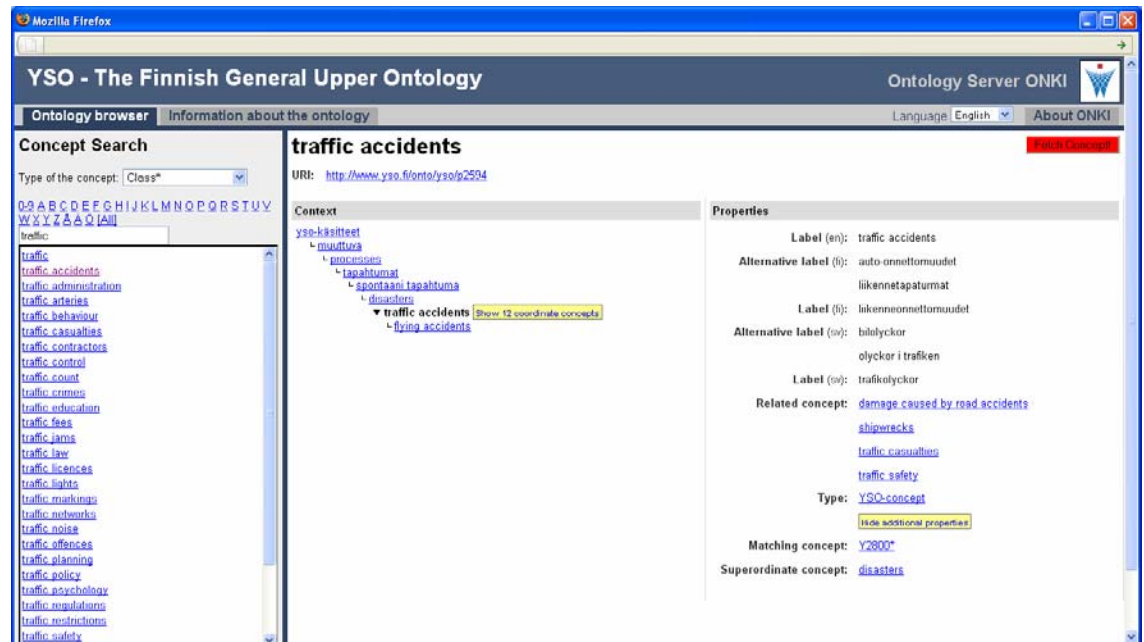
```
<input id="dc:subject" onkeyup="onki['yso'].search()"/>
```



Kuva 3.2. ONKI-valitsin (widget).

Tuloksena syntyy lomakkeelle kuvan 3.2 mukainen syötekenttä (vasemmalla), joka jakautuu neljään osaan vasemmalta oikealle: 1) Vasemalla olevasta ontologian valitsimesta voidaan valita käytettävä ontologia alavetovalikosta; oletusarvona on input-kentän parametrina annettu YSO. Haku voidaan kohdistaa myös samanaikaisesti kaikkiin eri ontologioihin. 2) Hakukenttään voidaan hakea käsitteitä valitu(i)sta ontologi(ois)ta semanttisella täydennyksellä (kuvassa oikealla ylhäällä). Kun valinta on tehty, käsitteeseen liittyvä tunniste ja nimike siirtyvät lomakkeelle hidden-kentän arvoksi, ja nimike näkyy käyttäjälle syötekentän yläpuolella käsitekeräimessä (collector) (oikealla alla kuvassa). Valittu käsite voidaan poistaa tai sitä muuttaa keräimen linkkejä painamalla. 3) kielivalitsemista voidaan valita käytettävä kieli – ONKI:n useimmat ontologiat ovat monikielisiä (jopa 12 eri kieltä). 4) Open ONKI Browser –napista voi käynnistää ONKI-selaimen, jolloin uuteen ikkunaan avautuu täysi selainäkymä ontologian käsitteiden hakuja ja selailua varten

”Nouda käsite” –painikkeella varustettua. Painiketta painamalla valittu käsite siirtyy selainikkunasta tietoinen lomakkeelle samalla tavalla kuin semanttista täydentäjää käyttämällä. Kaikki tämä keskitetyn palvelimen ajantasainen toiminnallisuus saadaan sovellukseen vain yhdellä rivillä koodia HTML -sivulla. Kuvassa 3.3. on esitetty tilanne, jossa käyttäjä on löytänyt ontologiasta liikenneonnettomuudet-käsitteen, nyt englannin kielellä. Nouda käsite –painike näkyy ikkunan oikeassa ylänurkassa.

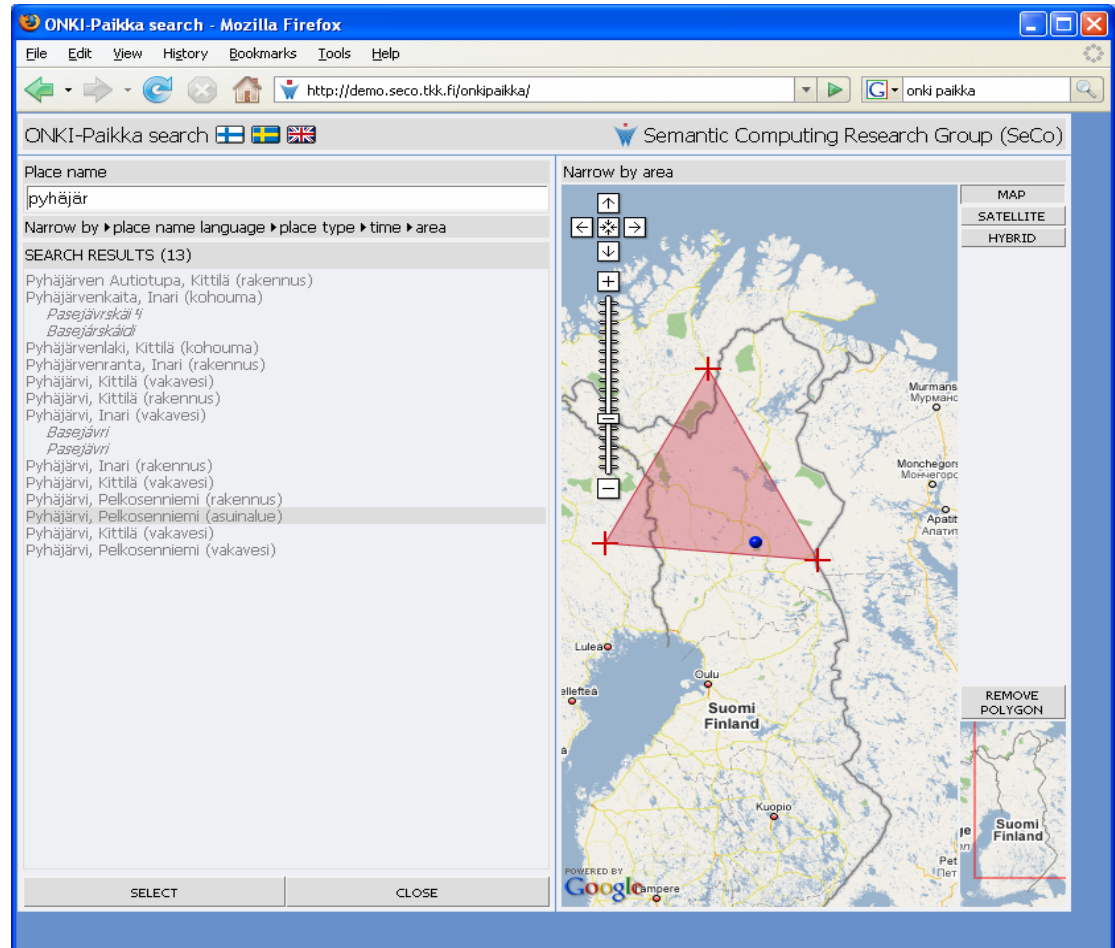


Kuva 3.3. ONKI selain. YSO-ontologiasta on selaimella löydetty käsite liikenneonnettomuudet, kielenä englanti. Nouda käsite –nappia painamalla (oikealla ylhäällä) käsitteen nimike ja URI siirtyvät verkosta asiakassovellukseen..

Vastaavalla tavalla voidaan luoda input-kenttä paikkatiedon lisäämiseksi kuvaan. Tällöin input-kentästä voidaan avata ONKI Geo -palvelimen karttapalvelu (Lindroos, 2008). Kirjoittamalla ”Pyhäjärvi” käyttäjä voi valita Suomen 49 Pyhäjärvestä oikean tulinnan tai vaihtoehtoisesti kertoa onnettomuuden tarkan paikan kartalla, jolloin lomakkeelle saadaan koordinaattitiedot. Kuvassa 3.4 näkyy tilanne, jossa käyttäjä on Pyhäjärvi-nimen ohella rajannut hakua piirtämällä Pohjois-Suomen yllä rajatun alueen. ONKI Geo -palvelu on ONKI valitsimen tapaan itsekin ns. mash-up-sovellus, joka hyödyntää Google Maps -palvelua karttanäyttöjen osalta.

Esimerkki osoittaa, että ONKI:n avulla voidaan helpottaa merkittävästi vaikkapa luettelointisovelluksen ohjelmointia, jos halutaan saada käyttöön ajantasaiset sanasto- ja paikkatietopalvelut. Sovelluksen oma asia on päättää, miten ONKI-palvelimelta saadut tunnistet, nimikkeet ja koordinaatit talletetaan esimerkiksi tietokantaan ja hyödynnetään. Yksilöivien käsitetunnisteiden kautta järjestelmällä kuvailut valokuvat voidaan automaattisesti kytkeä oikealla tavalla ilman sekaannuksia kansallisen semanttisen webin muihin sisältöihin.

ONKI-valitsinta voidaan sovellustilanteen tarpeiden mukaan konfiguroida monipuolisesti, esimerkiksi poistaa ontologian valintamahdollisuus, jos halutaan käyttää vain tiettyä ontologiaa, tai asettaa oletusarvo käytettävälle kielelle. Koodirivin automaattista muodostamista varten ONKI-palvelussa on erityinen interaktiivinen koodigeneraattori, jolla erilaisia valitsimia on helppo luoda ja kokeilla.



Kuva 3.4. ONKI Geo -palvelu on hakusovellus, joka yhdistää SUO paikkaontologian ja Google Maps -palvelun. Kuvassa käyttäjä etsii Suomen Pyhäjärvi-nimisiä paikkoja ja on rajoittanut hakuja Pohjois-Suomessa olevaan alueeseen. Tumma piste ilmaisee valitun paikan kartalla.

FinnONTO-ontologioita ja palveluita on alettu hyödyntää ensimmäisissä ”perinteisissä” verkkosovelluksissa, kuten Kirjasto.fi-portaalin Kysy kirjastohoitajalta -palvelussa³⁰ ja yhteisöllisessä Tilkut-portaalissa³¹.

3.3.2 ONKI Web Service: toiminta ja käyttö

ONKI palvelut voidaan ottaa käyttöön myös perinteisen Web Service -rajapinnan kautta. Tällöin asiakasovellus kutsuu ONKI-palvelimen WSDL-rajapintaa HTTP- ja

³⁰ <http://www.kirjastot.fi/fi-FI/tietopalvelu/>

³¹ <http://owela.vtt.fi/tilkut/>

SOAP-protokollia käyttäen. SOAP-kutsujen ja WSDL-rajapinnan määritysten generointiin on tarjolla oma interaktiivinen palvelu.

4 Yhteenveto

Semanttinen web perustuu yhteisiin sopimuksiin ontologioista ja metadatan esitystavoista. FinnONTO-hankkeessa on kehitetty malli ja prototyyppi kansallisen ontologiainfrastruktuurin rakentamiseksi maahamme. Sen tärkeimpänä tavoitteena on mahdollistaa yhteentoimivien ja älykkäiden verkkosovellusten kehittäminen maassamme kustannustehokkaasti semanttisen webin aikakaudella. Infrastruktuurin keskeisimmät ydinkomponentit ovat sisällön kuvailussa tarvittavat laajat ydinontologiat sekä palvelut, joilla nämä voidaan kytkeä osaksi asiakassovelluksia, ja joiden avulla ontologioita voidaan jatkossa kehittää ja ylläpitää. FinnONTO-infrastruktuurin ydinkomponentit on otettu avoimeen Living Laboratory -käyttöön Kansallisena ontologiakirjastopalveluna ONKI.

Kokemusten saamiseksi ja FinnONTO:n ratkaisumallin arvioimiseksi kaikki palaute järjestelmän ratkaisuihin ja toiminnasta otetaan kiitollisuudella vastaan. On selvää että vasta ensimmäiset askeleet pitkällä tiellä on nyt otettu.

Jos käytännön kokemukset FinnONTO-mallista osoittautuvat toimiviksi, seuraavana tavoitteena on toiminnan vakiinnuttaminen osana tietoyhteiskunnan muuta infrastruktuuria.

Kehitystiimi

FinnONTO-hankkeen ontologiajärjestelmän ja ONKI-ontologiapalvelun kehittäjäryhmään Semanttisen laskennan tutkimusryhmässä ovat osallistuneet eri vaiheissa lukuisat eri henkilöt Eero Hyvösen johdolla. KOKO-ontologioiden sisältötyötä vetää Katri Seppälä (YSO ym.) ja hankkeen ontologiatyöhön ovat osallistuneet eri vaiheissa mm. Anu Ylisalmi (YSO), Mirja Anttila (YSO), Eeva-Liisa Leppänen (YSO), Suvi Kettula (MAO), Päivi Lipsanen, Kimmo Koskinen (AFO), Satu Savia (VALO), Anne Isomursu (VALO), Ellen Karhulampi (TAO), Eetu Mäkelä (KOKO), Reetta Sinkkilä (muunnokset), Matias Frosterus (muunnokset), Tuukka Ruotsalo (yläontologiat) ja Osma Suominen (terveysalan ontologiat). Paikkaontologioita ovat kehittäneet Tomi Kauppinen (SAPO, SUO), Riikka Henriksson (SUO), Jari Väätäinen (SAPO) ja Robin Lindroos (SAPO ja SUO), ja toimijaontologioita Jussi Kurki sekä MuseoSuomen ja Kulttuurisammon kehittäjät. Kansallisen ontologiakirjastopalvelun kehitystä koordinoi Kim Viljanen ja ONKI SKOS selaimen ONKI-palveluineen kehitti Jouni Tuominen (Tuominen, 2008) aiemman Ville Komulaisen prototyypin (2007) pohjalta. ONKI Geon pääkehittäjä oli Robin Lindroos (2008) ja ONKI People -palvelun Jussi Kurki.

FinnONTO 2.0 -hanketta rahoittavat yritykset/organisaatiot sekä näiden edustajat projektin johtoryhmässä v. 2008 alkupuolella on esitetty taulukossa 4.1. Tutkimusosapuolina ovat Teknillinen korkeakoulu, Helsingin yliopisto ja Tampereen yliopisto. Hanke on pääosin Tekesin rahoittama. Kiitokset kaikille yhteistyöstä.

	Yritys / organisaatio	Jäsen	Varajäsen
1	AAC Global Oy	Janne Nevasuo	Arti Sinkkonen
2	Antikvaria-ryhmä	Merja Heiskanen	Hannu Mustonen
3	CelAmanzi Oy	Jyri Wuorisalo	Ville Luukkanen
4	Connexor Oy	Timo Järvinen	Pasi Tapanainen
5	Coronaria Oy	Teppo Linden	Teemu Puhakka
6	Eduskunta	Olli Mustajärvi	Antti Rautava
7	Ego Beta Oy	Esa Rauhala	
8	Elisa Oyj	Aimo Maanavilja	Mika Laurell
9	Espoon kaupungin museo	Liisa Ropponen	Sten Björkman
10	Geol. tutkimusk. GTK	Olli Rantala	Maija Pennanen
11	Hgin kaupunginkirjasto	Erkki Lounasvuori	Matti Sarmela
12	Hgin yliopiston Viikin tiedekirjasto	Kimmo Koskinen	Liisa Siipilehto
13	Kansallisarkisto	Markku Nenonen	
14	Kansalliskirjasto	Juha Hakala	Laila Heinemann
15	Kansanterveyslaitos	Eija Hukka	Jaason Haapakoski
16	Koulutuskeskus Dipoli	Lisbeth Edlund	
17	Maa- ja mesätalousministeriö	Antti Vertanen	
18	Maanmittauslaitos	Reino Ruotsalainen	Antti Saarikoski
19	M-Cult	Minna Tarkka	Maria Candia
20	Merenkulkulaitos	Jarmo Koistinen	
21	Museovirasto	Sirkka Valanto	Vesa Hongisto
22	Mysema Oy	Samppa Saarela	
23	Osuuspankkikeskus OSK	Jouni Lähteenmäki	Pekka Valta
24	PKT-Säätiö	Tommi Simon	
25	Profium Oy	Janne Saarela	
26	Rautaruukki Oyj	Jari Ahde	
27	Sanastokeskus TSK	Lena Jolkkonen	Katri Seppälä
28	Sininen Meteoriitti Oy	Aki Antman	Anniina Lehtinen
29	Stakes	Matti Ojala	Ani Kajander
30	Suomen Asiakastieto Oy	Teija Rantanen	Risto Kallio
31	Suomalaisen kirjallisuuden seura SKS	Lauri Harvilahti	Heli Kautonen
32	Suomen valokuvatait. museo	Elina Heikka	Anne Isomursu
33	Taideteollisen korkeakoulun kirjasto	Ellen Karhulampi	Marita Turpeinen
34	Tekes	Katja Ahola	
35	Terveyden edistämisen keskus (Tekry)	Tuula Kolari	Mika Pyykkö
36	TietoEnator Oyj	Jaroslav Skwarek	Lasse Akselin
37	Valtion taidemuseo	Perttu Rastas	Riitta Autere
38	Valtiovarainministeriö / ValtIT	Olli-Pekka Rissanen	
39	Wärtsilä Oyj	Jari Kallela	

Taulukko 4.1. FinnONTO 2.0 –hankkeen yritys/organisaatiojäsenedustajat ja näiden edustajat johtoryhmässä.

Viitteet

- J. Aitchison, A. Gilchrist, D. Bawden (2000): Thesaurus construction and use: a practical manual. Europa Publications, London.
- O. Alm (2007): Tekstidokumenttien automaattinen ontologiaperustainen annotointi. Pro Gradu, Helsingin yliopisto, 2007. <http://www.seco.tkk.fi/publications/2007/alm-gradu-2007.pdf>
- R. Baeza-Yates (1999): Modern information retrieval. Addison-Wesley, New-York.
- Jérôme Euzenat, Pavel Shvaiko (2007): Ontology matching. Springer-Verlag.
- C. Fellbaum (1998): WordNet. The MIT Press, Massachusetts.
- D. Fensel (2003): Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce, Springer-Verlag, Berlin, 2001. 2nd Edition, Springer-Verlag.
- A. Gangemi, N. Guarino, C. Masolo, A. Oltramari, L. Schneider (2002): Sweetening ontologies with DOLCE. In: Proc. of the 13th International Conference on Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web, Springer-Verlag.
- Mikko Haaramo: Iconclass-luokittelujärjestelmän ontologisointi ja soveltaminen. Diplomityö, Teknillinen korkeakoulu, 2006. <http://www.seco.tkk.fi/publications/>
- Riikka Henriksson, Tomi Kauppinen, Eero Hyvönen (2008): Core Geographical Concepts: Case Finnish Geo-Ontology. Location and the Web (LocWeb) 2008 workshop, 17th International World Wide Web Conference WWW 2008, ACM International Conference Proceeding Series; Vol. 300.
- E. Hyvönen (2005): Miksi asiasanastot eivät riitä vaan tarvitaan ontologioita. Signum, nro 5. <http://www.seco.tkk.fi/publications/2005/hyvonen-miksi-asiasanastot-eivat-riita-2005.pdf>
- E. Hyvönen, A. Valo, V. Komulainen, K. Seppälä, T. Kauppinen, T. Ruotsalo, M. Salminen, and A. Ylisalmi (2005a): Finnish National Ontologies for the Semantic Web - Towards a Content and Service Infrastructure. Proceedings of International Conference on Dublin Core and Metadata Applications (DC 2005), Madrid, Spain, short papers.
- E. Hyvönen, E. Mäkelä, M. Salminen, A. Valo, K. Viljanen, S. Saarela, M. Junnila, and S. Kettula (2005b): MuseumFinland - Finnish Museums on the Semantic Web. Journal of Web Semantics, Vol. 3, No. 2.
- Eero Hyvönen, Eetu Mäkelä (2006): Semantic Autocompletion. Proceedings of the first Asia Semantic Web Conference (ASWC 2006), Springer-Verlag.

Eero Hyvönen, Kim Viljanen and Osma Suominen (2007a): HealthFinland - Finnish Health Information on the Semantic Web. Proceedings of the 6th International Semantic Web Conference (ISWC 2007), Springer-Verlag.

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

Eero Hyvönen, Kim Viljanen, Jouni Tuominen and Katri Seppälä (2008a): Building a National Semantic Web Ontology and Ontology Service Infrastructure - The FinnONTO Approach. Proceedings of the European Semantic Web Conference ESWC 2008, Springer-Verlag.

Eero Hyvönen (2008b): Semantic Interoperability of Content and Services. A National Cross-domain Content Infrastructure for the Semantic Web. In: Phillip Sheu, Heather Yu, C. V. Ramamoorthy, Arvind K. Joshi, and Lotfi A. Zadeh (Eds.). Semantic Computing. IEEE Press / Wiley, New York, (ilmestyvä).

Ulla Karttunen: Valokuvakirjallisuuden asiasanasto. Suomen valokuvataiteen museo. http://www.fmp.fi/fmp_fi/muvieras/kirjasto/asiasan/index.htm

Tomi Kauppinen, Eero Hyvönen (2007): Modeling and Reasoning about Changes in Ontology Time Series. Ontologies. In Rajiv Kishore, Ram Ramesh, Raj Sharman (eds.): A Handbook of Principles, Concepts and Applications in Information Systems, Springer-Verlag.

Tomi Kauppinen, Jari Väätäinen, Eero Hyvönen (2008): Creating and Using Geospatial Ontology Time Series in a Semantic Cultural Heritage Portal. S. Bechhofer et al.(Eds.): Proceedings of the 5th European Semantic Web Conference 2008 ESWC 2008, Springer-Verlag.

Robin Lindroos (2008): Paikkatiedon ontologiapalvelu. Diplomityö, Teknillinen korkeakoulu. <http://www.seco.tkk.fi/publications/>

Ville Komulainen: Public Services for Ontology Library Systems. Pro gradu, Helsingin yliopisto, 2007. <http://www.seco.tkk.fi/publications/>

Tellervo Krogerus, Mirja Kälviäinen, Riitta-Liisa Leskinen ja Anne Mikkanen (toim.) (1997) Käsi- ja taideteollisuuden asiasanasto. Taitemia 6, Kuopio. Verkkoersio: <http://www.designkuopio.fi/kirjasto/asiasanasto/>

R. L. Leskinen (toim.) (1997): Museoalan asiasanasto. Museovirasto, Helsinki.

Tuukka Ruotsalo, Eero Hyvönen (2007): An Event-based Approach for Semantic Metadata Interoperability. Proceedings of the 6th International Semantic Web Conference (ISWC 2007), Springer-Verlag.

Reetta Sinkkilä (2008): Käsitteen kontekstiperustainen valinta semanttisessa webissä. Pro Gradu, Helsingin yliopisto. <http://www.seco.tkk.fi/publications/>

Reetta Sinkkilä, Eetu Mäkelä, Tomi Kauppinen and Eero Hyvönen (2008): Combining Context Navigation with Semantic Autocompletion to Solve Problems in Concept Selection. First International Workshop on Semantic Metadata Management and Applications, SeMMA 2008 (at (ESWC 2008). CEUR Workshop Proceedings, vol. 346, pp. 61-68, <http://CEUR-WS.org>.

S. Staab, R. Studer (toim.) (2003, 2008 (2nd edition)): Handbook on Ontologies. Springer-Verlag.

J. Sowa (2000): Knowledge Representation: Logical, Philosophical, and Computational Foundations. Brooks Cole Publishing Co., Pacific Grove, California.

Osma Suominen, Kim Viljanen, Eero Hyvönen: User-centric Faceted Search for Semantic Portals. Proceedings of the European Semantic Web Conference (ESWC 2007), Springer-Verlag.

Osma Suominen (2008): Käyttäjakeskeinen moninäkömähaku semanttisessa portaalissa. Pro Gradu, Helsingin yliopisto. <http://www.seco.tkk.fi/publications/>

H. Suonuuti (2001): Guide to Terminology. NordTerm Publication 8, Tekniikan Sanastokeskus TSK, Helsinki.

J. Tuominen. (2008): Helppokytkentäiset ontologiapalvelut semanttisessa webissä. Pro Gradu, Helsingin yliopisto, ilmestyvä.

Taideteollinen korkeakoulu (1995): Taideteollisuuden asiasanasto. Helsinki.

O. Valkeapää (2006): Verkkoresurssien ontologiaperustainen annotointi. Diplomityö, Teknillinen korkeakoulu, <http://www.seco.tkk.fi/publications/>.

O. Valkeapää, O. Alm and E. Hyvönen (2007): A Framework for Ontology-based Adaptable Content Creation on the Semantic Web. Journal of Universal Computer Science, Vol. 13, No. 12.

Antti Vehviläinen (2006): Ontologiapohjainen kysymys-vastauspalvelu (Ontology-based question-answer service). Diplomityö, Teknillinen korkeakoulu. <http://www.seco.tkk.fi/publications/>

Antti Vehviläinen, Eero Hyvönen, Olli Alm (2008): A Semi-automatic Semantic Annotation and Authoring Tool for a Library Help Desk Service. Emerging Technologies for Semantic Work Environments: Techniques, Methods, and Applications, IGI Group, Hershey, USA.

Kim Viljanen, Jouni Tuominen, Teppo Käsälä, Eero Hyvönen (2008a): Distributed Semantic Content Creation and Publication for Cultural Heritage Legacy Systems.

Proceedings of the 2008 IEEE International Conference on Distributed Human-Machine Systems, IEEE Press.

Kim Viljanen, Jouni Tuominen, Eero Hyvönen (2008b): Publishing and Using Ontologies as Mash-Up Services. Proceedings of the 4th Workshop on Scripting for the Semantic Web (SFSW2008), 5th European Semantic Web Conference 2008 (ESWC 2008), Tenerife, Spain.