

hyväksymispäivä arvosana

arvostelija

Helppokytkenmäiset ontologiapalvelut semanttisessa webissä

Jouni Tuominen

Helsinki 19.5.2010

Pro gradu -tutkielma

HELSINGIN YLIOPISTO

Tietojenkäsittelytieteen laitos

Tiedekunta — Fakultet — Faculty		Laitos — Institution — Department	
Matemaattis-luonnontieteellinen tiedekunta		Tietojenkäsittelytieteen laitos	
Tekijä — Författare — Author			
Jouni Tuominen			
Työn nimi — Arbetets titel — Title			
HelppokytKentäiset ontologiapalvelut semanttisessa webissä			
Oppiaine — Läroämne — Subject			
Tietojenkäsittelytiede			
Työn laji — Arbetets art — Level		Aika — Datum — Month and year	
Pro gradu -tutkielma		19.5.2010	
		Sivumäärä — Sidoantal — Number of pages	
		79 sivua + 13 liitesivua	
Tiivistelmä — Referat — Abstract			
<p>Ontologiat luovat semanttisen webin perustan: 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 datana, tiedostomuodossa. Tällöin jokainen toimija joutuu toteuttamaan itse toiminnallisuuksia ontologioiden hyödyntämiseen. Osa toiminnallisuuksista on yleisiä, useissa järjestelmissä toistuvia, kuten ontologian visualisointi, selaaminen ja käsitteily. On kuitenkin kustannustehokkaampaa toteuttaa yleisiä ontologiatoiminnallisuuksia valmiina palveluina. Palveluita voidaan tarjota ihmiskäyttäjille käyttöliittymäkomponentteina sekä ohjelmalliseen käyttöön rajapintoina, joita käyttämällä toiminnallisuudet voidaan integroida asiakasjärjestelmiin. Lisäksi käytettäessä ontologioita palveluina toimijoiden käytössä on aina ontologioiden ajantasaiset versiot.</p> <p>Tässä tutkielmassa kuvataan ontologioiden käyttäjäryhmien – ontologioiden kehittäjien, tiedon annotoijien, tiedon hakijoiden ja semanttisen webin sovellusten kehittäjien – tarpeita sekä esitellään ontologioiden hyödyntämiseen kehitettyjä sovelluksia. Yleisten ontologioiden käyttämiseen liittyvien toiminnallisuuksien tarjoamiseksi esitetään ontologiapalvelu ONKI, joka julkistettiin virallisesti käyttöön syyskuussa 2008.</p> <p>ACM Computing Classification System (CCS):</p> <p>H.3.5 [Information Storage and Retrieval]: Online Information Services</p> <p>H.5.4 [Information Interfaces and Presentation]: Hypertext/Hypermedia</p> <p>I.2.4 [Artificial Intelligence]: Knowledge Representation Formalisms and Methods</p>			
Avainsanat — Nyckelord — Key words			
ontologiapalvelut, ontologiapalvelin, semanttinen web, ontologiat			
Säilytyspaikka — Förvaringsställe — Where deposited			
Muita tietoja — Övriga uppgifter — Additional information			

Sisältö

1	Johdanto	1
2	Metatiedon yhteismitallisuus ja semanttisuus	3
2.1	Asiasanastot, taksonomiat ja tesaaurukset	5
2.2	Ontologiat	6
2.3	Semanttinen web	6
2.4	Ongelmat ja haasteet sanastojen käytössä	10
3	Ontologiapalvelut	12
3.1	Ontologioiden käyttäjäryhmät	12
3.1.1	Tiedon annotoija	12
3.1.2	Tiedon hakija	14
3.1.3	Ontologian kehittäjä	14
3.1.4	Semanttisen webin sovelluksen kehittäjä	14
3.2	Ontologiapalvelimet	15
3.2.1	Käyttöliittymätoteutukset	16
3.2.2	Ohjelmalliset rajapinnat	22
3.3	Yleiset vaatimukset	26
3.3.1	Ontologian visualisointi ja selaaminen	26
3.3.2	Ontologian hakutoiminnallisuudet	29
3.3.3	Ontologian ohjelmallinen käsitteleminen	30
3.3.4	Ontologioiden ajantasaisuus	30
4	Ontologiapalvelu ONKI	30
5	Ontologiapalvelin ONKI SKOS	35
5.1	Käyttöliittymä	36
5.2	Tekninen toteutus	39
5.3	Sanastojen käsittely	42

	iii
5.4 Olemassa olevan sanaston muuntaminen SKOS-muotoon	45
6 Ontologiapalveluiden kustannustehokas hyödyntäminen	47
6.1 Käsitevalitsin	47
6.2 Ohjelmalliset rajapinnat	56
7 Tulosten tarkastelu	59
7.1 Tutkimusongelmat	59
7.2 ONKI-palvelun vertailu muihin ontologiapalvelimiin	60
7.3 Jatkokehitys	61
8 Yhteenveto	63
Lähteet	64
Liitteet	
1 ONKI SKOS -palvelimen esimerkkikonfiguraatio	
2 ONKI SKOS -palvelimen HTTP-rajapinnan metodikuvaus	

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 (Hypertext Markup Language) [RHJ99] 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 [BHL01, FHH07] on webin laajennos, jossa tieto on koneymmärrettävässä muodossa tiedon aiempaa 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 hakusanaan täsmäävää paikkakuntaa, yritystä, eläintä 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 on kuvattu *ontologioiden* (engl. ontology) avulla. Ontologia on tietyn aihealueen jaettu käsitteistö, jossa on kuvattu formaalisti aihealueeseen kuuluvat käsitteet ja niiden väliset suhteet [Gru93]. Ontologia määrittelee eksplisiittisesti ihmisille ominaisen implisiittisen tietämyksen 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* (engl. annotation) [KPT04], ontologiat täytyy ensin luoda ja julkaista toimijoiden käyttöön. Ontologioiden kehittäjät ovat tyypillisesti alueensa asiantuntijoita ja vastaavat ontologioiden ylläpidosta. Tässä tutkielmassa keskitytään ontologioiden käyttöönottoon ja julkaisemiseen, ja ontologioiden varsinainen kehitystyö [Dev02] rajautuu tutkielman

ulkopuolelle. Ontologiaperustaisen tiedon kuvailun kannalta konkreettinen haaste on, miten saada metatietoa tuottavaan järjestelmään viittauksia ontologioiden käsitteisiin.

Ontologiat voidaan julkaista yksinkertaisimmillaan datana, 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 *ontologia-palveluina* (engl. ontology service) ja julkaista ne toimijoiden käyttöön. Tällainen yksinkertainen tapa käyttää ontologioita madaltaisi toimijoiden kynnystä ottaa ontologioita käyttöönsä ja aloittaa semanttisen tiedon tuottaminen.

Ontologioiden tyypillinen ominaisuus on niiden muuttuminen ajan kuluessa [SMM02]. Kun ontologioiden kehittäjien tietämys ontologian aihealueesta lisääntyy, esimerkiksi tutkijoiden löytäessä uuden eläinlajin, ontologiaa on tarpeen päivittää. Tieto ontologian muutoksista on välitettävä ontologian käyttäjille, jotta eri toimijoiden tuottaman tiedon yhteismitallisuus säilyy.

Tämä tutkielma pyrkii löytämään ratkaisuja seuraaviin tutkimusongelmiin:

- Minkälaisia toiminnallisuuksia toimijat tarvitsevat ontologioiden käyttämiseen?
- Miten ontologioita voidaan julkaista toimijoiden käyttöön?
- Miten voidaan varmistaa käytössä olevien ontologioiden ajantasaisuus?

Edellä mainittuja tutkimusongelmia on aiemmin käsitelty muun muassa ontologia-palveluita tarjoavia sovelluksia vertailevissa tutkimuksissa [DiF01, AhC07]. Ontologiapalveluihin liittyviä vaatimuksia on tutkittu sekä laajana kokonaisuutena [FFR97, BCC08, AqL09, RMK08, BaS09] että liittyen tiettyyn käyttötarkoitukseen, kuten ontologian löytämiseen [ABG07, DFJ04, BED04] tai hyödyntämiseen [BMG06, ERG02, BRM04, VCH05, DNK04].

Tässä työssä tehty tutkimus perustuu ontologioiden käyttäjäryhmien perusteella tehtyyn vaatimusmäärittelyyn, aiemmin tehdyn ontologiapalveluiden tutkimuksen käsittelyyn ja prototyypisovelluksen toteuttamiseen. Tutkielman luvussa 2 käsitellään tiedonhaun tehostamiseen käytettävän metatiedon ja kontrolloitujen sanastojen peruskäsitteitä sekä esitellään semanttisen webin perusteet, minkä jälkeen

kuvataan sanastojen käyttämiseen liittyviä ongelmia ja haasteita. Ontologioiden käyttäjäryhmien tarpeita sekä ontologioiden hyödyntämiseen kehitettyjä sovelluksia käsitellään luvussa 3. Luvun 3 havaintojen perusteella ontologioiden käyttämiseen liittyvien yleisten toiminnallisuuden tarjoamiseksi esitetään ontologiapalvelua, jonka prototyyppi ONKI esitellään luvussa 4. Luvussa 5 käsitellään tarkemmin ONKI-palvelun osana käytettyä ONKI SKOS -ontologiapalvelintoteutusta. ONKI-palveluiden kustannustehokkaita hyödyntämistapoja kuvataan luvussa 6. Tutkielman lopussa luvussa 7 tarkastellaan työn tuloksia sekä luvussa 8 esitetään tutkielman yhteenveto.

Tämä tutkielma perustuu osittain aiempiin ONKI-ontologiapalvelua käsitteleviin julkaisuihin [TFV09, VTH09, VTH08, TKV09]. ONKI-palvelun kehitystyössä on ollut mukana kirjoittajan lisäksi muita henkilöitä, jotka on mainittu tutkielman lopussa olevassa Kiitokset-osiossa. ONKI-ontologiapalvelu on osa Suomalaiset semanttisen webin ontologiat (FinnONTO, 2003-2012) -hankkeessa kehitettyä FinnONTO-infrastruktuuria [HVT08]. ONKI-ontologiapalvelu on kuvattu kokonaisuutena Viljasen ja kumppaneiden vuoden 2009 julkaisussa [VTH09]. Tuomisen ja kumppaneiden vuoden 2009 ensimmäisessä julkaisussa [TFV09] on tarkasteltu olemassa olevien sanastojen hyödyntämistä semanttisessa webissä ONKI SKOS -ontologiapalvelimen avulla. ONKI-ontologiapalvelussa julkaistujen ontologioiden käyttämistä sovelluksiin integroitavan käyttöliittymäkomponentin avulla on kuvattu Viljasen ja kumppaneiden vuoden 2008 julkaisussa [VTH08]. Käyttöliittymäkomponentin käyttämistä tiedonhaun tehostamisessa on käsitelty Tuomisen ja kumppaneiden vuoden 2009 toisessa julkaisussa [TKV09].

ONKI SKOS -ontologiapalvelimen HTTP-rajapinta on yleistetty Viljasen ja kumppaneiden vuoden 2010 julkaisussa [VTS10] yleiseksi ontologiapalvelinten rajapinnaksi, jonka avulla voidaan luoda hajautettu ontologiapalvelinten LOOS-verkosto (engl. Linked Open Ontology Services). Tuomisen ja kumppaneiden vuoden 2010 julkaisussa [TSV10] kuvataan erillisistä ONKI SKOS -ontologiapalvelimista koostuvaan LOOS-verkostoon kehitetty yhteinen käyttöliittymä ONKI2.

2 Metatiedon yhteismitallisuus ja semanttisuus

Metatiedolla (engl. metadata) tarkoitetaan tietoa tiedosta ja se voidaan määrittellä yleisesti miksi tahansa ilmaukseksi jostakin informaatio-objektista [Gar04]. 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 Core [DCM08], 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) [Ste98].

Tekstimuotoisiin objekteihin kohdistuvassa tiedonhaussa [MRS08] metatiedon hyödyllisyys perustuu siihen, että haku voidaan kohdistaa itse tekstisisällön sijaan tai sen lisäksi objektia kuvailevaan tietoon. Yksinkertainen pelkkään tekstisisältöön kohdistuva kokotekstihaku on haasteellista, koska hakukone ei erota tekstistä tietosisällön kannalta merkityksellisiä ja epämerkityksellisiä sanoja toisistaan, mikä johtaa haun *tarkkuuden* (engl. precision) heikentymiseen [Cat97]. Tarkkuuden heikentymistä voidaan kompensoida järjestämällä hakutulokset relevanssin perusteella [MRS08], jolloin hakutehtävän kannalta oletettavasti relevantimmat tulokset ovat hakutulosityoukon alussa. Toisaalta koska tekstissä ei välttämättä esiinny kaikkia tietosisällön kannalta merkityksellisiä sanoja, haun *saanti* (engl. recall) voi heikentyä [Hef01]. Eksplisiittisen metatiedon käyttämisen sijaan tiedonhakua voidaan pyrkiä tehostamaan luonnollisen kielen menetelmillä, esimerkiksi tunnistamalla tekstistä koneellisesti merkityksiä kontekstin perusteella [Cun99]. 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 [HaZ07]. Lisäksi metatieto-ominaisuuksia pitää käyttää yhdenmukaisella tavalla. Metatieto-ominaisuuksien arvot voidaan kuvata käyttämällä esimerkiksi luonnollista kieltä, kontrolloitua sanastoa tai tietyllä tavalla muodostettuja merkkijonoja (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 [Gar04]. Sanastossa voidaan esimerkiksi määrittää, että käsitteeseen “kummitus” viitataan aina termillä “kummitus”, eikä koskaan termillä “haamu”, ja että termi “Nokia” viittaa aina käsitteeseen “Nokia paikkakuntana”. 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 [AGB00].

Kontrolloitu sanasto voi olla rakenteeltaan yksinkertainen asiasanasto [Gar04] tai monimutkaisempi käsitelmä, kuten taksonomia [Gar04], tesaurus [AGB00] tai ontologia [Gru93]. Ontologiat ovat tesauruksia formaalimpia ja mahdollistavat älykkäämmän koneellisen käsittelyn. Toisaalta *yksinkertaiset ontologiat* (engl. *lightweight ontologies*) [FSH04] ovat tesaurusten kaltaisia.

2.1 Asiasanastot, taksonomiat ja tesaurukset

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 viittamaan 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 [Gar04]. Tesauruksessa on hierarkkisten suhteiden lisäksi yleisiä assosiativisia suhteita käsitteiden välillä [Gar04]. Tesauruksella kuvataan tyyppillisesti tietyn sovellusalueen sanasto. Esimerkiksi Art & Architecture Thesaurus (AAT) [AAT10] sisältää taiteen ja arkkitehtuurin aloilla käytettävää sanastoa. Toisaalta tesaurus voi myös koostua yleiskäsitteistä, kuten Yleinen suomalainen asiasanasto (YSA) [YSA10].

Tesaurusten ja muiden kontrolloitujen sanastojen pääasiallinen käyttötarkoitus on tiedonhaun tehostaminen [AGB00]. 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. Käsitteiden välisiä suhteita voidaan käyttää tiedonhakuaiheessa *kyselyiden muokkaamiseen* (engl. *query modification*) [Man07]. Tiedonhaun saantia voidaan pyrkiä parantamaan *laajentamalla kyselyitä* (engl. *query expansion*) lisäämällä niihin hakutermeihin liittyviä termejä. Haun tarkkuutta puolestaan voidaan pyrkiä kasvattamaan erottelemalla monimerkityksellisten termien merkitykset toisistaan ja *yksikäsitteistämällä* (engl. *disambiguation*) ne käyttökontekstissaan. Toissijaisia asiasanastojen käyttökohteita ovat muun muassa sovellusalueen ymmärryksen lisääminen käyttäjälle näyttämällä käsitteiden välisiä suhteita ja käsitteiden määritelmiä

sekä automaattisen tiivistelmän tuottaminen [AGB00].

2.2 Ontologiat

Sana ontologia on peräisin filosofiasta, jossa se tarkoittaa todellisuuden ja olemisen luonteen tutkimusta. Tietojenkäsittelytieteessä ontologialla tarkoitetaan tietyn aihealueen *käsitteistön eksplisiittistä mallia* (engl. explicit specification of conceptualization) [Gru93]. Ontologia kuvaa aihealueeseensa kuuluvat käsitteet ja niiden väliset suhteet formaalisti mahdollistaen aihealueen jaetun ymmärryksen ihmisten ja koneiden kesken [UsG96, Fen03].

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

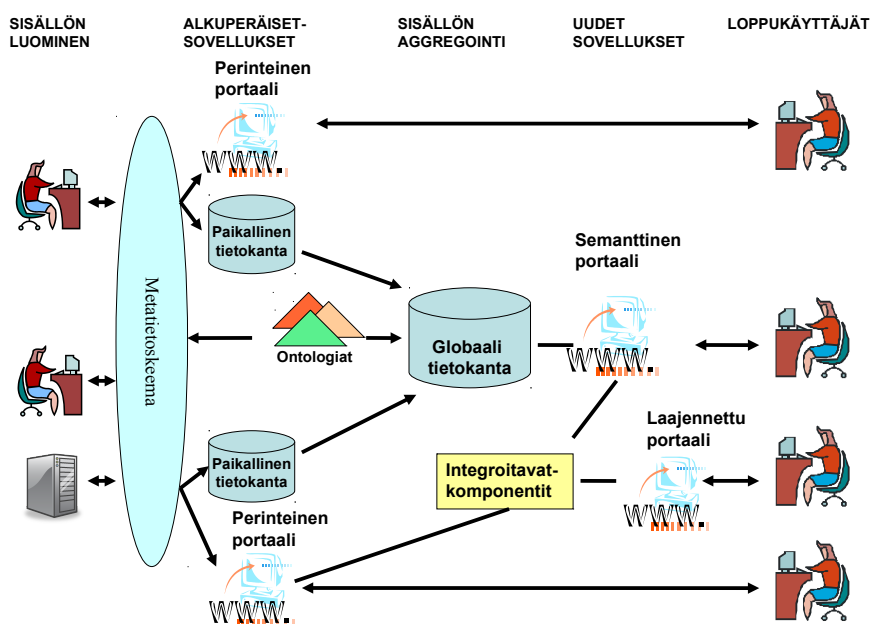
Tesaurusella kuvataan ainoastaan aihealueen sisältämät käsitteet ja niiden väliset suhteet erityisesti indeksoinnin ja tiedonhaun tarpeita varten. Ontologia puolestaan määrittelee aihealueen formaalimmin ja sitä voidaan käyttää esimerkiksi koneellisessa päättelyssä ja objektien luokittelussa. Toisaalta ontologia voi olla tesauksen kaltainen, yksinkertainen käsitelmä, joka muodostaa yksinkertaisen käsittehierarkian, mutta ei sisällä juuri muita käsitteiden välisiä suhdetyyppejä tai loogisia rakenteita [FSH04]. Tesaurus voidaan muuntaa ontologiaksi rikastamalla sitä semanttisesti [WSW01, HVT08].

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 [UsJ99]. Esimerkkejä yleiskäyttöisistä ontologioista ovat Yleinen suomalainen ontologia (YSO) [HSV07] ja Standard Upper Merged Ontology (SUMO) [NiP01].

2.3 Semanttinen web

Semanttisella webillä tarkoitetaan webiä, jossa tieto on koneymmärrettävässä muodossa tiedon tehokkaamman koneellisen käsittelyn mahdollistamiseksi. Tämä mah-

dollistaa muun muassa tiedon automaattisen koostamisen useasta lähteestä, koneellisen päättelyn ja tehokkaamman tiedonhaun. Semanttisen webin tekniikoilla voidaan toteuttaa esimerkiksi semanttisia informaatioportaaleja [RSC04, HMS05, HVS07], joissa yhdistetään tietoa useista lähteistä yhteisen käyttöliittymän avulla käytettäväksi. Hajautettuun tiedontuotantoon perustuvan portaalin toimintaperiaate on esitetty kuvassa 1 [HVS07].

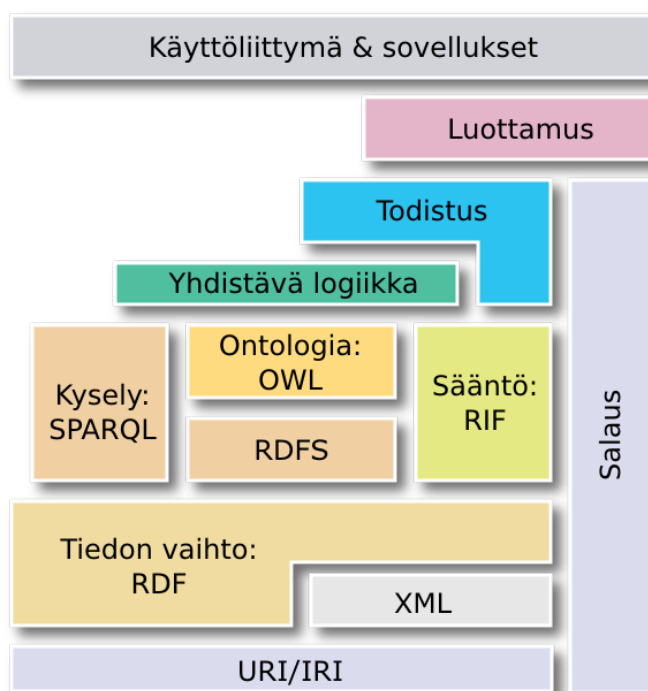


Kuva 1: Hajautettu tiedontuotanto semanttisessa informaatioportaalissa. Kuva on Hyvösen ja kumppaneiden julkaisusta [HVS07].

Semanttinen web perustuu tiedon kuvailuun eli metatiedon tuottamiseen. Metatietoskeemoilla voidaan määrittellä ominaisuudet, joita kuvattavista objekteista kuvataan. Metatietoskeemat mahdollistavat tiedon syntaktisen yhteentoimivuuden määrittämisen, 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.

Kuvassa 2 on esitetty W3C:n esittämä semanttisen webin kerroksittain rakentuva arkkitehtuuri [Her10]. Kuvasta nähdään, että ontologiakerros sijoittuu tiedon vaih-

tokerroksen päälle. Ontologiakerroksen päällä ovat logiikka-, todistus- ja luottamuskerrokset, joita ei ole vielä standardoitu. Varsinaiset sovellukset rakentuvat puolestaan edellä mainittujen kerroksien päälle.



Kuva 2: Semanttisen webin kerrosmalli. Kuva on W3C:n julkaisusta [Her10].

Semanttisen webin idea perustuu koneymmärrettävän tiedon hajautettuun sijoittamiseen, yhdistämiseen ja uudelleenkäyttöön. Myös ontologiat voivat viitata toisiinsa ja laajentaa toistensa käsitteistöjä. Ontologian käsitteiden välillä voidaan tehdä *siltauksia* (engl. mapping), jolloin eri ontologioilla kuvatut tiedot voidaan yhdistää toisiinsa automaattisesti [EhS04]. Näin on esimerkiksi mahdollista luoda yleiskäsitteistä koostuva yläontologia, jota spesifejä aihealueita kuvaavat ontologiat laajentavat.

Semanttisessa webissä objektit identifioidaan käyttäen URI-tunnisteita [BFM05]. Sekä kuvailtavaan tietoon, metatieto-ominaisuuksiin että kuvailussa käytettäviin ontologian käsitteisiin viitataan URI-tunnisteilla. Yksikäsitteisen objekteihin viittaamisen lisäksi URI-tunnisteita voidaan käyttää pääsymekanismina tietoon käyttämällä HTTP URI -skeeman mukaisia tunnisteita [SaC08].

Tässä tutkielmassa esiintyvät sanastojen nimiavaruuksien nimet ja niitä vastaavat nimiavaruuslyhenteet määritellään seuraavasti.

Resource Description Framework:

rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

RDF Schema:

rdfs: <http://www.w3.org/2000/01/rdf-schema#>

Web Ontology Language:

owl: <http://www.w3.org/2002/07/owl#>

Simple Knowledge Organization System:

skos: <http://www.w3.org/2004/02/skos/core#>

Dublin Core Metadata Element Set:

dc: <http://purl.org/dc/elements/1.1/>

RDF (Resource Description Framework) [K1C04] on semanttisessa webissä tietämyksen esittämiseen käytettävä tietomalli, jossa tietämys esitetään lausekkeina. Lausekkeet ovat kolmikoita koostuen subjektista, predikaatista ja objektista. Lausekkeet muodostavat RDF-verkon, jonka solmut ovat URI-tunnisteella identifioituja resursseja (URI-resurssi), anonyymejä resursseja tai literaaliarvoja, ja kaaret nimettyjä suhteita (predikaatti) solmujen välillä. Lausekkeen subjektina voi olla URI-resurssi tai anonyymi resurssi, predikaattina URI-resurssi ja objektina URI-resurssi, anonyymi resurssi tai literaaliarvo. Kolmikko kuvaa siten resurssin tietyn ominaisuuden. RDF-malli käsittää suppean sanaston, joka sisältää joitakin semantiikaltaan määritettyjä resurssityyppejä ja ominaisuuksia sekä ainoastaan syntaktisella tasolla määritettyjä nimiä. RDF määrittää muun muassa ominaisuuden *rdf:type*, jolla voidaan ilmaista tietyn resurssin tyyppi eli luokka, jonka yksilö resurssi on. RDF-verkon esittämiseen on useita vaihtoehtoisia syntakseja, joista yleisin on RDF/XML [Bec04]. Ihmisluettavampi muoto on esimerkiksi Turtle [BeB08].

RDF Schema (RDFS) [BrG04] on RDF-perustainen tietämyksen esityskieli yksinkertaisten ontologioiden kuvaamiseen. Esimerkiksi RDFS-kielen määrittelemää ominaisuutta *rdfs:subClassOf* käytetään luokkien (käsitteiden) välisen alaluokkasuhteen ilmaisemiseen ja vastaavasti ominaisuudella *rdfs:subPropertyOf* määritetään ominaisuuksien välinen hierarkkinen suhde.

OWL (Web Ontology Language) [DeS04] on RDF-perustainen ontologiakieli, jolla voidaan kuvata monimutkaisempia ontologisia rakenteita kuin RDFS-kielellä, kuten määrittää ominaisuuksia, joiden perusteella resurssin voidaan päätellä olevan tietyn luokan instanssi. OWL jakautuu kolmeen erilliseen alikielen, jotka eroavat toisistaan ilmaisuvoiman ja laskettavuuden perusteella. Alikielien ovat ilmaisuvoimaltaan heikoimmasta vahvimpaan OWL Lite, OWL DL ja OWL Full. OWL-kielestä on standardoitu myös uudempi versio OWL 2 [MPP09], joka laajentaa aiempaa versio-

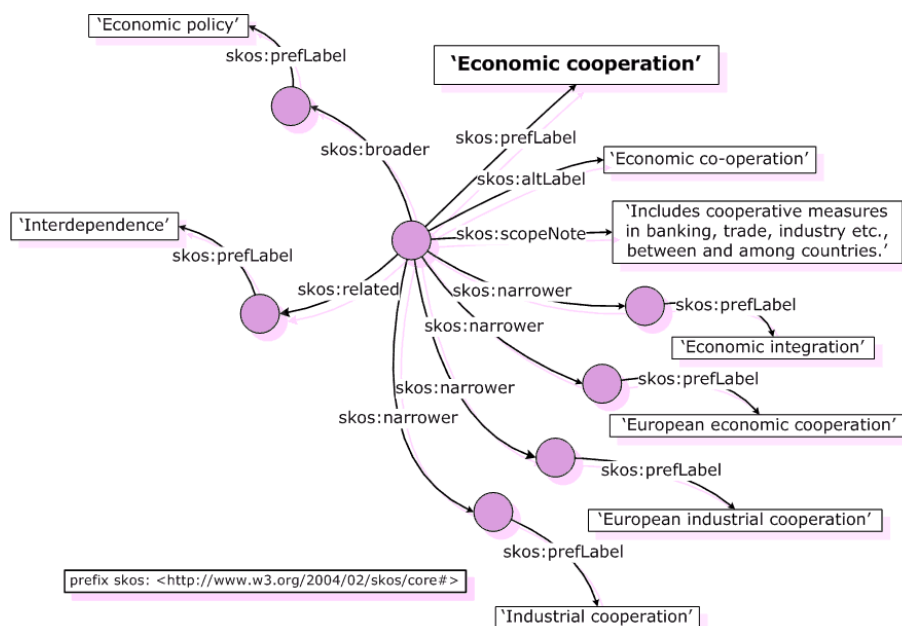
ta.

Aina ei ole tarvetta käyttää monimutkaisia ontologioita, vaan yksinkertaisempi asiasanasto riittää tiedon kuvailuun. Sanastojen esittämiseen semanttisessa webissä soveltuu SKOS – Simple Knowledge Organization System [MiB09, MMW05], joka on RDF-perustainen tietomalli käsiteskeemojen kuvaamiseen. Käsiteskeemoja ovat muun muassa tesaurukset, taksonomiat ja muut kontrolloidut sanastot. Esimerkiksi Yleinen suomalainen asiasanasto voidaan esittää käyttämällä taulukossa 1 kuvattuja SKOS-tietomallin luokkia ja ominaisuuksia. Tällaisen yleisen sanastojen esittämiseen soveltuvan mallin käyttäminen vähentää kustannuksia käytettäessä yhteistä sanastoa eri toimijoiden kesken, käytettäessä yksittäisessä sovelluksessa useita sanastoja sekä kehitettäessä yleisiä sanastojen käsittelyyn tarkoitettuja sovelluksia [AMM06].

Luokka/ominaisuus	Kuvaus
skos:ConceptScheme	Käsiteskeemaa edustavan resurssin tyyppi
skos:Concept	Sanaston käsitteen edustavan resurssin tyyppi
skos:Collection	Käsitteiden kokoelmaa edustavan resurssin tyyppi
skos:inScheme	Käsite kuuluu käsiteskeemaan
skos:hasTopConcept	Käsiteskeemalla on ylätasoinen käsite
skos:member	Kokoelmaan kuuluu käsite
skos:prefLabel	Resurssin ensisijainen nimike
skos:altLabel	Resurssin toissijainen nimike
skos:broader, skos:narrower	Käsitettä laajempi/suppeampi käsite
skos:related	Käsitteeseen liittyvä käsite
skos:note	Huomautus

Taulukko 1: SKOS-tietomallin keskeiset luokat ja ominaisuudet.

Esimerkkinä SKOS-mallin käytöstä on kuvassa 3 esitetty UK Archival Thesaurus (UKAT) -sanaston [UKA04] käsitteeseen “Economic cooperation” liittyvä käsitteistö ja terminologia. Kuvasta nähdään muun muassa, että käsitteen ensisijainen nimike on “Economic cooperation”, käsitteen yläkäsite on “Economic policy”, käsitteellä on neljä alakäsitettä ja käsitteeseen liittyvä käsite “Interdependence”.



Kuva 3: Käsitteeseen “Economic cooperation” liittyvä tietämys. Kuva on W3C:n julkaisusta [MiB05].

2.4 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 käyttäkkeen 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.

Tutkimushypoteesina on, että ontologioiden ja muiden sanastotyyppien käsittelyyn liittyy yleisiä toiminnallisuuksia, joita tarvitaan toistuvasti erilaisissa ontologioita hyödyntävissä sovelluksissa. Tällaisiksi toiminnallisuuksiksi esitetään ontologioiden selausta ja visualisointia sekä käsittehakua. 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 meta-tiedon tuottaminen todennäköisesti lisääntyisi, ja entistä älykkäämpien palvelujen toteuttaminen olisi mahdollista.

Muita hyötyjä ontologioiden käytössä palveluina tiedostomuodon sijaan on ontologioiden jatkuva ajantasaisuus. Jos ontologiaa käytetään sovelluksessa tiedostomuodossa, sovelluksen on säännöllisesti tarkistettava onko ontologiasta julkaistu uutta versiota ja otettava mahdollinen uusi versio käyttöön, mikäli ontologian ajantasaisuus on sovelluksen kannalta tärkeää.

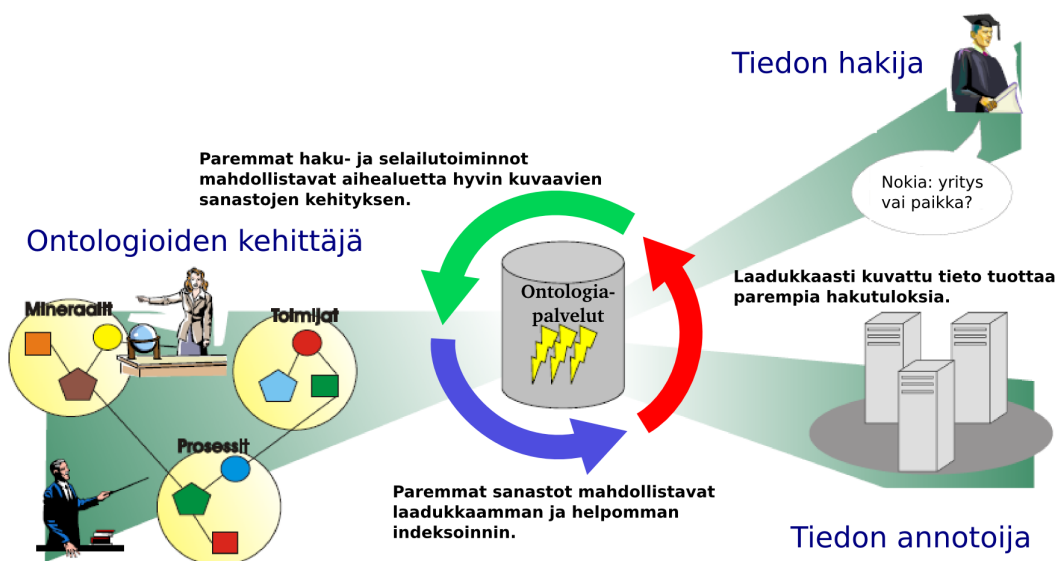
3 Ontologiapalvelut

Ontologiapalveluilla tarkoitetaan palveluita, joilla tarjotaan ontologioiden hyödyntämiseen liittyviä toiminnallisuuksia eri käyttäjäryhmille. Tämän luvun aliluvussa 3.1 käsitellään ontologiapalveluiden käyttäjäryhmiä ja heidän tarpeitaan. Ontologiapalveluiden tarjoamiseen kehitettyjä toteutuksia kuvataan aliluvussa 3.2. Näiden havaintojen perusteella esitetään yleisiä vaatimuksia ontologiapalvelun toiminnallisuksille aliluvussa 3.3.

3.1 Ontologioiden käyttäjäryhmät

FinnONTO-hankkeen alkuvaiheessa ontologioiden käyttäjäryhmiksi on tunnistettiin tiedon annotoijat, tiedon hakijat ja ontologioiden kehittäjät [HVK05]. Näiden ryhmien toiminta muodostaa vuorovaikutteisen prosessin, joka on kuvattu kuvassa 4. Kun ontologioiden kehittäjät tuottavat laadukkaita sanastoja, tiedon annotoijat voivat kuvata tietoa tarkoituksenmukaisella 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.

Ontologioiden käyttäjäryhmät tarvitsevat sekä keskenään samankaltaisia että erilaisia toiminnallisuuksia ontologioiden käsittelyyn. Useille käyttäjäryhmille yhteisiä toiminnallisuuksia ovat muun muassa käsitteiden haku, ontologioiden selaus ja visualisointi.



Kuva 4: Ontologioiden käyttäjäryhmät.

3.1.1 Tiedon annotoija

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 [TaL01, SDW01], jossa yksittäisten metatietoarvojen välillä on viittauksia. Esimerkiksi museoesineistä voidaan tallentaa esineen nimi, kuvaus, aihe, valmistusaika, -paikka, käyttöpaikka ja käyttäjä. Esineen nimi ja kuvaus esitetään tyypillisesti vapaana tekstinä. Esineen aihe voidaan esittää ennalta määritetyn sanaston asiasanoina (esimerkiksi “kello”).

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

Hildebrand ja kumppanit tunnistivat suorittamansa käyttäjätutkimuksen [HOH09] 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.

3.1.2 Tiedon hakija

Perinteisissä tiedonhakujärjestelmissä hakutulosten *relevanssi* (engl. relevancy) riippuu käyttäjän kyvystä kuvailla tiedontarpeensa kyselyn muodossa [Voo94]. Jos käyttäjän käyttämä termistö eroaa järjestelmän termistöstä, tai jos yhteistä termistöä käytetään eri tarkkuustasoilla, hakutulokset ovat yleensä huonoja. Näiden ongelmien ratkaisemiseksi on esitetty kontrolloidun sanaston käyttöä sekä kyselyn laajennosta. Kontrolloidun sanaston käyttö ohjaa käyttäjää käyttämään samaa termistöä kuin järjestelmä. Kyselyn laajentamisessa puolestaan lisätään käyttäjän muotoilemaan kyselyyn esimerkiksi hakutermeihin jollain tavalla liittyviä termejä. Kyselyn laajentaminen voi perustua (haun kohteena olevaan) aineistoon, esimerkiksi analysoimalla termien yhteisesiintymiä, tai tietämysmalleihin, kuten tesaarikseen [WVE85] tai ontologiaan [Voo94]. Tietämysmalleihin perustuvat tekniikat soveltuvat erityisesti tilanteisiin, joissa kyselylausekkeet ovat lyhyitä ja vaillinaisia, sisältäen vain vähän hakuindeksissä olevia termejä [WVE85, Voo94].

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

3.1.3 Ontologian kehittäjä

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 [FFR97]. Ontologia-

kehittäjien tarpeet ovat huomattavasti muita käyttäjäryhmiä vaativammat ja monitahoisemmat. Ontologiakehityksen työkalut rajataan tämän tutkielman ulkopuolelle.

3.1.4 Semanttisen webin sovelluksen kehittäjä

Monet semanttisen webin sovellukset tarvitsevat ontologiat käyttöönsä kokonaisuudessaan, jotta sovelluksissa voidaan suorittaa päättelyä, tai jotta päästään eroon verkkoviiveistä ontologioiden käytössä. Tällöin ontologian tiedostomuodossa tapahtuva julkaiseminen on tarpeen. Jos ontologiaa ei tarvita kokonaisuudessaan ja jos verkkoviiveet eivät aiheuta ongelmia, sovellukset voivat hyödyntää ontologioita ohjelmallisten rajapintojen avulla.

3.2 Ontologiapalvelimet

Ontologiapalveluihin liittyvä keskeinen käsite on *ontologiapalvelin* (engl. ontology server), jolla viitataan ontologioiden varastointiin ja ontologiapalveluiden tarjoamiseen käytettyyn tietojärjestelmään [FiF97, AhC07]. Kirjallisuudessa on käytetty myös termejä *ontologiakirjastojärjestelmä* (engl. ontology library system) [DiF01], joka korostaa ontologiapalvelimen roolia ontologioita kokoavana järjestelmänä, sekä *ontologiavarasto* (engl. ontology repository), joka painottaa ontologioiden varastointia. Tässä tutkielmassa näiden termien välillä ei nähdä eroa. Myös toteutukseen yksinkertaisia, ainoastaan ontologioiden visualisoimiseen ja selaamiseen tarkoitettuja *ontologiaselaimia* (engl. ontology browser) käsitellään ontologiapalveliminä. Ontologiapalvelimia on toteutettu useisiin erilaisiin tarkoituksiin, joten ontologiapalvelin on käsitteenä monimerkityksellinen ja laaja. Ontologiapalvelintoteutusten keskinäistä vertailua helpottaa niiden jaottelu erilaisin perustein.

Ontologian elinkaaren eri vaiheissa tarvitaan erilaisia työkaluja [AhC07]. Ontologian *suunnitteluvaiheessa* (engl. design) tarvitaan ontologiakehitystyökaluja, ontologioiden mallinnusapuvälineitä sekä versiointi-, julkaisu- ja siltaustoimintoja. *Sitoutumisvaiheessa* (engl. commit) etsitään sopivaa ontologiaa tai sen osaa käytettäväksi esimerkiksi annotoinnissa. Tällöin tarvitaan työkaluja ontologioiden hakuun, visualisointiin, selaukseen ja vertailuun. *Käyttövaiheessa* (engl. runtime) 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 [AhC07], ja siten sitoutumis- ja käyttövaiheen toteutukset ovat aliedustettuja ontologiapalvelinten joukossa. Ontologioiden hyödyntämisen kannalta kuitenkin juuri nämä vaiheet ovat tärkeitä. Koska ontologiapalvelintutkimus on keskittynyt voimakkaasti ontologioiden suunnitteluvaiheeseen, monet tutkijat määrittelevätkin ontologiapalvelimen ensisijaisesti ontologioiden ylläpidon ja uudelleenkäytön välineenä [DiF01]. Ahmad ja Colomb [AhC07] toteavat, että koska suurin osa kehitetyistä ontologiapalvelimistä 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 siltauksien 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 käyttötarpeeseen vastaavia selainperustaisia sovelluksia. Ohjelmallisia rajapintoja puolestaan voidaan toteuttaa esimerkiksi WWW-sovelluspalveluina (engl. web service), WSDL- ja SOAP-protokollalla [BHM04], tai REST-arkkitehtuurin mukaisesti pelkkää HTTP-protokollaa [FGM99] käyttäen [Fie00]. Rajapintojen avulla ontologiapalvelimen toiminnallisuuksia voidaan integroida toimijoiden omiin sovelluksiin.

3.2.1 Käyttöliittymätoteutukset

Tässä aliluvussa käsitellään ontologiapalvelintoteutuksia, jotka ovat ensisijaisesti ihmiskäyttäjille suunnattuja käyttöliittymätoteutuksia.

SUMO Browser

SUMO Browser [Sev03] on SUMO-yläontologian ja siihen sillatun WordNetin [Mil95] sekä mahdollisten muiden sillattujen ontologioiden selaukseen suunnattu ontologia-selain. SUMO Browser käyttää ontologian visualisoinnissa puumaista esitystapaa. Ontologiassa voidaan liikkua valitsemalla hierarkiapuusta käsite, jolloin valitun käsitteen ominaisuudet näytetään käsitesivulla. Aina kun valitun käsitteen ominaisuuksina on arvoja, jotka ovat viittauksia ontologian käsitteisiin, niistä tehdään hyperlinkki, jolloin ontologiaa voidaan selata näiden suhteiden avulla. Sovelluk-

sen erikoisuutena on muun muassa SUMO-ontologiassa käytettyjen KIF-esityskielen (Knowledge Interchange Format) [GeF92] mukaisten aksioomien esittäminen alkuperäisen esitysmuodon formaalin logiikan lisäksi luonnollisella kielellä.

GTAA Browser

GTAA Browser [BMG06] on hollantilaisen audiovisuaalisen alan tesauruksen (Gemeenschappelijke Thesaurus Audiovisuele Archieven – GTAA) visualisoimiseen ja selaamiseen toteutettu sovellus. GTAA Browser käyttää ontologian visualisointiin puumaista esitystapaa. GTAA-tesauruksen käsitteet on jaettu fasettien (asiasanat, genret, henkilöt, nimet, valmistajat, paikat) mukaan eri välilehtiin. Asiasana-fasetin käsitteisiin liittyy ryhmitteleviä kategorioita, ja asiasana-välilehdellä voidaan valita haluttu kategoria puuhierarkiasta. Valitun kategorian käsitteet näkyvät hakutuloksena, josta voidaan valita käsitteitä tarkasteltavaksi. Kategorioita voidaan valita useampia, jolloin hakutuloksena näytetään käsitteet, jotka kuuluvat kaikkiin valittuihin kategorioihin. Kategoriahierarkiassa näytetään aina vain kategoriat, joiden valinta jo valittujen kategorioiden lisäksi johtaa epätyhjään hakutulokseen. Hakutulosten lukumäärä on ilmaistu kategorianimen yhteydessä. Kun jokin käsite on valittu, sen ominaisuudet näytetään taulukkomaisena esityksenä. Muihin käsitteisiin viittaavat ominaisuusarvot toimivat hyperlinkkeinä kyseisen käsitteen valintaan. Ontologiaa voidaan selata myös käsittehierarkian avulla. Selaimen on myös toteutettu käsitteiden nimiin kohdistuva merkkijonohaku, jota varten tesaurusta on rikastettu käsitteiden nimien synonyymeillä. Lisäksi merkkijonohaussa käytetään oikoluku-toimintoa (engl. spelling suggest), jolla pyritään eliminoimaan väärinkirjoitetut tai sanastoon kuulumattomat hakutermit.

OntoRama

OntoRama [ERG02] on RDF-ontologioiden visualisoimiseen ja selaamiseen kehitetty Java-sovellus. OntoRama visualisoi ontologian hyperbolisella tasolla [LRP95] verkkona, jossa ontologian luokat ovat solmuja ja niiden väliset suhteet kaaria. Moniperintäiset hierarkiat visualisoidaan monistamalla solmu useaan kohtaan verkkoa, ja monistetut solmut yhdistetään keskenään viivalla. Sovellus käyttää verkon kaarina käyttäjän dynaamisesti valitsemia ontologian suhteita, ja erityyppiset suhdekaaret on esitetty verkossa eri väreillä. OntoRama näyttää hyperbolisen visualisoinnin rinnalla ontologian myös puuhierarkiana, jolloin visualisointitavat täydentävät toisiaan. Ontologian käsitteiden välillä voidaan liikkua valitsemalla käsite jommastakummasta edellä mainitusta näkymästä. Valitun käsitteen tiedot näytetään visualisointinäkymien alapuolella tekstimuodossa.

Ontolingua Server

Ontolingua Server [FFR97] on ontologioiden yhteisölliseen kehittämiseen, julkaisemiseen ja selaamiseen suunnattu ontologiapalvelin, joka tarjoaa kehitys- ja julkaisu-ympäristön KIF-kielen perustuvalla Ontolingua-kielillä [Gru93] kuvatuille ontologioille. Sovelluksessa julkaistut ontologiat on lueteltu hakemistosivulla, josta voidaan valita ontologia tarkasteltavaksi. Ontologian käsittehierarkian visualisointiin ja ontologian yleiskuvan antamiseen käytetään puumaista esitystapaa. Hierarkiasta voidaan valita käsite haluttaessa tarkastella käsitteen tietoja. Käsitesivulla näytetään sekä käsitteen eksplisiittiset ominaisuudet (engl. asserted properties) että päätellyt ominaisuudet (engl. inferred properties) ominaisuus-arvo-pareina. Tieto, jota ei voida esittää näin yksinkertaisesti, esitetään KIF-kielen aksiomina. Sovellus tarjoaa merkkijonohaun, joka kohdistuu kaikkiin sovelluksessa julkaistuihin ontologioiden käsitteiden nimiin.

Medical Ontology Server

Medical Ontology Server [GOP95] on ontologioiden selaamiseen, kehittämiseen ja ohjelmalliseen hyödyntämiseen tarkoitettu sovellus, joka tarjoaa Ontolingua Serverin tapaan kehitys- ja julkaisutoimintoja Ontolingua-kielen ontologioille. Ontologioita voidaan selata puumaisen esitystavan avulla. Kun puuhierarkiasta valitaan jokin käsite, käsitteen tiedot näytetään omassa näkymässään.

Neologism

Neologism [BCC08] on yksinkertaisten ontologioiden julkaisuun ja muokkaamiseen kehitetty Drupal-alustan¹ moduuli. Sovellus pyrkii tarjoamaan ontologian julkaisu- ja muokkaustoiminnallisuudet mahdollisimman yksinkertaisesti, pitäen kynnyksen sovelluksen käyttöön matalana. Neologism toteuttaa W3C:n suosittelemat käytännöt [BeP08] RDF-tiedon julkaisemiseen sovelluksen sisällä, mikä on päinvastainen ratkaisutapa kuin useimmissa RDF-julkaisujärjestelmissä, joita käytettäessä sisältöneuvottelu (engl. content negotiation) ja URI-tunnisteiden ohjaus URL-osoitteisiin toteutetaan HTTP-palvelimen asetuksia konfiguroimalla. Sovellus tarjoaa yleiskuvan ontologiasta verkkomaisena esityksenä käsitteistä ja niiden välisistä suhteista sekä yksinkertaisen näkymän ontologian metatietoihin ja yksittäisen käsitteen tietoihin.

Cupboard

Cupboard [AqL09] on ontologioiden julkaisemiseen, löytämiseen ja käyttöön ot-

¹<http://drupal.org>

tamiseen suunnattu yhteisöllinen ontologiapalvelu, johon käyttäjät voivat luoda omia ontologiatilojaan (engl. ontology space) omien ontologioiden julkaisemista varten. Ontologioista voidaan kuvata OMV-formaatin (Ontology Metadata Vocabulary) [HPS05] mukaista metatietoa, joka esitetään ontologian etusivulla. Ontologian etusivulla näytetään myös tilastotietoa ontologiasta sekä ontologian yleisnäkymä, joka pyrkii esittämään ontologian oleelliset käsitteet verkkomuodossa [PMA08]. Palvelussa julkaistuja ontologioita voidaan hakea, ja käyttötilanteeseen sopivan ontologian valinnassa voidaan hyödyntää muiden käyttäjien ontologioista tekemiä arvosteluja, jotka koostuvat sanallisista kommentteista sekä eri osa-alueista annetuista arvosanoista. Arvosanoja annetaan ontologian uudelleenkäytettävyydestä, oikeellisuudesta, kompleksisuudesta, peittävydestä ja mallinnuksesta. Lisäksi ontologioiden välille voidaan tehdä siltauksia palvelun avulla. Cupboard hyödyntää monia olemassa olevia ohjelmistokomponentteja ja sovelluksia, kuten Watson-hakukonetta [ABG07], jonka avulla ontologioita voidaan selata, hakea ja hyödyntää ohjelmallisesti.

BioPortal

BioPortal [RMK08, NSW09] on ontologioiden löytämiseen, hyödyntämiseen ja niiden välisten siltauksien tekemiseen tarkoitettu palvelu tarjoten yhteisen käyttöliittymän erilaisilla mallinnuskielillä (OBO [SAR07], Protégé Frames [NFM00], RDF, OWL) kuvatuille ontologioille. Ontologioita voidaan käsitellä myös ohjelmallisesti yhden rajapinnan kautta riippumatta kohteena olevan ontologioiden mallinnuskielistä. Palveluun on kerätty erityisesti lääketieteen ja biologian alueiden ontologioita. Ontologioista säilytetään versiohistoriaa ja metatietoja. Palvelussa julkaistut ontologiat esitetään aakkosellisena hakemistona sekä aihepiirin mukaan luokiteltuina.

Käsitteitä voidaan hakea merkkijonolla kaikista ontologioista tai yksittäisestä ontologiasta. Haku voidaan kohdistaa pelkästään käsitteisiin tai niiden lisäksi käsitteiden tietoihin. Ontologioita voidaan selata puu- ja verkkovisualisaatioiden avulla. Valittaessa käsite hakutuloksista tai visualisaatioista käsitteen tiedot näytetään taulukkomaisena esityksenä, jossa käsiteviittaukset eivät kuitenkaan toimi hyperlinkkeinä. Käsitteiden väliset siltaukset näkyvät erillisellä välilehdellä. Lisäksi siltauksia voidaan tutkia ontologiatasolla, jolloin nähdään mihin ontologioihin kyseisestä ontologiasta on siltauksia. Siltauksista on myös saatavilla metatietoa, joka ilmaisee siltauksen tekijän ja luontipäivämäärän. Siltauksia voidaan myös lisätä joko luomalla niitä käsiteselaimen avulla tai lataamalla palveluun siltaukset sisältävä tiedosto.

Palvelu tarjoaa käyttäjille mahdollisuuden antaa palautetta ontologioista joko lisäämällä vapaamuotoisia kommentteja tai antamalla ehdotuksia ontologioiden muuttamiseksi. Kommentit ja muutosehdotukset voidaan kohdistaa tiettyyn ontologian käsitteeseen, jolloin palautteen käsittely tehostuu. Lisäksi käyttäjät voivat muokata ontologioiden metatietoja, julkaista uusia ontologiaversiota sekä julkaista arvioita ontologioista. Palveluun tehtyjä muutoksia (kuten ontologian julkaisut ja kommentit) voidaan seurata RSS-syötteitä [BBD08] käyttäen.

BioPortaliin on toteutettu myös aineistojen automaattinen ontologiaperustainen annotointitoiminnallisuus. Palveluun voidaan syöttää tekstiaineisto, josta pyritään ohjelmallisesti tunnistamaan ontologioiden käsitteitä. Palveluun on ladattu valmiiksi webistä löytyviä lääketieteellisiä aineistoja, jotka on automaattisesti annotoitu tällä tavoin. Palveluun on toteutettu hakukäyttöliittymä, jolla näihin aineistoihin voidaan tehdä hakuja käyttämällä ontologioita hakutermien valintaan. Haussa hyödynnetään ontologioiden semanttisia suhteita, kuten alaluokkasuhdetta ja siltauksia.

Ontology Lookup Service

Ontology Lookup Service (OLS) [CJA06] on palvelu, joka tarjoaa pääsyn lääketieteellisen alan ontologioihin keskitetysti. Palvelu kokoaa lääketieteellisen alan ontologioita keskitettyyn palveluun, jolloin jokaiselle ontologialle ei tarvita omaa julkaisukanavaa. Samalla useaa ontologiaa hyödyntävän käyttäjän työskentely helpottuu, kun ontologiat löytyvät samasta palvelusta yhteisen käyttöliittymän avulla käytettävänä. Palvelussa julkaistut ontologiat esitetään aakkosellisena hakemistona, jossa on esitetty myös ontologian termien määrä sekä julkaisuajankohta. Käyttäjä voi etsiä käsitteitä merkkijonohauulla joko yksittäisestä ontologiasta tai kaikista ontologioista samanaikaisesti. Ontologioita voidaan selata puunäkymän avulla. Kun hakutuloksista tai puunäkymästä valitaan käsite, käsitteen tiedot näytetään taulukkomaisena esityksenä.

OLS2OWL

OLS2OWL [GGV09] on Protégé-ontologiaeditorin [NFM00] laajennos (engl. plugin), jonka avulla ontologiakehityksessä voidaan käsitellä useita ontologioita samanaikaisesti. Sovelluksen asiakasosa kommunikoi eri ontologiapalvelintoteutusten kanssa yhteisen rajapinnan avulla mahdollistaen erilaisten toteutusten käytön. Käyttöliittymällä voidaan tehdä hakuja useaan ontologiaan, vertailla ontologioita keskenään ja kehittää uusia ontologioita käyttäen niiden perustana olemassa olevia ontologioita. Sovellus visualisoi ontologiat verkkoina.

Watson

Watson [ABG07] indeksoi webistä löytyviä ontologioita ja RDF-tietämuskantoja ja tarjoaa niihin hakurajapintoja tiettyyn käyttötarkoitukseen sopivan ontologian löytämiseksi. Palvelu tarjoaa käyttäjälle mahdollisuuden merkkijonohakuihin ja SPARQL-kyselykielellä [PrS08] esitettyihin kyselyihin. Merkkijonohaku kohdistetaan sovelluksen indeksoimien RDF-resurssien nimiin, URI-tunnisteiden paikallisiin nimiin, kommentteihin ja muihin literaaliarvoihin. Hakutulokset näytetään ryhmiteltyinä tuloksen sisältävän RDF-tiedoston mukaan. Jos hakutuloksista valitaan RDF-resurssi, kyseisen resurssin kuvaus näytetään ominaisuusluettelona. Resurssityyppiset arvot toimivat hyperlinkkeinä kyseisten resurssien kuvauksiin. Jos hakutuloksista tai resurssin kuvauksesta valitaan RDF-tiedosto, tiedoston tilastotietoja näytetään taulukkona.

Swoogle

Swoogle [DFJ04] indeksoi webistä löytämiään semanttisen webin dokumentteja ja luokittelee ne ontologioiksi tai tietämuskannoiksi dokumentin sisällön perusteella. Tiedosto on ontologia silloin, kun sen sisältämistä RDF-kolmikoista merkittävä osa määrittää uusia resursseja (kuten luokkia tai ominaisuuksia) tai laajentaa muualla määriteltäjä resursseja uusilla ominaisuuksilla tai rajoituksilla. Muussa tapauksessa tiedosto on tietämuskanta. Tietämuskanta voi sisältää uusien yksilöiden määrittäjiä ja laajentaa muualla määriteltäjä yksilöitä.

Palvelu tarjoaa käyttäjälle merkkijonohaun, joka kohdistuu indeksoiduissa dokumenteissa esiintyviin URI-tunnisteisiin ja annotaatioihin (kuvattu esimerkiksi ominaisuudella *rdfs:comment*), mutta ei kaikkiin literaaliarvoihin, kuten käsitteiden nimiin. Tämän lisäksi hakua voidaan rajoittaa tietyn tyyppisiin dokumentteihin (ontologia, tietämuskanta), tietynkokoisiin dokumentteihin (luokkien, yksilöiden, ominaisuuksien lukumäärä), tietyllä mallinnuskielellä (RDF, RDFS, OWL) tai syntaksilla (RDF/XML, Turtle) kuvattuihin dokumentteihin tai tietyn tyyppisiin resursseihin (kuvattu ominaisuudella *rdf:type*). Swoogella voidaan hakea joko semanttisen webin dokumentteja tai yksittäisiä käsitteitä. Hakutuloksen yhteydessä näytetään siitä kerättyä metatietoa sekä URL-viittaus vastaavaan dokumenttiin tai käsitteen sisältävään dokumenttiin.

Hakutulosten järjestämiseen käytetään ontologiasijoitusta (engl. Ontology Rank, vrt. Googlen käyttämä PageRank [PBM99]). Ontologiasijoitus pyrkii kuvaamaan kuinka laajassa käytössä dokumentti on. Siinä missä PageRank perustuu sivujen välisiin hyperlinkkeihin, ontologiasijoitus käyttää hyväkseen dokumenttien vä-

lisiä suhteiden semantiikkaa. Esimerkiksi dokumenttien A ja B välinen suhde $\langle A \rangle$ *owl:imports* $\langle B \rangle$ kasvattaa ontologiasijoitusta voimakkaammin kuin jokin suhde, jonka avulla dokumentin A resurssi käyttää resurssia dokumentista B ilman, että koko dokumenttia B sisällytetään (engl. import) dokumenttiin A.

OntoSelect

OntoSelect [BED04] on esimerkiksi annotoinnissa käytettävän ontologian valintaan keskittyvä sovellus. Se kokoaa webistä löytyviä ontologioita ja suosittelee tietyn aineiston kuvaamistarkoitukseen parhaiten soveltuvia ontologioita. Sovellus on lähestymistavaltaan osittain Watsonin ja Swooglen kaltainen – se kerää webistä löytyviä ontologioita – mutta hakutoiminnallisuuden sijaan se listaa kaikki ontologiat taulukkomuodossa. Ontologioista esitetään meta- ja tilastotietoja, kuten ontologian nimi, formaatti, kieli sekä nimien, luokkien, ominaisuuksien ja ontologiaan sisällytettyjen ontologioiden lukumäärä.

Kun käyttäjä lataa sovellukseen aineiston, sovellus vertailee aineistoa webistä löytämiinsä ontologioihin ja laskee jokaiselle ontologialle kattavuusarvon, joka ilmaisee kuinka suureen osaan aineistoa ontologian luokkien ja ominaisuuksien nimet täsmäävät. Arvon laskennassa käytetään kieliteknologisia (ontologioiden luokkien ja ominaisuuksien nimien normalisointi, aineiston sanaluokkien tunnistaminen ja morfologinen analyysi) ja tilastollisia menetelmiä (tilastollisesti merkittävien termien tunnistaminen aineistosta). Aineistokohtaisen kattavuusarvon lisäksi ontologian yleistä laatua pyritään mittaamaan kahdella suurella: rakenne ja yhdistyneisyys. Rakenne ilmaisee ontologiassa käytettyjen ominaisuuksien lukumäärän suhteen ontologian luokkien lukumäärään ja kertoo, kuinka tarkasti kuvattu ontologia on. Yhdistyneisyys puolestaan ilmaisee, kuinka vahvasti ontologia on linkittynyt muihin ontologioihin. Mitä enemmän tietyssä ontologiassa on viitteitä muihin ontologioihin, ja mitä enemmän muut ontologiat viittaavat kyseiseen ontologiaan, sitä vahvemmin ontologia on linkittynyt. Ontologiat listataan kattavuusarvojen perusteella järjestettynä. Käyttäjälle suositellaan ontologiaa, jonka yhdistetty kattavuus-, rakenne- ja yhdistyneisyysarvo on korkein.

3.2.2 Ohjelmalliset rajapinnat

Tässä aliluvussa käsitellään ontologiapalvelintoteutuksia, joiden pääasiallinen käyttötapa on ohjelmallinen rajapinta. Myös käyttöliittymätoteutuksia sisältäviä ontologiapalvelimista ohjelmallisia rajapintoja tarjoavat SUMO Browser, GTAA

Browser, Ontolingua Server, Medical Ontology Server, Cupboard, BioPortal, Ontology Lookup Service, Watson ja Swoogle.

FIPA-standardi

IEEE:n standardointiorganisaatio FIPA (The Foundation for Intelligent Physical Agents) on määrittänyt alustavan spesifikaation ontologiapalvelun tarjoavan ontologia-agentin kommunikaatorajapinnasta [FIP01]. Spesifikaatio ei sisällä ontologiapalvelimen toteutusta, vaan ainoastaan määrittelee rajapinnan, jonka avulla toimijat voivat hyödyntää ontologiapalvelua ontologia-agentin kautta. Ontologia-agentit voivat käyttää erillisiä itsenäisiä ontologiapalvelimia, tai ne voivat sisältää ontologiapalvelimen. Spesifikaation mukainen ontologia-agentti tarjoaa ohjelmallisen rajapinnan ontologioiden löytämiseen, käyttämiseen ja ylläpitoon, ontologioiden ja ontologiakielten välisiin muunnoksiin, käsitteiden ja ontologioiden välisiin suhteisiin kohdistuvaan hakuun sekä sopivan ontologian löytämiseen toimijoiden väliseen kommunikaatioon. Ontologia-agentin on kyettävä edellä mainittuihin tehtäviin liittyvään kommunikaatioon, mutta sen ei ole välttämätöntä pystyä toteuttamaan kyseisiä tehtäviä. FIPA-standardia noudattavat ontologia-agentit rekisteröityvät hakemistopalveluun, jonka avulla ontologiapalveluita tarvitsevat sovellukset saavat yhteyden ontologia-agenttiin.

Agent Cities Ontology Service

Agent Cities Ontology Service (ACOS) [LTT03] on yhteisöllinen ontologiapalvelu keskittyen ontologioiden kehitykseen, niihin pääsyn tarjoamiseen sekä ontologioiden välisten siltausten tekemiseen. Palvelu määrittää FIPA-standardia vastaavan tavan ontologioiden ohjelmalliseen käsittelyyn, mutta erona FIPA-standardiin on yhteisöllisyyden ja ontologioiden uudelleenkäytön korostaminen. Kun ACOS-palveluun julkaistaan ontologia, sen julkaisesta käyttäjästä tulee ontologian omistaja, joka voi antaa muille käyttäjille käyttö- ja ylläpito-oikeuksia ontologiaan. Palvelussa julkaisuihin ontologioihin voidaan eksplisiittisesti sisällyttää muita ontologioita. Ontologian suosion mittarina käytetään kyseisen ontologian ja sen sisällyttäneiden ontologioiden käyttäjien yhteislukumäärää. Vastaavasti käyttäjät luokitellaan arvostuksen mukaan: mitä suositumpia ontologioita käyttäjä omistaa ja ylläpitää, sitä korkeampi hänen arvostuksensa on. Ontologioita voivat päivittää ylläpito-oikeuden saaneiden käyttäjien lisäksi käyttäjät, joiden arvostus on riittävän korkea. Aina kun ontologiaan ollaan tekemässä muutoksia, palvelu varmistaa että ontologioiden eheys (engl. consistency) säilyy muutoksen jälkeen.

Otago Ontology Repository

Otago Ontology Repository [PCC03] on ontologioiden julkaisemiseen, löytämiseen ja niihin pääsyn tarjoamiseen kehitetty sovellus, joka tarjoaa FIPA-standardia kevyemmän tavan käyttää ontologioita ohjelmallisesti. Sovelluksen kehittäjät uskovat, että ontologia- ja agenttiteknologian laajemman leviämisen saavuttamiseksi sovellusten pitää pystyä kommunikoimaan suoraan ontologiapalvelinten kanssa yksinkertaisella, tunnetulla protokollalla. Sovellus perustuukin yleisiin web-standardeihin: HTTP-protokollaan, RDF-tietomalliin ja URN-tunnisteisiin [Moa97] käyttäen REST-arkkitehtuuria. URN-tunnisteet mahdollistavat ontologioiden sijainnista riippumattoman identifioimisen, toisin kuin HTTP URI -skeemaa käyttävät tunnisteet. Ohjelmallisen rajapinnan lisäksi sovellus tarjoaa ihmiskäyttäjälle käyttöliittymän, jolla voi julkaista ontologioita ja muokata niiden metatietoja.

SKOS API

SKOS API [BRM04] on SKOS-muotoisten sanastojen ohjelmalliseen käsittelyyn tarkoitettu rajapintamäärittely. Rajapinnan testaamiseksi on toteutettu graafinen käyttöliittymä, Pilot SKOS API Browser [TuB05]. Rajapintatoteutuksen teknisten ongelmien vuoksi käyttöliittymä hyödyntää vain kahta rajapinnan metodia: `getConcept` ja `getAllConceptRelatives`. Käyttöliittymässä on tekstikenttä, johon voidaan syöttää käsitteen URI-tunniste. URI-tunnistetta vastaavan käsitteen nimetiedot, kuvausteksti (engl. scope note) sekä käsitteeseen liittyvät muut käsitteet näytetään tekstikentän alapuolella. Käsitteet toimivat linkkeinä, joiden avulla voidaan selata sanastoa. Rajapinnan käytöstä johtuvien viiveiden pienentämiseksi käyttöliittymää varten on toteutettu paikallinen välimuisti.

OCLC Terminology Web Services

OCLC Terminology Services -projektissa toteutettu Terminology Web Services [VCH05] on rajapintatoteutus erilaisissa formaateissa kuvattujen sanastojen (SKOS, MARC 21 [MAR00], Zthes [Tay06]) ohjelmalliseen käsittelyyn. Järjestelmän toteutuksessa erilaiset sanastot voivat olla tallennettuina järjestelmään koko teksti -tietokantoihin, SQL-tietokantoihin tai XML-tiedostoihin. Kommunikaatio järjestelmän sisällä tietokantoihin tapahtuu SRU- [SRU09], REST- tai SOAP-rajapintojen avulla. Ulkoinen rajapinta on toteutettu SRU-protokollalla ja tarjoaa metodeja muun muassa termien merkkijonohakuun sekä termin tietojen selvittämiseen eri formaateissa (HTML, MARCXML [MAR09], SKOS, Zthes). Termejä voidaan hakea myös SRU CQL -syntaksin [CQL08] mukaisilla kyselyillä. Järjestelmä tarjoaa ohjelmallisen rajapinnan lisäksi ulkoisiin järjestelmiin

integroitavan käyttöliittymäkomponentin, jolla voidaan hakea termejä sanastoisista. Komponentti toimii Microsoft Internet Explorer -selaimen MS Office 2003 -toimistosovelluspaketin mukana asennettavassa Research Pane -sivupalkissa. Kun sivupalkin hakukomponentin hakutuloksista valitaan termi, termin tiedot näytetään käyttöliittymässä, ja sanastoa voidaan selata termien välisien suhteiden avulla. Kun sopiva termi on löydetty, viite siihen voidaan syöttää kohdejärjestelmään painamalla valintapainiketta.

Ontology Web Services

Ontology Web Services (OWS) [DNK04] on ehdotus yleisistä toiminnallisuuksista, joita ontologioita hyödyntävät sovellukset tarvitsevat, ja joita olisi siten hyödyllistä tarjota sovelluksille WWW-sovelluspalveluna. Ehdotus sisältää toimintoja ontologioiden kyselyihin, näkymien luomiseen, ontologioiden muunnoksiin, päättelyyn sekä usean ontologian hallitsemiseen. Koska ontologioiden kyselykieliä on useita, yleistä kyselytoimintoa käytettäessä on välitettävä itse kyselyn lisäksi tieto kyselykielestä ja ontologiasta, jota kysely koskee. Näkymällä tarkoitetaan ontologian osajoukkoa, jonka käyttäminen on hyödyllistä esimerkiksi silloin, kun kohteena oleva ontologia on laaja, mutta siitä tarvitaan vain pieni osa käsittelyä varten. Muunnospalvelun avulla ontologia voidaan muuntaa mallinnuskielestä toiseen. Yksinkertaisen syntaktisen muunnoksen lisäksi voidaan tarjota erikoistunutta palvelua tietyille ontologialle, jolloin voidaan ottaa huomioon ontologian erikoispiirteet. Päättelyä voidaan käyttää esimerkiksi ontologian eheyden tarkistamiseen tai luokkien sisältyvyyshierarkian selvittämiseen. Usean ontologian hallintaan kuuluvat ontologioiden siltaukseen, yhdistämiseen ja versiointiin liittyvät toiminnot. WWW-sovelluspalveluille tyypillisesti ehdotuksen määrittämät ontologiapalvelut voivat käyttää toisia palveluita hyväksien pyyntöön vastatessaan.

KAON Server

KAON Server [OVM04] on modulaarinen semanttisen webin yleisiä ohjelmistokomponentteja sisältä järjestelmä, joka sisältää toiminnallisuuksia ontologioiden uudelleenkäyttöön ja hajautettuun kehitykseen [MMS03]. Järjestelmä tarjoaa käyttötarkoitukseen sopivan ontologian löytämiseen keskitetyn ontologiarekisterin, johon on tallennettu metatietoja hajautetun ontologiapalvelinympäristön ontologioista. Ontologioita voidaan hakea metatietojen, kuten julkaisijan, mallinnuskielen tai julkaisujankohdan, perusteella. Lisäksi ontologioiden sisältämät käsitteet täsmäytetään WordNetin käsitteisiin. Täsmäytyksen perusteella ontologian metatietoihin lisätään WordNet-käsitteevastineet, jolloin ontologian hakeminen sen sisällön perusteel-

la on mahdollista. Ohjelmallisen rajapinnan lisäksi ontologiarekisteriin on toteutettu käyttöliittymä OI-Modeler -ontologiaeditorin laajennoksena, jonka avulla ontologiaa kehitettäessä siihen voidaan sisällyttää muita palvelussa julkaistuja ontologioita. Toteutuksessa on kiinnitetty erityistä huomiota ontologiaan tehtyjen muutosten levittämiseen ontologioihin, joihin kyseinen ontologia on sisällytetty. Näin varmistetaan ontologioiden ajantasaisuus ja eheys.

Open Ontology Repository

Open Ontology Repository -hankkeen (OOR) [BaS09] tavoitteena on ontologioiden käytön, uudelleenkäytön ja yhteiskäytön lisääminen edistäen semanttisen webin leviämistä. Keinona tavoitteen saavuttamiseksi ehdotetaan kehitettäväksi infrastruktuuri, jolla varmistetaan ontologioiden saatavuus. Ehdotettuun infrastruktuuriin sisältyy toiminnallisuuksia ontologioiden kehittämiseen, yhteiskäyttöön, hakemiseen ja hallintaan sekä ontologioiden käyttämiseen metatiedon tuottamisessa. Infrastruktuuri on hajautettu, ja yksittäiset ontologiapalvelimet voivat erota toisistaan toteutustekniikaltaan, mutta ne kommunikoivat keskenään yhteisen rajapinnan avulla. Yhteisten perustoiminnallisuuksien lisäksi ontologiapalvelimet voivat toteuttaa spesifejä käyttötapauskohtaisia toiminnallisuuksia. Ontologiapalvelinten tiedot ovat saatavilla keskitetysti ontologiarekisterissä.

3.3 Yleiset vaatimukset

Lähes jokaisessa edellä esitellyn ontologiapalvelimen käyttöliittymätoteutuksessa 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.

3.3.1 Ontologian visualisointi ja selaaminen

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 semantiikkaa hahmotettaessa. Käsitteen merkityksen ymmärtämisessä auttaa käsitteen ontologisen kontekstin, eli käsitteeseen hierarkkisilla suhteilla sekä muilla tavoin liittyvien käsitteiden, näyttäminen. Visualisointia voidaan käyttää myös ontologian selailussa. Esi-

merkiksi tiedon annotoijan käyttötapauksessa, jossa käyttäjä ei tiedä entuudestaan sopivaa käsitettä tiedon kuvailuun, visualisointi voi auttaa käyttäjää tarkentamaan tai yleistämään alustavasti valittua käsitettä käsitehierarkian avulla tai löytämään muutoin käsitteeseen liittyviä käsitteitä, jotka sopivat käyttötarkoitukseen [HOH09].

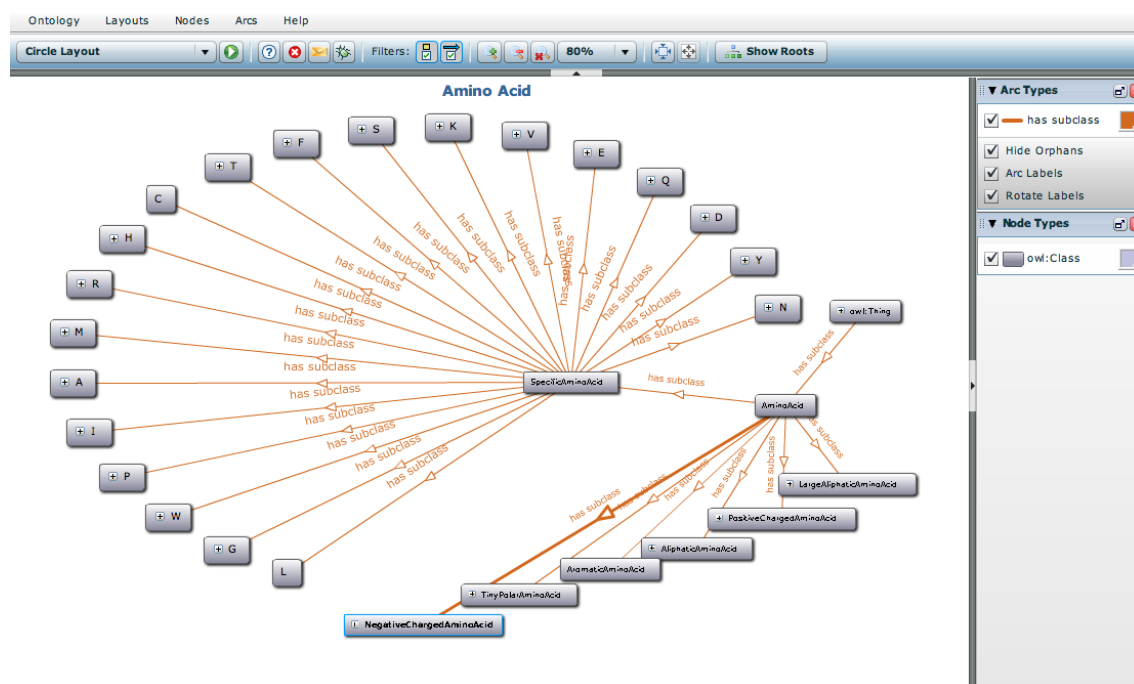
Ontologian hierarkkisen luonteen vuoksi sille intuitiivinen ja yksinkertainen visualisointitapa on käsitehierarkian esittäminen puumuodossa [KHL07]. Yksinkertainen tapa esittää ontologia puumuodossa on sisennetty lista, jossa alakäsitteet ovat yläkäsitteen alla sisennettyinä. Tämä esitysmuoto muistuttaa käyttöjärjestelmien käyttöliittymien tapaa visualisoida tiedostojärjestelmien hakemistorakenteita [ERG02, Sev03] ja on siten tuttu käyttäjille. Esimerkkinä sisennettyyn listaan perustuvasta visualisoinnista kuvassa 5 on esitetty Protégé-ontologiaeditorin visualisoima ontologian hierarkia. Sisennetty listaesitys on yleisin ontologian visualisointitapa edellä tarkastelluissa ontologiapalvelintoteutuksissa: sitä käytetään seitsemässä toteutuksessa kaikkiaan kolmestatoista käyttöliittymätoteutuksesta.



Kuva 5: Ontologian hierarkian visualisointi puumuodossa Protégéssa.

Neljässä ontologiapalvelintoteutuksessa ontologia visualisoidaan verkkona. Verkko on luonteva tapa semanttisen webin datan visualisointiin, koska RDF-tietomalli on verkkomuotoinen. Verkon solmut ovat resursseja ja kaaret niiden välisiä suhteita. Verrattuna puumaiseen esitykseen verkkoesityksessä voidaan näyttää enemmän informaatiota, koska hierarkkisen suhteiden lisäksi voidaan esittää muitakin suhteita. Suhteiden tyypit voidaan merkitä kaarien yhteyteen esimerkiksi värikoodaamalla tai nimeämällä ne. Verkkoesityksessä ontologian hierarkkinen luonne ei ilmene yhtä selkeästi kuin puumaisessa esityksessä. Ontologioiden mahdollisesti sisältämä suuri

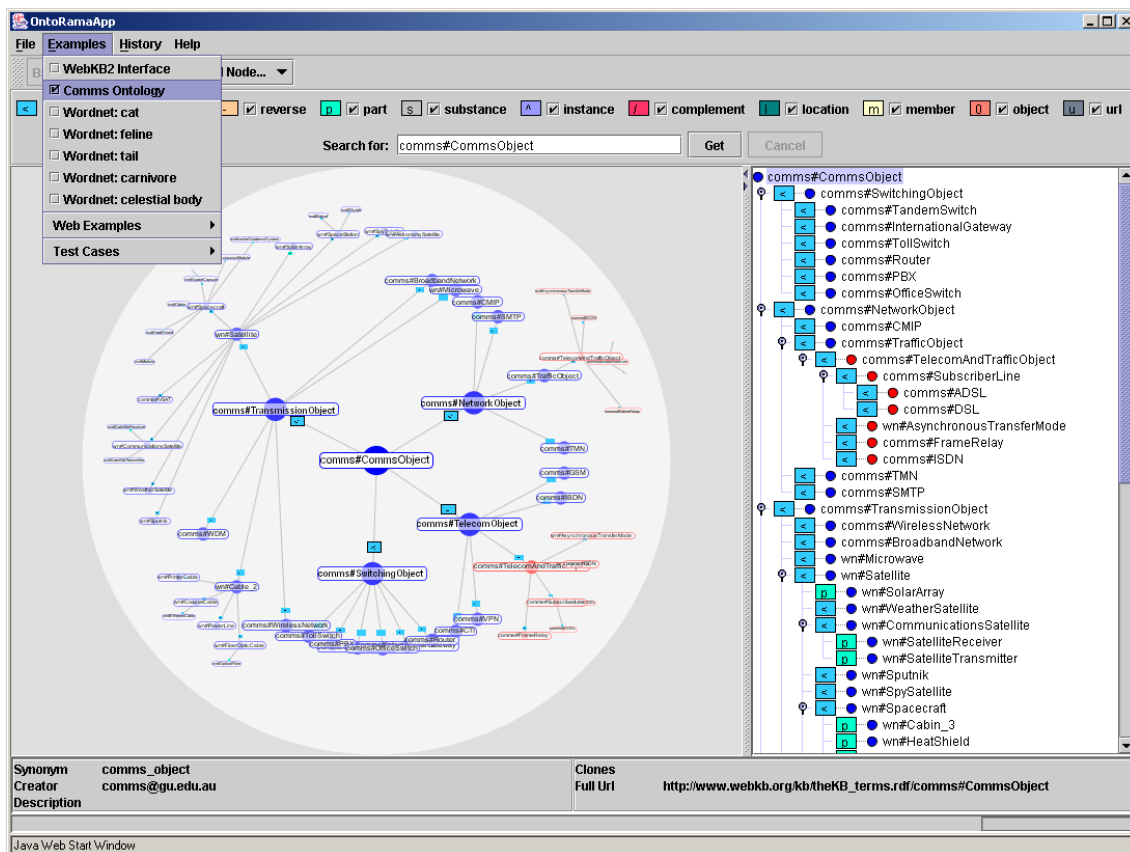
tietomäärä tekee verkoista suuria, jolloin käyttäjän voi olla vaikea hahmottaa niitä ja keskittää huomio tiettyyn kohtaan verkkoa [Son00]. Esimerkki verkkomuotoisesta visualisoinnista on kuvassa 6 esitetty näkymä BioPortalista.



Kuva 6: Ontologian visualisointi verkkomuodossa BioPortalissa.

Yhdessä tarkastelluista ontologiapalvelintoteutuksista käytetään hyperbolista tasoa, jossa ympyrän kehän koko kasvaa eksponentiaalisesti ympyrän säteen kasvaessa, minkä johdosta ympyrän keskipisteestä kauemmas siirryttäessä tilan määrä kasvaa eksponentiaalisesti [LRP95]. Samalla objektien koko ja etäisyys toisistaan pienenevät. Täten hyperbolinen visualisointi soveltuu tilanteeseen, jossa on tarvetta esittää suuria hierarkioita. Etuna puumaiseen ja yksinkertaiseen verkkomaiseen esitystapaan on tilan tehokkaampi hyödyntäminen, jolloin suurempien ontologioiden visualisoiminen on mahdollista rajatussa tilassa, esimerkiksi tietokoneen näytöllä. Ontologian kokonaisuuden hahmottaminen ja tietyn sen osan tarkempi tutkiminen on mahdollista samanaikaisesti. OntoRaman hyperbolinen visualisointi on esitetty kuvassa 7.

Lähes kaikissa käsitellyissä ontologiapalvelimissa ontologiaa voidaan selata visualisointinäkömman avulla. Näkymässä kuvatut käsitteet toimivat linkkeinä käsitteiden tarkempaan tarkasteluun. Monissa visualisointitekniikoissa näkymä päivittyy ontologiaa selatessa. Tyypillisesti puuesitystavoissa valitun käsitteen alakäsitteet tu-



Kuva 7: OntoRaman visualisoima ontologia hyperbolisella tasolla.

levat näkyviin puuhun, ja verkkomaisissa esitystavoissa valittu käsite asetetaan näkymän keskelle. Hyperbolisessa tasossa valittu käsite ja sen semanttisesti läheiset käsitteet visualisoidaan suurempina kuin semanttisesti kauempana olevat käsitteet. Kaikissa tarkastelluissa ontologiapalvelintoteutuksissa vallitseva toiminnallisuus on valitun käsitteen ominaisuuksien näyttäminen taulukkomaisena esityksenä. Taulukko sisältää tyypillisesti käsitteeseen liittyvät RDF-kolmikot (kolmikot, joiden subjektina on valittu käsite) siten, että taulukon ensimmäisessä sarakkeessa on kolmikon predikaatti ja toisessa sarakkeessa objekti. Joissakin toteutuksissa erotellaan toisistaan käsitteen eksplisiittiset ominaisuudet ja päätellyt ominaisuudet.

Valtaosassa usean ontologian julkaisukanavana toimivista ontologiapalvelimista on hakemistonäkymä, joka sisältää palvelussa julkaistut ontologiat. Hakemistoissa on ontologian nimen lisäksi ontologian metatietoa. Kun hakemistosta valitaan ontologia, ontologian tiedot näytetään omassa näkymässään, joka sisältää hakemistonäkymää laajemmin metatietoa, tilastotietoa sekä mahdollisia käyttäjien komment-

teja ja arvosteluja.

3.3.2 Ontologian hakutoiminnallisuudet

Käsitteen hakutoiminnallisuudet auttavat käyttäjää löytämään käyttötarkoitukseen sopivan käsitteen. Erityisesti automaattisesti täydentyvä haku on tehokas käyttötapaussessa, jossa tiedon annotoija tietää entuudestaan, millä käsitteellä haluaa kuvata sisällön. Käsitehaun tulosten yhteydessä on järkevää näyttää käsitteen nimien lisäksi muita tietoja käsitteistä, jotta käsitteiden merkitys olisi mahdollisimman selvä käyttäjälle käsitteen valinnan helpottamiseksi [HOH09]. Moneen ontologiaan kohdistuvaa hakua voidaan hyödyntää ontologioiden vertailuun.

Lähes kaikki edellä käsitellyt ontologiapalvelintoteutukset sisältävät toiminnallisuuden käsitteiden hakuun merkkijonolla. Käsitehaku kohdistuu yleensä käsitteiden nimiin ja joissakin toteutuksissa myös käsitteiden URI-tunnisteisiin, kommentteihin tai muihin ominaisuuksiin. GTAA Browserissa haku voidaan kohdistaa koko sanaston sijaan haluttuihin käsitteitä ryhmitteleviin kategorioihin. Lähes kaikissa usean ontologian julkaisuun suunnatuissa toteutuksissa käsitehaku voidaan kohdistaa samanaikaisesti kaikkiin palvelun ontologioihin. Suurin osa toteutuksista toimii julkaisijalähtöisesti tarjoten julkaisijalle kanavan ontologian julkaisemiseen. Toisenlaista lähestymistapaa edustavat sovellukset, jotka käyvät läpi webin tietosisältöä indeksoiden löytämänsä ontologiatiedostot. Tällaiset sovellukset toimivat semanttisen webin yleisinä hakukoneina tarjoten apuvälineitä koko webin laajuiseen tiedonhakuun.

3.3.3 Ontologian ohjelmallinen käsitteleminen

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

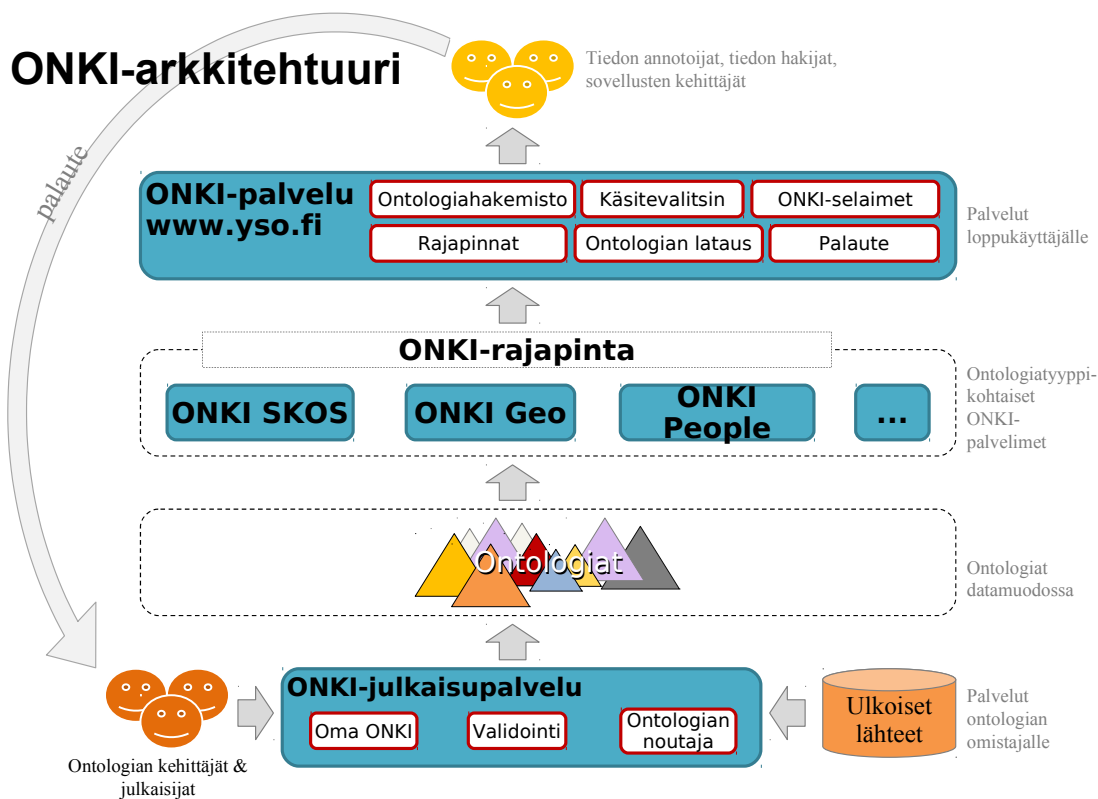
3.3.4 Ontologioiden ajantasaisuus

Käytettäessä ontologioita palveluina ontologioiden ajantasaisuuden suhteen ei ti- lannetta, jossa toimijoilla olisi tietyllä ajan hetkellä käytössä keskenään erilaisia on- tologian versioita. Palvelussa julkaistu uusi ontologiaversio on heti kaikkien toimi- joiden käytössä. Tiedostomuodossa tapahtuvassa ontologiajulkaisussa toimijoiden on huolehdittava ontologian uuden tiedostoversion käyttöönotosta. Ontologian julkaisu- tavasta riippumatta ongelmia voi kuitenkin syntyä ontologioiden muuttuessa. Esi- merkiksi jos tietyllä käsitteellä on annotoitu aineistoja, ja käsite poistetaan ontolo- giasta, annotaatiot muuttuvat epävalideiksi. Annotaatioiden validius täytyy varmis- taa esimerkiksi korvaamalla annotaatiokäsitteet joillakin voimassaolevilla käsitteil- lä tai säilyttämällä poistettu käsite ontologiassa, mutta merkitsemällä se käytöstä poistetuksi.

4 Ontologiapalvelu ONKI

Edellisessä luvussa käsiteltyjen ontologioiden hyödyntämiseen liittyvien toiminnal- lisuuksien tarjoamiseksi on toteutettu ontologiapalvelu ONKI [HVT08, VTH09], jo- ka on käytettävissä verkkopalveluna URL-osoitteessa <http://www.yso.fi/onki1/>. Palvelun tarkoituksena on tarjota käyttäjille ontologiapalveluita erilaisten ontologia- tyyppien käyttämiseen tarkoitettujen ontologiapalvelinten ja muiden ohjelmisto- komponenttien avulla. Palveluun on toteutettu ONKI SKOS -palvelin [TFV09] ylei- sille sanastotyyppisille käsiteontologioille, ONKI Geo [KHS08] paikkatiedon onto- logioille ja ONKI People [KuH09] toimijaontologioille. Ontologiatyyppikohtaisista ONKI-palvelimista tässä tutkielmassa keskitytään käsittelemään ONKI SKOS - palvelinta, joka kuvataan luvussa 5. ONKI-palvelun yleisarkkitehtuuri on kuvattu kuvassa 8. ONKI-palvelu tarjoaa ontologioiden hyödyntämiseen soveltuvia, ihmis- käyttäjille suunnattuja palveluita sekä ohjelmallisia rajapintoja, joita käsitellään tarkemmin luvussa 6.

ONKI-ontologiapalvelun etusivu on esitetty kuvassa 9. Palvelussa julkaistut ontolo- giat on esitetty etusivulla taulukkona, jonka sisältämät tiedot perustuvat ontologioi- den metatietoihin. Metatiedot on tuotettu tarkoitukseen määritettyä, Dublin Core -standardiin perustuvaa metatietoskeemaa käyttäen. Taulukko sisältää ontologia- kohtaisesti hyperlinkkejä ontologian palveluihin, kuten ontologiaselaimeen ja käsite- valitsimeen, sekä ontologian ulkoiseen dokumentaatioon. Lisäksi taulukossa on ku-

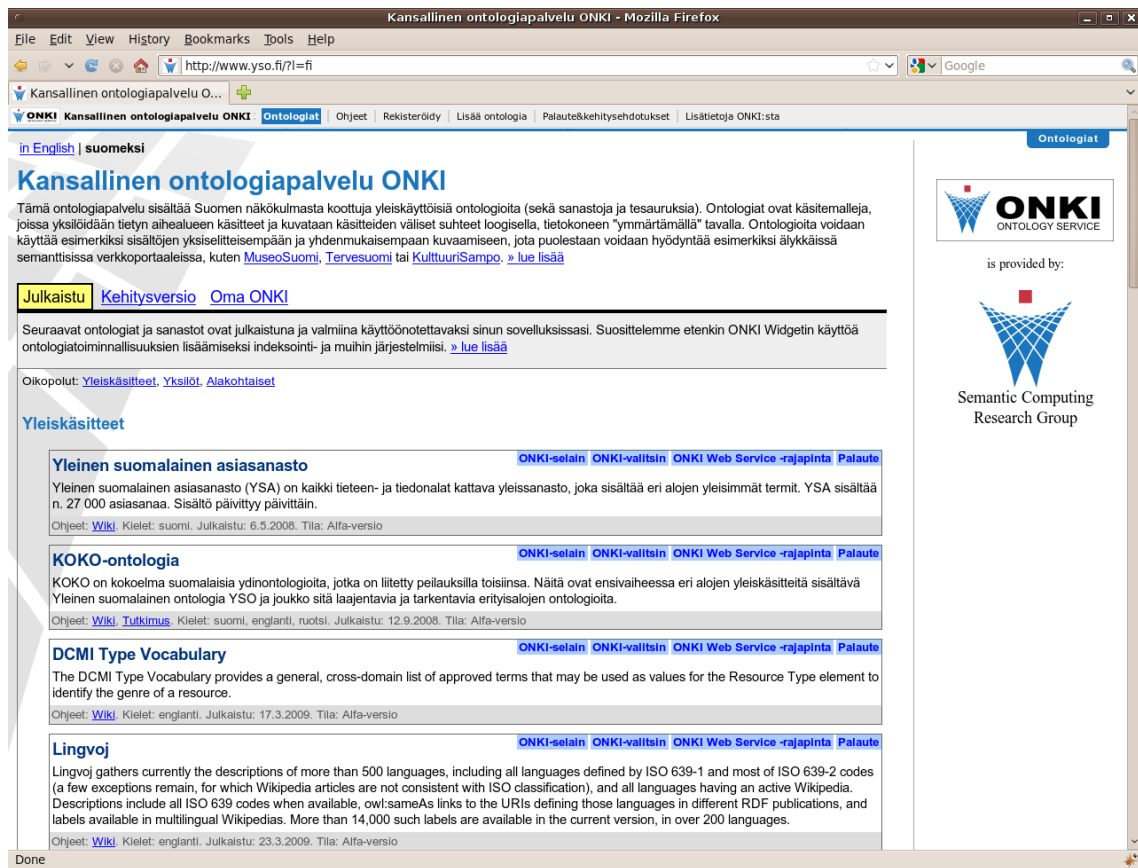


Kuva 8: ONKI-ontologiapalvelun yleisarkkitehtuuri. Kuva on Kim Viljasen konferenssiesitelmästä liittyen Viljasen ja kumppaneiden julkaisuun [VTH09].

vattu ontologioiden sisältämien käsitteiden nimien kielet sekä ontologian julkaisupäivämäärä ja tila. Ontologiat on luokiteltu ryhmiin niiden helpomman tarkastelun mahdollistamiseksi. Ylimmän tason ryhmiä ovat yleiskäsitteitä sisältävät ontologiat, alakohtaiset ontologiat ja yksilöistä koostuvat ontologiat. Alakohtaiset ontologiat jakautuvat edelleen kulttuurin, terveyden, luonnontieteiden, maa- ja metsätalouden, liike-elämän ja julkishallinnon alojen ontologioihin.

Palvelun etusivu toimii ontologiahakemistona tarjoten 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.

Tyypillisiä tapoja sanastokohtaisten julkaisujärjestelmien hyödyntämiseen sisällön



Kuva 9: ONKI-ontologiapalvelun etusivu.

kuvailussa ja haussa ovat 1) niiden selainkäyttöliittymän käyttäminen sopivan käsitteen löytämiseksi ja käsiteviittauksen (käsitteen nimi tai muu tunniste) välittäminen omaan järjestelmään manuaalisesti kopioi-liitä-menetelmällä tai käsin kirjoittamalla tai 2) kyselyiden tekeminen ohjelmallisen rajapinnan avulla [TuB05]. Kumpikaan tavoista ei ole ongelmaton. Ensimmäisessä tavassa käyttäjä joutuu jatkuvasti liikumaan kahden sovelluksen välillä ja välittämään käsiteviittauksen melko kömpelöllä tavalla. Jälkimmäisessä tavassa sanaston käyttäjätaho joutuu itse toteuttamaan käyttöliittymän sanaston käyttämistä varten.

ONKI-palvelua voi käyttää molemmilla yllä kuvatuilla tavoilla, mutta palvelu tarjoaa lisäksi käsitteiden valintaan kevyen verkkosivulle integroitavan käyttöliittymäkomponentin (engl. web widget) [MVA07, VTH08]. Käsitevalitsimen avulla yleiset ontologian käyttämiseen liittyvät toiminnallisuudet voidaan ottaa kustannustehokkaasti käyttöön verkkosovelluksessa.

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 ontologiatiedoston lisäksi ontologian metatietoja, joita ovat ontologian nimi, kuvaus, käyttöehdot ja omistajan tiedot. ONKI-palvelun ylläpito käy läpi palveluun lähetetyt ontologiat ja julkaisee laatukriteerit täyttävät ontologiat. Laatukriteereihin 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 HTTP-protokollalla noudattaen W3C:n ohjeita semanttisen webin URI-käytännöistä [SaC08] ja RDF-sanastojen julkaisemisesta [BeP08]. Näiden ohjeiden mukaan julkaistujen ontologioiden ja niiden sisältämien resurssien (käsitteiden) URI-tunnisteet toimivat URL-osoitteina mahdollistaen yksinkertaisen viittaamisen niiden määrittäisiin. URI-tunnistetta vastaavaan URL-osoitteeseen tehtävään HTTP-pyyntöön annetaan vastauksena URI-tunnisteen yksilöimän resurssin kuvaus, jonka muoto määrittyy HTTP-sisältöneuvottelua käyttäen. Esimerkiksi useimmat selaimet lähettävät HTTP-pyynnön *Accept*-kentässä arvon “text/html”, jolloin vastauksena annetaan HTML-esitys. Vastaavasti resurssin kuvauksen RDF-muodossa haluavat ohjelmat käyttävät *Accept*-kentän arvoa “application/rdf+xml” saaden vastauksen RDF-muodossa. Jos pyynnössä ei ole määritetty *Accept*-kentän arvoa, tai jos arvona on jokin muu kuin jompikumpi edellä mainituista arvoista, kuvaukset tarjotaan HTML-muodossa.

Jotta palvelussa julkaistujen ontologioiden ja niiden käsitteiden URI-tunnisteet toimivat URL-osoitteina, URI-tunnisteiden verkkotunnusosan (engl. domain) täytyy vastata sellaista verkkotunnusta, joka on palvelun ylläpitäjätahon hallinnassa, jotta internetin DNS-palvelimet voidaan konfiguroida ohjaamaan liikenne halutulle ONKI-palvelimelle. ONKI-palvelussa julkaistavien ontologioiden URI-tunnisteiksi suositellaan muotoa:

`http://www.yso.fi/onto/[ONTOLOGIA_ID]`

Esimerkiksi Yleisen suomalaisen ontologian URI-tunniste on:

`http://www.yso.fi/onto/yso`

Vastaavasti käsitteiden URI-tunnisteiksi suositellaan muotoa:

`http://www.yso.fi/onto/[ONTOLOGIA_ID]/[KÄSITE_ID]`

Esimerkiksi Yleisen suomalaisen ontologian käsitteen “kellot” URI-tunniste on:

<http://www.yso.fi/onto/yso/p3229>

Merkkijono “ONTOLOGIA_ID” on ontologian yksikäsitteinen tunniste ja “KÄSITE_ID” käsitteen tunniste. Tätä suositusta noudattamalla HTTP-pyyntöihin voidaan vastata kaikkien ontologioiden osalta yhtenevällä tavalla tekemällä tarvittavia uudelleenohjauksia ONKI-palvelimiin.

Jos URI-tunnistetta muutetaan, siihen viittaavat metatiedot muuttuvat epävalideiksi. Suositus ONKI-ontologiapalvelussa julkaistavien ontologioiden URI-tunnisteiden muodoksi on *pysyvä URI* (engl. persistent URI) [Ber98], jossa URI-tunnisteen paikallinen nimi (nimiavaruusosan jälkeinen osa) on numeerinen, eikä perustu käsitteen nimeen tietyllä kielellä. Tällöin URI-tunniste ei ole sidottu tiettyyn kieleen, joten paineita kielikohtaisten ihmisluettavien URI-tunnisteiden luomiseen ei synny. Muutospaineita ei myöskään synny, jos käsitteen nimi muuttuu tulevaisuudessa.

ONKI-palvelussa julkaistuilla ontologioilla on niiden käyttöehdot määrittävä lisenssi, joka on kuvattu ontologian metatiedoissa tekstimuodossa. Kun ontologian tai jonkin sen käsitteen tiedot pyydetään palvelusta RDF-muodossa, vastauksena annettavan RDF-datan alkuun liitetään kyseisen ontologian lisenssiteksti.

HTTP-sisältöneuvottelu ja siihen perustuva URL-uudelleenohjaus on toteutettu HTTP-palvelimella, jonka avulla HTTP-pyyntöt välitetään taustalla oleville ontologiapalvelimille. Poikkeuksen muodostavat staattiset RDF-tiedostot, jotka HTTP-palvelin hakee suoraan tiedostojärjestelmästä. Staattisia RDF-tiedostoja käytetään, kun ONKI-palvelusta pyydetään kokonainen ontologia RDF-muodossa.

ONKI-palvelun ontologian julkaisussa käyttämä HTTP-sisältöneuvottelu on testattu ja havaittu toimivaksi Vapour-palvelun [BFF08] avulla. Vapour tekee käyttäjän antamaan URL-osoitteeseen HTTP-pyyntöjä erilaisilla otsikkotiedoilla varustettuina ja tulkitsee vastausten perusteella noudattaako kyseisessä URL-osoitteessa julkaistu ontologia W3C:n ohjeita RDF-sanastojen julkaisemisesta.

ONKI-palvelun toteutuksessa käytettäviä teknologioita ovat Apache-HTTP-palvelin², Varnish-välimuistipalvelin³, PHP-ohjelmointikieli ja ARC-kirjasto⁴ RDF-tiedostojen käsittelyä varten. Palvelun käyttämät ONKI-palvelimet on toteutettu

²<http://httpd.apache.org>

³<http://varnish-cache.org>

⁴<http://arc.semsol.org>

itsenäisinä Java Servlet⁵ -perustaisina sovelluksina.

5 Ontologiapalvelin ONKI SKOS

Ontologiapalvelin ONKI SKOS on tesaurusten ja yksinkertaisten ontologioiden julkaisemiseen ja hyödyntämiseen suunnattu ontologiapalvelintoteutus. Se toteuttaa yleisen ONKI-rajapinnan ja on siten hyödynnettävissä ONKI-palveluiden avulla käyttöliittymäkomponenttina ja ohjelmallisena rajapintana.

Semanttisen webin sovellukset käyttävät tyypillisesti ontologioita, jotka ovat joko suoria muunnoksia laajassa käytössä olevista tesauruksista, sovelluskohtaisia sanastoja tai yksinkertaisia ontologioita, jotka voidaan esittää lähes samalla tavalla kuin tesaurokset [AMS04, FSH04, KPT04, HVT08]. Koska SKOS-standardi määrittää sopivan tietomallin tesauroksen esittämiseen, se valittiin ensisijaiseksi ONKI SKOS -palvelimen sanastojen esitystavaksi.

Kuvaamalla sanasto SKOS-muodossa ja julkaisemalla se ONKI SKOS -palvelimella saavutetaan useita etuja. Koska SKOS määrittää yleisen formaatin sanastojen esittämiseen, eri tesauroksia voidaan käsitellä samalla tavalla. Tällöin sovelluksissa ei tarvita sanastokohtaisia prosessointisääntöjä tai muunnostyökaluja formaattien välillä. ONKI SKOS tarjoaa pääsyn kaikkiin palvelimella julkaistuihin sanastoihin samalla tavalla, joten sanastojen käyttäjien ei tarvitse käyttää sanastokohtaisia selaimia ja muita työkaluja. Lisäksi yksi ONKI-ontologiapalvelun tavoitteista on tarjota pääsy laajaan valikoimaan sanastoja keskitetysti, jolloin sanastoja ei tarvitse etsiä useasta lähteestä. Sanaston kehittäjälle palvelu tarjoaa valmiin sanaston julkaisukanavan.

5.1 Käyttöliittymä

ONKI SKOS -palvelimen avulla voidaan selata ja visualisoida SKOS-tietomallin mukaisia sanastoja, sekä hakea niistä käsitteitä. Myös RDFS- ja OWL-ontologioita on mahdollista käyttää, mutta esimerkiksi monimutkaisten OWL-kielen rakenteiden käsittely on puutteellista. Sovellusta 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

⁵<http://java.sun.com/products/servlet/>

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 asiasanasto VNAS, 6 300 käsitettä) ja luonnontieteet (Maailman linnut -ontologia, 11 000 käsitettä).

ONKI SKOS -palvelimen käyttöliittymä ONKI SKOS -selain on esitetty kuvassa 10. Käyttöliittymä koostuu kolmesta osasta: 1) automaattisesti täydentyvä käsitte-haku, 2) käsittehierarkia ja 3) käsitteen ominaisuudet. Käsitte-haku perustuu *semanttiseen automaattisesti täydentyvään hakuparadigmaan* (engl. semantic auto-completion) [HyM06], jossa hakutulosta päivitetään käyttäjän kirjoittaessa hakuterminä tekstikenttään. Jos käyttäjä kirjoittaa tekstikenttään esimerkiksi merkki jonon “kis”, kaikki käsitteet, joiden nimi alkaa “kis”-merkkijonolla, palautetaan hakutuloksina. Automaattisesti täydentyvää käsitte-hakua käytettäessä käyttäjän ei tarvitse tuntea käytettävää ontologiaa entuudestaan, koska riittää, että hän kirjoittaa hakukenttään jotakin, ja järjestelmä ehdottaa sopivia käsitteitä. Haku täsmäytetään sekä ensisijaisiin että toissijaisiin käsitteiden nimiin. Jos haku täsmää käsitteen toissijaiseen nimeen, käyttöliittymässä näytetään toissijaisen nimen lisäksi käsitteen ensisijainen nimi sekä siihen osoittava nuoli ohjaamaan käyttäjää käyttämään käsitteestä suositeltua nimeä.

Käsittehaun hakutulokset toimivat linkkeinä käsitteen tarkastelua varten. Valitun käsitteen käsittehierarkiana näytetään tyypillisesti käsitteen ylä-, ala- ja vieruskäsitteet. Käsittehierarkia visualisoidaan puumaisena sisennettynä listana. Jos käsittehierarkiassa on moniperintää, eli jos jollakin sen käsitteistä on useampia yläkäsitteitä, puun moniperintäkohtiin muodostetaan haarautumia. Tämän seurauksena puun solmuja joudutaan monistamaan, joten tietty käsite voi esiintyä useamman kerran samassa hierarkiassa. Käsittehierarkian lisäksi valitun käsitteen ominaisuudet näytetään käyttöliittymässä taulukkomaisena esityksenä. Ominaisuudet on järjestetty ensisijaisesti literaalirvojen kielen mukaan ja toissijaisesti ominaisuuden nimen mukaan aakkosjärjestyksen perusteella. Tietyn ominaisuuden arvot on lueteltu aakkosjärjestyksessä. Käsittehierarkiassa ja ominaisuustaulukossa viittaukset muihin käsitteisiin toimivat linkkeinä viitatun käsitteen tarkastelua varten mahdollistaen ontologian selaamisen. Paikkatietoa sisältävien ontologioiden yhteydessä paikkainstanssit voidaan visualisoida kartalla käyttämällä Google Maps -palvelua⁶, mikäli paikkainstansseihin on liitetty koordinaattitieto.

Käsitteen kontekstin visualisoiminen auttaa käyttäjää ymmärtämään käsityksen

⁶<http://maps.google.com>

The screenshot shows the ONKI SKOS web interface in a Mozilla Firefox browser. The page title is "YSO - Yleinen suomalainen ontologia" and the URL is "http://www.yso.fi/onto/yso/p12732". The interface is divided into several sections:

- Käsitteiden haku (Concept Search):** Located on the left, it includes a search type dropdown (set to "Concept") and an alphabetical index (A-Z). A list of search results is shown, with "terveyden edistäminen" selected.
- Käsitteiden hierarkia (Concept Hierarchy):** A tree view showing the hierarchical structure of the concept. The path is: *yso-käsitteet* > *muuttuva* > *prosessit* > *toiminta* > *keskinäinen toiminta* > *hoito_1* > **terveyden edistäminen**.
- Käsitteen ominaisuudet (Concept Properties):** A table of properties for the selected concept:
 - Asiasana (en):** health promotion
 - Asiasana (fi):** terveyden edistäminen
 - Asiasana (sv):** hälsofrämjande
 - Lähikäsite:** [terveyskasvatus](#), [terveysneuvonta](#)
 - Tyyppi:** [Concept](#)
 - Vastaava käsite:** [terveyden edistäminen](#)
 - Yläkäsite:** [hoito_1](#)

Kuva 10: ONKI SKOS -selain.

merkityksen, erityisesti mikäli käsitteen 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 käsittehierarkiaa tutkimalla. Yläkäsitteet tarjoavat yleisempää ja alakäsitteet tarkempaa näkökulmaa.

ONKI SKOS -selainta voidaan käyttää yksittäisen ontologian visualisoinnin lisäksi toisiinsa sillattujen ontologioiden visualisointiin. Ekvivalenteiksi määriteltyjen käsitteiden käsittehierarkiat yhdistetään keskenään käsitettä visualisoidessa. Jos ontologiassa on määritetty siltaus "käsite A on ekvivalentti käsitteen B kanssa", käsitteen A yläkäsitteet lisätään käsitteen B yläkäsitteiksi, käsitteen A alakäsitteet lisätään käsitteen B alakäsitteiksi, ja päinvastoin. Kun käyttöliittymässä viitataan johonkin käsitteeseen, käsitteen nimen edessä näytetään käsitteen nimiavaruuden etuliite, jotta nähdään mihin ontologiaan käsite kuuluu ja miten ontologioiden "rajat" ilmenevät käsittehierarkiassa. Lisäksi käsittehierarkiassa näytetään käsitteen nimen perässä

ekvivalenttien käsitteiden nimiavaruuksien etuliitteet ontologioiden siltauksien tutkimisen helpottamiseksi.

Toisiinsa sillattujen ontologioiden visualisointia on hyödynnetty erityisesti Yleisen suomalaisen ontologian ja sitä laajentavien erikoisalueiden ontologioiden kehitystyössä ja käytössä. Esimerkiksi kun ontologiakehittäjä tekee ontologioiden välisiä siltauksia Protégé-editorilla [NFM00], hän ei näe editorin käyttöliittymästä suoraan miten tehdyt sillaukset vaikuttavat käsitteiden hierarkiaan, koska Protégé ei huomioi siltauksia käsittehierarkian visualisoinnissa.

Siltauksia tehtäessä usean ontologian perusteella muodostuvaan käsittehierarkiaan syntyy usein runsaasti moniperintää. ONKI SKOS poistaa käsittehierarkian visualisoinnista redundantteja haarautumia, jotta runsas moniperintä ei vaikeuta käsittehierarkian hahmottamista tarpeettomasti. Esimerkkinä tästä voidaan mainita sekä Taideteollisuusalan ontologiassa (TAO) [TAO10] että Museoalan ontologiassa (MAO) [MAO10] oleva käsite “albat”. Käsitteet on määritelty keskenään ekvivalenteiksi. TAO-ontologiassa käsitteellä on yläkäsite “kirkkotekstiilit”, kun taas MAO-ontologiassa yläkäsitteenä on “liturgiset vaatteet”, jonka yläkäsite puolestaan on “kirkkotekstiilit”. Käsittehierarkiaa visualisoitaessa ei ole tarpeen näyttää käsitteen olevan sekä käsitteen “kirkkotekstiilit” että “liturgiset vaatteet” alakäsite, vaan voidaan näyttää yläkäsitteenä vain “liturgiset vaatteet”. Moniperinnän selkeyttämisen lisäksi käsittehierarkiasta tunnistetaan syklit, jotta vältetään loputtomalta rekursiolta ja redundanttien käsitteiden lisäämisestä hierarkiaan.

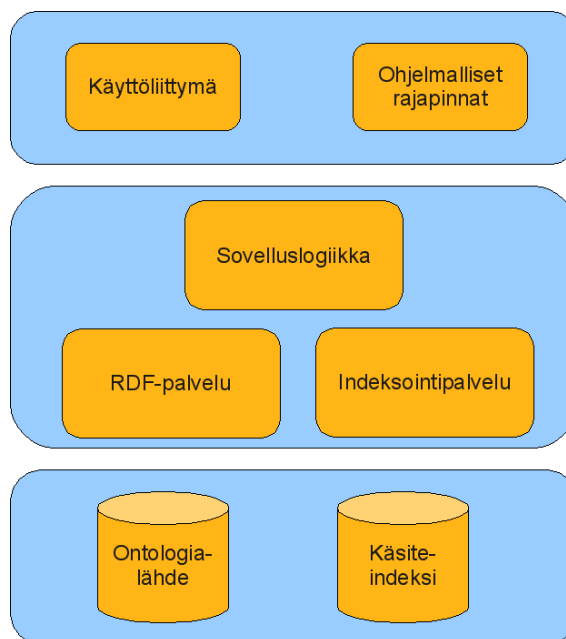
ONKI SKOS -selain on käytettävissä suomen, englannin ja ruotsin kielillä. Valittu käyttöliittymäkieli vaikuttaa käyttöliittymän elementtien kielen lisäksi käsittehakuun siten, että haku kohdistuu käyttöliittymäkielen mukaisiin käsitteisiin. Uusia käyttöliittymäkieliä voidaan lisätä luomalla uusi kielikohtainen tekstimuotoinen ominaisuustiedosto (engl. property file), joka sisältää käyttöliittymässä käytettävät tekstelementit ominaisuus-arvo-pareina halutulla kielellä.

5.2 Tekninen toteutus

ONKI SKOS -palvelimen arkkitehtuuri on toteutettu käyttäen MVC-ohjelmistoarkkitehtuurimallia (malli-näkymä-ohjain, engl. Model-View-Controller) [Try79] sekä DAO-suunnittelumallia (datan käsittelyobjekti, engl. Data Access Object) [Cor01]. MVC-arkkitehtuuri erottaa toisistaan sovellusalueen käsittelyn (malli), tietojen esitystavan (näkymä) sekä järjestelmään

kohdistuvien pyyntöjen käsittelyn (ohjain). MVC-arkkitehtuuri mahdollistaa erotettujen kokonaisuuksien kehittämisen irrallaan toisistaan ja selkiyttää ohjelmiston arkkitehtuuria. Datan käsittelyobjekti on osa MVC-arkkitehtuurin mallia. Koska datan käsittely on toteutettu keskitetysti, uusien datalähdetyyppien käyttämistä varten tarvittavien muutosten toteuttaminen rajoittuu datan käsittelyobjektiin aiheuttamatta muutoksia muuhun sovelluslogiikkaan.

Sovelluksen arkkitehtuuri on esitetty kuvassa 11. Sovelluslogiikka käyttää RDF-palvelua saadakseen ontologialähteiden (tiedosto, tietokanta tai URL-osoite) RDF-mallit ontologioiden käsittelyä varten ja Indeksointipalvelua käsittehaun suorittamiseen. Käyttöliittymä ja ohjelmalliset rajapinnat rakentuvat sovelluslogiikan päälle.



Kuva 11: ONKI SKOS -palvelimen arkkitehtuuri.

MVC-arkkitehtuurin mukainen malli koostuu seuraavista Java-pakkauksista:

- *fi.helsinki.cs.seco.onki.logic* - RDF-datan käsittely, käsitteiden esittäminen, konfiguraation käsittely,
- *fi.helsinki.cs.seco.onki.index* - käsiteindeksin käsittely,
- *fi.helsinki.cs.seco.onki.service* - objektien esittäminen ohjelmallisia rajapintoja varten,
- *fi.helsinki.cs.seco.onki.utils* - avustajaluokat ja

- *fi.helsinki.cs.seco.onki.stats* - ontologioiden tilastotietojen generointi.

Näkymä puolestaan koostuu HTML-mallineista ja staattisista tiedostoista seuraavissa hakemistoissa:

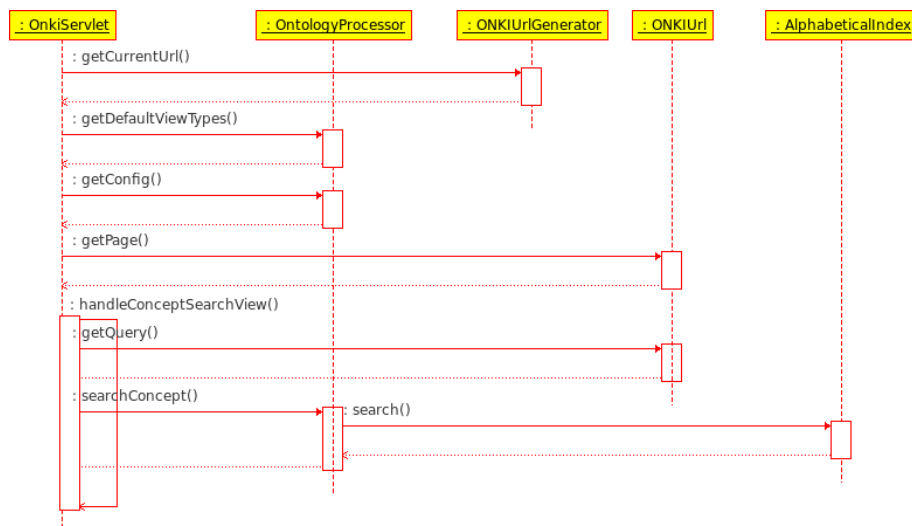
- *context/WEB-INF/app/templates* - HTML-mallineet,
- *context/app/css* - CSS-tiedostot ja
- *context/app/images* - kuvatiedostot.

Ohjain koostuu Java-pakkauksesta *fi.helsinki.cs.seco.onki.servlet* sisältäen muun muassa seuraavat luokat:

- *OnkiServlet.java* - käyttöliittymä,
- *OnkiQuery.java* - AJAX-/SOAP-rajapinnat ja
- *RestApplication.java* - HTTP-rajapinta.

Esimerkki käsitehakutilanteesta on kuvattu sekvenssikaaviona kuvassa 12. Käyttäjän syöttäessä ONKI SKOS -selaimen hakukenttään merkkijonon selain lähettää HTTP-pyyntöä palvelimelle, jossa pyyntö ohjataan *OnkiServlet*-luokalle. *OnkiServlet* pyytää pyynnön käsittelemistä varten *ONKIUrlGenerator*-luokalta URL-osoitteen, johon pyyntö kohdistui. Koska pyyntö ei sisältänyt tyyppirajoitusparametria, *OnkiServlet* pyytää *OntologyProcessor*-luokalta oletuksena käytettävän tyyppirajoituksen. Lisäksi *OnkiServlet* pyytää *OntologyProcessor*-luokalta sovelluksen konfiguraation, jota käytetään muun muassa oletushakukielen selvittämiseen. Kun *OnkiServlet* on pyytänyt *OnkiUrl*-luokalta tiedon nykyisestä sivusta (hakusivu), *OnkiServlet* kutsuu käsitehaun käsittelevää metodologiaan. Seuraavaksi *OnkiServlet* pyytää *OnkiUrl*-luokalta hakumerkkijonon ja välittää sen *OntologyProcessor*-luokalle hakua varten. *OntologyProcessor*-luokka puolestaan kutsuu *AlphabeticalIndex*-luokan metodologiaa, jossa varsinainen käsitehaku suoritetaan. Lopuksi *OnkiServlet* valitsee käsitehakua vastaavan HTML-mallineen, jonka mallineprosessori käsittelee, ja valmis HTML-sivu palautetaan selaimelle.

ONKI SKOS -selain lähettää kaikki ONKI SKOS -palvelimelle välitettävät tiedot HTTP GET -pyyntöinä, eikä mitään tietoa tallenneta sessioihin. Tietynsisältöistä sivua vastaa siten yksi URL-osoite, joten esimerkiksi sivujen tallennus palvelimen



Kuva 12: Käsithaku sekvenssikaaviona.

välimuistiin on tehokasta ja sivujen lisääminen selaimen kirjanmerkkeihin on mahdollista.

ONKI SKOS -palvelimen konfigurointiin käytetään seuraavia konfiguraatio-tiedostoja:

- *applicationContext.xml* - ontologian RDF-tiedoston polku tai RDF-tietokannan tiedot,
- *onki-services-conf.xml* - pääkonfiguraatitiedosto, muun muassa RDF-ominaisuuksien käsittely,
- *OnkiLocales_[KIELI].properties* - käyttöliittymän kielikohtaiset tekstelementit ja
- *onki-services-visualization.rdf* - yleisiä RDF-ominaisuuksien nimiä käyttöliittymäesitystä varten.

ONKI SKOS -palvelin on toteutettu Java Servlet -tekniikalla. RDF-tiedostojen käsittelyyn käytetään Jena Semantic Web Framework -kirjastoa⁷, käsithakuun Lucene-hakukonetta⁸, ladattavien ontologiatiedostojen konfigurointiin Spring-ohjelmointikehystä⁹ ja käyttöliittymäsivujen tekemiseen Velocity-mallinekieltä¹⁰.

⁷<http://jena.sourceforge.net>

⁸<http://lucene.apache.org>

⁹<http://www.springsource.org>

¹⁰<http://velocity.apache.org>

Vaatimukset ONKI SKOS -palvelimen ajamiseksi ovat:

- Java 2 -ohjelmistoalusta, Standard Edition 5.0 tai uudempi¹¹ ja
- Servlet-alusta (testattu Apache Tomcat 6.0 -alustalla)¹².

5.3 Sanastojen käsittely

SKOS-muotoisen sanaston julkaiseminen ONKI SKOS -palvelimella on yksinkertaista, sillä ainoastaan sanaston sisältävän RDF-tiedoston polku täytyy lisätä konfiguraatiotiedostoon. Kun palvelin käynnistetään, määritelty RDF-tiedosto ladataan, indeksoidaan ja julkaistaan käyttäjien saataville. Jos palvelimella halutaan julkaista jotakin muuta tietomallia kuin SKOS-standardia käyttävä sanasto, palvelimen asetukset pitää konfiguroida tietomallille sopiviksi, jotta sanasto voidaan käsitellä halutulla tavalla. Konfiguroitavia asetuksia ovat muun muassa käsittehierarkian muodostamisessa käytettävä ominaisuus, käsitteiden nimeämiseen käytettävä ominaisuus, oletuksena näytettävä käsite ja käsitehaun rajauksessa oletuksena käytettävä käsitetyyppi.

Kun ONKI SKOS -selaimen etusivu ladataan, käyttöliittymässä näytetään oletuksena näytettävän käsitteen hierarkia ja ominaisuudet. Useimmiten oletuksena halutaan näyttää sanaston juurikäsite, jos sanastossa on sellainen. Asiasanastoissa ei usein ole yhtä juurikäsitettä, joten SKOS-sanastojen tapauksessa näytetään oletuksena SKOS-käsiteskeemaa edustava resurssi (tyyppi *skos:ConceptScheme*). Käyttöliittymässä näytetään otsikkona konfiguraatiotiedostossa otsikoksi määritetty merkkijono, tai jos sellaista ei ole, ontologiaa (tyyppi *owl:Ontology*) tai käsiteskeemaa edustavan resurssin nimi.

Käsitteen käsittehierarkia muodostetaan kulkemalla konfiguraatiossa määritettyä ominaisuutta RDF-verkossa transitiivisesti. SKOS-tietomallissa käsitteiden väliset hierarkkiset suhteet kuvataan ominaisuudella *skos:broader* ja tämän käänteisominaisuudella *skos:narrower*. RDFS-ontologioissa luokkahierarkian määrittää ominaisuus *rdfs:subClassOf*. Lisäksi SKOS-tietomallia käytettäessä käsittehierarkian juurena näytetään ylätasoa käsitteisiin ominaisuudella *skos:hasTopConcept* liittyvä käsiteskeemaa edustava resurssi.

Toisiinsa sillattuja ontologioita käsiteltäessä keskenään ekvivalenttien käsitteiden

¹¹<http://java.sun.com/javase/>

¹²<http://tomcat.apache.org>

hierarkiat voidaan yhdistää keskenään. OWL-ontologioissa käsitteiden (luokkien) välinen ekvivalenttius määritellään ominaisuuden *owl:equivalentClass*. Ominaisuuden semantiikka pitää sisällään sen, että ekvivalentit käsitteet ovat toistensa alaluokkia. Tällä voidaan perustella käsitteiden hierarkioiden yhdistäminen.

Käsitteiden nimiä käytetään käsittehaun tuloslistauksen, käsittehierarkian ja käsitteen ominaisuustaulukon visualisoinnissa. SKOS-tietomallissa käsitteen nimi kuvataan ominaisuudella *skos:prefLabel*, RDFS-ontologioissa puolestaan ominaisuudella *rdfs:label*. Näiden lisäksi tyypillinen käsitteiden nimeämiseen käytetty ominaisuus on *dc:title*. Kun käyttöliittymässä viitataan johonkin käsitteeseen, sen käyttöliittymäkielen mukainen ensisijainen nimi näytetään. Jos käyttöliittymäkielen mukaista nimeä ei ole saatavilla, näytetään nimi jollakin saatavissa olevalla kielellä. Jos nimeä ei ole saatavilla millään kielellä, käyttöliittymässä näytetään käsitteen URI-tunnisteen paikallinen nimi. Vastaavasti käsitteen ominaisuuksien nimet näytetään käsitteen ominaisuustaulukossa niihin viitattaessa. Jotta yleisien ominaisuuksien nimiä ei tarvitse tallentaa jokaiseen ontologian datatiedostoon käyttöliittymäesitystä varten, ONKI SKOS -palvelimeen ladataan oletuksena RDF-tiedosto, joka sisältää yleisien standardien (RDF, RDFS, OWL, SKOS, Dublin Core) ominaisuuksien nimiä.

Käsiteviittaukset toimivat oletuksena linkkeinä kyseisen käsitteen tarkasteluun ONKI SKOS -selaimessa. Viitattavat käsitteet eivät kuitenkaan aina sisälly selailtavaan ontologiaan. Tällaisessa tapauksessa voidaan konfiguroida ominaisuudet, joilla tehtyjen käsiteviittauksien linkit osoittavat käsitteen URI-tunnisteen mukaiseen URL-osoitteeseen. Näin käyttäjä voidaan ohjata viitattavan käsitteen määrittämiseen, joka voi sijaita esimerkiksi jossakin toisessa ONKI-palvelussa julkaistussa ontologiassa tai missä tahansa webissä sijaitsevassa ontologiassa. Lisäksi literaaliarvot, jotka on tyypitetty XML Schema -suosituksen anyURI-tietotyyppiin [BiM04] mukaisiksi, toimivat linkkeinä kyseisen arvon mukaiseen URL-osoitteeseen.

Käsittehaun täsmäytyksessä käytetään käsitteiden ensisijaisia ja mahdollisia toissijaisia nimiä. Aiemmin mainittujen käsitteiden nimeämiseen käytettävien ominaisuuksien lisäksi SKOS-tietomalli sisältää ominaisuudet *skos:altLabel* ja *skos:hiddenLabel*, joilla esitetään käsitteiden toissijaisia nimiä. Nopeiden hakujen mahdollistamiseksi käsitteiden nimistä pidetään yllä hakuindeksiä.

Käsittehaaku voidaan rajoittaa tiettyntyyppisiin käsitteisiin (kuvattu ominaisuudella *rdfs:type*), jonka mahdollistamiseksi käsitteen tyyppi tallennetaan hakuindeksiin. Vastaavasti haku voidaan rajoittaa tiettyyn ontologian käsittehierarkian alipuuhun koh-

distamalla haku tietyn käsitteen alakäsitteisiin, joten myös käsitteiden yläkäsitteet tallennetaan hakuindeksiin. Hakuindeksiin tallennetaan oletuksena kaikki ladatun ontologian resurssit, joilla on nimi. Jos ontologian koko sisältöä ei kuitenkaan haluta haun kohteeksi, indeksoitavien resurssien joukkoa voidaan rajoittaa määrittämällä konfiguraatiossa nimiavaruudet, joihin kuuluvat resurssit indeksoidaan. Tämän lisäksi voidaan määrittää tyypit, joiden mukaiset resurssit indeksoidaan, tai vaihtoehtoiset tyypit, joiden mukaisia resursseja ei indeksoida.

Monet asiasanastot sisältävät rakenteita, joilla sanaston käsitteet ryhmitellään tiettyihin kategorioihin. Kategoriatiedon hyödyntämiseksi käsittehaussa ONKI SKOS-selaimessa on erillinen hakukenttä, jolla voidaan hakea kategorioita merkkijonohauulla. Kun hakutuloksista valitaan jokin kategoria, käsittehaun rajoitetaan kyseisen kategorian käsitteisiin, jolloin käsittehaun hakutuloksiksi päivitetään valittuun kategoriaan kuuluvat käsitteet. Tämän jälkeen käsittehaun voidaan edelleen jatkaa valitun kategorian sisällä. Käsitteiden kategoriat on tallennettu hakuindeksiin. SKOS-tietomalli sisältää tavan kuvata käsittekokoelmia (tyyppi *skos:Collection* ja ominaisuus *skos:member*), jota voidaan käyttää kategorioiden mallintamiseen. Kyseistä rakennetta tosin käytetään usein hieman eri tarkoitukseen, nimittäin vieruskäsittejoukon ryhmittelemiseen tietyllä perusteella (engl. node label). Esimerkki tällaisesta ryhmittelystä on Yleisen suomalaisen ontologian “aine käytön mukaan”. Tämän takia kyseistä rakennetta ei käytetä oletuksena kategorioiden muodostamiseen.

Koska ONKI SKOS -palvelin on suunniteltu yksinkertaisten SKOS-tyyppisten ontologioiden käsittelyyn, käyttöliittymässä näytetään valitun käsitteen ominaisuuksina vain sellaiset RDF-kolmikot, joissa valittu käsite on subjektina ja objektina on jokin literaali tai URI-resurssi. Usein käsitteen ominaisuuksia on kuitenkin esitetty monimutkaisempien RDF-rakenteiden avulla. Tällaisten ominaisuuksien näyttämiseksi voidaan toteuttaa tapauskohtaisesti *PropertyFiddler*-rajapinnan toteuttavia Java-luokkia, joilla käsitellään RDF-verkkoa ja poimitaan siitä tietyn ehdon toteuttavia kolmikoita käyttöliittymäesitystä varten. Esimerkiksi Yleinen suomalainen ontologia sisältää koostekäsitteitä, jotka ovat monimerkityksellisiä käsitteitä, mutta joita voidaan kuitenkin haluta käyttää tiedon kuvailussa, kuten koostekäsite “lapset (kooste)”, joka koostuu käsitteistä “lapset (ikään liittyvä rooli)”, “lapset (perheenjäsenet)” ja “lapset (sosioekonomiseen ryhmään liittyvä rooli)”. Koosteisuus on mallinnettu YSOssa siten, että koostekäsite on ekvivalentti (kuvattu ominaisuudella *owl:equivalentClass*) ominaisuuden *owl:unionOf* avulla esitetyn käsitteyhdistelmän kanssa. Java-luokan avulla voidaan tässä tapauksessa määrittää, että jos käsite on osa tällaista yhdistelmää, sen ominaisuustaulukossa näytetään viite

koostekäsitteeseen sopivalla ominaisuudella kuvattuna. Tämä on samantyyppinen lähestymistapa kuin Fresnel-sanastolla [PBK06], jolla voidaan määrittellä deklaratii-visesti miten RDF-verkko visualisoidaan käyttöliittymissä.

Käsitteen ominaisuustaulukosta voidaan myös jättää pois ominaisuuksia, joita ei haluta näyttää käyttäjälle. Tällaisia ovat tyypillisesti ontologian kehittäjien lisäämät sisäiset kommentit, käsitetunnisteet ja tiedot, joita ei ole vielä vahvistettu. Jos paikkatietoa sisältävien ontologioiden paikkainstansseja halutaan visualisoida kartalla, paikkojen koordinaattitiedot sisältävä ominaisuus voidaan määrittellä konfiguraatiossa.

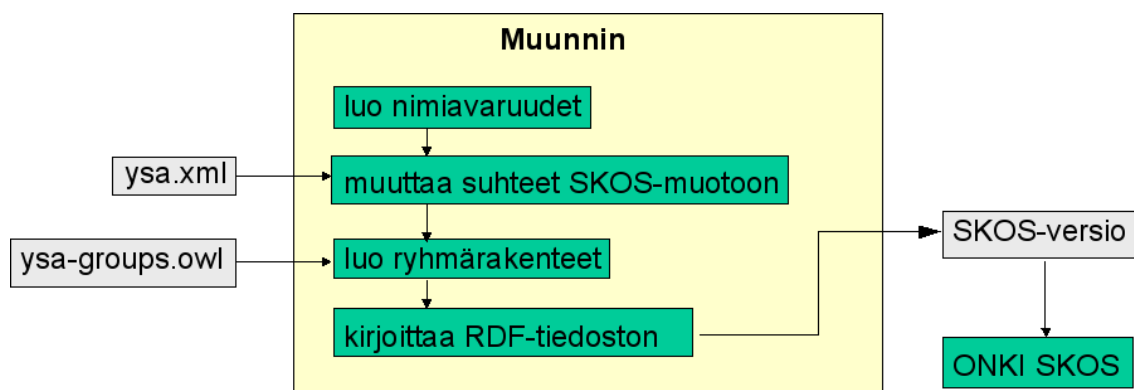
ONKI SKOS -palvelimen esimerkkikonfiguraatio SKOS-sanastolle on kuvattu liitteessä 1.

5.4 Olemassa olevan sanaston muuntaminen SKOS-muotoon

On olemassa paljon vakiintuneita sanalistoja, tesauroksia ja muita kontrolloituja sanastoja, joita ei ole julkaistu RDF-muodossa. Jotta näiden sanastojen kehittämiseen käytetty työ saataisiin hyödynnettyä semanttisessa webissä, tarvitaan muunnosprosesseja [AMM06, HaT07, SAR08, HVT08].

Esimerkkinä sanaston muuntamisesta SKOS-muotoon ja julkaisemisesta ONKI-palvelussa esitetään Yleisen suomalaisen asiasanaston (YSA) muunnos- ja julkaisuprosessi. YSA-sanaston muunnosprosessi alkaa sanaston julkaisemisella MARCXML-formaatissa [MAR09] Kansalliskirjaston käyttämästä asiasanaston hallintajärjestelmästä HTTP-palvelimelle. Tiedosto noudetaan Kansalliskirjaston HTTP-palvelimelta ONKI-palvelimelle päivittäisellä automatisoidulla ajolla käyttäen OAI-PMH-protokollaa [LSN08]. Kun tiedosto on noudettu, aloitetaan varsinainen SKOS-muunnosprosessi, joka on kuvattu kuvassa 13.

Java-sovelluksena toteutettu muunnosohjelma lataa MARCXML-muotoisen tiedoston (*ysa.xml*), luo Jena-kirjaston avulla uuden RDF-mallin ja lisää siihen nimiavaruusmääritykset SKOS-tietomallia ja YSA-sanastoa varten. RDF-malliin lisätään myös YSA-sanastoa edustava resurssi (tyyppi *skos:ConceptScheme*). Seuraavaksi MARCXML-määrityksen mukaiset rakenteet muunnetaan vastaaviksi SKOS-rakenteiksi. Käsitteen URI muodostetaan uniikin MARCXML-tiedostossa käytetävän tunnisteiden perusteella. Käsitteen ensisijainen nimi (*skos:prefLabel*), mahdolliset vaihtoehtoiset nimet (*skos:altLabel*) ja huomautustekstit (*skos:note*) muunnetaan sellaisenaan syntaksista toiseen. Käsitteet tyypitetään ominaisuudella *rdf:type*



Kuva 13: Yleisen suomalaisen asiasanaston muuntaminen SKOS-muotoon.

tyypiksi *skos:Concept* ja lisätään kuulumaan ominaisuudella *skos:inScheme* YSA-sanastoa edustavaan käsiteskeemaan. Koska MARCXML-formaatissa käsitteisiin viitataan niiden nimillä, eikä tunnisteilla, nimiä vastaavat tunnisteet haetaan URI-muunnoksia varten muunnettaessa käsitteiden välisiä suhteita (*skos:broader*, *skos:narrower*, *skos:related*). Kun syntaktinen muunnos on valmis, muunnosohjelma hakee YSA-käsitteiden kategorioille nimet erillisestä OWL-tiedostosta (*ysa-groups.owl*), koska MARCXML-tiedostossa kategorioihin viitataan tunnisteilla, eikä niiden nimiä ole sisällytetty MARCXML-tiedostoon. Lopuksi SKOS-sanasto kirjoitetaan RDF-tiedostoon ja ladataan ONKI SKOS -palvelimelle.

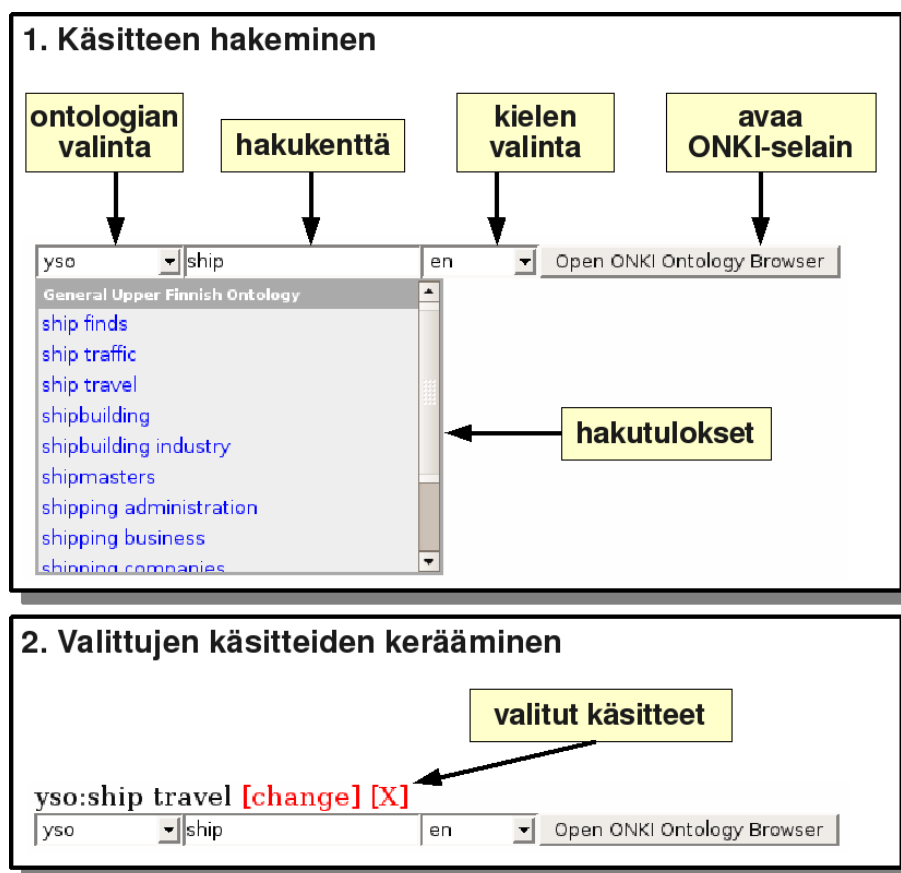
6 Ontologiapalveluiden kustannustehokas hyödyntäminen

ONKI-ontologiapalveluita voidaan käyttää kustannustehokkaasti asiakasjärjestelmissä käyttöliittymätasolla integroitavalla käsitevalitsimella sekä ohjelmallisilla rajapinnoilla. Tämän luvun aliluvussa 6.1 käsitellään käsitevalitsimen erityisesti käyttöä tiedon kuvailussa ja haussa. ONKI-palveluiden ohjelmallista käyttämistä kuvataan aliluvussa 6.2.

6.1 Käsitevalitsin

ONKI-ontologiapalvelut voidaan ottaa käyttöön verkkosovelluksissa kevyellä käyttöliittymätason integroinnilla samalla idealla kuin Google AdSense [Goo10], mutta

tässä tapauksessa asiakasjärjestelmään tuodaan mainosten sijaan ontologian hyödyntämiseen tarvittavia toimintoja. ONKI-palvelun käsitevalitsin on erityisesti tiedon annotoijan ja hakijan käyttötärpeisiin kehitetty käyttöliittymäkomponentti, jonka avulla asiakasjärjestelmään voidaan noutaa viitteitä sanastojen käsitteisiin. Käsitevalitsin on kuvattu kuvassa 14. Käsitevalitsimella voidaan tehdä merkkijonohakuja ontologiaan (kuvan 14 osa 1), selata ontologiaa ja kerätä käsiteviittauksia käsitekeräimeen (kuvan 14 osa 2). Samantyyppistä ideaa käyttöliittymätason käsitehakukomponentista on käytetty myös OCLC Terminology Web Services -palvelussa [VCH05], BioPortalissa (Ontology Widgets) [RMK08] sekä yleisesti RDF-varaston yhteydessä [HOA07].



Kuva 14: ONKI-palvelun käsitevalitsin.

Käsitevalitsimen käsitteen merkkijonohaku toimii samalla tavalla kuin ONKI SKOS-selaimessa täydentymällä automaattisesti käsitteisiin ja täsmäyttämällä hakutermin käsitteiden ensisijaisiin ja toissijaisiin nimiin. Valitun ontologian ONKI-selain voidaan avata painamalla käsitevalitsimen painiketta "Ava 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 kopioi-liitä-menetelmästä, jota joudutaan usein käyttämään perinteisten sanastoselainten yhteydessä.

Kun käsite valitaan käsitevalitsimella, se lisätään käsitekeräimen ihmisluettavan 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, viittausten yksilöivyyden kannalta on järkevää tallentaa nimen lisäksi käsitteen nimiavaruuden etuliite, joka toimii tyypillisesti ontologian tunnisteena. Jos ontologian käsitteiden nimet ovat uniikkeja ontologiassa, eivätkä ne muutu, tämä viittaustapa on kohtuullinen tiedon tallennuksen ja tulevaisuuden käytön kannalta.

Käsitevalitsin voidaan integroida verkkosivulle lisäämällä sivulle kaksi riviä JavaScript-koodia. Seuraava koodirivi lataa käsitevalitsimen käyttämiseen tarvittavan JavaScript-kirjaston selaimen. Koodirivi lisätään tyypillisesti HTML-sivun *HEAD*-osioon.

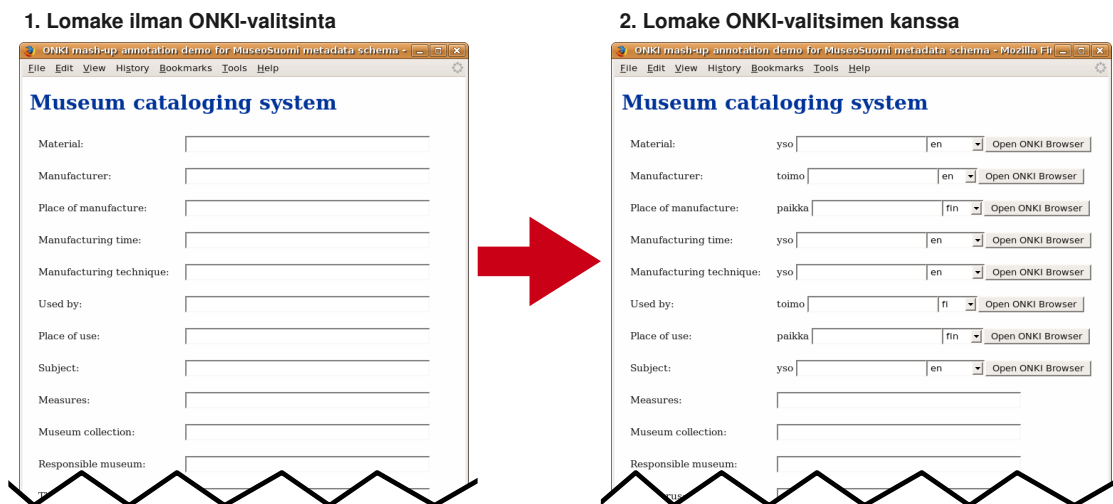
```
<script type="text/javascript"
      src="http://www.yso.fi/onki.js?KEY&l=LANG"></script>
```

Koodirivin merkkijono “KEY” viittaa ONKI-palvelun käyttämiseen tarvittavaan avaimen, jonka käyttäjä saa rekisteröityessään ONKI-palveluun. Jos avain jätetään tyhjäksi tai se on virheellinen, käsitevalitsimessa näytetään kehoitus rekisteröityä palvelun käyttäjäksi. Avain on verkkotunnuskohtainen, joten se on validi vain rekisteröidytessä ilmoitetun verkkotunnuksen alaisuudessa olevilla verkkosivuilla. Rekisteröitymiskäytännön avulla pyritään ylläpitämään tilastoja palvelun käyttäjämääristä ja -tahoista. Merkkijono “LANG” puolestaan ilmaisee käsitevalitsimen halutun käyttöliittymäkielen. Käyttöliittymä on toteutettu suomeksi ja englanniksi.

JavaScript-kirjaston avulla tavalliset HTML-lomakkeen tekstikentät voidaan kytkeä ontologiapalveluun. Lisäämällä tekstikenttään *onkeyup*-tapahtumankäsittelijä tekstikenttä muutetaan dynaamisesti käsitevalitsimeksi sivun latauksen yhteydessä. Seuraava koodirivi kytkee käsitevalitsimen tekstikenttään.

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

Koodirivin merkkijono “ONTOLOGIA_ID” viittaa ontologiaan, johon käsitevalitsimen haku- ja selailutoiminnallisuudet halutaan kohdistaa oletuksena. Esimerkiksi Yleisen suomalaisen ontologian tunniste on “yso”. Tekstikentällä on oltava *id*-attribuutti, jotta siihen voidaan viitata yksikäsitteisesti. Kuvassa 15 on esitetty kuvitteellinen museon annotointisovellus ennen ja jälkeen käsitevalitsimen kytkemistä.



Kuva 15: Museon luetteloitijärjestelmä ennen ja jälkeen käsitevalitsimen kytkemistä.

Käsitevalitsimen toimintaa voidaan konfiguroida erilaisilla parametreilla, jotka on kuvattu taulukossa 2. Parametreilla voidaan esimerkiksi asettaa hakuun lisärajoituksia tai poistaa käsitevalitsimesta käyttötärpeen kannalta tarpeettomia käyttöliittymäelementtejä. Lisäksi käsitevalitsimen käsitekeräimeen voidaan lisätä esivalittuja käsitteitä, esimerkiksi aiemmin asiakasjärjestelmään noudettuja käsitteitä, jotka halutaan näyttää käyttäjälle. Tämä on mahdollista lisäämällä HTML-sivulle piilotettuja tekstikenttiä, joiden nimeäminen ja arvojen (käsiteviittauksien) syöttäminen on toteutettu samalla tavalla kuin käsitevalitsimen omissa kentissä. Kun käsitevalitsin luodaan sivun latauksen yhteydessä, tällaiset piilokentät käydään läpi ja poistetaan, ja niiden perusteella käsitteet lisätään käsitekeräimeen.

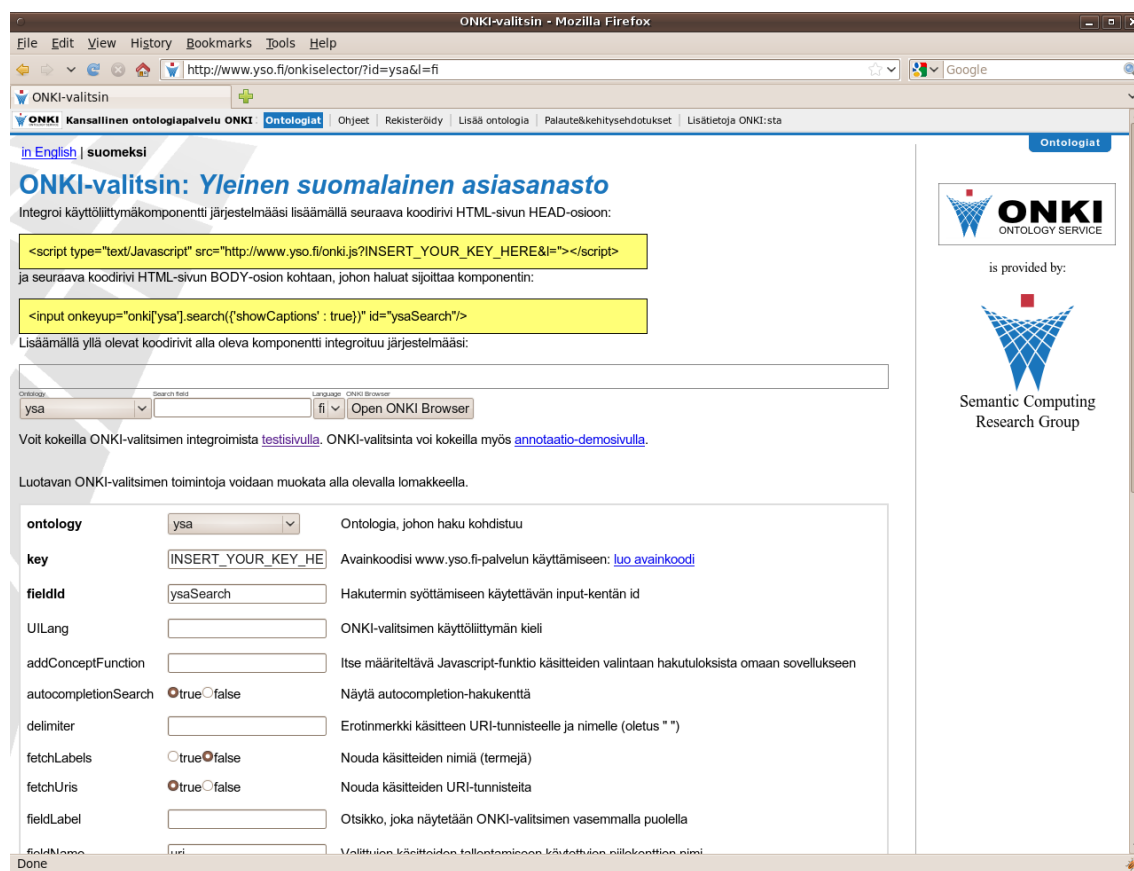
Käsitevalitsimen käyttöönoton helpottamiseksi ONKI-palveluun on toteutettu palvelu¹³, jolla voidaan luoda käsitevalitsimen integrointiin tarvittava JavaScript-koodi.

¹³<http://www.yso.fi/onkiselector/>

Parametrin nimi	Tietotyyppi	Kuvaus
addConceptFunction	viite funktioon	JavaScript-funktio käsitteiden noutoon hakutuloksista omaan sovellukseen
autocompleteSearch	totuusarvo	Näytetäänkö hakukenttää
delimiter	merkkijono	Noudettujen käsitteiden URI-tunnisteiden ja nimien erotin
fetchLabels	totuusarvo	Noudetaanko käsitteiden nimiä
fetchUris	totuusarvo	Noudetaanko käsitteiden URI-tunnisteita
fieldLabel	merkkijono	Otsikko, joka näytetään käsittevalitsimen vasemmalla puolella
fieldName	merkkijono	Valittujen käsitteiden tallentamiseen käytettyjen piilokenttien nimi
groupRestriction	merkkijono	Haun rajaus tietyn ryhmän käsitteisiin (URI)
lang	merkkijono	Haun rajaus tietyllä kielellä esitettyihin käsitteiden nimiin
languageMenu	totuusarvo	Näytä valitun ontologian vaihtoehtoiset kielet valikossa
maxHits	kokonaisluku	Hakutulosten enimmäismäärä
maxSelect	kokonaisluku	Valittavien käsitteiden enimmäismäärä
onkimenu	totuusarvo	Näytetäänkö valintaluettelo palvelun ontologioista
openonkibutton	totuusarvo	Näytetäänkö "Avaa ONKI-selain" -painike
openonkilabel	merkkijono	"Avaa ONKI-selain" -painikkeessa näytettävä teksti
parentRestriction	merkkijono	Haun rajaus käsitteen (URI) alipuuhun
prefix	totuusarvo	Liitetäänkö hakutuloksista valittujen käsitteiden nimiin nimiavaruuseliite
showCaptions	totuusarvo	Näytetäänkö käyttöliittymäelementeille kuvaustekstit
showOnkiName	totuusarvo	Jos ontologioiden valintaluettelo on asetettu pois näkyvästä, näytetäänkö valitun ontologian nimi
typeRestriction	merkkijono	Haun rajaus tietyn tyyppiin käsitteisiin (URI)

Taulukko 2: Käsittevalitsimen parametrit.

Käsitevalitsimen halutut parametrit asetetaan lomakkeella, jonka lähettämällä tarvittava koodi sekä parametrisoitu käsitevalitsin näytetään käyttäjälle. Palvelun käyttöliittymä on esitetty kuvassa 16.



Kuva 16: Käsitevalitsimen integrointikoodin generointilomake.

Jotta ONKI-selaimia voidaan hyödyntää käsitteen valintaan käsitevalitsinta käytettäessä, käsitevalitsimen toteutuksessa on otettu huomioon selainten tietoturvakäytäntöihin liittyvät rajoitukset. Keskenään eri verkkotunnuksista ladatut verkkosivut eivät pysty käsittelemään toistensa sisältöjä, eli tässä tapauksessa ONKI-selain ei pysty suoraan lisäämään asiakasjärjestelmän käytössä olevaan käsitevalitsimeen käyttäjän valitsemaa käsitettä. Rajoitus on ratkaistu käyttämällä tekniikkaa, joka perustuu siihen että *iframe*-elementtiin ladattu sivu voi kommunikoida isäntäsivun (sivun, jonka sisällä elementti on) kanssa muuttamalla selainikkunan URL-osoitetta, vaikka verkkosivut olisivatkin ladattu eri verkkotunnuksista. ONKI-selain ladataan siten uuden selainikkunan sisällä *iframe*-elementtiin. Kun jokin käsite valitaan, ONKI-selain lisää selainikkunan URL-osoitteeseen merkkijonon muotoa *#/[KÄSITE-URI]*. Tällöin selainikkunan sisältöä ei kuitenkaan ladata uudelleen,

koska vain URL-osoitteen loppuosa muuttuu. Isäntäsivun JavaScript-koodi tutkii tasaisin väliajoin selainikkunan URL-osoitteen arvon, ja havaitessaan sen muuttuneen välittää käsitteen URI-tunnisteen käsitevalitsimelle ja sulkee ONKI-selainmen. Muuhun kommunikointiin ONKI-palvelimien kanssa käsitevalitsin käyttää ohjelmallista ONKI-rajapintaa, joka esitellään myöhemmin tässä luvussa.

Käsitevalitsinta voidaan käyttää tiedonhakua varten sopivien hakutermien löytämiseen ontologioista. Lisäksi hakutermeistä muodostettua kyselyä voidaan laajentaa ontologisen tietämyksen perusteella [TKV09]. 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 luokkatäi osa-kokonaisuushierarkiaa, tai käsitteiden välisiä assosiatiivisia suhteita. Kyselyn laajentamiseen käytettävät suhteet määritetään ontologiakohtaisesti ontologia-palvelimen konfiguraatiossa.

Luokka- ja osa-kokonaisuushierarkian käyttäminen ratkaisee ongelmia, joissa tiedon hakija käyttää termistöä eri tarkkuustasolla kuin millä kyseistä termistöä käytetään tiedonhakujärjestelmän hakuindeksissä. Esimerkiksi jos käyttäjä hakee työkaluja käsitteleviä dokumentteja, kysely voidaan laajentaa kirveisiin ja sahoihin. Tällöin haun tuloksina saadaan kirveitä ja sahoja käsitteleviä dokumentteja, vaikka kyseisten dokumenttien kuvauksissa ei olisikaan mainittu työkaluja. Toisaalta käyttäjän hakies-sa esimerkiksi pohjoismaiden karttaa, osa-kokonaisuushierarkiaa käyttävän kyselyn laajennoksen avulla hakutuloksena saatava Suomen kartta saattaa olla epärelevantti tulos.

Muiden kuin hierarkkisten suhteiden käyttäminen hierarkkisten suhteiden sijaan tai niiden lisäksi kyselyn laajennoksessa voi olla hyödyllistä. Geneeristä assosiatiivista suhdetta käytettäessä on syytä huomioida, että sen käyttäminen voi johtaa hakutulospöytäkirjojen liialliseen kasvuun ja siten tarkkuuden heikkenemiseen [TAJ01, HSW07]. Assosiatiivisen suhteen käyttäminen lisää hakutuloksiin jollain tavalla alkuperäisiin hakutermeihin liittyviä tuloksia, mutta nämä tulokset eivät välttämättä ole tiedonhaktehtävän kannalta relevantteja. Epärelevanttien hakutulosten vaikutusta voidaan vähentää järjestämällä hakutulokset relevanssin perusteella asiakasjärjestelmässä. Paikkaontologioita käytettäessä paikkoja voidaan laajentaa esimerkiksi maantieteellisten aluetietojen ja niiden ajallisten muutosten perusteella. Esimerkiksi kunta voidaan laajentaa naapurikuntiinsa tai sellaisiin kuntiin, jotka ovat aiemmin sisältäneet alueita, jotka nykyhetkellä kuuluvat valittuun

kuntaan.

Hollinkin ja kumppaneiden tutkimuksessa [HSW07] alakäsitteiden käyttäminen kyselyn laajennoksessa paransi haun saantia heikentämättä tarkkuutta. Alakäsitteiden lisäksi osa-kokonaisuussuhdetta käytettäessä haun saanti parani tarkkuuden pysyessä hyväksyttävällä tasolla. Vastaavasti yläkäsitteiden käyttäminen paransi hakutuloksia, mutta yläkäsitteiden lisäksi muiden käsitteiden välisten suhteiden käyttäminen heikensi tarkkuutta enemmän kuin paransi saantia.

Kyselyn laajennos voidaan tehdä käsitevalitsimella joko käsitteiden URI-tunnisteita tai nimiä käyttäen. Jos tiedonhakupöytäkirja perustuu termeihin URI-tunnisteiden sijaan, käsitteiden toissijaisten nimien (synonyymit) käyttäminen kyselyn laajentamisessa voi olla hyödyllistä. Jos ontologia sisältää käsitteiden nimiä useilla kielillä, käsitevalitsinta voidaan käyttää kieltenvälisessä tiedonhaussa. Hakutermejä voidaan hakea esimerkiksi suomen kielellä, mutta valittuja hakutermejä ja niiden laajennoksia käytetään varsinaisessa tiedonhakupöytäkirjässä englannin kielellä. Tämä on hyödyllistä silloin, kun käyttäjä ei ymmärrä kieltä, jota on käytetty haettavan sisällön kuvailussa.

Kyselyn laajennokset välitetään asiakasjärjestelmään samalla tavalla kuin esimerkiksi tiedon kuvailua varten valitut käsitteet eli käyttämällä HTML-lomakkeen piilotettuja tekstikenttiä. Jokaista valittua käsitettä kohden luodaan piilokenttä, johon tallennetaan valittu käsite ja sitä laajentavat käsitteet. Tyypillisessä tapauksessa käsitettä laajentavat käsitteet halutaan lisätä keskenään vaihtoehtoisina hakutermeinä. Asiakasjärjestelmissä käytetään erilaisia syntakseja kyselyiden muodostamiseen. Käsitevalitsin tarjoaa käyttäjälle mahdollisuuden määrittää sovelluskohtaisia JavaScript-funktioita, jotka muodostavat yksittäisistä hakutermeistä halutunlaisen hakulausekkeen. Esimerkiksi Boolean algebran mukainen hakulauseke (*työkalut OR kirveet OR sahat*) AND *puu* esitetään Googlen syntaksin mukaan muodossa *“työkalut” OR “kirveet” OR “sahat” “puu”*. Lainausmerkkejä käytetään varmistamaan, että mahdolliset välilyönneistä sisältävät hakutermit esiintyvät tuloksessa kokonaisuudessaan. Käsitevalitsimen tiedonhakupöytäkirjan parametrit on kuvattu taulukossa 3.

Käyttäjän viedessä hiiren cursorin hakutermit valitsemansa käsitteen päälle selitetekstinä näytetään käsitteet, joihin valittu käsite on laajennettu. Käyttäjä voi halutessaan poistaa kyselyn laajennoksen poistamalla rastin valintaruudusta “laajenna”, mikä mahdollistaa hakutulosten liiallisen kasvun välttämisen.

Koska käsitevalitsimen käyttäminen perustuu kevyeen integrointiin asiakas-

Parametrin nimi	Tietotyyppi	Kuvaus
generateQueryANDPartition-Function	viite funktioon	JavaScript-funktio usean valitun hakutermijoukon yhdistämiseen
generateQueryORPartition-Function	viite funktioon	JavaScript-funktio toisilleen vaihtoehtoisten hakutermien joukon muotoiluun
generateQueryTermFunction	viite funktioon	JavaScript-funktio yksittäisen hakutermien muotoiluun
initializeWithFieldValue	totuusarvo	Esivalitaanko käsittekeräimeen hakukentässä mahdollisesti valmiina olevaa merkkijonoa vastaava käsite
maxQueryExpandResults	kokonaisluku	Käsitettä laajentavien käsitteiden enimmäismäärä
queryExpansion	totuusarvo	Laajennetaanko valittuja käsitteitä
queryExpansionCaption	merkkijono	Käyttöliittymäteksti valitun käsitteen laajennosten luettelon yhteydessä
resultLang	merkkijono	Kieli jonka mukaisia käsitteiden nimiä käytetään hakutermeinä (kieltenvälinen tiedonhaku)

Taulukko 3: Käsitevalitsimen tiedonhakukäytön parametrit.

järjestelmään, eikä hakutermien valinnassa ja laajentamisessa käytetä järjestelmän tietosisältöä, käsitevalitsimella on mahdollista valita hakutermiä, jotka palauttavat tyhjän hakutulospöytäkirjan. Tällaisten hakutermien ehdottamista käyttäjälle olisi järkevää välttää. Ongelman voisi ratkaista poistamalla hakutermien valintaan käytettävästä ontologiasta käsitteet, joita (tai joita laajentavia käsitteitä) ei ole käytetty tiedon kuvailussa, ja julkaisemalla karsittu ontologia ONKI-palvelussa erillisenä sovelluskohtaisena ontologiana.

Jos hakutermiksi valittu käsite laajenee suureen määrään käsitteitä, muodostettavasta hakukyselystä tulee huomattavan pitkä. Tämä saattaa aiheuttaa teknisiä ongelmia, koska kyselyiden enimmäispituus voi olla rajoitettu asiakasjärjestelmässä, ja rajoituksen ylittyessä sovellus voi palauttaa hakutulosten sijaan virheilmoituksen.

1. Haun muodostaminen

2. Haun tulokset

Kuva 17: Käsitevalitsimen käyttäminen Kantapuu-järjestelmässä.

- *search(merkkijono, kieli, tulostenEnimmäismäärä, tyyppirajaus, yläkäsiterajaus, ryhmärajaus)* - ontologian käsitteiden hakuun. Metodi palauttaa hakuparametreihin täsmäyvät käsitteet (URI ja nimi).
- *getLabel(URI, kieli)* - käsitteen nimen selvittämiseen. Metodi palauttaa parametrina annettua URI-tunnistetta vastaavan käsitteen nimen annetulla kielellä.
- *getAvailableLanguages()* - ontologian sisältämien kielten selvittämiseen. Metodi palauttaa kielet (lista kielikoodeista), joilla on esitetty käsitteiden nimiä.
- *getAvailableTypeUris()* - ontologian sisältämien käsitetyyppien selvittämiseen.

Metodi palauttaa käsitetyypit (lista URI-tunnisteita), joiden mukaisia käsitteitä ontologia sisältää.

Lisäksi ONKI SKOS -palvelimen rajapinta sisältää seuraavat metodit:

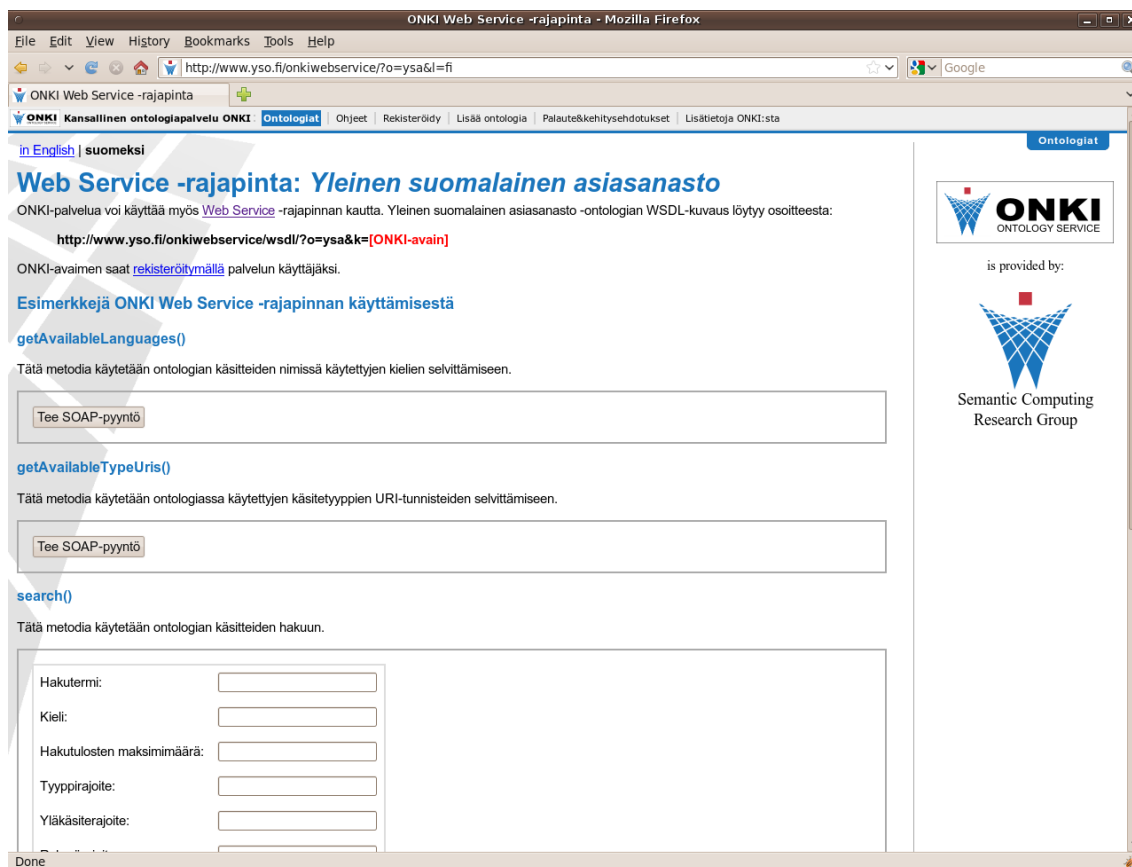
- *getConceptTree(URI, kieli)* - käsitteen käsittehierarkian selvittämistä varten. Metodi palauttaa parametrina annettua URI-tunnistetta vastaavan käsitteen käsittehierarkiaan kuuluvat käsitteet.
- *getProperties(URI, ominaisuusURI, kieli)* - käsitteen ominaisuuksien selvittämistä varten. Metodi palauttaa parametrina annettua URI-tunnistetta vastaavan käsitteen ominaisuudet.
- *expandQuery(URI, kieli)* - käsitteen laajentamiseen kyselyn laajennosta varten. Metodi palauttaa parametrina annettua URI-tunnistetta vastaavaa käsitettä laajentavat käsitteet.

ONKI-palvelun rajapinta on julkaistu sekä AJAX-käyttöön (engl. Asynchronous JavaScript And XML) että W3C:n WWW-sovelluspalveluarkkitehtuurin mukaisesti WSDL- ja SOAP-protokollalla käytettäväksi. Laajennetun rajapinnan lisäksi ONKI SKOS -palvelun toteuttaa myös toisen ohjelmallisen rajapinnan. Kyseinen rajapinta on toteutettu pelkkää HTTP-protokollaa käyttäen, ja sen vastaukset noudattavat JSON-muotoa [Cro06]. Rajapinnan metodit on kuvattu liitteessä 2.

Verrattuna yleiseen ONKI-palvelun rajapintaan ONKI SKOS -palvelimen HTTP-rajapinta tarjoaa käsitteiden tiettyjen ominaisuuksien kyselyyn tarkoitettuja metodeita, esimerkiksi käsitteen toissijaisten nimien, kuvausten tai ekvivalenttien käsitteiden selvittämiseen. Lisäksi HTTP-rajapinnan avulla voidaan pyytää ontologiasta metatietoja tai käsittehakemisto. Käsittehakemisto voi olla aakkosellinen, ryhmiin perustuva tai yläkäsitteisiin perustuva luettelo käsitteistä riippuen ontologiasta ja sen julkaisussa määritetystä konfiguraatiosta.

ONKI-palvelimien WWW-sovelluspalveluna käytettävän rajapinnan testaamista varten on tehty demonstraationsivu¹⁴, joka on esitetty kuvassa 18. Sivulla voidaan tehdä SOAP-protokollan mukaisia pyyntöjä valittuun ONKI-palvelimeen. Ihmisluettavan vastauksen lisäksi lähetetty pyyntö sekä ONKI-palvelimen vastaus näytetään myös SOAP-standardin käyttämässä XML-muodossa.

¹⁴<http://www.yso.fi/onkiwebservice/?o=yso>



Kuva 18: ONKI-palvelimien WWW-sovelluspalvelun demonstraationsivu.

ONKI SKOS -palvelimen AJAX-rajapinta on toteutettu Direct Web Remoting -kirjaston¹⁵ avulla, WWW-sovelluspalvelu (SOAP) Codehaus XFire -kirjastolla¹⁶ ja HTTP-rajapinta Restlet-kirjastoa¹⁷ käyttäen. AJAX-rajapinnan toimivuus eri verkkotunnusten välillä (asiakasjärjestelmä ja ONKI-palvelin) on ratkaistu käytämällä tekniikkaa, jossa generoidaan dynaamisesti *script*-elementtejä, joiden sisällöksi palvelin palauttaa JavaScript-koodin, joka kutsuu AJAX-pyyntöissä määriteltäviä *takaisinkutsumetodia* (engl. callback method) varsinaisella datalla. JavaScript-koodin lataaminen tällä tavoin on sallittu selaimissa eri verkkotunnusten välillä, toisin kuin esimerkiksi AJAX-toteutuksissa yleisen XMLHttpRequest-tekniikan [Kes09] selainten tällä hetkellä käyttämässä versiossa (versio 1).

¹⁵<http://directwebremoting.org>

¹⁶<http://xfire.codehaus.org>

¹⁷<http://www.restlet.org>

7 Tulosten tarkastelu

Tässä luvussa tarkastellaan tutkielman tuloksia. Aliluvussa 7.1 käsitellään tutkielman tutkimusongelmia ja aliluvussa 7.2 vertaillaan ONKI-palvelua luvussa 3.2 käsitelyihin aiemmin toteutettuihin ontologiapalvelimiin. Aihepiirin jatkotutkimusta pohditaan aliluvussa 7.3.

7.1 Tutkimusongelmat

Tutkielman johdannossa kuvattuihin tutkimusongelmiin liittyen esitetään seuraavat tulokset.

- Minkälaisia toiminnallisuuksia toimijat tarvitsevat ontologioiden käyttämiseen?

Yleisiä, usein toistuvia toiminnallisuuksia useiden ontologioiden käyttäjäryhmien – ontologioiden kehittäjien, tiedon annotoijien ja tiedon hakijoiden – keskuudessa ovat ontologioiden selaus, visualisointi ja niihin kohdistuva haku. Edellä mainittujen toiminnallisuuksien tarjoamiseksi kehitettiin prototyyppi-sovellus ONKI-ontologiapalvelu.

- Miten ontologioita voidaan julkaista toimijoiden käyttöön?

Yksinkertaisin tapa on julkaista ontologiat tiedostomuodossa, jolloin ontologian käyttäjä joutuu toteuttamaan ontologian hyödyntämiseen tarvittavia toiminnallisuuksia itse. Kustannustehokkaampaa on toteuttaa yleisiä, usein toistuvia toiminnallisuuksia valmiina palveluina käyttöliittymätasolla integroitavina komponentteina ihmiskäyttäjille sekä ohjelmallisina rajapintoina konekäyttäjille. ONKI-palvelu tarjoaa edellisen tutkimusongelman tuloksena tunnistettuja yleisiä toiminnallisuuksia, jotka ovat käytettävissä muun muassa ONKI SKOS -palvelimen, käsitevalitsimen sekä ohjelmallisten rajapintojen avulla.

- Miten voidaan varmistaa käytössä olevien ontologioiden ajantasaisuus?

Käyttämällä ontologioita palveluina tiedostomuodon sijaan toimijan käytössä on aina ontologioiden ajantasaiset versiot. Tällöin toimijan ei tarvitse huolehtia siitä, että käytössä on ontologian uusin tiedostoversio. Lisäksi ONKI-palvelun ontologioiden uusien versioiden julkaisuprosessi voidaan automatisoida, kuten esimerkiksi Yleisen suomalaisen asiasanaston tapauksessa on tehty.

7.2 ONKI-palvelun vertailu muihin ontologiapalvelimiin

ONKI-palvelu soveltuu tiettyyn käyttötarkoitukseen sopivan ontologian löytämiseen tarjoamalla hakemistonäkymän palvelussa julkaistuista ontologioista, toisin kuin yhden ontologian käyttämiseen suunnatut ontologiapalvelimet SUMO Browser, GTAA Browser ja OntoRama. Ontologiapalvelimia Ontolingua Server ja Medical Ontology Server voidaan käyttää vain Ontolingua-kielellä mallinnettujen ontologioiden julkaisemiseen, kun taas ONKI-palvelussa voidaan julkaista millä tahansa RDF-tietomalliin perustuvalla mallinnuskielellä kuvattuja ontologioita. Siten ONKI-palvelu on suoraviivaisesti hyödynnettävissä semanttisen webin sovelluksissa.

RDF-datan hakukoneet Watson ja Swoogle eivät tarjoa erityisesti ontologioiden selaamiseen ja visualisointiin suunnattuja toiminnallisuuksia, vaan käsittelevät ontologioita samalla tavalla kuin muutakin RDF-dataa. Kuten nämä RDF-datan hakukoneet, myös OntoSelect on suunnattu ontologian valintaan liittyviin käyttötapauksiin, eikä se siten sovellu ontologioiden varsinaiseen käyttöön kovin hyvin.

Sovelluksiin integroitavat ontologiapalveluita tarjoavat käyttöliittymäkomponentit mahdollistavat tyypillisesti joko ontologioihin kohdistuvan merkkijonohaun tai selaustoiminnon. BioPortal sisältää molemmat toiminnallisuudet, mutta toisistaan erillisinä komponentteina. OCLC Terminology Web Services tarjoaa ainoastaan merkkijonohaun, mutta toimii vain Internet Explorer -selaimessa. Myös Hildebrandin ja kumppaneiden RDF-varastoihin liitettävä hakukomponentti [HOA07] sisältää vain merkkijonohaun ilman ontologiaselainta. Toisaalta komponentti näyttää hakutulosten yhteydessä käsitteiden kontekstietoa, jonka avulla voidaan valita hakutulosta tarkentavia käsitteitä. OLS2OWL tarjoaa merkkijonohaun, mutta on käytettävissä vain Protégé-ontologiaeditorin laajennoksena. ONKI-palvelun käsitevalitsin sisältää yhdistetyn ontologian haku- ja selailutoiminnallisuuden. Näin käyttäjä voi tehdä käsitehaun ja valita hakutulosten joukosta käsitteen, josta aloittaa ontologian selaamisen. Käsitevalitsimen käyttäminen ei ole rajoitettu yksittäiseen sovellukseen tai selaimen, vaan se on käytettävissä verkkosovelluksissa yleisesti.

Ohjelmallisista rajapinnoista SKOS API soveltuu vain SKOS-tietomallilla kuvattujen sanastojen käyttämiseen. Verrattuna yleisiin RDF-kyselykieliin, kuten SPARQL-kieleen, ONKI-palvelun rajapinta tarjoaa pääsyn ontologiaan korkeammalla abstraktiotasolla. Rajapinnan käyttäjän ei tarvitse tuntea kyselyn kohteena olevan ontologian RDF-esityksen yksityiskohtia, kuten sitä, että käsitteen nimi on kuvattu ominaisuudella *rdfs:label* ja yläkäsitteet ominaisuudella *rdfs:subClassOf*, vaan rajapinnalta voidaan pyytää yleisesti käsitteen nimeä ja yläkäsitteitä. FIPA-

standardi, Agent Cities Ontology Service, Otago Ontology Repository, Ontology Web Services, KAON Server ja Open Ontology Repository tarjoavat laajan valikoiman palveluita ontologioiden käyttämiseen, kun taas ONKI-palvelu keskittyy ontologioiden käyttämiseen erityisesti tiedon annotoinnissa ja tiedonhaussa pitäen rajapinnan suhteellisen yksinkertaisena.

7.3 Jatkokehitys

Ontologioiden eri käyttäjäryhmien tarpeiden syvällisemmäksi ymmärtämiseksi olisi hyödyllistä tehdä käyttäjätutkimus, jossa selvitetään miten käyttäjät hyödyntävät ontologioita, ja miten nykyiset työkalut heidän vaatimuksiin. Toisaalta ONKI-palvelun jatkokehittämisen kannalta myös palvelun formaali evaluointi, esimerkiksi käytettävyydestin muodossa, olisi perusteltua.

Ontologian kehittäjän työhön tarvittaisiin kattavampia työkaluja. Ontologian varsinainen kehitystyö tapahtuu järjestelmän ulkopuolella, tyypillisesti Protégé-editorissa, ja editorin tuottama RDF-tiedosto ladataan ONKI-palveluun, kun ontologiasta halutaan julkaista uusi versio. Erityisesti siltauksia sisältävien ontologioiden kehittämisessä tarvitaan kehitysprosessia tehostavia välineitä.

Ontologian julkaisuvaiheessa olisi mielekästä tuottaa raportti ontologiaan tehdyistä muutoksista vertailemalla ontologian uutta ja edellistä versiota keskenään, jotta ontologian käyttäjät pystyisivät helpommin seuraamaan ontologian muutoksia. Ontologiaversioiden ohjelmallisen vertailun ongelmia on muun muassa se, että muutoksien syiden ja atomisiin muutoksiin sisältyvien osamuutosten päättelyminen on haasteellista. Muutosten kommunikoiminen on erityisen tärkeää siltauksia sisältävien, toisistaan erillään kehitettävien ontologioiden kehitystyössä, jotta ontologioiden eheys säilyisi.

ONKI-palvelun käsitevalitsin näyttää hakutuloksiin kuuluvista käsitteistä vain nimen. Pelkkä käsitteen nimi ei kuitenkaan aina riitä käsitteen merkityksen ymmärtämiseen, esimerkiksi jos käsitteen nimenä on monimerkityksellinen tai käyttäjälle vieras sana. Käsitteen merkityksen ymmärtämisessä auttaa käsitteen konteksti, jota käyttäjä pääsee tutkimaan avaamalla ONKI-selaimen. Turhan navigoinnin välttämiseksi olisi kuitenkin järkevämpää visualisoida konteksti erillisen ONKI-selaimen sijaan käsitevalitsimessa. Tällainen ratkaisu on toteutettu esimerkiksi IRMA-käsitevalitsimessa [SMK08]. Lisäksi jos ontologiasta ei löydy sopivaa käsitettä, olisi mielekästä tarjota käyttäjälle mahdollisuus lisätä ontologiaan uusi

käsite.

ONKI-palvelun ohjelmalliset rajapinnat ovat keskittyneet käsitteiden hakuun, käsitteen tietojen selvittämiseen ja ontologian yleiskuvan tarjoamiseen. Vastatakseen laajempaan joukkoon mahdollisia käyttötarpeita palvelu voisi tarjota lisäksi yleisen SPARQL-kyselykielellä käytettävän rajapinnan. Tällöin on huomioitava ontologioiden julkaisussa käytettävät lisenssiehdot – osa ONKI-palvelun ontologioista on julkaistu lisenssillä, joka ei salli ontologian julkaisua kokonaisuudessaan datamuodossa. SPARQL-kielellä on kuitenkin mahdollista tehdä kysely, joka palauttaa koko ontologian RDF-muodossa.

ONKI-palvelu tarjoaa ihmiskäyttäjille keinoja löytää sopivia ontologioita muun muassa ontologistauksen avulla. Tällaista palvelua ei kuitenkaan ole käytettävissä koneymmärrettävässä muodossa. Koneille voitaisiin tarjota rajapinta, joka tarjoaa kaikkien palvelussa julkaistujen ontologioiden metatiedot. Metatietojen kuvaamisessa on järkevää käyttää tarkoitukseen soveltuvia yleisiä standardeja, kuten Dublin Corea ja mahdollisesti OMV-sanastoa [HPS05].

ONKI SKOS -palvelin prosessoi ontologioita Jena-kirjaston avulla RDF-malleina, joiden käsittely ei ole niin tehokasta ja nopeaa kuin erillisen, tiettyyn tarkoitukseen toteutetun hakemiston. Käsittehaun suorittamiseen käytetään Lucene-indeksiä, mutta esimerkiksi käsitteen hierarkia ja muut ominaisuudet haetaan RDF-mallista. Erityisesti suurissa ontologioissa, joissa käsittehierarkia on monimutkainen runsaan moniperinnän takia ja joissa käsitteillä on paljon vieruskäsitteitä, käsittehierarkian prosessointi on melko hidasta. Erillisen hakemiston toteuttaminen käsittehierarkian selvittämistä varten olisi siten järkevää.

8 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 tutkielmassa tunnistettiin ontologioiden käyttäjäryhmien tarpeiden ja ontologiapalvelinten toteutusten perusteella 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 Kansallinen ontologiapalvelu ONKI, joka perustuu erilaisten ontologiatyyppien julkaisemiseen suunnattuihin ontologiapalvelintoteutuksiin. ONKI SKOS on tesaurusten ja yksinkertaisten ontologioiden hyödyntämiseen tarkoitettu ontologiapalvelin, jonka avulla voidaan julkaista esimerkiksi olemassa olevia tesauroksia RDF-muunnosprosessien avulla. ONKI-palvelu tarjoaa valmiita käyttöliittymäkomponentteja ja ohjelmallisia rajapintoja muun muassa tiedon annotointiin ja hakuun sekä työkaluja ontologioita hyödyntävien sovellusten kehittäjille.

ONKI-palvelu julkistettiin virallisesti käyttöön syyskuussa 2008, mutta oli jo sitä ennen pilottikäytössä. Esimerkkinä palvelun käyttäjämääristä Google Analytics¹⁸-tilaston perusteella helmikuun 2010 aikana ONKI-palvelulla oli n. 15 900 uniikkia kävijää, jotka tekivät yhteensä n. 247 000 sivulatausta. Lukumäärissä ei ole mukana käsitevalitsimella tehtyjä hakuja, eikä ohjelmallisten rajapintojen käyttöä. Käsitevalitsimen ja ohjelmallisten rajapintojen käyttöön tarvittavan ONKI-avaimen on rekisteröinyt itselleen n. 200 käyttäjää.

Kiitokset

Tämä tutkielma on tehty Aalto-yliopiston teknillisen korkeakoulun ja Helsingin yliopiston Semanttisen laskennan tutkimusryhmässä (SeCo) osana Suomalaiset semanttisen webin ontologiat (FinnONTO, 2003-2012) -hanketta.

Haluan kiittää professori Eero Hyvöstä ja Miro Lehtosta työni ohjaamisesta. Kiitän Kim Viljasta ja Krista Tuomista kannustamisesta työn tekemiseen sekä rakentavista ehdotuksista tutkielman sisällön, rakenteen ja kieliäsun parantamiseksi. Haluan kiittää myös koko Semanttisen laskennan tutkimusryhmää – erityiskiitokset Kim Viljalle, Osma Suomiselle, Eetu Mäkelälle, Tomi Kauppiselle, Matias Frosterukselle ja Mikko Salonojalle yhteistyöstä ONKI-ontologiapalvelun toteuttamisessa.

ONKI SKOS -palvelimen toteutus perustuu osittain Ville Komulaisen ja kumppaneiden työhön ONKI-palvelimen ensimmäiseen versioon [Kom07, VHK05] liittyen.

¹⁸<http://www.google.com/analytics/>

Lähteet

- AAT10 Art & architecture thesaurus online – about the AAT, 2010. http://www.getty.edu/research/conducting_research/vocabularies/aat/about.html. [17.3.2010]
- ABG07 d’Aquin, M., Baldassarre, C., Gridinoc, L., Sabou, M., Angeletou, S. ja Motta, E., Watson: Supporting next generation semantic web applications. *Proceedings of the IADIS International Conference on WWW/Internet (ICWI’07)*, Vila Real, Portugal, lokakuu 2007.
- AGB00 Aitchison, J., Gilchrist, A. ja Bawden, D., *Thesaurus Construction and Use: A Practical Manual*. Europa Publications, neljäs painos, 2000.
- AhC07 Ahmad, M. N. ja Colomb, R. M., Managing ontologies: a comparative study of ontology servers. *Proceedings of the Eighteenth Conference on Australasian Database (ADC’07)*, Ballarat, Victoria, Australia, tammi-kuu 2007, Australian Computer Society, sivut 13–22.
- AMM06 van Assem, M., Malaisé, V., Miles, A. ja Schreiber, G., A method to convert thesauri to SKOS. *Proceedings of the 3rd European Semantic Web Conference (ESWC’06)*, Budva, Montenegro, kesäkuu 2006, Springer-Verlag, sivut 95–109.
- AMS04 van Assem, M., Menken, M. R., Schreiber, G., Wielemaker, J. ja Wielinga, B., A method for converting thesauri to RDF/OWL. *Proceedings of the 3rd International Semantic Web Conference (ISWC’04)*, Hiroshima, Japan, marraskuu 2004, Springer-Verlag, sivut 17–31.
- AqL09 d’Aquin, M. ja Lewen, H., Cupboard – a place to expose your ontologies to applications and the community. *Proceedings of the 6th European Semantic Web Conference (ESWC’09)*, Heraklion, Greece, kesäkuu 2009, Springer-Verlag, sivut 913–918.
- BaS09 Baclawski, K. ja Schneider, T., The open ontology repository initiative: Requirements and research challenges. *Proceedings of the Workshop on Collaborative Construction, Management and Linking of Structured Knowledge (CK’09), the 8th International Semantic Web Conference (ISWC’09)*, Washington, D.C., USA, lokakuu 2009.

- BBD08 Begeed-Dov, G., Brickley, D., Dornfest, R., Davis, I., Dodds, L., Eizenszopf, J., Galbraith, D., Guha, R., MacLeod, K., Miller, E., Swartz, A. ja van der Vlist, E., RDF site summary (RSS) 1.0, 2008. <http://web.resource.org/rss/1.0/spec>. [18.3.2010]
- BCC08 Basca, C., Corlosquet, S., Cyganiak, R., Fernández, S. ja Schandl, T., Neologism: Easy vocabulary publishing. *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.
- BeB08 Beckett, D. ja Berners-Lee, T., Turtle - terse RDF triple language, W3C team submission, 2008. <http://www.w3.org/TeamSubmission/2008/SUBM-turtle-20080114/>. [15.3.2009]
- Bec04 Beckett, D., RDF/XML syntax specification (revised), W3C recommendation, 2004. <http://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/>. [28.4.2009]
- BED04 Buitelaar, P., Eigner, T. ja Declerck, T., OntoSelect: A dynamic ontology library with support for ontology selection. *Proceedings of the Demo Session, the 3rd International Semantic Web Conference (ISWC'04)*, Hiroshima, Japan, marraskuu 2004.
- BeP08 Berrueta, D. ja Phipps, J., Best practice recipes for publishing RDF vocabularies, W3C working group note, 2008. <http://www.w3.org/TR/2008/NOTE-swbp-vocab-pub-20080828/>. [28.12.2009]
- Ber98 Berners-Lee, T., Cool URIs don't change, W3C style guide for online hypertext, 1998. <http://www.w3.org/Provider/Style/URI>. [28.12.2009]
- BFF08 Berrueta, D., Fernández, S. ja Frade, I., Cooking HTTP content negotiation with Vapour. *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.
- BFM05 Berners-Lee, T., Fielding, R. ja Masinter, L., Uniform resource identifier (URI): Generic syntax, 2005. <http://www.ietf.org/rfc/rfc3986.txt>. [12.3.2010]
- BHL01 Berners-Lee, T., Hendler, J. ja Lassila, O., The semantic web. *Scientific American*, 284,5(2001), sivut 34–43.

- BHM04 Booth, D., Haas, H., McCabe, F., Newcomer, E., Champion, M., Ferris, C. ja Orchard, D., Web services architecture, W3C working group note, 2004. <http://www.w3.org/TR/2004/NOTE-ws-arch-20040211/>. [28.4.2009]
- BiM04 Biron, P. V. ja Malhotra, A., XML schema part 2: Datatypes second edition, W3C recommendation, 2004. <http://www.w3.org/TR/2004/REC-xmlschema-2-20041028/>. [19.1.2010]
- BMG06 Brugman, H., Malaisé, V. ja Gazendam, L., A web based general thesaurus browser to support indexing of television and radio programs. *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy, toukokuu 2006.
- BrG04 Brickley, D. ja Guha, R. V., RDF vocabulary description language 1.0: RDF schema, W3C recommendation, 2004. <http://www.w3.org/TR/2004/REC-rdf-schema-20040210/>. [28.4.2009]
- BRM04 Beckett, D., Rogers, N. ja Miles, A., SKOS API, SWAD-Europe thesaurus activity, W3C, 2004. <http://www.w3.org/2001/sw/Europe/reports/thes/skosapi.html>. [6.1.2010]
- Cat97 Cathro, W., Metadata: an overview. *Proceedings of the Standards Australia Seminar: Matching Discovery and Recovery*, Sydney and Melbourne, Australia, elokuu 1997.
- CJA06 Côté, R. G., Jones, P., Apweiler, R. ja Hermjakob, H., The ontology lookup service, a lightweight cross-platform tool for controlled vocabulary queries. *BMC Bioinformatics*, 7,97(2006).
- Cor01 Core J2EE patterns - data access object, 2001. <http://java.sun.com/blueprints/corej2eepatterns/Patterns/DataAccessObject.html>. [14.3.2010]
- CQL08 CQL: Contextual query language (SRU version 1.2 specifications), 2008. <http://www.loc.gov/standards/sru/specs/cql.html>. [17.3.2010]
- Cro06 Crockford, D., The application/json media type for javascript object notation (JSON), 2006. <http://www.ietf.org/rfc/rfc4627.txt>. [19.3.2010]

- Cun99 Cunningham, H., Information extraction - a user guide (revised version). Tekninen raportti CS-99-07, Institute for Language, Speech and Hearing (ILASH) and Department of Computer Science, University of Sheffield, Sheffield, UK, 1999. URL <http://www.dcs.shef.ac.uk/~hamish/IE/>.
- DCM08 DCMI metadata terms, 2008. <http://dublincore.org/documents/2008/01/14/dcmi-terms/>. [28.12.2009]
- DeS04 Dean, M. ja Schreiber, G., OWL web ontology language reference, W3C recommendation, 2004. <http://www.w3.org/TR/2004/REC-owl-ref-20040210/>. [2.1.2010]
- Dev02 Devedžić, V., Understanding ontological engineering. *Communications of the ACM*, 45,4(2002), sivut 136–144.
- DFJ04 Ding, L., Finin, T., Joshi, A., Pan, R., Cost, R. S., Peng, Y., Reddivari, P., Doshi, V. ja Sachs, J., Swoogle: a search and metadata engine for the semantic web. *Proceedings of the Thirteenth ACM International Conference on Information and Knowledge Management (CIKM'04)*, Washington, D.C., USA, marraskuu 2004, ACM Press, sivut 652–659.
- DiF01 Ding, Y. ja Fensel, D., Ontology library systems: The key to successful ontology reuse. *Proceedings of the 1st Semantic Web Working Symposium (SWWS'01)*, Stanford University, California, USA, elokuu 2001, sivut 93–112.
- DNK04 Dameron, O., Noy, N. F., Knublauch, H. ja Musen, M. A., Accessing and manipulating ontologies using web services. *Proceedings of the Semantic Web Services Workshop, the 3rd International Semantic Web Conference (ISWC'04)*, Hiroshima, Japan, marraskuu 2004.
- EhS04 Ehrig, M. ja Sure, Y., Ontology mapping – an integrated approach. *Proceedings of the 1st European Semantic Web Symposium (ESWS'04)*, Heraklion, Greece, toukokuu 2004, Springer-Verlag, sivut 76–91.
- ERG02 Eklund, P., Roberts, N. ja Green, S., OntoRama: Browsing RDF ontologies using a hyperbolic-style browser. *Proceedings of the 1st International Symposium on Cyber Worlds (CW'02)*, Tokyo, Japan, marraskuu 2002, IEEE Press, sivut 405–411.

- Fen03 Fensel, D., *Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce*. Springer-Verlag, toinen painos, 2003.
- FFR97 Farquhar, A., Fikes, R. ja Rice, J., The Ontolingua Server: a tool for collaborative ontology construction. *International Journal of Human-Computer Studies*, 46,6(1997), sivut 707–727.
- FGM99 Fielding, R., Gettys, J., Mogul, J., Frystyk, H., Masinter, L., Leach, P. ja Berners-Lee, T., Hypertext transfer protocol – HTTP/1.1, 1999. <http://www.ietf.org/rfc/rfc2616.txt>. [18.3.2010]
- FHH07 Feigenbaum, L., Herman, I., Hongsermeier, T., Neumann, E. ja Stephens, S., The semantic web in action. *Scientific American*, 297,6(2007), sivut 90–97.
- Fie00 Fielding, R. T., *Architectural Styles and the Design of Network-based Software Architectures*. Väitöskirja, University of California, Irvine, 2000.
- FiF97 Fikes, R. ja Farquhar, A., Large-scale repositories of highly expressive reusable knowledge. Tekninen raportti KSL-97-02, Stanford Knowledge Systems Laboratory, Stanford, California, USA, 1997. URL http://www.ksl.stanford.edu/KSL_Abstracts/KSL-97-02.html.
- FIP01 FIPA ontology service specification, 2001. <http://www.fipa.org/specs/fipa00086/XC00086D.html>. [28.4.2009]
- FSH04 Fluit, C., Sabou, M. ja van Harmelen, F., Supporting user tasks through visualisation of light-weight ontologies. Teoksessa *Handbook on Ontologies*, Staab, S. ja Studer, R., toimittajat, Springer-Verlag, 2004, sivut 415–434.
- Gar04 Garshol, L. M., Metadata? thesauri? taxonomies? topic maps! making sense of it all. *Journal of Information Science*, 30,4(2004), sivut 378–391.
- GeF92 Genesereth, M. R. ja Fikes, R. E., Knowledge interchange format, version 3.0 reference manual. Tekninen raportti Logic-92-1, Computer Science Department, Stanford University, Stanford, California, USA, 1992. URL <http://www-ksl.stanford.edu/knowledge-sharing/papers/kif.ps>.

- GGV09 García-Castro, A., García-Castro, L., Villaveces, J. M., Calderón, G. ja Hepp, M., OLS2OWL – a repository management facility. *Proceedings of the 11th International Protégé Conference*, Amsterdam, Netherlands, kesäkuu 2009.
- Goo10 Google AdSense, 2010. <http://www.google.com/adsense/>. [20.3.2010]
- GOP95 Gennari, J. H., Oliver, D. E., Pratt, W., Rice, J. ja Musen, M. A., A web-based architecture for a medical vocabulary server. *Proceedings of the 19th Annual Symposium on Computer Applications in Medical Care*, New Orleans, Louisiana, USA, lokakuu 1995, sivut 275–279.
- Gru93 Gruber, T. R., A translation approach to portable ontology specification. *Knowledge Acquisition*, 5,2(1993), sivut 199–220.
- HaT07 Harper, C. A. ja Tillett, B. B., Library of Congress controlled vocabularies and their application to the semantic web. *Cataloging & Classification Quarterly*, 43,3/4(2007).
- HaZ07 Hawking, D. ja Zobel, J., Does topic metadata help with web search? *Journal of the American Society for Information Science and Technology*, 58,5(2007), sivut 613–628.
- Hef01 Heflin, J., *Towards the Semantic Web: Knowledge Representation in a Dynamic, Distributed Environment*. Väitöskirja, University of Maryland, College Park, 2001.
- Her10 Herman, I., W3C semantic web activity, 2010. <http://www.w3.org/2001/sw/>. [15.3.2010]
- HMS05 Hyvönen, E., Mäkelä, E., Salminen, M., Valo, A., Viljanen, K., Saarela, S., Junnila, M. ja Kettula, S., MuseumFinland – Finnish museums on the semantic web. *Journal of Web Semantics*, 3,2(2005), sivut 224–241.
- HOA07 Hildebrand, M., van Ossenbruggen, J., Amin, A., Aroyo, L., Wielemaker, J. ja Hardman, L., The design space of a configurable auto-completion component. Tekninen raportti INS-E0708, Centrum voor Wiskunde en Informatica (CWI), Amsterdam, Netherlands, 2007. URL <http://www.cwi.nl/ftp/CWIreports/INS/INS-E0708.pdf>.

- HOH09 Hildebrand, M., van Ossenbruggen, J., Hardman, L. ja Jacobs, G., Supporting subject matter annotation using heterogeneous thesauri: A user study in web data reuse. *International Journal of Human-Computer Studies*, 67,10(2009), sivut 887–902.
- HPS05 Hartmann, J., Palma, R., Sure, Y., del Carmen Suárez-Figueroa, M., Haase, P., Gómez-Pérez, A. ja Studer, R., Ontology metadata vocabulary and applications. *Proceedings of the Workshop on Web Semantics (SWWS), the International Conference on Ontologies, Databases and Applications of Semantics (ODBASE'05)*, Agia Napa, Cyprus, lokakuu 2005, Springer-Verlag, sivut 906–915.
- HRH08 Hyvönen, E., Ruotsalo, T., Häggström, T., Salminen, M., Junnila, M., Virkkilä, M., Haaramo, M., Mäkelä, E., Kauppinen, T., ja Viljanen, K., Culturesampo – Finnish culture on the semantic web: The vision and first results. Teoksessa *Information Technology for the Virtual Museum: Museology and the Semantic Web*, Robering, K., toimittaja, LIT Verlag, Berlin, Germany, 2008.
- HSV07 Hyvönen, E., Seppälä, K., Viljanen, K. ja Frosterus, M., Yleinen suomalainen ontologia YSO – kohti suomalaista semanttista webiä. *Tietolinja*, 26,1(2007).
- HSW07 Hollink, L., Schreiber, G. ja Wielinga, B., Patterns of semantic relations to improve image content search. *Journal of Web Semantics*, 5,3(2007), sivut 195–203.
- HVK05 Hyvönen, E., Valo, A., Komulainen, V., Seppälä, K., Kauppinen, T., Ruotsalo, T., Salminen, M. ja Ylissalmi, A., 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)*, Madrid, Spain, syyskuu 2005, Dublin Core Metadata Initiative, sivut 219–222.
- HVS07 Hyvönen, E., Viljanen, K. ja Suominen, O., HealthFinland – Finnish health information on the semantic web. *Proceedings of the 6th International Semantic Web Conference (ISWC'07)*, Busan, Korea, marraskuu 2007, Springer-Verlag, sivut 778–791.

- HVT08 Hyvönen, E., Viljanen, K., Tuominen, J. ja Seppälä, K., Building a national semantic web ontology and ontology service infrastructure—the FinnONTO approach. *Proceedings of the 5th European Semantic Web Conference (ESWC'08)*, Tenerife, Spain, kesäkuu 2008, Springer-Verlag, sivut 95–109.
- HyM06 Hyvönen, E. ja Mäkelä, E., Semantic autocompletion. *Proceedings of the 1st Asian Semantic Web Conference (ASWC'06)*, Beijing, China, elokuu 2006, Springer-Verlag, sivut 739–751.
- Hyv08 Hyvönen, E., FinnONTO-malli kansallisen semanttisen webin sisältöinfrastruktuurin perustaksi – visio ja sen toteutus, julkaistu Kansallinen ontologiapalvelu ONKI -julkistustilaisuudessa, syyskuu 2008.
- KaH07 Kauppinen, T. ja Hyvönen, E., Modeling and reasoning about changes in ontology time series. Teoksessa *Ontologies: A Handbook of Principles, Concepts and Applications in Information Systems*, Kishore, R., Ramesh, R. ja Sharman, R., toimittajat, Springer-Verlag, 2007, sivut 319–338.
- Kan10 Kantapuu – museotietokanta, 2010. <http://www.kantapuu.fi>. [20.3.2010]
- Kes09 van Kesteren, A., XMLHttpRequest, W3C working draft, 2009. <http://www.w3.org/TR/2009/WD-XMLHttpRequest-20091119/>. [9.2.2010]
- KHL07 Katifori, A., Halatsis, C., Lepouras, G., Vassilakis, C. ja Giannopoulou, E., Ontology visualization methods—a survey. *ACM Computing Surveys*, 39,4(2007).
- KHS08 Kauppinen, T., Henriksson, R., Sinkkilä, R., Lindroos, R., Väättäinen, J. ja Hyvönen, E., 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.
- KIC04 Klyne, G. ja Carroll, J. J., Resource description framework (RDF): Concepts and abstract syntax, W3C recommendation, 2004. <http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/>. [28.4.2009]

- Kom07 Komulainen, V., Public services for ontology library systems. Pro gradu, Helsingin yliopisto, Tietojenkäsittelytieteen laitos, 2007.
- KPT04 Kiryakov, A., Popov, B., Terziev, I., Manov, D. ja Ognyanoff, D., Semantic annotation, indexing, and retrieval. *Web Semantics: Science, Services and Agents on the World Wide Web*, 2,1(2004), sivut 49–79.
- KuH09 Kurki, J. ja Hyvönen, E., 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.
- LRP95 Lamping, J., Rao, R. ja Pirolli, P., A focus+context technique based on hyperbolic geometry for visualizing large hierarchies. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'95)*, Denver, Colorado, USA, toukokuu 1995, ACM Press/Addison-Wesley Publishing, sivut 401–408.
- LSN08 Lagoze, C., de Sompel, H. V., Nelson, M. ja Warner, S., The open archives initiative protocol for metadata harvesting, 2008. <http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>. [13.3.2010]
- LTT03 Li, Y., Thompson, S., Tan, Z., Giles, N. ja Gharib, H., Beyond ontology construction: Ontology services as online knowledge sharing communities. *Proceedings of the 2nd International Semantic Web Conference (ISWC'03)*, Sanibel, Florida, USA, lokakuu 2003, Springer-Verlag, sivut 469–483.
- Man07 Mangold, C., A survey and classification of semantic search approaches. *International Journal of Metadata, Semantics and Ontologies*, 2,1(2007), sivut 23–34.
- MAO10 Museolan ontologia (MAO), 2010. <http://www.seco.tkk.fi/ontologies/mao/>. [18.3.2010]
- MAR00 MARC 21 specifications for record structure, character sets, and exchange media, 2000. <http://www.loc.gov/marc/specifications/spechome.html>. [17.3.2010]

- MAR09 MARCXML – MARC 21 XML schema, 2009. <http://www.loc.gov/standards/marcxml/>. [17.3.2010]
- MiB05 Miles, A. ja Brickley, D., SKOS core guide, W3C working draft, 2005. <http://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20051102/>. [15.3.2010]
- MiB09 Miles, A. ja Bechhofer, S., SKOS simple knowledge organization system reference, W3C recommendation, 2009. <http://www.w3.org/TR/2009/REC-skos-reference-20090818/>. [2.1.2010]
- Mil95 Miller, G. A., WordNet: A lexical database for English. *Communications of the ACM*, 38,11(1995), sivut 39–41.
- MMS03 Maedche, A., Motik, B., Stojanovic, L., Studer, R. ja Volz, R., An infrastructure for searching, reusing and evolving distributed ontologies. *Proceedings of the 12th International Conference on World Wide Web (WWW'03)*, Budapest, Hungary, toukokuu 2003, ACM Press, sivut 439–448.
- MMW05 Miles, A., Matthews, B., Wilson, M. ja Brickley, D., SKOS core: Simple knowledge organisation for the web. *Proceedings of the International Conference on Dublin Core and Metadata Applications (DC'05)*, Madrid, Spain, syyskuu 2005, Dublin Core Metadata Initiative, sivut 3–10.
- Moa97 Moats, R., URN syntax, 1997. <http://www.ietf.org/rfc/rfc2141.txt>. [17.3.2010]
- MPP09 Motik, B., Patel-Schneider, P. F. ja Parsia, B., OWL 2 web ontology language structural specification and functional-style syntax, W3C recommendation, 2009. <http://www.w3.org/TR/2009/REC-owl2-syntax-20091027/>. [6.5.2010]
- MRS08 Manning, C. D., Raghavan, P. ja Schütze, H., *Introduction to Information Retrieval*. Cambridge University Press, 2008.
- MVA07 Mäkelä, E., Viljanen, K., Alm, O., Tuominen, J., Valkeapää, O., Kauppinen, T., Kurki, J., Sinkkilä, R., Käsälä, T., Lindroos, R., Suominen, O., Ruotsalo, T. ja Hyvönen, E., 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.
- NFM00 Noy, N. F., Ferguson, R. ja Musen, M. A., The knowledge model of Protégé-2000: Combining interoperability and flexibility. *Proceedings of the 12th International Conference on Knowledge Engineering and Knowledge Management (EKAW'00)*, Juan-les-Pins, France, lokakuu 2000, Springer-Verlag, sivut 17–32.
- NiP01 Niles, I. ja Pease, A., Towards a standard upper ontology. *Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS'01)*, Ogunquit, Maine, USA, lokakuu 2001, ACM Press, sivut 2–9.
- NSW09 Noy, N. F., Shah, N. H., Whetzel, P. L., Dai, B., Dorf, M., Griffith, N., Jonquet, C., Rubin, D. L., Storey, M.-A., Chute, C. G. ja Musen, M. A., BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Research*, 37, Web Server issue(2009), sivut 170–173.
- OVM04 Oberle, D., Volz, R., Motik, B. ja Staab, S., An extensible ontology software environment. Teoksessa *Handbook on Ontologies*, Staab, S. ja Studer, R., toimittajat, Springer-Verlag, 2004, sivut 311–333.
- PBK06 Pietriga, E., Bizer, C., Karger, D. ja Lee, R., Fresnel: A browser-independent presentation vocabulary for RDF. *Proceedings of the 5th International Semantic Web Conference (ISWC'06)*, Athens, Georgia, USA, marraskuu 2006, Springer-Verlag, sivut 158–171.
- PBM99 Page, L., Brin, S., Motwani, R. ja Winograd, T., The PageRank citation ranking: Bringing order to the web. Tekninen raportti 422, Stanford University InfoLab, Stanford, California, USA, 1999. URL <http://ilpubs.stanford.edu:8090/422/>.
- PCC03 Pan, J., Cranefield, S. ja Carter, D., A lightweight ontology repository. *Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'03)*, Melbourne, Australia, heinäkuu 2003, ACM Press, sivut 632–638.

- PMA08 Peroni, S., Motta, E. ja d'Aquin, M., Identifying key concepts in an ontology, through the integration of cognitive principles with statistical and topological measures. *Proceedings of the 3rd Asian Semantic Web Conference (ASWC'08)*, Bangkok, Thailand, helmikuu 2009, Springer-Verlag, sivut 242–256.
- PrS08 Prud'hommeaux, E. ja Seaborne, A., SPARQL query language for RDF, W3C recommendation, 2008. <http://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/>. [15.1.2010]
- RHJ99 Raggett, D., Le Hors, A. ja Jacobs, I., HTML 4.01 specification, W3C recommendation, 1999. <http://www.w3.org/TR/1999/REC-html401-19991224/>. [17.3.2010]
- RMK08 Rubin, D. L., Moreira, D. A., Kanjamala, P. P. ja Musen, M. A., 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.
- RSC04 Reynolds, D., Shabajee, P. ja Cayzer, S., Semantic information portals. *Proceedings of the 13th International Conference on World Wide Web (WWW'04)*, New York City, New York, USA, toukokuu 2004, ACM Press, sivut 290–291.
- SaC08 Sauermann, L. ja Cyganiak, R., Cool URIs for the semantic web, W3C interest group note, 2008. <http://www.w3.org/TR/2008/NOTE-cooluris-20081203/>. [28.12.2009]
- SAR07 Smith, B., Ashburner, M., Rosse, C., Bard, C., Bug, W., Ceusters, W., Goldberg, L. J., Eilbeck, K., Ireland, A., Mungall, C. J., The OBI Consortium, Leontis, N., Rocca-Serra, P., Ruttenberg, A., Sansone, S.-A., Scheuermann, R. H., Shah, N., Whetzel, P. L. ja Lewis, S., The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*, 25,11(2007), sivut 1251–1255.
- SAR08 Summers, E., Isaac, A., Redding, C. ja Krech, D., LCSH, SKOS and linked data. *Proceedings of the International Conference on Dublin Core and Metadata Applications (DC'08)*, Berlin, Germany, syyskuu 2008, Dublin Core Metadata Initiative, sivut 25–33.

- SDW01 Schreiber, A. T. G., Dubbeldam, B., Wielemaker, J. ja Wielinga, B., Ontology-based photo annotation. *IEEE Intelligent Systems*, 16,3(2001), sivut 66–74.
- Sev03 Ševčenko, M., Online presentation of an upper ontology. *Proceedings of the Znalosti 2003*, Ostrava, Czech Republic, helmikuu 2003, sivut 153–162.
- SMK08 Sinkkilä, R., Mäkelä, E., Kauppinen, T. ja Hyvönen, E., Combining context navigation with semantic autocompletion to solve problems in concept selection. *Proceedings of the First International Workshop on Semantic Metadata Management and Applications (SeMMA'08), the 5th European Semantic Web Conference (ESWC'08)*, Tenerife, Spain, kesäkuu 2008, sivut 61–68.
- SMM02 Stojanovic, L., Maedche, A., Motik, B. ja Stojanovic, N., User-driven ontology evolution management. *Proceedings of the European Conference on Knowledge Engineering and Knowledge Management (EKAW'02)*, Sigüenza, Spain, lokakuu 2002, Springer-Verlag, sivut 285–300.
- Son00 Song, M., Visualization in information retrieval: a three-level analysis. *Journal of Information Science*, 26,1(2000), sivut 3–19.
- SRU09 SRU: Search/retrieval via URL, 2009. <http://www.loc.gov/standards/sru/>. [17.3.2010]
- Ste98 Stenvall, J., Dublin core -formaatin käyttöopas. Tekninen raportti, Helsingin yliopiston kirjasto, Helsinki, Suomi, 1998. URL http://www.kansalliskirjasto.fi/julkaisuala/dublincore/Files/liitetiedosto2/dc_opas.pdf.
- TAJ01 Tudhope, D., Alani, H. ja Jones, C., Augmenting thesaurus relationships: Possibilities for retrieval. *Journal of Digital Information*, 1,8(2001).
- TaL01 Tam, A. M. ja Leung, C. H. C., Structured natural-language descriptions for semantic content retrieval of visual materials. *Journal of the American Society for Information Science and Technology*, 52,11(2001), sivut 930–937.

- TAO10 Taideteollisuusalan ontologia (TAO), 2010. <http://www.seco.tkk.fi/ontologies/tao/>. [18.3.2010]
- Tay06 Taylor, M., The Zthes specifications for thesaurus representation, access and navigation, 2006. <http://zthes.z3950.org/>. [17.3.2010]
- TFV09 Tuominen, J., Frosterus, M., Viljanen, K. ja Hyvönen, E., ONKI SKOS server for publishing and utilizing SKOS vocabularies and ontologies as services. *Proceedings of the 6th European Semantic Web Conference (ESWC'09)*, Heraklion, Greece, kesäkuu 2009, Springer-Verlag, sivut 768–780.
- TKV09 Tuominen, J., Kauppinen, T., Viljanen, K. ja Hyvönen, E., 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.
- Try79 Trygve, R., Models-views-controllers, Xerox PARC technical note, joulukuu 1979.
- TSV10 Tuominen, J., Salonoja, M., Viljanen, K. ja Hyvönen, E., A user interface for ontology repositories. *Proceedings of the Workshop on Ontology Repositories and Editors for the Semantic Web (ORES'10), the 7th Extended Semantic Web Conference (ESWC'10)*, Heraklion, Greece, kesäkuu 2010.
- TuB05 Tudhope, D. ja Binding, C., Towards terminology service: experiences with a pilot web service thesaurus browser. *Proceedings of the International Conference on Dublin Core and Metadata Applications (DC'05)*, Madrid, Spain, syyskuu 2005, Dublin Core Metadata Initiative, sivut 241–244.
- UKA04 About UKAT, 2004. <http://www.ukat.org.uk/about/>. [17.3.2010]
- UsG96 Uschold, M. ja Grüninger, M., Ontologies: principles, methods, and applications. *Knowledge Engineering Review*, 11,2(1996), sivut 93–155.
- UsJ99 Uschold, M. ja Jasper, R., 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.

- VAH07 Valkeapää, O., Alm, O. ja Hyvönen, E., An adaptable framework for ontology-based content creation on the semantic web. *Journal of Universal Computer Science*, 13,12(2007), sivut 1835–1853.
- VCH05 Vizine-Goetz, D., Childress, E. ja Houghton, A., Web services for genre vocabularies. *Proceedings of the International Conference on Dublin Core and Metadata Applications (DC'05)*, Madrid, Spain, syyskuu 2005, Dublin Core Metadata Initiative, sivut 197–200.
- VHK05 Valo, A., Hyvönen, E. ja Komulainen, V., A tool for collaborative ontology development for the semantic web. *Proceedings of the International Conference on Dublin Core and Metadata Applications (DC'05)*, Madrid, Spain, syyskuu 2005, Dublin Core Metadata Initiative, sivut 209–212.
- Voo94 Voorhees, E. M., Query expansion using lexical-semantic relations. *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Dublin, Ireland, heinäkuu 1994, Springer-Verlag, sivut 61–69.
- VTH08 Viljanen, K., Tuominen, J. ja Hyvönen, E., 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.
- VTH09 Viljanen, K., Tuominen, J. ja Hyvönen, E., Ontology libraries for production use: The Finnish ontology library service ONKI. *Proceedings of the 6th European Semantic Web Conference (ESWC'09)*, Heraklion, Greece, kesäkuu 2009, Springer-Verlag, sivut 781–795.
- VTS10 Viljanen, K., Tuominen, J., Salonoja, M. ja Hyvönen, E., Linked open ontology services. *Proceedings of the Workshop on Ontology Repositories and Editors for the Semantic Web (ORES'10), the 7th Extended Semantic Web Conference (ESWC'10)*, Heraklion, Greece, kesäkuu 2010.

- WSW01 Wielinga, B., Schreiber, G., Wielemaker, J. ja Sandberg, J. A. C., From thesaurus to ontology. *Proceedings of the International Conference on Knowledge Capture*, Victoria, Canada, lokakuu 2001, ACM Press, sivut 194–201.
- WVE85 Wang, Y.-C., Vandendorpe, J. ja Evens, M., Relational thesauri in information retrieval. *Journal of the American Society for Information Science*, 36,1(1985), sivut 15–27.
- YSA10 YSA (Yleinen suomalainen asiasanasto), 2010. <http://www.kansalliskirjasto.fi/kirjastoala/asiasanastot/ysa.html>.
[17.3.2010]

Liite 1. ONKI SKOS -palvelimen esimerkkikonfiguraatio

ONKI SKOS -palvelimen pääkonfiguraatiodosto sisältää muun muassa RDF-ominaisuuksien käsittelyyn liittyviä asetuksia. Alla olevaan esimerkkikonfiguraatioon on määritelty tarvittavat asetukset SKOS-sanastoja varten.

```
<?xml version="1.0" encoding="UTF-8" ?>

<!-- ===== -->
<!-- Configuration file for Ontology Server ONKI SKOS -->
<!-- Uses Apache Commons configuration framework. -->
<!-- For details: http://commons.apache.org/configuration/ -->
<!-- ===== -->
<onki-services-conf>

  <!-- === 1. Define used root classes, properties, labels etc. = -->
  <!-- Configures the properties that are used in rendering the -->
  <!-- hierarchies. Also defines labels that are shown from the -->
  <!-- concept. -->
  <!-- ===== -->
  <classes>

    <!-- === 1.1 Name of the ontology / ONKI instance ===== -->
    <!-- The text which is shown in top of the UI. -->
    <!-- If not defined, the label of the ontology resource or -->
    <!-- ConceptScheme is used. -->
    <!-- ===== -->
    <!-- <ontology-label xml:lang="fi"></ontology-label> -->

    <!-- === 1.2 Property used in building hierarchy ===== -->
    <!-- The property which creates the class-hierarchy. -->
    <!-- ===== -->
    <follow type="subClassOf">
      http://www.w3.org/2004/02/skos/core#broader</follow>
```

```

<follow type="superClassOf">
    http://www.w3.org/2004/02/skos/core#narrower</follow>

<!-- === 1.3 Labels of concepts ===== -->
<!-- Labels-node can have 1..n label-elements. -->
<!-- Having multiple labels is useful, if browser is -->
<!-- constructed from many ontologies, which have -->
<!-- different label describing the human readable label. -->
<!-- ===== -->
<labels>
    <label>http://www.w3.org/2004/02/skos/core#prefLabel</label>
    <label>http://www.w3.org/2000/01/rdf-schema#label</label>
</labels>

<!-- === 1.4 Additional properties ===== -->
<!-- Properties that aren't shown in the concept view by -->
<!-- default. They are shown when the "show all properties" -->
<!-- button is pushed. -->
<!-- ===== -->
<additional>
    <!-- <additional></additional> -->
</additional>

<!-- === 1.5 Hidden properties ===== -->
<!-- Properties that aren't shown in the concept view. -->
<!-- ===== -->
<hidden-properties>
    <property>
        http://www.w3.org/2004/02/skos/core#hiddenLabel</property>
</hidden-properties>

<!-- === 1.6 Alphabetical index configuration ===== -->
<!-- Alphabetical concept index is generated from this -->
<!-- comma separated list. -->
<!-- The default alphabet in the search view can be -->
<!-- configured, otherwise it is the first alphabet of the -->

```

```

<!-- list. -->
<!-- ===== -->
<alpha-index>0-9,A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,
        Y,Z,Å,Ä,Ö,*</alpha-index>

<!-- === 1.6.1 The default alphabet selected in the search == -->
<!-- <alpha-index-default></alpha-index-default> -->

<!-- === 1.7 Namespaces to index ===== -->
<!-- Concepts in these namespaces are indexed. -->
<!-- -->
<!-- lang-spec: (boolean true/false) -->
<!-- Whether to use language specification of indexed -->
<!-- labels (xml:lang) or not: -->
<!-- if true, the search is performed only to the labels of -->
<!-- current language, otherwise to all labels. -->
<!-- ===== -->
<index-namespaces lang-spec="true">
  <!-- <namespace></namespace> -->
</index-namespaces>

<!-- === 1.8 Alternative labels used in index ===== -->
<!-- Define alternative labels of concepts which are used -->
<!-- in the index. -->
<!-- ===== -->
<index-alt-labels>
  <label>http://www.w3.org/2004/02/skos/core#altLabel</label>
  <label>http://www.w3.org/2004/02/skos/core#hiddenLabel</label>
</index-alt-labels>

<!-- === 1.9 PropertyFiddlers ===== -->
<!-- Classes that generate simple properties from complex -->
<!-- RDF structures. -->
<!-- ===== -->
<property-fiddlers>
  <fiddler>fi.helsinki.cs.seco.onki.logic.propertyfiddlers.

```

```

        SkosMemberPropertyFiddler</fiddler>
</property-fiddlers>

<!-- === 1.10 Group configuration ===== -->
<!--   Define properties for the group search           -->
<!--   -----                                         -->
<!--   enabled:      (boolean true/false)             -->
<!--   Is the group search enabled or not.           -->
<!--   -----                                         -->
<!--   isGroupProperty:                               -->
<!--   Property used to identify group properties     -->
<!--   type: ("normal" / "reverse")                  -->
<!--   Is a group resource a subject ("normal") or   -->
<!--   a object ("reverse") of the property.         -->
<!--   -----                                         -->
<!--   isGroupValue:                                  -->
<!--   RDFNode related to the isGroupProperty, either object -->
<!--   (if isGroupProperty is "normal") or subject  -->
<!--   (if isGroupProperty is "reverse")           -->
<!--   type: ("resource" / "literal")                -->
<!--   The type of a value (if isGroupProperty is "reverse", -->
<!--   it can only be a resource).                  -->
<!--   -----                                         -->
<!--   belongsToGroupProperty:                       -->
<!--   Property used to express a concept belonging to a group -->
<!--   type: ("normal" / "reverse")                  -->
<!--   Is a concept a subject ("normal") or a object -->
<!--   ("reverse") of the property.                 -->
<!--   =====                                         -->
<group-search enabled="true" searchable="true" default_query="#">
  <isGroupProperty type="normal">
    http://www.w3.org/1999/02/22-rdf-syntax-ns#type
  </isGroupProperty>
  <isGroupValue type="resource">
    http://www.w3.org/2004/02/skos/core#Collection
  </isGroupValue>

```



```

    <belongsToGroupProperty type="inverse">
      http://www.w3.org/2004/02/skos/core#member
    </belongsToGroupProperty>
</group-search>

<!-- === 1.11 Type restriction configuration ===== -->
<!--   Define properties for the type restriction in search   -->
<!--   -----   -----   -----   -----   -----   -->
<!-- enabled:      (boolean true/false)                      -->
<!-- Is the group restriction search menu enabled or not.    -->
<!--   -----   -----   -----   -----   -----   -->
<!-- inclusive:   (boolean true/false)                       -->
<!-- Are the listed types inclusive or exclusive            -->
<!-- (to be/not to be shown in the type restriction menu    -->
<!--   -----   -----   -----   -----   -----   -->
<!-- type: (URI)                                           -->
<!-- type URIs to be/not to be shown in the type restriction -->
<!-- menu                                                 -->
<!-- =====   =====   =====   =====   ===== -->
<type-restriction enabled="true" inclusive="false">
  <type>http://www.w3.org/2004/02/skos/core#ConceptScheme</type>
  <type>http://www.w3.org/2002/07/owl#DatatypeProperty</type>
  <type>http://www.w3.org/2002/07/owl#ObjectProperty</type>
  <type>
    http://www.w3.org/1999/02/22-rdf-syntax-ns#Property</type>
</type-restriction>

<!-- === 1.12 Properties acting as URLs ===== -->
<url-properties>
  <!-- <property></property> -->
</url-properties>

<!-- === 1.13 Related properties ===== -->
<related-properties>
  <property>
    http://www.w3.org/2004/02/skos/core#related</property>

```

```

</related-properties>

<!-- === 1.14 Description properties ===== -->
<description-properties>
  <property>http://www.w3.org/2004/02/skos/core#note</property>
</description-properties>

<!-- === 1.15 Equivalent class properties ===== -->
<equivalent-class-properties>
  <property>
    http://www.w3.org/2004/02/skos/mapping#exactMatch
  </property>
</equivalent-class-properties>

<!-- === 1.16 HTTP API configuration ===== -->
<rest-api>
  <get-properties>
    <property-order>
      <!-- <property></property> -->
    </property-order>
    <hidden-properties>
      <!-- <property></property> -->
    </hidden-properties>
  </get-properties>
  <get-full-presentation>
    <hierarchy-size>
      <parents>-1</parents>
      <children>1</children>
      <siblings>1</siblings>
    </hierarchy-size>
    <empty-methods>
      <!-- <method></method> -->
    </empty-methods>
  </get-full-presentation>
</rest-api>

```

```

</classes>

<!-- === 2. Define visualization of ontologies ===== -->
<!--   Defines localization.                               -->
<!-- ===== -->
<visualization>

    <!-- === Define available locales and default for the UI ==== -->
    <!--   ResourceBundles containing the locales are stored in -->
    <!--   WEB-INF/classes/OnkiLocales_*.properties. Each locale -->
    <!--   should have own bundle.                               -->
    <!-- -->
    <!-- id: -->
    <!--   Language-code for the locale, OnkiLocales_id-file -->
    <!--   should contain the localization.                     -->
    <!-- value: -->
    <!--   Locale name shown in the UI.                         -->
    <!-- ===== -->
    <locales default="fi">
        <locale id="fi">Suomi</locale>
        <locale id="en">English</locale>
        <locale id="sv">Svenska</locale>
    </locales>

</visualization>

<!-- === 3. MultiONKI functionality ===== -->
<!--   For multiple ontologies with mappings, uses -->
<!--   owl:equivalentClass in building of concept hierarchy. -->
<!-- ===== -->
<multionki enabled="false"/>

<!-- === 4 Default view of the ONKI server ===== -->
<!--   concept = the uri of the concept shown -->
<!--   type = the type selection -->
<!-- ===== -->

```

```

<defaultview>
  <!-- <concept></concept> -->
  <type>http://www.w3.org/2004/02/skos/core#Concept</type>
</defaultview>

<!-- === 5 Lucene index settings ===== -->
<!--   Index can be memory- or file-based -->
<!-- -->
<!-- type: (ram/file) -->
<!-- path: path to the disk-based index directory, -->
<!--       can be absolute or relative. -->
<!-- ===== -->
<index type="file" path="lucene-index"/>

<!-- === 6 Link to an ontology info page ===== -->
<!--   Link to an external page describing ontology. -->
<!-- -->
<!-- enabled:   (boolean true/false) -->
<!-- ===== -->
<ontology-info-page enabled="true">
  http://www.example.com</ontology-info-page>

<!-- language settings - preference order -->
<languages>
  <preference>fi,en</preference>
</languages>

</onki-services-conf>

```

Liite 2. ONKI SKOS -palvelimen HTTP-rajapinnan metodikuvaus

ONKI SKOS -palvelimen ohjelmallinen HTTP-rajapinta sisältää alla kuvatut metodit. Metodit on ryhmitelty käyttötarkoituksen mukaan.

Palvelimen tilatiedot

ping

kuvaus: testataan onko palvelin käynnissä

parametrit: -

vastaus: "1"

Ontologian metatiedot

getMetadata

kuvaus: metatietoa ontologiasta

parametrit:

l - kieli

vastaus: metatietoa ontologiasta

getAvailableLanguages

kuvaus: kielet joilla ontologian käsitteiden nimiä on saatavissa

parametrit: -

vastaus: lista kielikoodista

getAvailableTypeUris

kuvaus: ontologiassa olevien käsitteiden tyyppien URI-tunnisteet

parametrit: -

vastaus: lista käsitetyyppien URI-tunnisteita

Käsitteen haku

search

kuvaus: käsitteen haku parametrit:

q - hakumerkkijono

l - hakukieli

qt - haun rajausta tietyn tyyppiin käsitteisiin (URI)

qp - haun rajausta tietyn käsitteen alakäsitteisiin transitiiivisesti (URI)

qg - haun rajausta tiettyyn ryhmään kuuluviin käsitteisiin (URI)

h - hakutulosten enimmäismäärä

s - aloita tulosjoukko hakutuloksesta nro (hakutulosten sivutusta varten)

ac - sisällyttä tulosjoukkoon hakumerkkijonolla alkavat käsitteet (totuusarvo: "true"/"false")

vastaus: metatietoa hausta ja lista hakutuloksista, joilla nimi johon haku täsmäsi (prefLabel/altLabel) ja URI

Käsitteen hierarkia

getConceptHierarchy

kuvaus: käsitteen hierkia (ylä-/ala-/vieruskäsitteet)

parametrit:

u - käsitteen URI

l - kieli

p - yläkäsitteiden tasojen lkm: 0..n; 1 == vain suorat yläkäsitteet, 2 == suorat yläkäsitteet ja niiden suorat yläkäsitteet, jne.

c - alakäsitteiden tasojen lkm: 0..n

s - vieruskäsitteet: 0..1

vastaus: lista hierarkiapuun solmuista, joilla URI, nimi, yläkäsitteen URI, sisennystaso (käyttöliittymän tekemisen helpottamiseksi)

getHierarchyRoots

kuvaus: käsitteen hierarkian juuret, tai jos URIa ei anneta, kaikki ontologian käsitteet, joilla ei yläkäsitettä

parametrit:

u - käsitteen URI

l - kieli

vastaus: lista käsitteistä, joilla URI ja nimi

Käsitteen tiedot

getProperties

kuvaus: käsitteen tiedot (RDF properties)

parametrit:

u - käsitteen URI

l - kieli

p - ominaisuuden URI, jos halutaan vain tietty ominaisuus

vastaus: lista ominaisuus-arvo-pareista, joilla ominaisuuden URI, nimi sekä arvon URI ja nimi (jos kyseessä on resurssiviittaus) tai arvo literaalina ja kielikoodi (jos kyseessä literaali)

getLabels

kuvaus: käsitteen nimet

parametrit:

u - käsitteen URI

l - kieli

pl - ominaisuuden nimen kieli

vastaus: lista nimistä, joilla itse nimen lisäksi kielikoodi, ominaisuuden URI ja ominaisuuden nimi

getAltLabels

kuvaus: käsitteen vaihtoehtoiset (korvatut) nimet

parametrit:

u - käsitteen URI

l - kieli

pl - ominaisuuden nimen kieli

vastaus: lista vaihtoehtoisista nimistä, joilla itse vaihtoehtoisen nimen lisäksi kielikoodi, ominaisuuden URI ja ominaisuuden nimi

getDescriptions

kuvaus: käsitteen kuvaukset

parametrit:

u - käsitteen URI

l - kieli

pl - ominaisuuden nimen kieli

vastaus: lista kuvauksista, joilla itse kuvauksen lisäksi kielikoodi, ominaisuuden URI ja ominaisuuden nimi

getEquivalentConcepts

kuvaus: ekvivalentit käsitteet

parametrit:

u - käsitteen URI

l - kieli

vastaus: lista ominaisuus-käsite-pareista, joilla ominaisuuden URI, nimi sekä käsitteen URI ja nimi

getRelatedConcepts

kuvaus: assosiatiiviset käsitteet / rinnakkaiskäsitteet

parametrit:

u - käsitteen URI

l - kieli

vastaus: lista ominaisuus-käsite-pareista, joilla ominaisuuden URI, nimi sekä käsitteen URI ja nimi

getGroups

kuvaus: ryhmät joihin käsite kuuluu

parametrit:

u - käsitteen URI

l - kieli

vastaus: lista ominaisuus-ryhmä-pareista, joilla ominaisuuden URI, nimi sekä ryhmän URI ja nimi

getTypes

kuvaus: käsitteen tyypit

parametrit:

u - käsitteen URI

l - kieli

vastaus: lista ominaisuus-tyyppi-pareista, joilla ominaisuuden URI, nimi sekä tyyppin URI ja nimi

getFullPresentation

kuvaus: kaikki oleellinen tieto käsitteestä (yhdistää metodit *getProperties*, *getLabels* jne. yhteen metodikutsuun)

parametrit:

u - käsitteen URI

l - kieli

pl - ominaisuuden nimen kieli

vastaus: lista metodinimi-vastaus-pareista

Ontologian hakemisto / yleisnäkymä

getDirectory

kuvaus: ontologian hakemisto - aakkosellinen hakemisto, ryhmät (esim. YSA) tai ylätasoon käsitteet (SKOS-sanastot)

parametrit:

t - hakemiston tyyppi: "groups", "topconcepts" tai "alphabetical"; oletus on "groups", jos ontologiassa on määritelty ryhmät; "topconcepts", jos ontologia on SKOS-sanasto ja ylätasoon käsitteitä on enintään 50; muutoin "alphabetical"

m - kuinka monta hakemistoelementteihin (ryhmä/ylätasoon käsite/kirjain) kuuluvaa käsitettä palautetaan: 0..n

l - kieli

vastaus: lista hakemistoelementeistä (ryhmä/ylätason käsite/kirjain), joilla URI (ryhmä/ylätason käsite), nimi, hakemistoelementtiin kuuluvien käsitteiden lkm sekä lista siihen kuuluvista käsitteistä