

Miksi asiasanastot eivät riitä vaan tarvitaan ontologioita?

Eero Hyvönen

Tesaurusia on perinteisesti laadittu ihmisten, erityisesti tiedon indeksoijien tarpeita silmällä pitäen. Tällöin termien merkitysten kuvaus on voitu jättää ihmisen tulkinnan varaan. Yhä enenevässä määrin tesaurusia käyttävät kuitenkin tietokonesovellukset, joille sanastot tulisi aiempaa täsmällisemmin määritellä ontologioina. Uusi tärkeä sovellusalue on mm. semanttinen web ja web-palvelut.

Tässä artikkelissa esitetään yksikertainen malli perinteisen tesauksen muuttamiseksi semanttisen webin ontologiaksi. Mallia on kehitetty mm. Museoalan ontologian kehitystyössä ja ontologisoinnin tuloksia on pilotoitu menestyksellisesti useissa semanttisissa portaalihankkeissa. Mallia testataan parhaillaan FinnONTO-hankkeessa mm. Yleisen suomalaisen asiasanaston muutos-työssä yksinkertaiseksi ontologiaksi.

Usko ontologioiden avulla saavutettavista sovelluseduista on viime vuosina kasvanut voimakkaasti. Yksi merkittävä syy tähän on ollut webin isänä tunnetun Tim Berners-Leen visio semanttisesta webistä, jonka kehittämiseksi webin infrastruktuuria kansainvälisesti kehittävä W3C-järjestö käynnisti vuonna 2001 erityisen Semantic Web Activity -ohjelman [1].

Semanttisen webin ehkä keskeisimpänä teknologisenä elementtinä ovat ontologiat. Sovellusten kesken jaettujen ontologioiden avulla voidaan parantaa tietojärjestelmien yhteentoimivuutta (interoperability), mikä on ollut jatkuvasti yksi tietotekniikan suurimpia käytännön haasteita.

Koneet asiasanastojen käyttäjiksi

Tesaurusta voidaan käyttää indeksoinnin (asiasanoituksen, luokituksen) apuvälineenä, tiedon haun apuvälineenä tai molemmissa tehtävissä.

Näkökulma on keskeinen informaatiotutkimuksen ja tiedonhaun (Baeza-Yates, 1999) alueella. Tieto- ja tietämystekniikan (knowledge engineering) näkökulmasta tarkasteltuna tesauukset ovat eräänlaisia ontologisia kuvauksia maailmasta ja niitä voidaan käyttää hyväksi tiedon haussa, päättelyssä ja esittämisessä monin eri tavoin.

Sanastojen sähköiset versiot ovat lisääntyneet. Esimerkiksi YSA julkaistaan nykyisin ajantasaisena vain verkkoversiona [2]. Google -hakupalvelun yhteydessä on käytössä Google Suggest -versio [3], joka osaa täydentää käyttäjän kirjoituksen mielekkäiksi hakusanoiksi, mikä mm. estää kirjoitusvirheitä. Googlen hakemisto-osa taas hyödyntää Open Directory Project -hankkeen [4], (ODP) 590 000 kategorian luokittelua.

Viime aikoina monia asiasanastoja on alettu kehittää ns. ontologioiksi (Fensel, 2003; Staab, Studer, 2003). Esimerkiksi ODP-hankkeessa ol-

laan indeksoimassa n. 70 000 vapaaehtoisen voimin koko webin sisältöä. Indeksoituna on jo yli 5 miljoonaa sivua. Järjestelmä perustuu semanttisen webin avoimeen RDF(S)-formaattiin, joka on yksinkertainen kieli ontologioiden ja niihin liittyvän metadatan esittämiseksi.

Monien nykyisten asiasanastojen ongelmana on, että niitä on kehitetty lähinnä ihmiskäyttäjän, erityisesti indeksoijan ehdoilla. Koneelle yksittäisellä termillä ei ole merkitystä, vaan merkitys syntyy termien välisten suhteiden kautta, esimerkiksi siitä, että termi ”Suomi” on hierarkkisessa osa-kokonaisuussuhteessa termiin ”Pohjoismaat”. Suhteiden täsmällinen kuvaaminen on siksi ensiarvoisen tärkeää koneiden kannalta.

MuseoSuomi esimerkkinä

Esimerkki semanttisen webin ontologioiden mahdollisuuksista Suomessa on kansainvälisesti ja kotimaassa palkittu [5] MuseoSuomi -järjestelmä [6], (Hyvönen et al. 2005b). MuseoSuomi -järjestelmässä eri museoiden tietosisällöt voitiin yhdistää sisällöllisesti seitsemän yhteisen ontologian avulla yhdeksi virtuaaliseksi kokoelmaksi: käyttäjä löytää ”yhden luukun takaa” eri museoiden aineistot saumattomasti toisiinsa liitettyinä. Pilottijärjestelmässä ovat mukana Kansallismuseo, Espoon kaupunginmuseo sekä Lahden kaupunginmuseo, joiden kokoelmat käyttävät erilaisia museotietojärjestelmiä ja luettelointikäytäntöjä.

MuseoSuomi tarjoaa käyttäjilleen myös älykkään ”semanttisen hakukoneen” sekä mahdollisuuden ”semanttiseen samoiluun”, jossa konepäättelysääntöjen avulla voi suositella käyttäjälle linkkejä eri tavoin toisiinsa liittyviin tietosisältöihin MuseoSuomessa hyödynnettiin n. 6000 termin Museoalan asiasanastoa MASA:a (Leskinen, 1997), mutta ilman sen ontologisointia Museoalan ontologiaksi MAO ei näitä järjestelmiä olisi voitu toteuttaa. Yleisradio Oy:lle tehdyssä semanttisessa Orava-portaalissa [7] MuseoSuomen tietosisällöt voitiin linkittää vielä yli 2000 videoklipin opetusaineistoon.

Tesaurusten ontologisoitkysymys on tärkeä asiasanastojen laajan käytön ja webin käytön nopean yleistymisen takia. Aihe on otettu käsiteltäväksi mm. W3C:n Semantic Web Best Practices -työryhmässä [8] joka on valmistellut tähän liittyvää ohjeistusta ja SKOS-suositusta.

Seuraavassa esitetään, millaisia muutoksia tesauraukseen käytännössä joudutaan tekemään suoritettaessa kevyttä muunnosta ontologiaksi. Esimerkkinä käytetään n. 23 000 termin YSA:a, jota TKK:ssa ja Helsingin yliopistolla toimiva Semanttisen laskennan tutkimusryhmä [9] on parhaillaan ontologisoimassa Yleiseksi Suomalaiseksi Ontologiaksi YSO.

Asiasanaston muuttaminen ontologiaksi

Asiasanastot 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.

Sanaston semanttiset suhteet jakaantuvat kolmeen ryhmään (Aitchison et al., 2000): Ekvivalenssisuhteilla ilmaistaan termien synonyymit ja näille käytettävä suositeltava termi (preferred term). Nämä tiedot kuvataan KÄYTÄ/KORVAA-suhteilla. Synonyymien osalta voidaan tehdä ero aidosti eri termeille ja leksikaalisille varianteilla (lexical variant), jotka ovat saman termin erilaisia ilmiäsuja. Kvasisynonyymilla (quasi-synonym) tarkoitetaan lähes samaa merkitsevää termiä.

Hierarkkisia suhteita kuvataan LT- ja ST-suhteilla. Tärkeimmät suhdetyypit ovat osakokonaisuus (part of), kuten ja ala-yläkäsité (subclass of). Muita kuin ekvivalenssi- ja hierarkkisia suhteita kuvataan RT-suhteella.

Tesaurus voidaan muuttaa yksinkertaiseksi ontologiaksi seuraavissa vaiheissa:

1. Syntaktinen muunnos ontologiaformaattiin
2. Termistön erottaminen ontologiasta
3. Termien päämerkitysten erottelu
4. Luokkien ja yksilöiden erottelu

5. Hierarkkisten suhteiden erottelu
6. Hierarkkisten suhteiden systematisointi hierarkiaksi
7. Assosiativisten suhteiden esittäminen

Vaihe 1 on luonteeltaan syntaktinen. Vaiheet 2–3 liittyvät semanttisiin ekvivalenssisuhteisiin, vaiheet 4–6 hierarkkisten suhteiden täydentämiseen, korjaamiseen ja täsmentämiseen ja vaihe 7 assosiativisten suhteiden tarkempaan määrittelyyn.

Syntaktinen muunnos ontologiaformaattiin

Ensin asiasanasto muutetaan syntaktisesti tavoitteena olevalle ontologiakielelle, joka voi olla esimerkiksi RDF(S) [10] tai OWL [11]. Näitä kieliä varten on saatavilla erityisiä ontologiaeditoreita, joiden avulla muunnoksen seuraavat vaiheet on helpompi suorittaa. Käytetyin ontologiaeditori on Stanfordin yliopistossa kehitetty Pretege-2000 [12] lisäohjelmiseen (plug-in).

YSA:n tapauksessa sanasto oli saatavilla MARC-formaatissa [13], josta se muutettiin ensiksi XML-muotoon ja sitten RDF(S)- ja OWL-kielelle. Näin sanasto voitiin ottaa edelleen kehitettäväksi Protege-2000-editorissa.

Termistön erottaminen ontologiasta

Tesauukset ovat normatiivisia, termien käyttöä ohjaavia termistöjä. Ontologioissa lähtökohtana eivät niinkään ole termit kuin niiden alla olevat koneen ymmärtämät merkitykset. Harkittavaksi tulee, miten termit ja merkitykset esitetään ontologiassa. Tässä voidaan käyttää kahta periaatteellista tapaa.

Suoraviivainen tapa on esittää kunkin käsitteen termit aina käsitteen yhteydessä jonkin ominaisuuden arvona. Esimerkiksi RDF-kehysesä voidaan käyttää `rdfs:label`-ominaisuutta ja tämän tarkentamiseen kielen ilmaisevaa `xml:lang`-attribuuttia.

Lähestymistavan ongelmana on, että käsitteeseen voi liittyä hyvin paljon termejä ja näi-

den leksikaalisia variantteja. Esimerkiksi henkilötesauruksessa Union List of Artist Names [14] (ULAN) löytyy Ivan Aivazovskille = Ivan Aivazovsky (Valtion taidemuseon käyttämät nimet) syntaktisina variantteina 13 translitteroitua muotoa.

Terminologiset määrittelyt voidaan erottaa itse ontologiasta mutkikkaiden terminologisten määrittelyiden toteuttamiseksi ja käyttäjäryhmäkohtaisten termistöjen luomiseksi. Lähestymistapa on luonteva esimerkiksi monikielisten ontologioiden toteuttamiseksi.

Lähestymistapaa kehitettiin ja kokeiltiin MuseoSuomi-hankkeen yhteydessä. Siinä jokainen semanttiseen portaaliin aineistoa tuottanut museo saattoi käyttää omaa termistöään ilmoittamalla kullekin termille sitä vastaavan ontologisen käsitteen yhteisissä ontologioissa.

Menetelmän käytännöllisenä hankaluutena on, että nykyisten ontologiaeditorien ja ontologian toimittajan kannalta on usein luontevampaa ylläpitää asiasanastoja käsitteiden yhteen kerättyinä ominaisuuksina eikä erillisinä termiontologioina. Ongelmaa voidaan lähestyä kehittämällä erityisesti asiasanastotyötä tukevia ontologiaeditoreita.

Termien päämerkitysten erottelu

Sanat ovat monimerkityksisiä. Tesauksia ja ontologioita luotaessa keskeinen tehtävä on eri merkitysten erottaminen termeiksi ja käsitteiksi. Ilmeinen ongelma ovat homonyymiset termit, joissa sama sana tarkoittaa kahta aivan eri käsitettä. Esimerkiksi:

Nokia (paikkakunta)

Nokia (yritys)

Polyseemisillä termeillä taas on useita toisiinsa liittyviä merkityksiä. Esimerkiksi sanalla ”johtaminen” on mm. seuraavia merkityksiä, joista kaksi ensimmäistä liittyvät toisiinsa melko läheisesti:

johtaminen (musiikki)

johtaminen (liiketalous)

johtaminen (sähkötekniikka)
johtaminen (matematiikka)

Asiasanaston ontologisoinnin yhteydessä jokaiselle käsitteelle annetaan yksilöivä tunniste. Webmaailmassa tunnisteina käytetään yleisesti URI-osoitteita [15] (Universal Resource Identifier). Muitakin tunnistejärjestelmiä on kehitetty, kuten ITU-järjestön OID [16] (Object Identifier). Tunnisteen ideana on erottaa käsite tai muu olio toisista ja antaa tapa viitata siihen.

Tunnisteen tulee olla ajallisesti pysyvä (persistent), sillä tunnisteen muuttamisen jälkeen pitäisi muuttaa myös kaikki tunnisteen ilmentymät aiemmin indeksoituissa aineistoissa. Esimerkiksi sanaston laajentaminen uusilla termeillä ei saa vaikuttaa aiemmin nimettyjen käsitteiden tunnisteisiin.

Ilmiasuun perustuva yksilöinti on ongelmallista myös silloin, kun merkitykseen pitää liittää useita ilmiäsuja. Esimerkiksi monikielisissä asiasanastoissa yhteen käsitteeseen pitäisi voida viitata erikielistä symmetrisesti, jolloin viittaus kieli-riippumattomaan tunnisteeseen eikä jonkun tietyn kielen ilmaukseen on luontevaa. Näin menettellään mm. luokitusjärjestelmissä kuten ICON-CLASS [17].

Globaalin käsitteiden erottelukyvyn aikaansaamiseksi URI-tunniste on muodoltaan web-osoitteen kaltainen, esimerkiksi:

<http://www.sahko.fi/sanasto#johtaminen>

Ideana on, että osoitteen aluenimi (domain), tässä www.sahko.fi, rekisteröidään webissä normaaliin tapaan sanastoa hallinnoivan organisaation toimesta, jolloin kenelläkään muulla ei voi olla samannimistä aluenamea. Joku toinen organisaatio voi ottaa vapaasti käyttöön nimen johtaminen toisella alueella esimerkiksi seuraavan URI:n avulla:

<http://www.yritys.fi/sanasto#johtaminen>

URI-tunnisteen tärkeä ero muihin tunniste-koodauksiin on, että sillä voidaan paitsi tunnistaa

identiteetti myös kertoa, missä päin webiä käsite on määritelty. Esimerkiksi liikkeenjohtamisen käsite tulisi määritellä tiedostossa <http://www.yritys.fi/> -sanasto, josta sovellukset voivat käydä lukemassa sen. Tiedostossa oleva RDF- tai OWL-kielinen kuvaus määrittelee liikkeenjohtamisen merkityksen siihen liittyvien semanttisten suhteiden verkoston avulla.

Luokkien ja yksilöiden erottelu

Tesauruset eivät yleensä erottele yksilöitä luokista. Esimerkiksi YSA:ssa on määrittely: Halley'n komeetta LT komeetat Halley'n komeetta on kuitenkin semanttiselta olemukseltaan yksilö eikä yksilöiden luokka. Kone ei tätä erottelua yleensä pysty tesauksen perusteella tekemään.

Esimerkiksi Pohjoismaat-termillä on suppeamat termit Suomi, Ruotsi jne. Tässä tapauksessa kaikki paikannimet viittavat yksilöihin. Yksilöiden ja näitä määrittävien luokkien erottelu on keskeinen idea tietotekniikassa; se on perustana mm. koko modernille oliokeskeiselle ohjelmoinnin paradigmalle. Ajatuksena on, että luokka määrittelee yksilöiden yhteiset ominaisuudet.

Hierarkkisten suhteiden erottelu

Hierarkkiset suhteet esitetään tesauksissa ST/LT-suhteella. Tämä suhdetyyppi jakautuu käytännössä kolmeen suhdetyyppiin: edellä mainittuun yksilö-luokka suhteeseen, ala-yläluokkasuhteeseen (hyponymia) ja osa-kokonaisuus-suhteeseen (meronymia).

Lisäksi luokalla voi olla koko luokkaan liittyviä ns. luokkaominaisuuksia. Esimerkiksi osa-kokonaisuussuhde voi jakaantua vielä moneen semanttisesti erilaiseen tapaukseen (Fellman, 1998). Laajasti käytetty englanninkielinen WordNet-tesaurus [18] (Fellman, 1998) käyttää kolmea meronymiasuhdetta: component part of, member of, made form stuff.

WordNetissä tunnistetaan hierarkkisten luokka- ja osa-kokonaisuussuhteiden lisäksi verbien välinen troponymiasuhde, joka on eräänlainen toimintojen ja tapahtumien ala-yläluokkasuh-

de. Esimerkiksi ”kulkea” käsitteen troponyminen alakäsite voisi olla ”kävellä”, jolla taas voisi olla alakäsite ”ontua”. YSA-YSO-muunnoshankkeessa ei erilaisia osa-kokonaisuussuhteita ainakaan hankkeen alkuvaiheessa erotella toisistaan.

Asiasanastossa hierarkkiset suhteet on usein kehitetty vain osittain, jolloin kaikkia käsitteitä ei ole kytketty toisiinsa eikä käsitteistä muodostu systemaattisista hierarkkista rakennetta. Esimerkiksi MASA:n 6000 käsitteestä n. 2600:lla ei ole yläkäsitettä ja tilanne on vastaava YSA:ssa.

Sanaston hyödyntämisen kannalta käsitteet olisi hyvä järjestää kattavasti hierarkkioihin. Nämä ovat tarpeen mm. termejä lavennettaessa, sanaston esittelemisessä loppukäyttäjälle sovelluksissa sekä hakutulosten ryhmittelyssä. Systematiikka helpottaa myös sanaston ylläpitoa.

MASAn ontologisoinnin yhteydessä hierarkkioiden täydentäminen johti lukuisten uusien käsitteiden luomiseen. Samalla kun hierarkkiset luokka- ja osa-kokonaisuussuhteet erotellaan ja täydennetään systemaattisiksi hierarkkioiksi, pitää vielä tarkastaa, että näiden transitiivisuusominaisuudet ovat kunnossa. Tämä tarkoittaa sitä, että luokkahierarkiassa kaikki yläluokan ominaisuudet ovat myös kaikkien – ei ainoastaan seuraavan – alaluokkien ominaisuuksia ja käänteisesti että yksilöt ovat kaikkien yläluokkiensa yksilöitä.

Tesauruksissa näin ei aina ole, mikä johtaa mm. virheellisiin hakutuloksiin termejä lavennettaessa. Vastaavasti on tarkistettava osa-kokonaisuussuhteiden toimivuus läpi koko hierarkian.

Assosiativisten suhteiden esittäminen

Hierarkkisia suhteita suuremman semanttisen mallittamisen haasteen asettavat erilaiset assosiativiset suhteet. Sovelluksista ja kuvaustarkkuudesta riippuen assosiativisten suhteiden kirjo muuttuu helposti hankalan suureksi. Esimerkiksi lääketieteellisen UMLS-tesauruksen [19] (Unified Medical Language System) semanttisessa verkossa käytetään 54 eri semanttista suhdetta, joista

suurin osa on luonteeltaan assosiativisia.

YSA-YSO-muunnostyössä assosiativisten suhteiden tarkempi mallittaminen jätetään ensivaiheessa eri sovellusten kontolle. Tällainen on esimerkiksi MuseoSuomesta kehitettävä kulttuuri-sisältöjä yhdistävä semanttinen KulttuuriSampoportaali [20]. Siinä ajatuksena on soveltaa kehysperustaista (frame) lähestymistapaa, jossa olennaista ovat tapahtumat sekä näihin eri semanttisissa rooleissa (semantic role) liittyvät tekijät.

Tietämyksen esittäminen kehystyyppisinä rakenteina on tekoälytutkimuksessa ja tietämyksen esittämisessä laajalti käytetty menetelmä (Sowa, 2000). Kehysmallissa esimerkiksi myyminen-tapahtuma voitaisiin kuvata kehyksellä, johon liittyy roolit toimija (myyjä), osallistuja (myytävä asia) ja kohde (ostaja). Ostaminen voitaisiin kuvata samalla semanttisella kehyksellä roolien täyttäjii vaihtamalla. Tunnettu kehyksiin perustuva järjestelmä on esimerkiksi amerikkalainen FrameNet [21].

Tarkastellaan esimerkkinä YSA:ssa olevaa seuraavaa assosiativista suhdetta, joka ilmaisee syytä ja seurausta:

aurinkotuuli RT revontulet

Tämä voitaisiin kehysmallissa esittää aiheuttaminen-tapahtuman yhtenä alityyppinä seuraavasti:

aiheuttaminen1 toimija aurinkotuuli
tulos revontulet

Määrittelyjen semantiikka kehysjärjestelmissä on täsmällisempi kuin perinteisissä assosiativisissa suhteissa. Mekanismi erottaa selkeästi toiminat ja tapahtumat ja ilmaisee roolit näiden suhteen. Roolien valinta riippuu sovelluksesta ja siitä, miten tarkkoja semanttisia roolierotteluja tarvitaan halutun toiminnallisuuden toteuttamiseksi. YSO:ssa tavoitteena on mahdollisimman monelle sovellusalueelle soveltuva yleisontologia, joten muunnostyön yhteydessä on mielekästä ottaa mukaan vain hyvin yleiskäyttöisiä rooleja.

Taulukossa 1. on hahmoteltu yksi mahdollinen sovelluksessa käytettävä roolijärjestelmä:

Rooli	Selite	Esimerkki
toimija (agent)	Toiminnan aktiivinen alullepanija tai suorittaja.	laulaminen – lintu laulaminen – Kirka
väline (instrument)	Toiminnassa käytettävä väline tai toiminnan edistäjä	naulaaminen – vasara
tulos (goal)	Toiminnan tulos	rakentaminen – talo opettaminen – oppiminen
kohde (patient)	Toiminnan kohteena oleva objekti.	opettaminen – oppilaat
paikka	Toiminnan paikka, jolle voidaan antaa koordinaatit.	maraton – Helsinki
ympäristö	Toiminnan ympäristötyyppi.	puunhakkuu – metsä
aika (time)	Toiminnan ajankohta, joka voidaan sijoittaa aikajanelle	tuhoutuminen – 2002-09-11
ajankohta	Toiminnan syklinen ajankohta.	iltarukous – ilta

Ontologia luo luokkahierarkian

Keuyen ontologisoinnin lopputuloksena syntyy yhtenäinen, kaikki käsitteet systemaattisesti kattava luokkahierarkia, jossa on eroteltu yksilöt luokista ja jossa ominaisuuksia voidaan periä läpi koko luokkahierarkian. Käsitteillä on yksilöivät URI-tunnisteet, joiden määrittelyt löytyvät webistä RDF-muotoisina.

Osa-kokonaisuussuhteet on erotettu ja näistä on mahdollisesti koostettu omat paikka-, aika- ja tms. ontologiansa. Tuloksena syntyneisiin taksonomiatyyppeihin ontologioihin voidaan eri sovel-

luksissa alkaa luoda tarkempia ontologisia kuvauksia esimerkiksi kehysmallin avulla.

Muunnoksen aikana ei ole syytä hävittää tarpeettomasti tesauruksessa olevia semanttisia suhteita tai muuta informaatiota, joilla on luonnollisesti oma arvonsa. LT-, ST-, RT- ym. suhteet voidaan jättää sellaisenaan ontologiaan ja hyödyntää niitä siltä osin kuin se on mahdollista. Esimerkiksi MuseoSuomi-hankkeessa ontologioissa säilytettiin kaikki MASA-tesauruksen informaatio ja suhteet ja käytettiin assosiativisia RT-suhteita eri tavoin hyväksi tietosisältöjä toisiinsa yhdistettäessä.

Verkkosivut

- [1] <http://www.w3.org/2001/SW>
- [2] <http://vesa.lib.helsinki.fi>
- [3] <http://www.google.com/webhp?complete=1&hl=en>
- [4] <http://dmoz.org>
- [5] MuseoSuomi sai v. 2004 semanttisen webin kansivälisen tutkijayhteisön Semantic Web Challenge Award -sovelluspalkinnon ja pääministerin kunniamaininnan innovatiivisimmasta sovelluksesta Tietoyhteiskuntaohjelman Laatusa verkkoon -kilpailussa.
- [6] MuseoSuomi on vapaasti käytettävissä osoitteessa <http://www.museosuomi.fi>. Pilottijärjestelmää ylläpitää Teknillisen korkeakoulun viestintätekniiikan laboratorio ja Semanttisen laskennan tutkimusryhmä (SeCo).
- [7] Orava-portaali on kokeiltavissa osoitteessa <http://www.museosuomi.fi/orava>.
- [8] <http://www.w3.org/2001/sw/BestPractices/>
- [9] <http://www.seco.tkk.fi>
- [10] <http://www.w3.org/RDF/>
- [11] <http://www.w3.org/2004/OWL/>
- [12] <http://protege.stamford.edu>
- [13] <http://www.loc.gov/marc/>
- [14] http://www.getty.edu/research/conducting_research/vocabularies/ulan/
- [15] <http://www.w3.org/Addressing/>
- [16] <http://www.itu.int/ITU-T/asn1/uuid.html>
- [17] <http://www.iconclass.nl/>
- [18] <http://wordnet.princeton.edu/>
- [19] <http://www.nlm.nih.gov/research/umls/>
- [20] <http://www.cs.helsinki.fi/group/seco/ontologies/kulttuurisampo/>
- [21] <http://framenet.icsi.berkeley.edu/>

Viitteet

- J. Aitchison, A. Gilchrist, D. Bawden (2000): Thesaurus construction and use: a practical manual. Europa Publications, London.
- R. Baeza-Yates (1999): Modern information retrieval. Addison-Wesley, New-York.
- D. Fensel (2003): Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce, Springer-Verlag, Berlin, 2001. 2nd Edition, Springer-Verlag, Berlin.
- 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, 2005.
- 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.
- R. L. Leskinen (toim.) (1997): Museoalan asiasanasto. Museovirasto, Helsinki.
- S. Staab, R. Studer (toim.) (2003): Handbook on Ontologies. Springer-Verlag, Berlin.
- J. Sowa (2000): Knowledge Representation: Logical, Philosophical, and Computational Foundations. Brooks Cole Publishing Co., Pacific Grove, California.
- H. Suonuuti (2001): Guide to Terminology. NordTerm Publication 8, Tekniikan Sanastokeskus TSK, Helsinki.
- C. Fellbaum (1998): WordNet. The MIT Press, Massachusetts.

Tietoa kirjoittajasta:

Eero Hyvönen toimii TKK Viestintäteknikassa professorina ja Helsingin yliopiston tietojenkäsittelytieteen laitoksella dosenttina. Email. eero.hyvonen@tkk.fi. http://www.tkk.fi/~eahyvone