



Ontologiapalvelut semanttisessa webissä

Jouni Tuominen, Kim Viljanen & Eero Hyvönen

Aalto-yliopisto, mediatekniikan laitos ja

Helsingin yliopisto, tietojenkäsittelytieteen laitos

{jouni.tuominen, kim.viljanen, eero.hyvonen}@aalto.fi

Tiivistelmä

Ontologiat ovat keskeinen osa semanttista webiä: ne toimivat yhteisinä jaettuina käsitteistöinä, joiden avulla tietokoneet voivat käsitellä tietoa älykkäämmin. Jotta eri toimijat voivat hyödyntää yhteisiä käsitteistöjä sovelluksissaan, ontologiat on julkaistava heidän käyttöönsä. Yksinkertaisimmillaan ontologiat voidaan julkaista tiedostomuodossa. Tällöin jokainen toimija joutuu toteuttamaan itse toiminnallisuuksia ontologioiden hyödyntämiseen. Koska osa toiminnallisuuksista on yleisiä, useissa järjestelmissä toistuvia, niiden toteuttaminen valmiina palveluina on mielekästä. Palveluita voidaan tarjota ihmis-käyttäjille käyttöliittymäkomponentteina sekä ohjelmalliseen käyttöön rajapintoina, joita käyttämällä toiminnallisuudet voidaan integroida asiakasjärjestelmiin. Tässä artikkelissa kuvataan ontologioiden käyttäjäryhmien tarpeita sekä ontologiapalveluiden toteutuksia. Yleisten ontologioiden käyttämiseen liittyvien toiminnallisuuksien tarjoamiseksi esitetään ontologiapalvelu ONKI, joka on osa Suomalaiset semanttisen webin ontologiat-hankesarjassa (FinnONTO, 2003–2012) kehitettyä ontologiainfrastruktuuria. Artikkelin perustuu Jouni Tuominen pro gradu -työhön, jolle Tietojenkäsittelytieteen Seura ry myönsi lukuvuoden 2009–2010 pro gradu -palkinnon. Tutkimustyöhön ovat osallistuneet myös Kim Viljanen ja Eero Hyvönen.

1 Johdanto

Tietojärjestelmiin tallennettujen tietomäärien kasvaessa käyttäjän on yhä vaikeampaa löytää niistä tarpeeseensa sopivaa tietoa. Erityisesti tämä ongelma esiintyy maailmanlaajuisesti hajautetussa internetin World Wide Webissä. Verkkosivujen kuvaamiseen käytettävä HTML-kieli on tarkoitettu pääosin dokumenttien ulkoasun määrittämiseen, eikä se kuvaa dokumenttien sisältämän tiedon merkityksiä. Koneymmärrettävän semantiikan puuttuessa koneet käsittelevät verkkosivuja syntaktisella tasolla, mutta eivät ymmärrä niiden sisältöjä. Semanttinen

web [4, 13] on webin laajennos, jossa tieto on koneymmärrettävässä muodossa tiedon aiempaa yhteentoimivamman ja älykkäämmän koneellisen käsittelyn mahdollistamiseksi. Semanttisen webin tekniikoita voidaan hyödyntää myös webin ulkopuolella, esimerkiksi yritysten intranet-ympäristöissä.

Esimerkkinä nykyisen webin ongelmista voidaan esittää tiedonhakutilanne, jossa haetaan tietoa ”Nokia”-nimisestä paikkakunnasta yksinkertaiseen merkkijonojen vertailuun perustuvalla hakukoneella. Hakusanalla ”nokia” tehdyn haun tuloksissa ei erotella toisistaan hakusa-

naan täsmäävää paikkakuntaa, yritystä, eläintä (soopeli) tai Frans Eemil Sillanpään teoksen ”Ihmiset suviyössä” päähenkilöä. Ihmisen on vaikea löytää juuri Nokia-paikkakuntaa käsittelevät dokumentit kaikkien hakutulosten joukosta. Tällaisten monimerkityksellisten sanojen lisäksi merkkijonojen vertailuun perustuvien hakumenetelmien kannalta ongelmallisia tapauksia ovat muun muassa synonyymit. Esimerkiksi hakusanalla ”diabetes” ei löydetä dokumentteja, joissa on käytetty sanaa ”sokeritauti” sanan ”diabetes” sijaan.

Semanttisessa webissä tieto voidaan kuvata ontologioiden avulla. Ontologia on tietyn aihealueen jaettu käsitteistö, jossa on kuvattu formaalisti aihealueeseen kuuluvat käsitteet ja niiden väliset suhteet [17]. Ontologia määrittelee eksplisiittisesti ihmisille ominaista implisiittistä tietämystä maailmasta, mikä mahdollistaa tiedon koneellisen käsittelemisen aiempaa älykkäämmin. Semanttisen webin vision toteutumisen kannalta semanttisen tiedon määrän lisääntyminen on ensisijaisen tärkeää.

Jotta ontologioita voidaan käyttää tiedon kuvailussa eli annotoinnissa [31], ontologiat täytyy ensin luoda ja julkaista toimijoiden käyttöön. Ontologioiden kehittäjät ovat tyypillisesti alueensa asiantuntijoita ja vastaavat ontologioiden ylläpidosta. Tässä artikkelissa keskitytään ontologioiden käyttöönottoon ja julkaisemiseen, ontologioiden varsinaisen kehitystyön [9] jäädessä tarkastelun ulkopuolelle. Ontologiaperustaisen tiedon kuvailun kannalta konkreettinen haaste on, miten saada metatietoa tuottavaan järjestelmään viittauksia ontologioiden käsitteisiin.

Ontologiat voidaan julkaista yksinkertaisimmillaan tiedostomuodossa. Tällöin jokainen ontologian käyttöönsä haluava

toimija joutuu toteuttamaan toiminnallisuuksia ontologian käyttöön. Ontologioiden käyttämiseen liittyy monia yleisiä toiminnallisuuksia, joita tarvitaan ontologioita hyödyntävissä sovelluksissa. Toimijat toteuttavat keskenään samankaltaisia toiminnallisuuksia omissa sovelluksissaan, vaikka olisi kustannustehokkaampaa toteuttaa toiminnallisuudet yleisinä ontologiapalveluina ja julkaista ne toimijoiden käyttöön. Tällainen yksinkertainen tapa käyttää ontologioita madaltaisi toimijoiden kynnystä ottaa ontologioita käyttöönsä ja aloittaa semanttisen tiedon tuottaminen.

2 Yhteismitallinen ja semanttinen metatieto

Metatiedolla tarkoitetaan tietoa tiedosta ja se voidaan määritellä yleisesti miksi tahansa ilmaukseksi jostakin informaatio-objektista [16]. Informaatio-objekti voi olla esimerkiksi verkkosivu, tekstidokumentti tai fyysinen esine. Metatiedon käyttötarkoituksia ovat muun muassa tiedon löydettävyyden tehostaminen ja tiedon hallinta. Tunnetuin metatiedon esittämiseen käytetty standardi on Dublin Core¹, joka määrittää tiedon kuvailussa käytettävän ominaisuusjoukon. Dublin Core sisältää elementtejä kohteen sisällön kuvaamiseen (mm. nimeke, aihe, kuvaus), tekijätietojen esittämiseen (mm. tekijä, julkaisija) ja tunnistamiseen (mm. identifikaatiotunnus, formaatti).

2.1 Metatiedon hyödyntäminen

Tekstimuotoisiin objekteihin kohdistuvassa tiedonhaussa metatiedon hyödyllisyys perustuu siihen, että haku voidaan kohdistaa itse tekstisisällön sijaan tai sen lisäksi objektia kuvailevaan tietoon. Yksinkertainen pelkkään tekstisisältöön kohdistuva kokotekstihaku on haasteellista, kos-

¹<http://dublincore.org/documents/2008/01/14/demi-terms/>

ka hakukone ei erota tekstistä tietosisällön kannalta merkityksellisiä ja epämerkityksellisiä sanoja toisistaan, mikä johtaa haun tarkkuuden heikentymiseen [7]. Tarkkuuden heikentymistä voidaan kompensoida järjestämällä hakutulokset relevanssin perusteella [34], jolloin hakutehtävän kannalta oletettavasti relevanteimmat tulokset ovat hakutulosjoukon alussa. Toisaalta koska tekstissä ei välttämättä esiinny kaikkia tietosisällön kannalta merkityksellisiä sanoja, haun saanti voi heikentyä [19].

Eksplisiittisen metatiedon käyttämisen sijaan tiedonhakua voidaan pyrkiä tehostamaan luonnollisen kielen menetelmillä, esimerkiksi tunnistamalla tekstistä koneellisesti merkityksiä kontekstin perusteella. Ei-tekstimuotoisten objektien, kuten kuvien ja videoiden, yhteydessä metatiedon käyttäminen on perusteltua, koska aineistoon ei voida muutoin kohdistaa tekstihakua.

Jotta metatieto olisi hyödyllistä esimerkiksi tiedonhaun kannalta, sen pitää olla oikeellista, kattavaa, riittävän spesifiä ja tarkkaa [18]. Lisäksi metatiet ominaisuuksia pitää käyttää yhdenmukaisella tavalla. Metatieto-ominaisuuksien arvot voidaan kuvata käyttämällä esimerkiksi luonnollista kieltä, kontrolloitua sanastoa tai tietyllä tavalla muodostettuja merkijonoja (esim. päivämäärät).

Kontrolloitua sanastoa käytettäessä metatieto on yhteismitallista, koska tiettyyn käsitteeseen (merkitykseen) viitataan johdonmukaisesti aina samalla termillä ja monimerkityksisiä termejä käytetään viittaamaan vain tiettyyn käsitteeseen [16]. Sanastossa voidaan esimerkiksi määrittää, että käsitteeseen ”kummitus” viitataan aina termillä ”kummitus”, eikä koskaan termillä ”haamu”, ja että termi ”Nokia” viittaa aina käsitteeseen ”Nokia paik-

kakuntana”. Jälkimmäisessä tapauksessa voisi tosin olla selkeyden vuoksi perusteltua käyttää esimerkiksi termiä ”Nokia (paikkakunta)”. Sekä kokotekstihakua että luonnolliseen kieleen perustuvaa metatietoa hyödyntävässä tiedonhaussa monimerkitykselliset termit heikentävät haun tarkkuutta. Synonyymitermit puolestaan heikentävät tiedonhaun saantia [2].

Kontrolloitu sanasto voi olla rakenteeltaan yksinkertainen asiasanasto tai monimutkaisempi käsittemalli, kuten taksonomia, tesaurus tai ontologia. Ontologiat ovat tesaurusmuotoisempia ja mahdollistavat älykkäämmän koneellisen käsitteilyn.

2.2 Asiasanastot, taksonomiat ja tesaurusukset

Asiasanasto on yksinkertaisimmillaan termien luettelo, joka sisältää tietyn aihealueen kannalta oleelliset käsitteet. Jokaisella käsitteellä on yksi sitä kuvaava termi, josta käytetään usein nimitystä asiasana. Asiasanaksi valitun termin synonyymejä voidaan käyttää ohjaustermeinä, joiden tarkoituksena on ohjata käyttäjät viittaamaan käsitteisiin valituilla asiasanoilla. Asiasanasto voi sisältää myös muuta tietoa termeistä, kuten kuvaustekstejä ja lähdetietoja.

Taksonomia puolestaan on hierarkkisesti järjestetty kontrolloitu sanasto tai luokittelumalli, joka ryhmittelee toisiinsa liittyvät käsitteet helpottaen tarkoitukseen sopivan termin valintaa esimerkiksi tiedon kuvailussa [16]. Tesaurusuksessa on hierarkkisten suhteiden lisäksi yleisiä assosiativisia suhteita käsitteiden välillä. Tesaurusella kuvataan tyypillisesti tietyn sovellusalueen sanasto. Esimerkiksi Art & Architecture Thesaurus (AAT)² sisältää taiteen ja arkkitehtuurin aloilla käytettävää sanastoa. Toisaalta tesaurus voi

²http://www.getty.edu/research/conducting_research/vocabularies/aat/about.html

myös koostua yleiskäsitteistä, kuten Yleinen suomalainen asiasanasto (YSA)³.

Tesaurusten ja muiden kontrolloitujen sanastojen pääasiallinen käyttötarkoitus on tiedonhaun tehostaminen [2]. Sanastoa voidaan käyttää tiedon kuvailu- ja hakuvaiheessa tai ainoastaan jommassakummassa. Kun aineiston kuvailussa ja haussa käytetään yhteistä sanastoa, haun täsmäytys aineistoon helpottuu. Sanastoa käyttämällä päästään eroon luonnollisen kielen monitulkintaisuudesta ohjaamalla tiedon kuvailijat ja/tai hakijat käyttämään samoja termejä samoista käsitteistä.

Jos aineiston kuvailussa ja haussa käytetään yhteistä termistöä eri tarkkuustasoilla, hakutulokset ovat yleensä huonoja. Tällöin tiedonhaun saantia voidaan pyrkiä parantamaan laajentamalla kyselyitä lisäämällä niihin hakutermeihin liittyviä termejä. Tietämysmalleihin, kuten tesauruksiin tai ontologioihin, perustuvat tekniikat soveltuvat erityisesti tilanteisiin, joissa kyselyausekkeet ovat lyhyitä ja vaillinaisia, sisältäen vain vähän hakuindeksissä olevia termejä [52, 53].

Haun tarkkuutta puolestaan voidaan pyrkiä kasvattamaan erottelemalla monimerkityksellisten termien merkitykset toisistaan ja yksikäsitteistämällä ne käyttökontekstissaan. Toissijaisia asiasanastojen käyttökohteita ovat muun muassa sovelusalueen ymmärryksen lisääminen käyttäjälle näyttämällä käsitteiden välisiä suhteita ja käsitteiden määritelmiä sekä automaattisen tiivistelmän tuottaminen [2].

2.3 Ontologiat

Sana ontologia on peräisin filosofiasta, jossa se tarkoittaa todellisuuden ja oleminen luonteen tutkimusta. Tietojenkäsittelytieteessä ontologialla tarkoitetaan tietyn aihealueen käsitteistön eksplisiittistä mallia [17]. Ontologiaa kuvaa aihealueeseen-

sa kuuluvat käsitteet ja niiden väliset suhteet formaalisti mahdollistaen aihealueen jaetun ymmärryksen ihmisten ja koneiden kesken.

Tesaurukseen verrattuna ontologia on semanttisesti tarkemmin määritelty. Tesauruksessa käsitteiden väliset hierarkkiset suhteet eivät ole aina tarkkaan määriteltyjä ja ne voivat tarkoittaa esimerkiksi alaluokkasuhdetta, osa-kokonaisuussuhdetta tai luokkayksilösuhdetta. Ontologiassa suhteiden tyyppi on tarkkaan määritelty, ja esimerkiksi edellä mainittujen suhteiden kuvaamiseen käytetään eri suhdetyyppejä.

Tesauruksella kuvataan ainoastaan aihealueen sisältämät käsitteet ja niiden väliset suhteet erityisesti indeksoinnin ja tiedonhaun tarpeita varten. Ontologia puolestaan määrittelee aihealueen formaalimmin ja sitä voidaan käyttää esimerkiksi koneellisessa päättelyssä ja objektien luokittelussa. Toisaalta ontologia voi olla tesauruksen kaltainen, yksinkertainen käsittemalli, joka muodostaa yksinkertaisen käsittehierarkian, mutta ei sisällä juuri muita käsitteiden välisiä suhdetyyppejä tai loogisia rakenteita [15]. Tesaurus voidaan muuntaa ontologiaksi rikastamalla sitä semanttisesti [28, 54].

Ontologiaa voidaan käyttää järjestelmien välisen yhteentoimivuuden saavuttamiseksi, jolloin järjestelmien keskinäinen kommunikaatio perustuu ontologian määrittämään yhteiseen käsitteistöön. Muita ontologian käyttötarkoituksia ovat aineistojen yhdistäminen ja sisältöjen saavuttamisen tehostaminen käyttämällä eri sisältöjen kuvaamiseen yhteismitallista käsitteistöä [47]. Esimerkkejä yleiskäyttöisistä ontologioista ovat Yleinen suomalainen ontologia (YSO) [25] ja Standard Upper Merged Ontology (SUMO) [38].

³<http://www.kansalliskirjasto.fi/kirjastoala/asiasanastot/ysa.html>

2.4 Semanttinen web

Semanttisella webillä tarkoitetaan webiä, jossa tieto on koneymmärrettävässä muodossa tiedon tehokkaamman koneellisen käsittelyn mahdollistamiseksi. Tämä mahdollistaa muun muassa tiedon automaattisen koostamisen useasta lähteestä, koneellisen päättelyn ja tehokkaamman tiedonhaun. Semanttisen webin tekniikoilla voidaan toteuttaa esimerkiksi semanttisia informaatioportaaleja [23, 27, 39], joissa yhdistetään tietoa useista lähteistä yhteisen käyttöliittymän avulla käytettäväksi. Hajautettuun tiedontuotantoon perustuvan semanttisen portaalin toimintaperiaate on esitetty kuvassa 1, jossa perinteisten portaalien tietosisällöt on yhdistetty ontologiaperustaisen metatiedon avulla ja julkaistu semanttisessa portaalisessa.

Semanttinen web perustuu tiedon kuvailuun eli metatiedon tuottamiseen. Metatietoskeemoilla voidaan määrittellä ominaisuudet, joita kuvattavista objekteista kuvataan. Metatietoskeemat mahdollistavat tiedon syntaktisen yhteentoimivuuden määrittäen sen, mitä metatietoja kuvataan, ja missä muodossa nämä esitetään. Jos metatietona käytetään luonnollista kieltä, kone ei pysty käsittelemään metatietoa kovinkaan älykkäästi, koska koneelta puuttuu ihmiselle ominainen kielen ja aihealueen tietämys. Tämän takia käsitteitä ja niiden välisiä suhteita on vaikea tunnistaa, eikä monimutkaista päättelyä pystytä tekemään. Semanttisessa webissä semanttinen yhteentoimivuus mahdollistetaan käyttämällä ontologioita tiedon kuvailussa.

Ontologioiden esittämiseen käytettyjä RDF-tietomalliin [32] perustuvia kieliä ovat RDFS [6] ja OWL [8], ja SKOS [35] puolestaan soveltuu yksinkertaisempien sanastojen määrittelyyn.

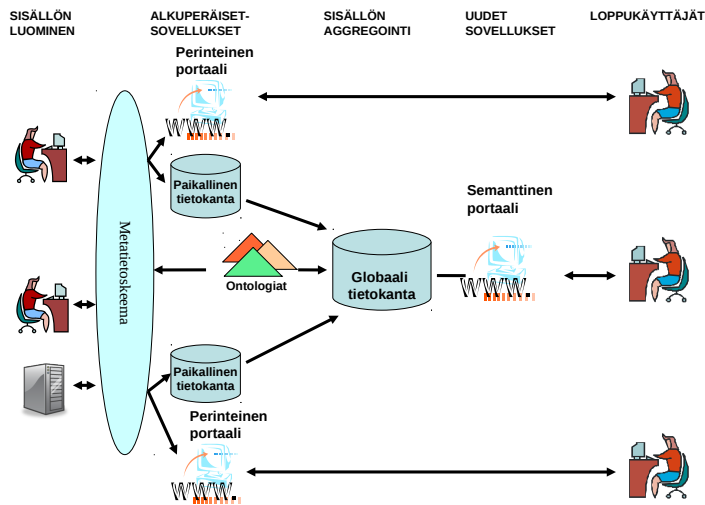
Semanttisessa webissä koneymmärrettävää hajautettua tietoa voidaan yhdistää ja uudelleenkäyttää. Myös ontologiat voivat viitata toisiinsa ja laajentaa toistensa käsiteistöjä. Ontologian käsitteiden välillä voidaan tehdä siltauksia, jolloin eri ontologioilla kuvatut tiedot voidaan yhdistää toisiinsa automaattisesti [11]. Näin on esimerkiksi mahdollista luoda yleiskäsitteistä koostuva yläontologia, jota spesifejä aihealueita kuvaavat ontologiat laajentavat.

Semanttisessa webissä objektit identifoidaan käyttäen URI-tunnisteita. Sekä kuvailtavaan tietoon, metatiet ominaisuuksiin että kuvailussa käytettävään ontologian käsitteisiin viitataan URI-tunnisteilla. Yksikäsitteisen objekteihin viittaamisen lisäksi URI-tunnisteita voidaan käyttää pääsymekanismina tietoon käyttämällä HTTP URI -skeeman mukaisia tunnisteita [41].

2.5 Ongelmat ja haasteet sanastojen käytössä

Jotta toimijat pystyvät käyttämään sanastoja, tarvitaan keinoja niiden julkaisemiseksi, löytämiseksi ja hyödyntämiseksi. Moni webissä julkaistu sanasto on käytettävissä sanaston kehittäjän verkkopalvelussa sanastoa varten kehitetyn käyttöliittymän avulla. Sanastoa käyttääkseen käyttäjän on tällöin löydettävä kyseinen palvelu ja tutustuttava sen käyttöliittymään. Joissain tapauksissa sanasto julkaistaan ainoastaan tiedostona, tai peräti paperidokumenttina. Semanttisessa webissä ontologiat julkaistaan tyypillisesti RDF-tiedostoina. Jos sanasto on saatavilla ainoastaan tiedostomuodossa, toimijan on itse toteutettava tarvitsemansa toiminnallisuudet sanaston käsittelyyn tai hankittava sovellus, joka pystyy hyödyntämään kyseistä sanastoformaattia.

Ontologioiden ja muiden sanastotyyppien käsittelyyn liittyy yleisiä toiminnal-



Kuva 1: Hajautettu tiedontuotanto semanttisessa informaatiportaalissa.

lisuuksia, joita tarvitaan toistuvasti erilaisissa ontologioita hyödyntävissä sovelluksissa. Tällaisiksi toiminnallisuuksia ovat esimerkiksi ontologioiden selaus ja visualisointi sekä käsittehuaku. Yleisten toiminnallisuuksien toteuttaminen ja tarjoaminen toimijoille valmiina palveluina on järkevämpää kuin toiminnallisuuksien toteuttaminen erikseen jokaisen toimijan omassa sovelluksessa. Lisäksi mahdollisuus käyttää eri ontologioita yhteisen käyttöliittymän avulla helpottaa usean ontologian samanaikaista käyttöä. Jos ontologioiden käyttäminen sovelluksissa olisi nykyistä helpompaa, ontologioiden hyödyntäminen ja semanttisen metatiedon tuottaminen todennäköisesti lisääntyisi, ja entistä älykkäämpien palvelujen toteuttaminen olisi mahdollista.

Käyttäessään ontologiaa palveluna toimijalla on aina käytössään ontologian ajantasainen versio. Jos ontologia on palvelun sijaan julkaistu ainoastaan tiedostomuodossa, toimijan on säännöllisesti tar-

kistettava onko ontologiasta julkaistu uusi versio ja otettava se käyttöön, mikäli ontologian ajantasaisuus on toimijan kannalta tärkeää.

3 Ontologiapalvelu

Ontologiapalvelulla tarkoitetaan palvelua, jolla tarjotaan ontologioiden hyödyntämiseen liittyviä toiminnallisuuksia eri käyttäjäryhmille. Seuraavassa käsitellään ontologiapalveluiden käyttäjäryhmiä, heidän käyttötarpeitaan sekä palveluiden tarjoamiseen toteutettuja ontologiapalvelimia.

3.1 Ontologioiden käyttäjäryhmät

FinnONTO-hankkeen alkuvaiheessa ontologioiden käyttäjäryhmiksi tunnistettiin tiedon annotoijat, tiedon hakijat ja ontologioiden kehittäjät [26]. Näiden ryhmien toiminta muodostaa vuorovaikutteisen prosessin, joka on kuvattu kuvassa 2. Kun ontologioiden kehittäjät tuottavat laadukkaita sanastoja, tiedon annotoijat voivat kuvata tietoa tarkoituksenmu-

kaisella tavalla. Yhteismitallisesti ja riittävän tarkasti kuvattu tieto mahdollistaa tiedon hakijoiden löytävän paremmin tietoa. Tehokkaista hakumenetelmistä on hyötyä myös ontologioiden kehittäjien työssä. Kuvassa ontologioiden kehittäjät ylläpitävät mineraalien, toimijoiden ja prosessien aihealueiden ontologioita. Lisäksi ontologioita hyödyntävien sovellusten kehittäjät tarvitsevat palveluita ontologioiden käyttöön ottamiseksi sovelluksissa.

Tiedon annotoija kuvaa sisältöä, kuten museoesineistöä, ja tallentaa sisällön kuvauksen tietojärjestelmään. Yleensä annotointi perustuu tiettyyn metatietoskeemaan, joka määrittää kuvattavalle tiedolle formaalin muodon ja ominaisuudet, joita sisällöstä kuvataan. Metatietoskeema voi olla yksinkertaisimmillaan kokoelma ominaisuus-arvo-pareja, kuten Dublin Core, tai semanttisesti rikkaampi, rakenteinen malli [42, 43], jossa yksittäisten metatietoarvojen välillä on viittauksia.

Annotoijan tehtävänä on löytää oikeat käsitteet kuvaamaan annotoitavaa sisältöä. Tähän tarkoitukseen tarvitaan ontologioiden haku-, selaus- ja visualisointitoimintoja.

Hildebrand ja kumppanit tunnistivat suorittamansa käyttäjätutkimuksen [21] perusteella annotoijan kolme erilaista käyttötapausta:

- Käyttäjä tietää jo entuudestaan, millä käsitteellä haluaa kuvata sisällön, ja haluaa löytää tämän käsitteen nopeasti sanastoista.
- Käyttäjä ei tiedä entuudestaan sopivaa käsitettä sisällön kuvailuun ja hänen pitää tutkia sanastoja löytääkseen sellaisen.
- Käyttäjä epäilee, ettei sanastoista löydy hänen tarvitsemaansa käsitettä, ja hän haluaa varmistaa asian, ennen kuin lisää käsitteen johonkin sanastoista.

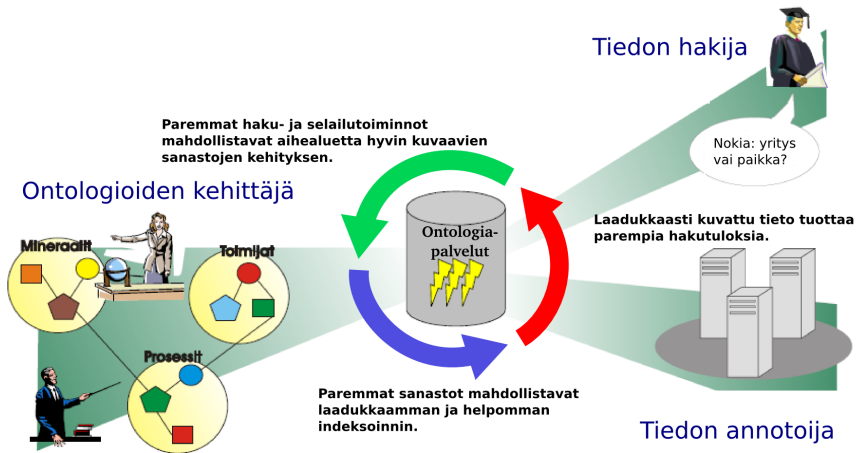
Kuten annotoija myös tiedon hakija tarvitsee ontologioiden haku-, selaus- ja visualisointitoimintoja. Tiedon hakijan tyypillinen käyttötapa on tiedontarpeen kuvaileminen kyselyn muodossa. Hyödyllisiä toiminnallisuuksia ovat tällöin muun muassa käyttäjän ohjaaminen käyttämään tiedon kuvailussa käytettyä sanastoa, sopivien termien löytämisessä avustaminen ja monimerkityksellisten termien merkitysten toisistaan erottaminen. Lisäksi hakukyselyjä voidaan tarvittaessa laajentaa.

Ontologioiden kehittäjät tarvitsevat edellisten käyttäjäryhmien tarvitsemien palveluiden lisäksi ontologioiden kehitys- ja hallintatyökaluja. Työkaluja tarvitaan muun muassa ontologioiden muokkaamiseen, versiointiin, vertailuun, ontologioiden välisten siltauksien tekemiseen sekä yhteisölliseen ontologiakehitykseen [12]. Ontologiakehittäjien tarpeet ovat huomattavasti muita käyttäjäryhmiä vaativammat ja monitahoisemmat. Ontologiakehityksen työkaluja ei käsitellä tässä artikkelissa tarkemmin.

3.2 Ontologiapalvelimet

Ontologiapalveluihin liittyvä keskeinen käsite on ontologiapalvelin, jolla viitataan ontologioiden varastoimiseen ja ontologiapalveluiden tarjoamiseen käytettyyn tietojärjestelmään [1, 14]. Kirjallisuudessa on käytetty myös termejä ontologiakirjastojärjestelmä [10], joka korostaa ontologiapalvelimen roolia ontologioita kokoavana järjestelmänä, sekä ontologiavarasto, joka painottaa ontologioiden varastointia. Myös toteutukseltaan yksinkertaisia, ainoastaan ontologioiden visualisointiin ja selaamiseen tarkoitettuja ontologiaselaimia voidaan ajatella ontologiapalveliminä.

Ontologian elinkaaren eri vaiheissa tarvitaan erilaisia työkaluja [1]. Ontologian suunnitteluvaiheessa tarvitaan ontologiakehitystyökaluja, ontologioiden mallin-



Kuva 2: Ontologioiden käyttäjäryhmät.

nusapuvälineitä sekä versiointi-, julkaisu- ja siltaustoimintoja. Sitoutumisvaiheessa etsitään sopivaa ontologiaa tai sen osaa käytettäväksi esimerkiksi annotoinnissa. Tällöin tarvitaan työkaluja ontologioiden hakuun, visualisointiin, selaukseen ja vertailuun. Käyttövaiheessa ontologioita käytetään varsinaiseen tehtävään, esimerkiksi aineiston annotointiin tai tiedonhakuun. Tällöin oleellisia ovat toiminnallisuudet käsitteiden hakuun, ontologioiden selaukseen ja visualisointiin.

Suurin osa kehitetyistä ontologiapalvelimista on suunnattu ontologioiden suunnitteluvaiheeseen [1], ja siten sitoutumis- ja käyttövaiheen toteutukset ovat aliedustettuja ontologiapalvelinten joukossa. Ontologioiden hyödyntämisen kannalta juuri nämä vaiheet ovat kuitenkin tärkeitä. Koska ontologiapalvelintutkimus on keskittynyt voimakkaasti ontologioiden suunnitteluvaiheeseen, monet tutkijat määrittelevätkin ontologiapalvelimen ensisijaisesti

ontologioiden ylläpidon ja uudelleenkäytön välineenä [10].

Ahmad ja Colomb [1] toteavat, että koska suurin osa kehitetyistä ontologiapalvelimista on suunnattu ontologioiden kehittäjille, niistä puuttuu tärkeitä semanttisessa webissä tarvittavia toiminnallisuuksia. He erottelevat ontologiapalvelinten käyttötarpeet kahteen erilliseen kategoriaan: ontologioiden uudelleenkäyttö ja ontologioiden käyttö. Ontologian uudelleenkäytöllä tarkoitetaan muun muassa ontologian käyttöä uusien ontologioiden perustana ja ontologioiden välisten siltausten tekoa. Ontologian käytöllä puolestaan viitataan ontologian varsinaiseen käyttöön esimerkiksi tiedon kuvailussa tai tiedonhaussa.

Ontologiapalvelin voi tarjota palveluita ihmis- ja konekäyttäjille, käyttöliittyminä ja ohjelmallisina rajapintoina. Ihmiskäyttäjille suunnatut ontologiapalvelut ovat usein tiettyyn käyttötärpeeseen vastaavia

selainperustaisia sovelluksia. Ohjelmallisia rajapintoja puolestaan voidaan toteuttaa esimerkiksi WWW-sovelluspalveluina WSDL- ja SOAP-protokollalla, tai REST-arkkitehtuurin mukaisesti pelkkää HTTP-protokollaa käyttäen. Rajapintojen avulla ontologiapalvelimen toiminnallisuuksia voidaan integroida toimijoiden omiin sovelluksiin.

Tyypillisiä ontologiapalvelimien käyttöliittymätoteutuksissa toistuvia elementtejä ovat ontologian visualisointi, selaus ja käsitteen haku. Näitä voidaan perustellusti pitää oleellisina ontologian käyttämiseen liittyvinä toimintoina, joita tarvitaan muun muassa ontologian kehittämisessä, ontologian käytössä annotoinnissa ja tiedonhaussa.

Ontologian visualisointi auttaa ontologiaan tutustuvaa käyttäjää hahmottamaan ontologian sisältöä, rakennetta, laajuutta ja muita ominaisuuksia. Tästä on hyötyä esimerkiksi valittaessa ontologiaa tiettyyn käyttötarkoitukseen ja vertaillessa ontologioita. Visualisointi on tärkeää myös tietyn ontologian käsitteen semanttikkaa hahmotettaessa. Käsitteen merkityksen ymmärtämisessä auttaa käsitteen ontologisen kontekstin, eli käsitteeseen hierarkkisilla suhteilla sekä muilla tavoin liittyvien käsitteiden näyttäminen. Visualisointia voidaan käyttää myös ontologian selailussa. Esimerkiksi tiedon annotoijan käyttötapauksessa, jossa käyttäjä ei tiedä entuudestaan sopivaa käsitettä tiedon kuvailuun, visualisointi voi auttaa käyttäjää tarkentamaan tai yleistämään alustavasti valittua käsitettä käsittehierarkian avulla tai löytämään muutoin käsitteeseen liittyviä käsitteitä, jotka sopivat käyttötarkoitukseen [21].

Ontologian hierarkkisen luonteen vuoksi sille intuitiivinen ja yksinkertainen visualisointitapa on käsittehierarkian esittäminen puumuodossa [29]. Toinen tyy-

pillinen ontologian visualisointitapa on verkkomainen esitys, joka on luonteva tapa semanttisen webin tiedon visualisointiin RDF-tietomallin verkkomuotoisuuden vuoksi.

Tyypillisesti ontologiapalvelintoteutuksissa ontologiaa voidaan selata visualisointinäkömällä avulla. Näkymässä kuvattut käsitteet toimivat linkkeinä käsitteiden tarkempaan tarkasteluun. Monissa visualisointitekniikoissa näkömää päivittyy ontologiaa selatessa. Tyypillisesti puuesitystavoissa valitun käsitteen alakäsitteet tulevat näkyviin puuhun, ja verkkomaisissa esitystavoissa valittu käsite asetetaan näkömää keskelle. Hyperbolisessa tasossa valittu käsite ja sen semanttisesti läheiset käsitteet visualisoidaan suurempina kuin semanttisesti kauempana olevat käsitteet.

Käsitteen hakutoiminnallisuudet auttavat käyttäjää löytämään käyttötarkoitukseen sopivan käsitteen. Erityisesti automaattisesti täydentyvä haku on tehokas käyttötapauksessa, jossa tiedon annotoija tietää entuudestaan, millä käsitteellä haluaa kuvata sisällön. Käsittehaun tulosten yhteydessä on järkevää näyttää käsitteen nimien lisäksi muita tietoja käsitteistä, jotta käsitteiden merkitys olisi mahdollisimman selvä käyttäjälle käsitteen valinnan helpottamiseksi [21]. Moneen ontologiaan kohdistuvaa hakua voidaan hyödyntää ontologioiden vertailuun.

Ontologiapalvelintoteutusten tyypillisiä ontologian ohjelmallisen käsittelyyn tarkoitettuja toiminnallisuuksia ovat käsitteen haku, käsitteen ominaisuuksien selvittäminen, ontologian haku metatietojen perusteella ja ontologian metatietojen selvittäminen. Jotkin ontologiapalvelimet mahdollistavat ontologioiden käytön päätelyssä, tarjoavat toiminnallisuuksia ontologioiden kehitykseen tai sisältävät palveluja ontologioiden välisten muunnosten tekemiseen.

4 Ontologiapalvelu ONKI

Ontologioiden hyödyntämiseen liittyvien yleisten toiminnallisuuksien tarjoamiseksi on toteutettu ontologiapalvelu ONKI⁴ [28, 50], joka on osa FinnONTO-hankkeessa kehitettyä ontologiainfrastruktuuria. Palvelun tarkoituksena on tarjota käyttäjille ontologiapalveluita erilaisten ontologiatyyppien käyttämiseen tarkoitettujen ontologiapalvelinten ja muiden ohjelmistokomponenttien avulla. Palveluun on toteutettu ONKI SKOS -palvelin [45] yleisille sanastotyyppisille käsiteontologioille, ONKI Geo [30] paikkatiedon ontologioille ja ONKI People [33] toimijaontologioille. Ontologiatyyppikohtaisista ONKI-palvelimista tässä tutkielmassa keskitytään käsittelemään ONKI SKOS -palvelinta. ONKI-palvelun yleisarkkitehtuuri on esitetty kuvassa 3. ONKI-palvelu tarjoaa ontologioiden hyödyntämiseen soveltuvia, ihmiskäyttäjille suunnattuja palveluita sekä ohjelmallisia rajapintoja.

4.1 Mihin ONKIa käytetään?

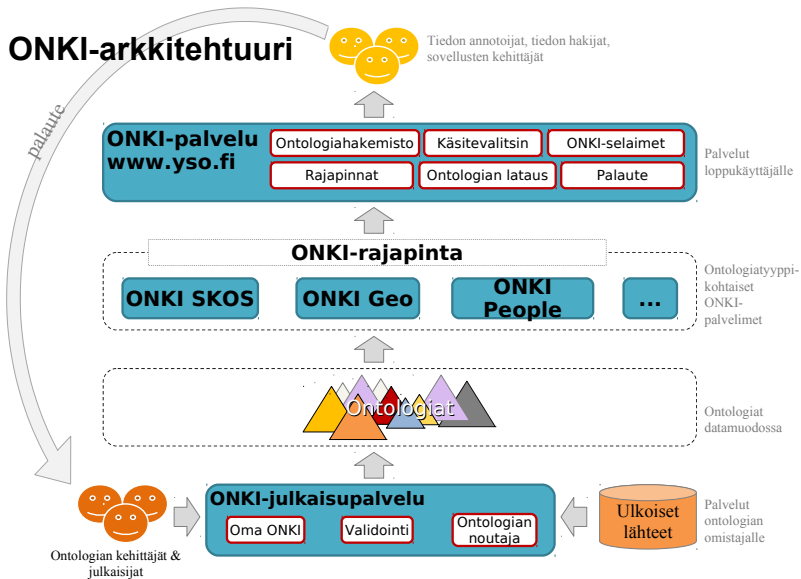
ONKI-palvelua on pilotoitu julkaisemalla useita sanastoja ja ontologioita erilaisilta aihealueilta. Aihealueita ovat muun muassa kulttuuri (Art & Architecture Thesaurus AAT, 28 000 käsitettä), terveyden edistäminen (Medical Subject Headings MeSH, 24 000 käsitettä), liiketalous (United Nations Standard Products and Service Code UNSPSC, 21 000 käsitettä), maantiede (The Getty Thesaurus of Geographic Names TGN, 143 000 käsitettä), yleistieto (Yleinen suomalainen ontologia YSO, 21 000 käsitettä; Library of Congress Subject Headings, 370 000 käsitettä), hallinto (Valtioneuvoston asiansasto VNAS, 6 300 käsitettä) ja luonnontieteet (Maailman linnut -ontologia, 11 000 käsitettä). Palvelussa on julkaistu tällä hetkellä yli 70 ontologiaa.

ONKI-palvelu on ollut julkisessa käytössä syyskuusta 2008 lähtien. Palvelulla on kuukausittain yli 10 000 kävijää, jotka tekevät yhteensä n. 200 000 sivulatausta. Lukumäärissä ei ole mukana integroitavilla käyttöliittymäkomponenteilla tai ohjelmallisilla rajapinnoilla tehtyjä hakuja. Käyttöliittymäkomponenttien ja ohjelmallisten rajapintojen käyttöön tarvittavan ONKI-avaimen on rekisteröinyt itselleen n. 300 käyttäjää.

Palvelussa julkaistut ontologiat on esitetty ”Ontologiat ja sanastot” -näkyvässä luettelona, jonka sisältämät tiedot perustuvat ontologioiden metatietoihin. Metatiedot on tuotettu tarkoitukseen määritettyä – Dublin Core -standardiin perustuvaa – metatietoskeemaa käyttäen. Ontologioista on kerrottu niiden nimi, mahdollinen kuvausteksti ja lisenssiehdot. Luettelo sisältää ontologiakohtaisesti hyperlinkkejä ontologian palveluihin, kuten ontologiaselaimen ja ontologian RDF-muotoisen lähdetiedoston lataamiseen. Ontologiat on luokiteltu ryhmiin niiden helpomman tarkastelun mahdollistamiseksi. Ontologialuettelo voi suodattaa muun muassa aiheen, tyyppin tai julkaisijan perusteella. Aiheluokittelu jakautuu yleiskäsitteitä sisältäviin ontologioihin sekä alakohtaisiin ontologioihin, kuten julkishallinnon, kulttuurin, liike-elämän, luonnon, paikkatiedon ja terveyden ontologioihin.

Palvelun ontologialuettelo tarjoaa tiettyyn tarpeeseen sopivaa ontologiaa etsivälle käyttäjälle yleiskuvan palvelussa julkaistuista ontologioista helpottaen ontologian valintaprosessia. Käyttötarpeita voivat olla esimerkiksi tietyn aineiston annotointi ontologialla tai ontologian käyttö tiedonhaun apuna. Käyttäjä voi vertailla tietyn aihealueen ontologioita keskenään tutkimalla niiden metatietoja sekä sisältöjä ontologiaselainten avulla. Sopivaa

⁴<http://onki.fi>



Kuva 3: ONKI-ontologiapalvelun yleisarkkitehtuuri.

ontologiaa etsiessään käyttäjä voi myös hyödyntää palvelun hakutoiminnallisuuksia, joilla voidaan hakea käsitteitä yhdestä tai useammasta palvelun ontologiasta. Esimerkiksi kaikkiin ontologioihin kohdistuvalla haulla voidaan selvittää löytyykö palvelusta ontologiaa tietyllä hakusamalla.

Ontologian kehittäjille ONKI-palvelu tarjoaa kustannustehokkaan julkaisukanavan ”Oma ONKI” -palvelun muodossa. Ontologian kehittäjä voi lähettää ontologiansa julkaistavaksi ONKI-palveluun tarkoitusta varten tehdyllä lomakkeella. Lomakkeeseen täytetään lisäksi ontologian metatietoja, kuten ontologian nimi, käyttöehdot ja omistajan tiedot. ONKI-palvelun ylläpito käy läpi palveluun lähetetyt ontologiat ja julkaisee laatuksiteerit täyttävät ontologiat. Laatuksiteereihin kuuluu muun muassa syntaktinen validius ja lisensointiehtojen avoimuus.

ONKI-palvelun ontologiat julkaistaan HTML- ja RDF-muodossa ihmis- ja konekäyttäjille HTTP-protokollan avulla sekä erilaisissa rakenteisissa muodoissa ohjelmallisten rajapintojen kautta. Ontologioiden HTML- ja RDF-julkaisu tapahtuu noudattaen W3C:n ohjeita semanttisen webin URI-käytännöistä [41] ja RDF-sanastojen julkaisemisesta [5].

W3C:n ohjeiden mukaan julkaistujen ontologioiden ja niiden sisältämien resurssien (käsitteiden) URI-tunnisteet toimivat URL-osoitteina mahdollistaen yksinkertaisen viittaamisen niiden määrittelyyn. URI-tunnistetta vastaavaan URL-osoitteeseen tehtävään HTTP-pyyntöön annetaan vastauksena URI-tunnisteen yksilöimän resurssin kuvaus, jonka muoto määrittyy HTTP-sisältöneuvottelua käytetään. Esimerkiksi useimmat selaimet lähettävät HTTP-pyyntöön *Accept*-kentässä arvon ”text/html”, jolloin vastauksena an-

netaan HTML-esitys. Vastaavasti resurs- sin kuvauksen RDF-muodossa haluavat ohjelmat käyttävät *Accept*-kentän arvoa ”application/rdf+xml” saaden vastauksen RDF-muodossa.

4.2 Ontologiapalvelin ONKI SKOS

Ontologiapalvelin ONKI SKOS on te- saurusten ja yksinkertaisten ontologioiden julkaisemiseen ja hyödyntämiseen suun- nattu ontologiapalvelintoteutus. Se toteut- taa yleisen ONKI-rajapinnan ja on sit- ten hyödynnettävissä ONKI-palveluiden avulla käyttöliittymäkomponenttina ja oh- jelmallisena rajapintana.

Semanttisen webin sovellukset käyttä- vät tyypillisesti ontologioita, jotka ovat joko suoria muunnoksia laajassa käytös- sä olevista tesauksista, sovelluskohtai- sia sanastoja tai yksinkertaisia ontologioi- ta, jotka voidaan esittää lähes samalla ta- valla kuin tesaukset [3, 15, 28, 31]. Kos- ka SKOS-standardi määrittää sopivan tie- tomallin tesauksen esittämiseen, se va- littiin ensisijaiseksi ONKI SKOS -palveli- men sanastojen esitystavaksi. Palvelimel- la voidaan julkaista myös RDFS- ja OWL- ontologioita, mutta esimerkiksi monimut- kaisten OWL-kielen rakenteiden käsittely on puutteellista.

Kuvaamalla sanasto SKOS-muodossa ja julkaisemalla se ONKI SKOS -palve- limella saavutetaan useita etuja. Koska SKOS määrittää yleisen formaatin sanas- tojen esittämiseen, eri tesauksia voidaan käsitellä samalla tavalla. Tällöin sovelluk- sissa ei tarvita sanastokohtaisia proses- sointisääntöjä tai muunnostyökaluja for- maattien välillä. ONKI SKOS tarjoaa pää- syn kaikkiin palvelimella julkaistuihin sa- nastoihin samalla tavalla, joten sanastojen käyttäjien ei tarvitse käyttää sanastokoh- taisia selaimia ja muita työkaluja. Lisäksi yksi ONKI-ontologiapalvelun tavoitteista on tarjota pääsy laajaan valikoimaan sa-

nastoja keskitetysti, jolloin sanastoja ei tarvitse etsiä useasta lähteestä. Sanaston kehittäjälle palvelu tarjoaa valmiin sanas- ton julkaisukanavan.

4.3 Ontologiaselain

ONKI-palveluun on toteutettu ontologia- selain, jonka avulla voidaan selata ja vi- sualisoida ONKI SKOS -ontologiapalve- limella julkaistuja sanastoja, sekä hakea niistä käsitteitä. Ontologiaselaimen käyt- töliittymä on esitetty kuvassa 4. Käyttö- liittymä koostuu kolmesta pääosasta: 1) automaattisesti täydentyvä käsitehaku, 2) käsitehierarkia ja 3) käsitteen muut omi- naisuudet.

Käsitehaku perustuu semanttiseen au- tomaattisesti täydentyvään hakuparadig- maan [22], jossa hakutulosta päivitetään käyttäjän kirjoittaessa hakutermiä teks- tikenttään. Jos käyttäjä kirjoittaa teks- tikenttään esimerkiksi merkkijonon ”te- levi”, kaikki käsitteet, joiden nimi al- kaa ”televi”-merkkijonolla, palautetaan hakutuloksina. Automaattisesti täydenty- vää käsitehakua käytettäessä käyttäjän ei tarvitse tuntea käytettävää ontologiaa en- tuudestaan, koska riittää, että hän kirjoit- taa hakukenttään jotakin, ja järjestelmä ehdottaa sopivia käsitteitä. Haku täsmäy- tetään sekä ensisijaisiin että toissijaisiin käsitteiden nimiin. Jos haku täsmää kä- sitteen toissijaiseen nimeen, käyttöliitty- mässä näytetään toissijaisen nimen lisäk- si käsitteen ensisijainen nimi sekä siihen osoittava nuoli ohjaamaan käyttäjää käyt- tämään käsitteestä suositeltua nimeä.

Käsitehaun hakutulokset toimivat link- keinä käsitteen tarkastelua varten. Valitun käsitteen käsitehierarkiana näytetään tyy- pillisesti käsitteen ylä-, ala- ja vieruskäsit- teet. Käsitehierarkia visualisoidaan puu- maisena sisennettynä listana. Käsitehie- rarkian lisäksi valitun käsitteen ominai- suudet näytetään käyttöliittymässä tauluk-

The screenshot shows the ONKI ontology browser interface. On the left, there is a search sidebar with a search bar containing 'televi' (labeled 1) and a list of search results including 'televisio'. The main area displays the 'televisio' concept (labeled 2) with a hierarchical tree structure showing related terms like 'yso-käsitteet', 'pysyvä', 'järjestelmät', 'muut järjestelmät', 'viestimet', 'joukkoviestimet', and 'televi'. Below the tree, there is a table of metadata for 'televisio' (labeled 3), including 'Lähde', 'Rinnakkaistermi', 'Tyypit', 'URI', 'Nimikkeet ja vastaavat käsitteet', and 'Jaa'.

Kuva 4: ONKI-palvelun ontologiaselain. Käyttäjä on tehnyt käsitehaun merkkijonolla ”televi” (1). Tarkasteltavaksi valitun käsitteen ”televisio” käsitehierarkia on visualisoitu puumodossa (2) ja muut ominaisuudet taulukkona (3).

komaisena esityksenä. Käsitehierarkiassa ja ominaisuustaulukossa viittaukset muihin käsitteisiin toimivat linkkeinä viitattun käsitteen tarkastelua varten mahdollistaen ontologian selaamisen. Viitattujen käsitteiden niminä käyttöliittymässä näytetään valitun käyttöliittymäkielen mukainen nimi, jos sellainen on ontologiassa saatavilla. Valitun käsitteen erikieliset nimet sekä semanttisesti vastaaviksi määritellyt käsitteet muissa ontologioissa on ryhmitelty omaksi osioksi käsitteen ominaisuustaulukossa.

Paikkatietoa sisältävien ontologioiden yhteydessä paikkainstanssit voidaan visualisoida kartalla käyttämällä Google Maps -palvelua⁵, mikäli paikkainstansseihin on liitetty koordinaattitieto.

Käsitteen ontologisen kontekstin visualisoiminen auttaa käyttäjää ymmärtä-

mään käsitteen merkityksen, erityisesti mikäli merkitys ei käy suoraan ilmi käsitteen nimestä. Lisäksi nimeltään samankaltaisten käsitteiden merkitykset voidaan erottaa toisistaan käsitteiden konteksteja tutkimalla. Käyttötilanteeseen alunperin valittua paremmin soveltuvan käsitteen löytäminen on mahdollista esimerkiksi käsitteen käsitehierarkiaa tutkimalla. Yläkäsitteet tarjoavat yleisempää ja alakäsitteet tarkempaa näkökulmaa.

Käsitehaun lisäksi ontologian selaaminen voidaan aloittaa ontologian hakemistonäkymästä. Ontologian rakenteesta riippuen ontologialla voi olla seuraavia hakemistonäkymiä: 1) aakkosellinen hakemisto, 2) ryhmähakemisto, ja 3) juurikäsitehakemisto. Kaikilla ontologioilla oleva aakkosellinen hakemisto listaa ontologian käsitteitä aakkostettuna. Ryhmä-

⁵<http://maps.google.com>

hakemisto puolestaan sisältää ontologias-
sa olevat käsitteitä ryhmittelevät katego-
riat. Juurikäsitehakemistossa on lueteltu
ontologian ylimmän tason käsitteet. Ha-
kemistonäkymien tarkoituksena on myös
helpottaa yleiskäsityksen saamista ontolo-
gian sisällöstä.

4.4 Käsitevalitsin ja ohjelmalliset rajapinnat

Tyypillisiä tapoja sanastokohtaisten jul-
kaisujärjestelmien hyödyntämiseen sisäl-
lön kuvailussa ja haussa ovat 1) niiden se-
lainkäyttöliittymän käyttäminen sopivan
käsitteen löytämiseksi ja käsiteviittauksen
(käsitteen nimi tai muu tunniste) välittä-
minen omaan järjestelmään manuaalisesti
kopioiduista menetelmällä tai käsin kirjoit-
tamalla tai 2) kyselyiden tekeminen oh-
jelmallisen rajapinnan avulla [44]. Kum-
pikaan tavoista ei ole ongelmaton. Ensim-
mäisessä tavassa käyttäjä joutuu jatkuvasti
liikkumaan kahden sovelluksen välillä ja
välittämään käsiteviittauksen melko kömpelöllä tavalla. Jälkimmäisessä tavassa sa-
naston käyttäjätaho joutuu itse toteutta-
maan käyttöliittymän sanaston käyttämistä varten.

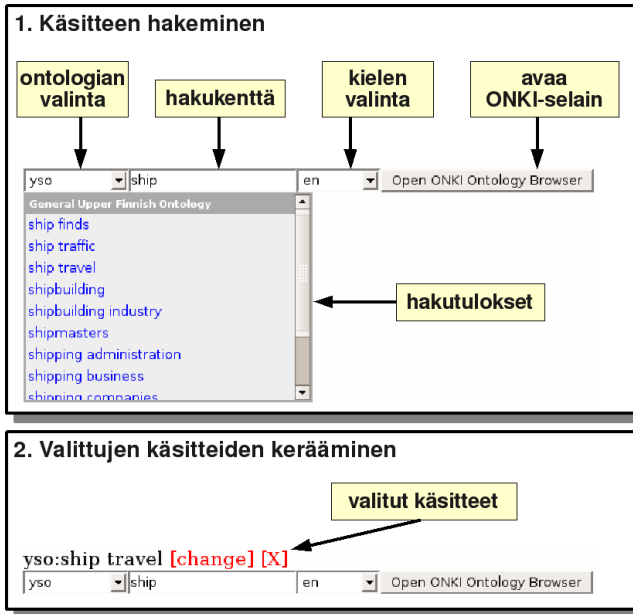
ONKI-palvelua voidaan käyttää mo-
lemmilla yllä kuvatuilla tavoilla, mut-
ta palvelu tarjoaa lisäksi käsitteiden va-
lintaan kevyen verkkosivulle integroitava-
n käyttöliittymäkomponentin [37, 49].
Käsitevalitsimen avulla yleiset ontologian
käyttämiseen liittyvät toiminnallisuudet
voidaan ottaa kustannustehokkaasti käyt-
töön verkkosovelluksessa.

ONKI-palvelun käsitevalitsin on erityi-
sesti tiedon annotoijan ja hakijan käyttö-
tarpeisiin kehitetty käyttöliittymäkompo-
nentti, jonka avulla asiakasjärjestelmään
voidaan noutaa viitteitä sanastojen kä-
sitteisiin. Käsitevalitsin on kuvattu ku-
vassa 5. Käsitevalitsimella voidaan tehdä
merkkijonohakua ontologiaan (kuvan
5 osa 1), selata ontologiaa ja kerätä käsi-

teviittauksia käsitekeräimeen (kuvan 5 osa
2). Samantyyppistä ideaa käyttöliittymä-
tason käsitehakukomponentista on käytet-
ty myös OCLC Terminology Web Ser-
vices -palvelussa [51], BioPortalissa (On-
tology Widgets) [40] sekä yleisesti RDF-
varaston yhteydessä [20].

Käsitevalitsimen käsitteen merkkijonohaku toimii samalla tavalla kuin ONKI-selaimessa täydentymällä automaattisesti käsitteisiin ja täsmäyttämällä hakutermin käsitteiden ensisijaisiin ja toissijaisiin nimiin. Valitun ontologian ONKI-selain voidaan avata painamalla käsitevalitsimen painiketta ”Avaa ONKI-selain”. ONKI-selaimella voidaan selata ja tarkastella ontologiaa. Sopivan käsitteen löytyessä voidaan painaa painiketta ”Nouda käsite”, jolloin viite valittuun käsitteeseen lisätään käsitevalitsimen käsitekeräimeen. Tällä tavoin päästään eroon kömpelöstä kopioiduista menetelmästä, jota joudutaan usein käyttämään perinteisten sanastose-
lainten yhteydessä.

Kun käsite valitaan käsitevalitsimella, se lisätään käsitekeräimen ihmisluet-
tavan muodon lisäksi HTML-sivun piilotettuun tekstikenttään. Käsiteviittaukset voidaan välittää asiakasjärjestelmän taustapalvelimelle HTML-lomakkeella, jonka mukana piilokenttien tiedot kulkevat. Käsiteviittauksina voidaan käyttää käsitteiden URI-tunnisteita, nimiä tai molempia samanaikaisesti. Näin käsitevalitsinta voidaan käyttää sekä semanttisen webin sovelluksissa että järjestelmissä, jotka eivät käytä URI-tunnisteita. Jos käsiteviittauksina käytetään vain käsitteiden nimiä, viittauksien yksilöivyyden kannalta on järkevää tallentaa nimen lisäksi käsitteen sisältävän ontologian tunniste. Jos ontologian käsitteiden nimet ovat uniikkeja ontologiassa, eivätkä ne muutu, tämä viittausta-
pa on kohtuullinen tiedon tallennuksen ja tulevaisuuden käytön kannalta.



Kuva 5: ONKI-palvelun käsitevalitsin.

Käsitevalitsinta voidaan käyttää tiedonhakuun varten sopivien hakutermin löytämiseen ontologioista. Lisäksi hakutermeistä muodostettua kyselyä voidaan laajentaa ontologisen tietämyksen perusteella [46]. Kyselyn laajentamistoiminnallisuus on tarkoitettu erityisesti järjestelmiin, jotka eivät käytä haun täsmäytyksessä ontologioita, vaan perustuvat yksinkertaiseen merkkijonojen vertailuun. Kyselyyn voidaan lisätä valittuihin käsitteisiin liittyviä käsitteitä käyttäen esimerkiksi käsitteiden välisiä hierarkkisia tai assosiatiivisia suhteita.

Käsitevalitsimen käyttöönoton helpottamiseksi ONKI-palveluun on toteutettu palvelu⁶, jolla voidaan luoda käsitevalitsimen integrointiin tarvittava JavaScript-koodi. Käsitevalitsimen halutut parametrit asetetaan lomakkeella, jonka lähettämisen

⁶<http://onki.fi/widget/selector/>

jälkeen tarvittava koodi sekä parametrisoitu käsitevalitsin näytetään käyttäjälle.

ONKI-palvelun ohjelmallisia rajapintoja voidaan käyttää ONKI-palveluiden hyödyntämiseksi tilanteissa, joihin käsitevalitsimen tarjoamat toiminnallisuudet eivät riitä. Toinen rajapintojen käyttökohde on muut kuin selainperustaiset järjestelmät, joihin käsitevalitsinta ei voida integroida. Esimerkiksi Saha-annotaatioeditorissa [48] käytetään ONKI-palvelun rajapintaa käsitehakuun ja käsitteiden tietojen näyttämiseen. Saha-annotaatioeditoria käytetään muun muassa TerveSuomi- [27], Kulttuurisampo- [24] ja Kirjasampo-portaalien [36] tiedontuotantoon.

ONKI-palvelun rajapinnat sisältävät metodeita käsitteiden hakuun ja käsitteiden tietojen sekä ontologioiden

metatietojen ja käsittehakemistojen kyselyyn. ONKI tarjoaa rajapintoja sekä WWW-sovelluspalveluna WSDL- ja SOAP-protokollalla, että pelkällä HTTP-protokollalla käytettäväksi⁷.

ONKI-palvelimien WWW-sovelluspalveluna käytettävän rajapinnan testaamista varten on tehty demonstraatio sivu⁸. Sivulla voidaan tehdä SOAP-protokollan mukaisia pyyntöjä valitun ontologian tietoihin. Ihmisluettavan vastauksen lisäksi lähetetty pyyntö sekä ONKI-palvelimen vastaus näytetään myös SOAP-standardin käyttämässä XML-muodossa.

5 Yhteenveto

Semanttisen webin mahdollistamat älykkäät palvelut perustuvat formaalisti määritettyyn metatietoon. Ontologiat esittävät tietämyksen eksplisiittisinä käsitteistöinä, jotka kuvaavat käsitteiden semantiikan koneymmärrettävässä muodossa mahdollistaen muun muassa entistä tehokkaamman tietojen yhdistämisen ja haun.

Tässä artikkelissa käsiteltiin ontologioiden käyttäjäryhmien tarpeiden ja ontologiapalvelinten toteutusten perusteella tunnistettuja ontologioiden hyödyntämiseen liittyviä yleisiä, usein toistuvia toiminnallisuuksia. Ontologioiden hakuun, selaamiseen ja ymmärtämiseen liittyviä toiminnallisuuksia esitetään tarjottavaksi palveluina, jolloin jokaisen toimijan ei tarvitse toteuttaa yleisiä toiminnallisuuksia itse. Lisäksi käytettäessä ontologiaa palveluina toimijan käytössä on aina ontologian ajantasainen versio.

Yleisen ontologiapalvelun prototyyppitoteutuksena esiteltiin ontologiapalvelu ONKI, joka perustuu erilaisten ontologiatyyppien julkaisemiseen suunnattuihin ontologiapalvelintoteutuksiin. ONKI

SKOS on tesaurusten ja yksinkertaisten ontologioiden hyödyntämiseen tarkoitettu ontologiapalvelin. ONKI-palvelu tarjoaa valmiita käyttöliittymäkomponentteja ja ohjelmallisia rajapintoja muun muassa tiedon annotointiin ja hakuun sekä työkaluja ontologioita hyödyntävien sovellusten kehittäjille.

Kiitokset

Tämä työ on tehty osana FinnONTO – Suomalaiset semanttisen webin ontologiat (2003–2012) -hankesarjaa⁹, jota ovat rahoittaneet Teknologian ja innovaatioiden kehittämiskeskus Tekes ja eri aikoina yli 40 julkisen alan organisaatiosta ja yrityksestä koostuva konsortio. Työtä hyödynnetään Linked Data Finland (2012–2014) -hankkeessa¹⁰, jota rahoittavat Teknologian ja innovaatioiden kehittämiskeskus Tekes ja 22 julkisen alan organisaatiosta ja yrityksestä koostuva konsortio.

Viitteet

1. Mohammad Nazir Ahmad ja Robert M. Colomb. Managing ontologies: a comparative study of ontology servers. *Proceedings of the Eighteenth Conference on Australasian Database (ADC'07)*, s. 13–22, Ballarat, Victoria, Australia, tammi-kuu 2007. Australian Computer Society.
2. Jean Aitchison, Alan Gilchrist ja David Bawden. *Thesaurus Construction ja Use: A Practical Manual*. Europa Publications, 4. painos, 2000.
3. Mark van Assem, Maarten R. Menken, Guus Schreiber, Jan Wielemaker ja Bob Wielinga. A method for converting thesauri to RDF/OWL. *Proceedings of the 3rd International Semantic Web Conference (ISWC'04)*, s. 17–31, Hiroshima, Japan, marraskuu 2004. Springer-Verlag.

⁷<http://onki.fi/api/>

⁸<http://onki.fi/api/v1/soap/>

⁹<http://www.seco.tkk.fi/projects/finnonto/>

¹⁰<http://www.seco.tkk.fi/projects/ldf/>

4. Tim Berners-Lee, James Hendler ja Ora Lassila. The semantic web. *Scientific American*, 284(5):34–43, 2001.
5. Diego Berrueta ja Jon Phipps. Best practice recipes for publishing RDF vocabularies, W3C working group note, 2008. <http://www.w3.org/TR/2008/NOTE-swbp-vocab-pub-20080828/> (haettu 15.3.2012).
6. Dan Brickley ja R. V. Guha. RDF vocabulary description language 1.0: RDF schema, W3C recommendation, 2004. <http://www.w3.org/TR/2004/REC-rdf-schema-20040210/> (haettu 15.3.2012).
7. Warwick Cathro. Metadata: an overview. *Proceedings of the Standards Australia Seminar: Matching Discovery and Recovery*, Sydney and Melbourne, Australia, elokuu 1997.
8. Mike Dean ja Guus Schreiber. OWL web ontology language reference, W3C recommendation, 2004. <http://www.w3.org/TR/2004/REC-owl-ref-20040210/> (haettu 15.3.2012).
9. Vladan Devedžić. Understanding ontological engineering. *Communications of the ACM*, 45(4):136–144, 2002.
10. Ying Ding ja Dieter Fensel. Ontology library systems: The key to successful ontology reuse. *Proceedings of the 1st Semantic Web Working Symposium (SWWS'01)*, s. 93–112, Stanford University, California, USA, elokuu 2001.
11. Marc Ehrig ja York Sure. Ontology mapping – an integrated approach. *Proceedings of the 1st European Semantic Web Symposium (ESWS'04)*, s. 76–91, Heraklion, Greece, toukokuu 2004. Springer-Verlag.
12. Adam Farquhar, Richard Fikes ja James Rice. The Ontolingua Server: a tool for collaborative ontology construction. *International Journal of Human-Computer Studies*, 46(6):707–727, 1997.
13. Lee Feigenbaum, Ivan Herman, Tonya Hongsermeier, Eric Neumann ja Susie Stephens. The semantic web in action. *Scientific American*, 297(6):90–97, 2007.
14. Richard Fikes ja Adam Farquhar. Large-scale repositories of highly expressive reusable knowledge. Technical Report KSL-97-02, Stanford Knowledge Systems Laboratory, Stanford, California, USA, 1997. http://www.ksl.stanford.edu/KSL_Abstracts/KSL-97-02.html (haettu 15.3.2012).
15. Christiaan Fluit, Marta Sabou ja Frank van Harmelen. Supporting user tasks through visualisation of light-weight ontologies. Steffen Staab ja Rudi Studer (toim.), *Handbook on Ontologies*, s. 415–434. Springer-Verlag, 2004.
16. Lars Marius Garshol. Metadata? thesauri? taxonomies? topic maps! making sense of it all. *Journal of Information Science*, 30(4):378–391, 2004.
17. Thomas R. Gruber. A translation approach to portable ontology specification. *Knowledge Acquisition*, 5(2):199–220, 1993.
18. David Hawking ja Justin Zobel. Does topic metadata help with web search? *Journal of the American Society for Information Science and Technology*, 58(5):613–628, 2007.
19. Jeff Heflin. *Towards the Semantic Web: Knowledge Representation in a Dynamic, Distributed Environment*. PhD thesis, University of Maryland, College Park, 2001.
20. Michiel Hildebrand, Jacco van Ossensbruggen, Alia Amin, Lora Aroyo, Jan Wielemaker ja Lynda Hardman. The design space of a configurable autocompletion component. Technical Report INS-E0708, Centrum voor Wiskunde en Informatica (CWI), Amsterdam, Netherlands, 2007. <http://www.cwi.nl/ftp/CWIreports/INS/INS-E0708.pdf> (haettu 15.3.2012).
21. Michiel Hildebrand, Jacco van Ossensbruggen, Lynda Hardman ja Geertje Jacobs. Supporting subject matter annotation using heterogeneous thesauri: A user study in web data reuse. *International Journal of Human-Computer Studies*, 67(10):887–902, 2009.
22. Eero Hyvönen ja Eetu Mäkelä. Semantic autocompletion. *Proceedings of the 1st Asian Semantic Web Conference*

- (ASWC'06), s. 739–751, Beijing, China, elokuu 2006. Springer-Verlag.
23. Eero Hyvönen, Eetu Mäkelä, Mirva Salminen, Arttu Valo, Kim Viljanen, Samp-pa Saarela, Miikka Junnila ja Suvi Kettula. MuseumFinland – Finnish museums on the semantic web. *Journal of Web Semantics*, 3(2):224–241, 2005.
 24. Eero Hyvönen, Tuukka Ruotsalo, Thomas Häggström, Mirva Salminen, Miikka Junnila, Mikko Virkkilä, Mikko Haaramo, Eetu Mäkelä, Tomi Kauppinen ja Kim Viljanen. Culturesampo – Finnish culture on the semantic web: The vision and first results. Klaus Robering (toim.), *Information Technology for the Virtual Museum: Museology and the Semantic Web*. LIT Verlag, Berlin, Germany, 2008.
 25. Eero Hyvönen, Katri Seppälä, Kim Viljanen ja Matias Frosterus. Yleinen suomalainen ontologia YSO – kohti suomalaista semanttista webiä. *Tietolinja*, 26(1), 2007.
 26. Eero Hyvönen, Arttu Valo, Ville Komulainen, Katri Seppälä, Tomi Kauppinen, Tuukka Ruotsalo, Mirva Salminen ja Anu Ylisaalmi. Finnish national ontologies for the semantic web – towards a content and service infrastructure. *Proceedings of the International Conference on Dublin Core and Metadata Applications (DC'05)*, s. 219–222, Madrid, Spain, syyskuu 2005. Dublin Core Metadata Initiative.
 27. Eero Hyvönen, Kim Viljanen ja Osmo Suominen. HealthFinland – Finnish health information on the semantic web. *Proceedings of the 6th International Semantic Web Conference (ISWC'07)*, s. 778–791, Busan, Korea, marraskuu 2007. Springer-Verlag.
 28. Eero Hyvönen, Kim Viljanen, Jouni Tuominen ja Katri Seppälä. Building a national semantic web ontology and ontology service infrastructure—the FinnONTO approach. *Proceedings of the 5th European Semantic Web Conference (ESWC'08)*, s. 95–109, Tenerife, Spain, kesäkuu 2008. Springer-Verlag.
 29. Akrivi Katifori, Constantin Halatsis, George Lepouras, Costas Vassilakis, and Eugenia Giannopoulou. Ontology visualization methods—a survey. *ACM Computing Surveys*, 39(4), 2007.
 30. Tomi Kauppinen, Riikka Henriksson, Reetta Sinkkilä, Robin Lindroos, Jari Väättäinen ja Eero Hyvönen. Ontology-based disambiguation of spatiotemporal locations. *Proceedings of the 1st International Workshop on Identity and Reference on the Semantic Web (IRSW'08), the 5th European Semantic Web Conference 2008 (ESWC'08)*, Tenerife, Spain, kesäkuu 2008.
 31. Atanas Kiryakov, Borislav Popov, Ivan Terziev, Dimitar Manov ja Damyan Ognyanoff. Semantic annotation, indexing, and retrieval. *Web Semantics: Science, Services and Agents on the World Wide Web*, 2(1):49–79, 2004.
 32. Graham Klyne ja Jeremy J. Carroll. Resource description framework (RDF): Concepts and abstract syntax, W3C recommendation, 2004. <http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/> (haettu 15.3.2012).
 33. Jussi Kurki ja Eero Hyvönen. Authority control of people and organizations on the semantic web. *Proceedings of the International Conference for Digital Libraries and the Semantic Web 2009 (ICSD'09)*, Trento, Italy, syyskuu 2009.
 34. Christopher D. Manning, Prabhakar Raghavan ja Hinrich Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.
 35. Alistair Miles ja Sean Bechhofer. SKOS simple knowledge organization system reference, W3C recommendation, 2009. <http://www.w3.org/TR/2009/REC-skos-reference-20090818/> (haettu 15.3.2012).
 36. Eetu Mäkelä, Kaisa Hypén ja Eero Hyvönen. Booksampo—lessons learned in creating a semantic portal for fiction literature. *Proceedings of the 10th International Semantic Web Conference (ISWC'11)*, s. 173–188, Bonn, Germany, nov 2011. Springer-Verlag.
 37. Eetu Mäkelä, Kim Viljanen, Olli Alm, Jouni Tuominen, Onni Valkeapää, Tomi

- Kauppinen, Jussi Kurki, Reetta Sinkkilä, Teppo Käsälä, Robin Lindroos, Osmo Suominen, Tuukka Ruotsalo ja Eero Hyvönen. Enabling the semantic web with ready-to-use web widgets. *Proceedings of the First Industrial Results of Semantic Technologies Workshop, the 6th International Semantic Web Conference (ISWC'07)*, Busan, Korea, marraskuu 2007.
38. Ian Niles ja Adam Pease. Towards a standard upper ontology. *Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS'01)*, s. 2–9, Ogunquit, Maine, USA, lokakuu 2001. ACM Press.
39. Dave Reynolds, Paul Shabajee ja Steve Cayzer. Semantic information portals. *Proceedings of the 13th International Conference on World Wide Web (WWW'04)*, s. 290–291, New York City, New York, USA, toukokuu 2004. ACM Press.
40. Daniel L. Rubin, Dilvan A. Moreira, Pradip P. Kanjamala ja Mark A. Musen. BioPortal: A web portal to biomedical ontologies. *Proceedings of the Symbiotic Relationships between Semantic Web and Knowledge Engineering, AAAI Spring Symposium Series*, Stanford, California, USA, maaliskuu 2008.
41. Leo Sauermann ja Richard Cyganiak. Cool URIs for the semantic web, W3C interest group note, 2008. <http://www.w3.org/TR/2008/NOTE-cooluris-20081203/> (haettu 15.3.2012).
42. A. Th. (Guus) Schreiber, Barbara Dubbel-dam, Jan Wielemaker ja Bob Wielinga. Ontology-based photo annotation. *IEEE Intelligent Systems*, 16(3):66–74, 2001.
43. Audrey M. Tam ja Clement H. C. Leung. Structured natural-language descriptions for semantic content retrieval of visual materials. *Journal of the American Society for Information Science and Technology*, 52(11):930–937, 2001.
44. Douglas Tudhope ja Ceri Binding. Towards terminology service: experiences with a pilot web service thesaurus browser. *Proceedings of the International Conference on Dublin Core and Metadata Applications (DC'05)*, s. 241–244, Madrid, Spain, syyskuu 2005. Dublin Core Metadata Initiative.
45. Jouni Tuominen, Matias Frosterus, Kim Viljanen ja Eero Hyvönen. ONKI SKOS server for publishing and utilizing SKOS vocabularies and ontologies as services. *Proceedings of the 6th European Semantic Web Conference (ESWC'09)*, s. 768–780, Heraklion, Greece, kesäkuu 2009. Springer-Verlag.
46. Jouni Tuominen, Tomi Kauppinen, Kim Viljanen ja Eero Hyvönen. Ontology-based query expansion widget for information retrieval. *Proceedings of the 5th International Workshop on Scripting and Development for the Semantic Web (SFSW'09), the 6th European Semantic Web Conference (ESWC'09)*, Heraklion, Greece, kesäkuu 2009.
47. Mike Uschold ja Robert Jasper. A framework for understanding and classifying ontology applications. *Proceedings of the Workshop on Ontologies and Problem-Solving Methods, the International Joint Conference on Artificial Intelligence (IJCAI'99)*, Stockholm, Sweden, elokuu 1999.
48. Onni Valkeapää, Olli Alm ja Eero Hyvönen. An adaptable framework for ontology-based content creation on the semantic web. *Journal of Universal Computer Science*, 13(12):1835–1853, 2007.
49. Kim Viljanen, Jouni Tuominen ja Eero Hyvönen. Publishing and using ontologies as mash-up services. *Proceedings of the 4th Workshop on Scripting for the Semantic Web (SFSW'08), the 5th European Semantic Web Conference (ESWC'08)*, Tenerife, Spain, kesäkuu 2008.
50. Kim Viljanen, Jouni Tuominen ja Eero Hyvönen. Ontology libraries for production use: The Finnish ontology library service ONKI. *Proceedings of the 6th European Semantic Web Conference (ESWC'09)*, s. 781–795, Heraklion, Greece, kesäkuu 2009. Springer-Verlag.

51. Diane Vizine-Goetz, Eric Childress ja Andrew Houghton. Web services for genre vocabularies. *Proceedings of the International Conference on Dublin Core and Metadata Applications (DC'05)*, s. 197–200, Madrid, Spain, syyskuu 2005. Dublin Core Metadata Initiative.
52. Ellen M. Voorhees. Query expansion using lexical-semantic relations. *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, s. 61–69, Dublin, Ireland, heinäkuu 1994. Springer-Verlag.
53. Yih-Chen Wang, James Vandendorpe ja Martha Evens. Relational thesauri in information retrieval. *Journal of the American Society for Information Science*, 36(1):15–27, 1985.
54. Bob Wielinga, Guus Schreiber, J. Wielemaker ja J. A. C. Sandberg. From thesaurus to ontology. *Proceedings of the International Conference on Knowledge Capture*, s. 194–201, Victoria, Canada, lokakuu 2001. ACM Press.