

# **Käsitteen kontekstiperustainen valinta semanttisessa webissä**

Reetta Sinkkilä

Helsinki 13.5.2008

Pro gradu -tutkielma

HELSINGIN YLIOPISTO

Tietojenkäsittelytieteen laitos

|   |  |                                   |   |
|---|--|-----------------------------------|---|
| Tiedekunta/Osasto — Fakultet/Sektion — Faculty  |  | Laitos — Institution — Department |   |
| Matemaattis-luonnontieteellinen   |  | Tietojenkäsittelytieteen laitos   |   |
| Tekijä — Författare — Author  |  |                                   |   |
| Reetta Sinkkilä   |  |                                   |   |
| Työn nimi — Arbetets titel — Title  |  |                                   |   |
| Käsitteen kontekstiperustainen valinta semanttisessa webissä  |  |                                   |   |
| Oppiaine — Läroämne — Subject   |  |                                   |   |
| Tietojenkäsittelytiede  |  |                                   |   |
| Työn laji — Arbetets art — Level  |  | Aika — Datum — Month and year     | Sivumäärä — Sidoantal — Number of pages |
| Pro gradu -tutkielma  |  | 13.5.2008                         | 68 sivua                                |
| Tiivistelmä — Referat — Abstract  |  |                                   |   |
| <p>Semanttisen webin ideana on kuvailla tietoa siten, että koneet pystyvät “ymmärtämään” sitä, ja käyttämään älykkäitä tekniikoita tiedon hyödyntämiseksi. Tiedon kuvailemiseen käytetään ontologioita jotka muodostavat laajoja käsiteverkkoja. Kuvaailua varten ontologiasta on löydettävä tarkimmin kohdetta kuvaavat käsitteet.</p> <p>Tässä työssä tutkittiin minkälaisia menetelmiä käsitteervalintaan on kehitetty ja millä tavoin käsitteervalintaa voidaan tukea visuaalisin keinoin. Lisäksi perehdyttiin joihinkin semanttisen webin sovelluksiin niiden käsitteevalinnan osalta. Käsitevalitsimissa havaittiin puutteita liittyen käsitteiden merkityksen disambiguointiin ja siihen, kuinka helppoa niiden avulla on valita paras käsite kuvaamaan tietoa. Ongelmia oli myös tavoissa esittää useiden ontologioiden yhdistelmiä.</p> <p>Näiden havaintojen pohjalta suunniteltiin ja toteutettiin yleinen ja monentyyppisen aineiston käsittelyyn soveltuva käsitevalitsin <i>IRMA</i>.</p> <p>ACM Computing Classification System (CCS):</p> <p>H.3.3. [Information Search and Retrieval]: Selection process,</p> <p>I.2.4 [Knowledge Representation Formalisms and Methods]: Semantic networks,</p> <p>H.5.2 [User Interfaces]</p> |  |                                   |   |
| Avainsanat — Nyckelord — Keywords   |  |                                   |   |
| semanttinen web, käsitteen valinta, ontologian visualisointi, metadata, annotointi  |  |                                   |   |
| Säilytyspaikka — Förvaringsställe — Where deposited   |  |                                   |   |
| Kumpulan tiedekirjasto, sarjanumero C-  |  |                                   |   |
| Muita tietoja — Övriga uppgifter — Additional information   |  |                                   |   |

# Sisältö

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Johdanto</b>  | <b>1</b>  |
| <b>2</b> | <b>Käsitevalitsimet ja ontologioiden visualisointi</b> | <b>4</b>  |
| 2.1      | Ontologioiden visualisointi . . . . .                  | 4         |
| 2.2      | Moniontologiavisualisointi . . . . .                   | 15        |
| 2.3      | Käsitevalitsimet . . . . .                             | 16        |
| 2.3.1    | Automaattitäydennys . . . . .                          | 17        |
| 2.3.2    | Puuselaimet . . . . .                                  | 20        |
| <b>3</b> | <b>Käsitevalitsimien arviointi</b>                     | <b>22</b> |
| 3.1      | Automaattitäydentävät valitsimet . . . . .             | 22        |
| 3.2      | Puuselaimet . . . . .                                  | 24        |
| 3.3      | Yhteenvedo käsitevalitsimista . . . . .                | 25        |
| <b>4</b> | <b>IRMA kontekstipohjainen käsitevalitsin</b>          | <b>26</b> |
| 4.1      | Malli . . . . .  | 29        |
| 4.2      | Ohjain . . . . .                                       | 31        |
| 4.3      | Näkymä . . . . .                                       | 33        |
| <b>5</b> | <b>IRMA käytännössä</b>                                | <b>34</b> |
| 5.1      | Kuvataiteen ontologia ICONCLASS . . . . .              | 36        |
| 5.2      | Lääketieteen ontologia MeSH . . . . .                  | 41        |
| 5.3      | Suomen ajallinen paikkaontologia SAPO . . . . .        | 44        |

|                     |           |
|---------------------|-----------|
|                     | iii       |
| <b>6 Arviointi</b>  | <b>50</b> |
| <b>7 Yhteenveto</b> | <b>56</b> |
| <b>Lähteet</b>      | <b>59</b> |

# 1 Johdanto

Semanttisella webillä viitataan World Wide Webin päälle rakennettuun metatietokerrokseen, joka määrittelee esitettävän tiedon merkityksen koneen käsiteltäväsä muodossa [BLHL01]. Abstraktimmin semanttisen webin ajatuksen esittää Guha [GMM03]: semanttinen web ei ole dokumenttiverkko, vaan verkko suhteista todellista maailmaa kuvaavien objektien, kuten ihmisten ja tapahtumien välillä.

Webissä olevat dokumentit on tarkoitettu ihmisen luettaviksi, niiden merkitystä ei voida käsitellä koneella. Semanttinen web pyrkii esittämään ihmismaailman semantiikkaa formaalisti *ontologioilla*. Ontologialla tarkoitetaan koneluettavaa käsitelmää, jolla tietämys esitetään semanttisessa webissä; ontologia on käsitteiden formaali määrittely [Gru93]. Ontologia voidaan mallintaa esimerkiksi RDF-kielellä [Fen03]. RDF perustuu kolmikkomalliin, jossa käsitteiden suhteet esitetään kolmikkona: subjekti, predikaatti ja objekti [KC04]. Kolmikoista voidaan muodostaa laajoja verkkoja, joissa subjektit ja objektit toimivat solmuina ja predikaatit niiden välisinä kaarina. Ontologioiden avulla voidaan koneellisesti viitata tiettyyn käsitteeseen monimerkityksellisten sanojen sijaan.

Semanttinen web edellyttää aineistojen kuvaamista semanttisesti liittämällä niihin metatietoa. Metatiedon liittämistä kutsutaan annotoinniksi. Esimerkiksi museoesineitä luetteloitaessa tallennetaan muun muassa esineen tekijä, tekotapa ja tekoaika [Sal06]. Sisällön kuvailua on perinteisesti tehty valitsemalla sanastojen ja thesaurusten avulla kuvailtavaa kohdetta luonnehtivat asiasanat. Semanttisessa webissä luettelointi eli annotointi tehdään valitsemalla literaaleja ja käsitteitä ontologiasta [SDWW01].

Luetteloinnissa noudatetaan tyypillisesti metatietoskeemaa, joka määrittää mitä asioita kohteesta luetteloitaan [SDWW01]. Metatietoskeeman avulla luetteloidusta aineistosta voidaan koneellisesti tunnistaa eri elementtejä. Eräs tunnettu metatie-

toskeema on Dublin Core<sup>1</sup> [dcs08]. Se sisältää perustasollaan 15 metatietoelementtiä kuten otsikko, tekijä sekä aihe. Annotoitaessa esimerkiksi Ateneumin taidemuseon maalausta *Haavoittunut enkeli* Dublin Coren mukaisesti, liitettäisiin otsikoksi merkijonona teoksen nimi. Tekijään voidaan liittää ontologinen viite toimijaontologian [KH07] instanssiin, jonka ominaisuuden ”nimi” arvona on Hugo Simberg. Vastaa- vasti aiheeksi voidaan liittää viitteet esimerkiksi Iconclass-ontologian [Gor88] käsitteisiin ”lapsi” ja ”maisema”. Ontologisten viitteiden käyttöä annotaatioiden arvona kutsutaan ontologiaperustaiseksi annotoinniksi [SDWW01]. Näin annotoitua aineistoa voidaan koneellisesti yhdistää. Voidaan esimerkiksi ymmärtää Akseli Gallen-Kallelan ja Axel Gallenin teoksien tekijäksi sama henkilö, kun tekijästä on viite samaan ontologian käsitteeseen, jolle on määritetty molemmat nimet.

Ontologioiden laajuuden vuoksi oikean käsitteen löytäminen ei ole itsestäänselvää. Jos käyttäjä ei tunne tarkasti ontologian sisältöä, on vaikea tietää, mikä käsitteen synonyymeistä sanastossa esiintyy. Ongelmia käsitteen valintaan aiheutuu myös homonymiasta: tarkoitetaanko käsitteellä ”nokia” yritystä vai kaupunkia? Entä onko valittu käsite tarkin kuvaamaan kohdetta?

Tässä työssä selvitetään, millä tavoin käsite voidaan valita ontologiasta ja millä tavoin oikean käsitteen löytymistä voidaan edistää. Kirjoituksessa perehdytään tapoihin esittää ontologian sisältö visuaalisessa muodossa ja selvitetään, miten näillä käytännöillä voitaisiin tukea käsitteen valintaa. Lisäksi tutustutaan toteutettuihin käsitevalitsimiin. Katsauksen perusteella eritellään mahdollisuuksia ja ongelmia käsitteenvalinnassa ja kehitetään tämän perusteella uudenlainen ratkaisu käsitevalitsimeksi.

Tarkoituksena on kehittää käsitevalitsin, jolla voidaan sekä helpottaa hakua että nopeuttaa annotointia. Ajatuksena on, että oikeanlaisella käsitevalitsimella voidaan ontologiasta löytää osuvin käsite kuvaamaan annotoitavaa kohdetta ja tätä kaut-

---

<sup>1</sup><http://dublincore.org/>

ta luetteloinnista saadaan täsmällinen. Tavoitteena on myös, että käsitevalitsimen avulla pystytään erottamaan samannimisten käsitteiden merkitys toisistaan.

Työ rajoittuu ontologiaperustaiseen käsitevalintaan annotoinnissa. Toteutetussa käsitevalitsimessa on erityisesti kiinnitetty huomiota tilanteisiin, joissa käytössä on useita ontologioita. Nykyisillä käsitevalitsimilla ei yleensä ole mahdollisuutta visualisoida useampaa ontologiaa ja haut joudutaan tekemään erikseen eri aineistoihin. Tuotettua käsitevalitsinta esitellään kolmessa käyttökontekstissa. Ensimmäisessä tapauksessa käsitevalitsin on konfiguroitu käyttämään kahta eri ontologiaa Valtion taidemuseon käyttötarpeisiin luetteloinnissa. Toisen käyttötapauksen avulla esitellään miten lääketieteen ontologiaa voidaan visualisoida, ja viimeiseksi esitellään käsitteen valintaa Suomen ajallisesta paikkaontologiasta.

Tutkimuskysymyksiksi voidaan muotoilla seuraavat:

- Miten ontologiasta saadaan valittua tarkin kohdetta kuvaava käsite?
- Miten käsitteen merkitys saadaan välitettyä valinnan tekijälle?
- Minkälaisilla visualisoinneilla käsitteen valintaa voidaan helpottaa?

Tutkielma rakentuu siten, että aluksi luvussa 2 käsitellään käsitevalintaa semanttisen webin sovelluksissa sekä esitellään tapoja visualisoida ontologioita. Luvussa 3 käsitellään ongelmia esiteltyjen järjestelmien käsitevalitsimissa. Luvussa 4 esitellään toteutettu ohjelmallinen ratkaisu kontekstipohjaiseksi käsitevalitsimeksi. Luvussa 5 esitellään kolme tapausta kehitetyn järjestelmän käytöstä. Luvussa 6 arvioidaan tuotettu ratkaisua ja esitetään suunnitelmia jatkokehitykseen. Viimeisessä luvussa tehdään yhteenveto tutkielman keskeisestä sisällöstä.

## 2 Käsitevalitsimet ja ontologioiden visualisointi

Tässä luvussa esitellään tapoja visualisoida ontologioita ja muita toisiinsa liittyviä verkkomaisia aineistoja. Lisäksi tutkitaan millä tavoin voidaan visualisoida useita ontologioita samanaikaisesti. Lopuksi tarkastellaan, miten käsitevalinta on toteutettu eräissä semanttisen webin sovelluksissa.

### 2.1 Ontologioiden visualisointi

Edward Tuften mukaan hyvä informaation visualisointi koostuu monimutkaisten asioiden esittämisestä selvästi, tarkasti ja tehokkaasti [Tuf01]. Ontologioiden kohdalla haasteena on saada monimutkainen verkko näkymään edustavasti. Oikean käsitteen löytämiseksi pelkän käsitehierarkian lisäksi voi olla aiheellista visualisoida käsitteisiin liittyviä ominaisuuksia ja käsitteiden välisiä suhteita.

Visualisointimenetelmistä on esitetty useita luokitteluja. Keim [Kei02] luokittelee visualisointitavat ensiksikin sen aineiston mukaan, mitä niillä voidaan visualisoida (kuten tekstidokumentit, kuvat, prosessit). Toiseksi hän jakaa tekniikat graafeihin, pikselöintiin perustuviin menetelmiin, geometrisiin menetelmiin, sekä pinovisualisointimenetelmiin. Viimeinen aspekti Keimin luokittelussa on interaktiivisuus: tästä hän erottelee aineiston suodatuksen, tarkentamisen, vääristymän (distortion) ja linkityksen. Suodatuksella tarkoitetaan menetelmää, jossa voidaan valita aineiston osajoukko tarkasteltavaksi. Tarkentamisella viitataan visualisointitekniikkaan, jossa aineistosta esitetään yleiskuva jota voidaan tarkentaa. Näkymän vääristämiseen perustuvilla tekniikoilla osa aineistosta esitetään yksityiskohtaisesti ja osa epätarkasti. Linkityksellä Keim viittaa useiden tekniikoiden yhdistelmiin. Shneiderman [Shn96] jaottelee visualisointimenetelmät sekä visualisoitavan tiedon tyypin mukaan (1-, 2- ja 3-ulotteiset, ajalliset, moniulotteiset, puu, verkko), että tehtävätyypin mukaan (yleiskuva, tarkennus, suodatus, yksityiskohdat, relaatiot, historia, otokset). Katifo-



ri [KHL<sup>+</sup>07] jaottelee visualisointimenetelmät menetelmän tyyppin mukaisesti kuu-teen luokkaan:

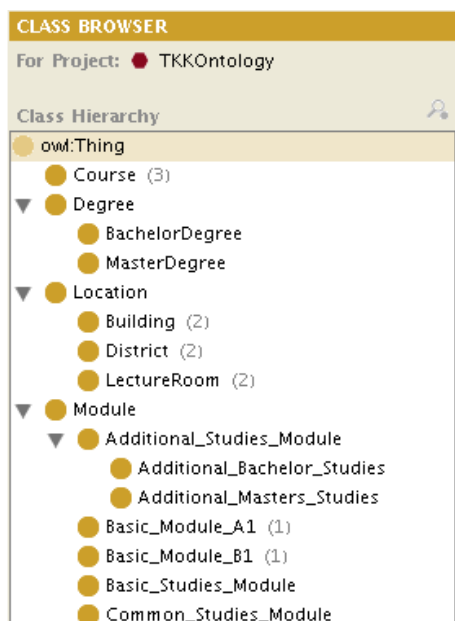
- sisennetyt listat
- puurakenteet
- tarkennettavat visualisoinnit
- alueperustaiset
- kohde ja ympäristö

Seuraavaksi esitellään tarkemmin Katiforin luokitukseen perustuen eri visualisointimenetelmiä ja miten niitä voidaan käyttää ontologioiden yhteydessä. Lisäksi kiinnitetään huomiota ontologisen aineistojen visualisoinnissa esiintyviin erikoistapauksiin.

### **Sisennetyt listat**

*Sisennetyt listat* ovat käytössä monessa ontologiavisualisointijärjestelmässä. Menetelmä on sama kuin monen käyttöjärjestelmän tiedostoselaimessa. Menetelmä sopii hyvin hierarkkisen aineiston esittämiseen, sillä ontologian luokkasuhteista saadaan hyvin muodostettua sisennys. Tällöin taksonomia voidaan esittää puuna, kuten Protege-ontologiaeditorissa [NSD<sup>+</sup>01] (Kuva 1). Kuvassa on osa esitetty korkeakouluun liittyvää ontologiaa, esimerkiksi nähdään, että maisterin tutkinto (Master-Degree) on sisennetty tutkinnon (Degree) alle.

Semanttisen webin järjestelmistä sisennettyä listaa ontologian visualisointikeinona käytetään muun muassa EKOSS-selaimessa [KGKN06], CIRI-tiedonhakujärjestelmässä [AJS<sup>+</sup>04] sekä ONKI-ontologiaselaimessa [VTH08]. Myös järjestelmissä, joissa pääasiallinen visualisointitapa on jokin muu, käytetään usein sisennettyä listavisuali-



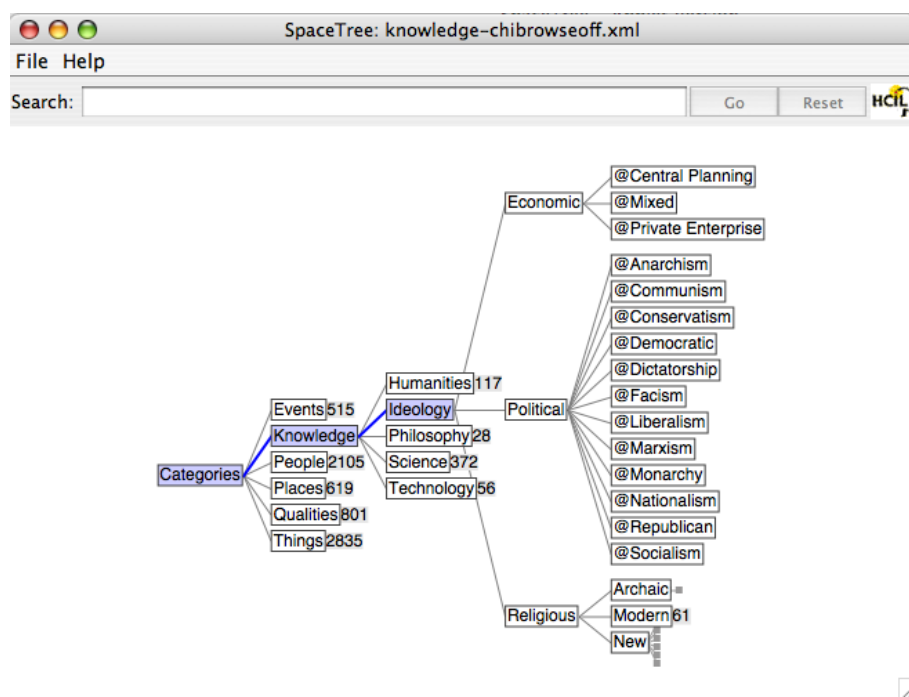
Kuva 1: Protegen luokkaselain.

sointia pääasiallisen esitystavan rinnalla. Tällaisia järjestelmiä ovat muun muassa ontologiaselain ja -editori Swoop [KPS<sup>+</sup>06], OntoEdit-ontologiaeditori [SEA<sup>+</sup>02] sekä Ontorama [ERG02].

## Puurakenteet

*Puurakenteisilla visualisointimenetelmillä* (node-link trees) tarkoitetaan verkkomaisesta esitystä, joka muodostuu toisiinsa linkittyvistä solmuista. Taksonomia voidaan esittää ylhäältä alas tai vasemmalta oikealle. Visualisointi voi olla myös laajennettavissa siten, että kaikkia solmuja ei oletusarvoisesti näytetä, vaan käyttäjä voi avata niitä tarpeen mukaan. Puurakenteista ontologiavisualisointia käyttävät muun muassa Protegen laajennokset OntoViz [Sin] ja OwlViz [Hor]. Java-appletina toteutettu SpaceTree [PGB02] lisää verkkovisualisointiin interaktiivisuutta. Näytettävä hierarkia laajenee käyttäjän valintojen mukaan aina kaksi tasoa käyttäjän valitsemasta solmusta ja graafi keskitetään näytölle (kuva 2). SpaceTreessä erotellaan visuaalisesti myös laajennettavissa olevat puun haarat merkitsemällä ne kuten kuvassa 2

alimmaisena näkyvät kategoriat “Archaic” ja “New”. Jos alipuita on vähän, niiden lukumäärä ilmaistaan pisteinä. Useampia alipuita sisältävien solmujen yhteydessä lukumäärä ilmoitetaan numerona kuten kuvassa kategorian “Modern” yhteydessä. Hyvin samantyylinen visualisointi on myös TreePlus-järjestelmässä [LPP<sup>+</sup>06].



Kuva 2: SpaceTree ontologiaselain.

Verkkovisualisoinneista on myös kolmi- ja useampiulotteisia versioita. OntoSpheren [BBP05] lähtökohtana on esittää visualisointi kolmiulotteisessa avaruudessa ja lisätä tähän uusia ulottuvuuksia kuten solmujen värikooodaus, muotoilu ja läpinäkyvyysaste. Ajatuksena tässä on vähentää graafisia elementtejä, kuten graafiin merkittyjä elementtien nimiä, ja tehdä esityksestä selkeämpi. Toinen moniulotteinen visualisointitapa on hierarkkisen aineiston visualisointiin kehitetty Cone Tree [RMC91]. Siinä solmut esitetään hierarkkisten kartioiden muodostamana kolmiulotteisena puuna siten, että yhdessä kartiossa on juurisolmu ja sen lapsisolmut. Kartioita voidaan dynaamisesti lähentää ja loitontaa ja saada näin näkyviin yksityiskohtainen näkymä käsitteistä tai yleiskuva ontologiapuusta. Kun kartiopuuta lähennetään riittävästi,

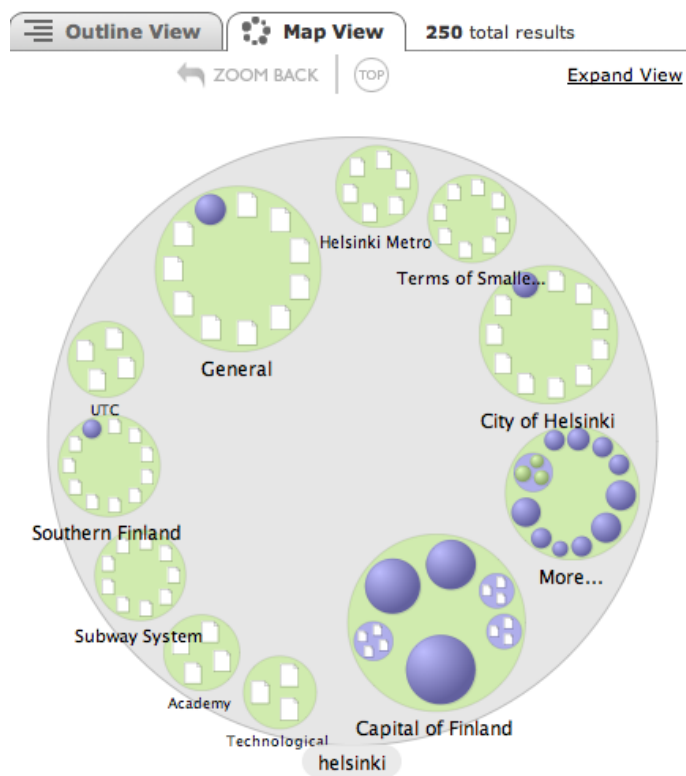
sen reunoilla olevat kartiot pyörähtävät näkymättömiin kolmanteen ulottuvuuteen.

### **Tarkennettavat menetelmät**

Kolmantena tyyppinä Katiforin [KHL<sup>+</sup>07] visualisointiluokittelussa on *tarkennettavat visualisointimenetelmät* (zoomable user interfaces, ZUI [BMG00]). Äskeisen kapaleen ConeTree- ja OntoSphere-visualisointisovelluksissa oli niiden pääasiallisen visualisointimenetelmän lisäksi mahdollista täsmentää näkymää ontologiasta lähentämällä sitä. Tästä on kyse Katiforin tarkennettavissa visualisointimenetelmissä. Niillä tarkoitetaan kaikkia sellaisia menetelmiä, joissa hierarkiassa kauempana olevat solmut piilotetaan tai esitetään vanhempiensa sisällä ja joita käyttäjä voi halutessaan tarkentaa. Näkyvillä on siis kerralla vähemmän dataa. Tarkennettavien visualisointien etuna on graafin selkeytyminen, sillä kaaret ja niiden nimet tekevät kuviosta sekavan ja heikentävät luettavuutta, jos niitä on nähtävillä kovin runsaasti [TH06].

Grokker [RB03] perustuu käsitteiden klusterointiin ja klusterien laajentamiseen. Grokker on suunniteltu minkä tahansa tekstimuotoisen aineiston visualisointiin, ei niinkään ontologioille. Grokkerissa aineisto esitetään sisäkkäisinä Vennin diagrammeina, siten että hierarkiassa alemmalla tasolla olevat käsiteklusterit näkyvät pienempinä diagrammeina isomman sisällä. Alemman tason käsiteklustereita voidaan tarkastella lähentämällä niitä käyttöliittymästä. Kuvassa 3 on Grokkerin web-käyttöliittymä, jossa on visualisoituna web-haun tulokset hakusanalla ”Helsinki”. Siinä esimerkiksi tarkentamalla diagrammiin ”Capital of Finland” päästään tarkastelemaan tähän klusteriin liittyviä dokumentteja.

Muita tarkennettavia visualisointijärjestelmiä ovat muun muassa Protegen Jambalaya-lisäosa [SMS<sup>+</sup>01] sekä Crop Circles [PWG05]. Näiden toimintaperiaate on sama kuin Grokkerin: aineistosta muodostetaan jollakin perusteella klustereita, joiden sisältöä voidaan tarkastella tarkentamalla sisään klusteriin.



Kuva 3: web-hakutulos sanalle “Helsinki” visualisoituna Grokker-työkalulla.

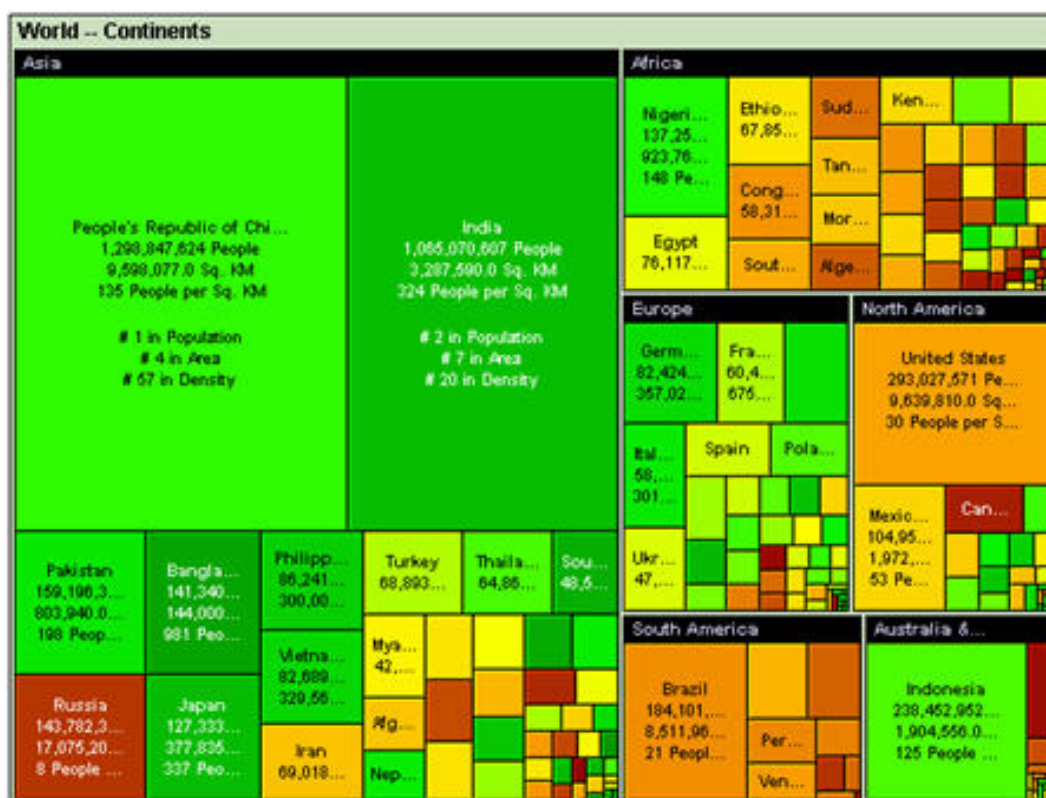
### Aluepohjaiset visualisoinnit

*Aluepohjainen visualisointi* (space-filling) perustuu ajatukseen, että koko käytettävissä oleva visualisointitila jaetaan osiin solmujen kesken siten, että ylempänä hierarkiassa oleva solmu saa suhteessa suuremman osan pinta-alasta kuin alempana olevat.

Eräs suosittu aluepohjainen visualisointimenetelmä on Treemap, jonka on alunperin kehittänyt Ben Shneiderman [Shn92]. Shneidermanin tekniikassa alue jaetaan hierarkkisiin neliskulmioihin, joiden koolla, värillä ja järjestyksellä voidaan havainnollistaa hierarkian kategorioiden ominaisuuksia ja niiden arvoja. Kuvassa 4<sup>2</sup> on treemap-esitys maailman väestön jakautumisesta maanosittain ja maittain. Koko alueen pinta-ala merkitsee maailman koko väestöä. Alueen pinta-ala on jaettu sa-

<sup>2</sup>Kuva Hive Group:n sovelluksesta [http://www.hivegroup.com/gallery/galleryapps\\_worldpop.html](http://www.hivegroup.com/gallery/galleryapps_worldpop.html)

massa suhteessa kuin maailman väestö jakautuu maanosittain. Kuvasta voidaan esimerkiksi nähdä Aasiaan kuuluvan yli puolet maailman väestöstä. Vastaavasti maanosan sisällä pinta-ala on jaettu maille niiden asukasluvun mukaan. Ontologioiden visualisoinnissa Treemap-tekniikkaa on käytetty muun muassa geeniontologian visualisoinnissa [BDBS04]. Toinen aluepohjainen visualisointimenetelmä on Information slices [AH98]. Sen ajatuksena on käyttää puoliympyrän muotoisia levyjä, jossa yhdellä levyllä esitetään 5 tai useampia hierarkiatasoa.



Kuva 4: Treemap-visualisointi maailman väestön jakautumisesta alueittain.

## Kohde ja ympäristö

*Kohde ja ympäristö* (Focus + context) on informaation visualisoinnissa käytetty periaate, jossa tärkein tieto näytetään tarkasti ja yksityiskohtaisesti visualisoinnin keskellä ja keskipisteen ympäristö (konteksti) sen ympärillä. Kauempana ole-



misen puussa ja keskittää valitun solmun aina visualisointitilan keskelle. Puun saa hyperboliseen muotoon siten, että sen kauempana keskussolmusta olevat solmut näytetään pienempänä ja ne näyttävät olevan “kauempana”.

## Aikajanavisualisointi

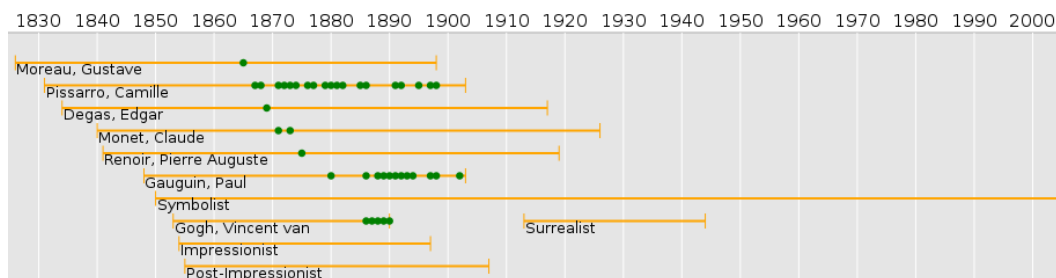
Katiforin esittämän luokittelun lisäksi ontologioiden visualisointimenetelmiä voidaan tarkastella visualisoitavan aineiston sisällön perusteella. Joidenkin ontologioiden kohdalla aineisto on järkevää esittää kronologisessa muodossa. Näin esimerkiksi historiallisten tapahtumien ontologiaa [HAK07] voidaan visualisoida aikajanan avulla [CK91]. Näin on tehtykin KulttuuriSampo-portaalissa [HRH<sup>+</sup>08], jossa on annotoitu 220 tapahtumaa Suomen historiasta vuosilta 1850–1920 käsitteillä, jotka on otettu historiallisten tapahtumien ontologiasta sekä Yleisestä suomalaisesta ontologiasta YSO [HSVF07]. KulttuuriSammossa havainnollistetaan myös museokoelmissa esiintyvien kulttuuriobjektien valmistusaikaa aikajanan avulla.

Toinen aikajanoja hyväksikäyttävä sovellus on MultimediaN E-Culture Demonstrator [SAvA<sup>+</sup>06]. Se on kulttuurialan portaali, jonka eräänä osana maalauksia, taidesuuntien ajoituksia sekä maalareiden elinaikoja visualisoidaan aikajanalla. Kuvassa 6 on tehty haku sanalla “symbolist” aineistoon, jossa on muun muassa Getty-säätiön taiteilijasanasto Union List of Artists Names (ULAN)<sup>3</sup> muunnettuna RDF-muotoon. Hakutulokset esitetään aikajanan avulla siten, että eri tyylikausia ja taitelijoiden elinaikoja visualisoidaan viivalla vuosilukujen välillä, ja taiteilijoiden töitä pisteinä aikajanalla. Pisteet toimivat linkkeinä tarkempaan tietoon maalauksista.

---

<sup>3</sup>[http://www.getty.edu/research/conducting\\_research/vocabularies/ulan/](http://www.getty.edu/research/conducting_research/vocabularies/ulan/)





Kuva 6: MultimediaN E-Culture Demonstrator -sovelluksen aikajanavisualisointi.

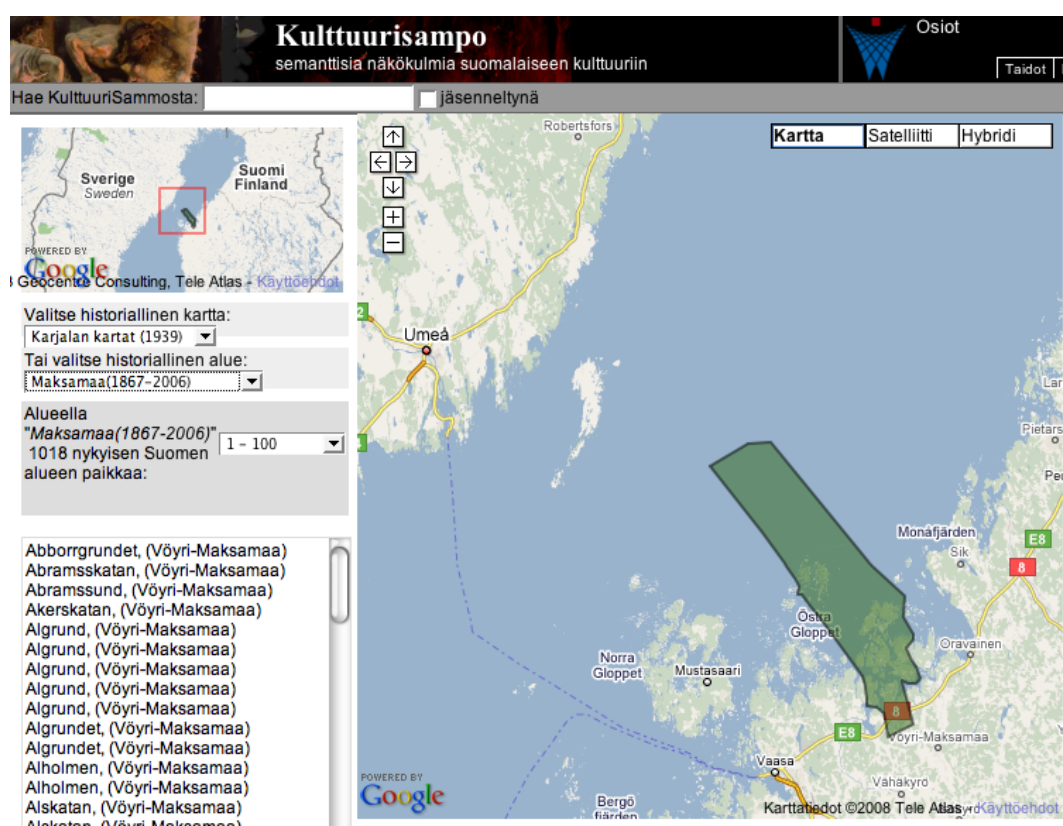
## Paikka-aineiston visualisointi kartalla

Paikkatietoa sisältävät ontologiat voidaan visualisoida on kartalla. Esimerkiksi suomalainen paikkaontologia SUO [KHS<sup>+</sup>08] koostuu topologisista, geometrisista sekä osa-kokonaisuus-suhteista. Siinä on yli 800 ontologista luokkaa maantieteellisistä paikoista, kuten kaupungeista, lääneistä sekä joista. Ontologia sisältää lisäksi noin 800 000 paikkainstanssia Suomesta. Kuvassa 7 on ONKI-Paikka-sovelluksessa [KHS<sup>+</sup>08] tehty haku sanalla “Helsinki”. Tuloksista on valittu Helsinki, Taivassalossa oleva asuinalue. Se on visualisoitu kartalle sinisellä ympyrällä.

Kuva 7: SUO-ontologian visualisointi kartalla.

Karttavisualisoinnin voi tehdä myös tarkemmalla tasolla. KulttuuriSampo-portaalissa [HRH<sup>+</sup>08] on visualisoitu historiallisia kuntia niiden kattaman alueen mukaan, kuten kuvassa 8 Maksamaan kunnan alue vuosina 1867–2006.

OntoWiki [ADR06] on web-pohjainen ontologiaeditori ja tietämyksenhallintajärjestelmä. Siinä on erilaisia visualisointimenetelmiä erityyppiselle datalle. Jos valittu käsite esimerkiksi sisältää ominaisuuksia, joiden arvona on koordinaattitietoa, se visualisoidaan kartalle. Toisaalta myös kartalla olevat koordinaattipisteet voidaan laajentaa ja näyttää instanssiin liittyvät tiedot kartalla.



Kuva 8: SAPO-ontologian kunnan Maksamaa visualisointi kartalla.

## 2.2 Moniontologiavisualisointi

Moniontologiavisualisointia (multi-ontology visualization) eli useamman toisiinsa linkittyvän ontologian visualisointia ei ole juurikaan tutkittu, vaan moniontologiaympäristöihin liittyvä tutkimus on kohdistunut useampien ontologioiden yhdistämiseen ja yhteiskäyttöön [DMD<sup>+</sup>03, GPS04]. Tässä kappaleessa esitellään joitakin semanttisen webin sovelluksia, joissa on käytössä useampia ontologioita, ja eritellään miten visualisointi on niissä toteutettu.

Onki-ontologiaselaimessa [VTH08] voidaan esittää useampia ontologioita samanaikaisesti. Sen visualisointimenetelmä on hierarkiaselain. Kuvassa 9 ONKI-selaimessa on yhdistettynä Yleinen suomalainen ontologia YSO, Museoalan ontologia MAO<sup>4</sup> sekä Taideteollisuusalan ontologia TAO<sup>5</sup>. Kuvassa näkyy hierarkiapuu käsitteestä "yso:rakennuspuutavara" ylöspäin sekä lähimmät alakäsitteet. Eri ontologioihin kuuluvat käsitteet erotetaan toisistaan ontologian etuliitteen perusteella, kuten yso:hirret(mao, tao). Tällöin kaikissa kolmessa ontologiassa on käsite "hirret".

Protege-ontologiaeditori tukee moniontologiakehitystä, eli siihen voidaan ladata useampia ontologioita samanaikaisesti ja niiden käsitteiden välille voidaan luoda suhteita. Protegen luokkahierarkiassa, joka on siis tavanomainen puunäkymä, eri ontologioihin kuuluvia käsitteitä ei eroteta toisistaan visuaalisin keinoin. Mikäli ontologiat ovat eri nimiavaruuksissa, näkyy käsittehierarkiassa käsitteiden edessä niiden nimiavaruuden tunniste. Kuvassa 10a on osa YSO:n ja TAO:n luokkahierarkiasta. Siinä TAO on ripustettu YSOon, siten että TAO:n luokat täydentävät YSO:n rakennetta. TAO:n luokat visualisoidaan puuhierarkiassa YSO-yläluokan alle ja TAO-käsitteet erottaa YSO-käsitteistä nimiavaruuden tunniste.

Protegen visualisointityökalu OwlViz [Hor] toimii eri ontologioiden käsitteiden erotelun suhteen samalla tavoin. Kuvassa 10b on visualisaatio kolmen ontologian yh-

---

<sup>4</sup><http://www.seco.tkk.fi/ontologies/mao>

<sup>5</sup><http://www.seco.tkk.fi/ontologies/tao>

# yso:rakennuspuutavara

URI: <http://www.yso.fi/onto/yso#p1739>



Kuva 9: Monta ontologiaa ONKI-ontologiaselaimessa.

distelmästä. Erivärisinä näkyvät solmut eivät visualisoi eri ontologioita, vaan Owl-Viz:ssä väreillä esitetään solmujen ekvivalenttisuutta, toisin sanoen oranssi solmu on sama jonkin muun solmun kanssa, keltainen ei ole.

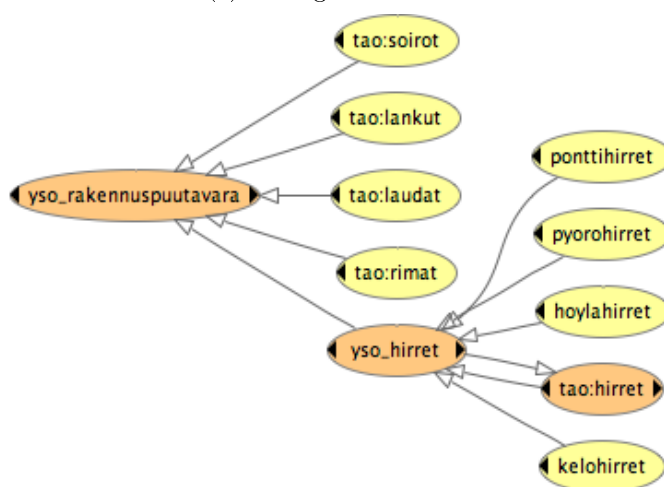
Useissa sovelluksissa, joissa on käytössä monta ontologiaa, niitä käsitellään erikseen. Esimerkiksi Swoop-ontologiaeditorissa [KPS<sup>+</sup>06] voidaan käsitellä useaa ontologiaa samassa näkymässä, mutta niiden hiarkiapuut ovat erilliset. Samalla tavoin moniontologiavisualisointi on järjestetty MIA-ontologiaeditorissa [LAWB03] .

## 2.3 Käsitevalitsimet

Semanttisen webin sovelluksissa käsitevalinta on toteutettu pääasiassa kahdella tavoin: luokkahierarkiaa esittävällä puuselaimella tai automaattisella täydennyksellä. Seuraavaksi esitellään joitakin käsitevalitsimia ja niiden ominaisuuksia.



(a) Protegen luokkaselain



(b) Protegen OwlViz

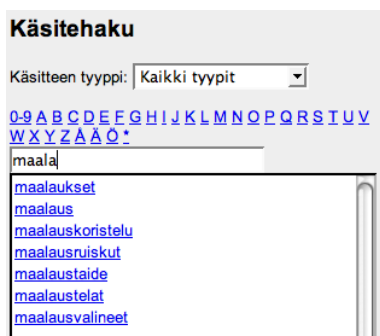
Kuva 10: Protegen moniontologiavisualisointi.

### 2.3.1 Automaattitäydennys

*Automaattitäydennyksellä* (autocompletion) tarkoitetaan menetelmää, jossa käyttäjä syöttää syötekenttään sanan alun ja järjestelmä täydentää automaattisesti siihen täsmäivät kokonaiset sanat käyttäjän kirjoittaessa [HM06]. Jos syötteeseen sopii useita täydennyksiä, useat käyttöliittymät listaavat kaikki mahdolliset vaihtoehdot [HvOA<sup>+</sup>07]. Tällöin käyttäjä voi joko valita jonkin tarjotuista vaihtoehdoista tai jatkaa kirjoittamista rajatakseen tulosjoukkoa. Menetelmä on käytössä useiden

sähköpostiohjelmien ja web-selainten osoitekentässä sekä ohjelmointieditoreissa. Esimerkiksi Applen Mail-ohjelmassa<sup>6</sup> lähetettävän viestin osoitekenttään täydennetään automaattisesti syötteeseen sopivat sähköpostiosoitteet käyttäjän osoitekirjasta sekä aiemmin lähetettyjen viestien vastaanottajista. Semanttisessa automaattitäydennyksessä syöte täydennetään yhteensopiviin ontologian käsitteisiin. Toisin kuin tavallisessa automaattitäydennyksessä syötteeseen voidaan palauttaa termejä, jotka eivät kirjoitetussa muodossa vastaa syötettä. Esimerkiksi syötteeseen “Presiden...” voidaan täydentää “Tarja Halonen”, koska Tarja Halonen on luokan presidentti ilmentymä [HM06].

Semanttisen webin sovelluksissa käsitteenvalintamenetelmänä automaattitäydennystä käytetään muun muassa KulttuuriSampo-portaalin [HRH<sup>+</sup>08] henkilöhaussa, jossa käyttäjän syötettä täydennetään TOIMO-toimijaontologian [KH07] henkilöinsansseilla. ONKI-ontologiaselaimessa [VTH08] käsitteitä voidaan hakea eri ontologioista, kuvan 11 käyttötapauksessa käsitteet haetaan Museoalan ontologia MAOsta<sup>7</sup>. Käyttäjä on kuvan tilanteessa syöttänyt merkkijonon “maala”, johon täsmäävät ontologian käsitteet tulostetaan syötekentän alle.



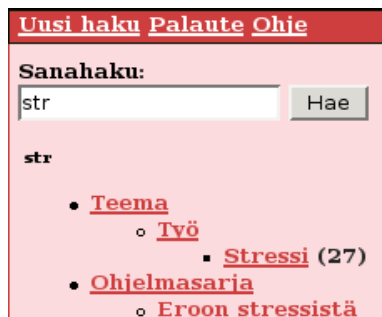
Kuva 11: ONKI-ontologiaselaimen automaattitäydentävä haku.

Orava-portaalissa [KH06] aineistona olevaa opetusohjelmien video- ja äänileikkeitä voidaan hakea automaattitäydentävällä *näkymäpohjaisella haululla* (view-based search)

<sup>6</sup><http://www.apple.com/macosx/features/mail.html>

<sup>7</sup><http://www.seco.tkk.fi/ontologies/mao/?l=fi>

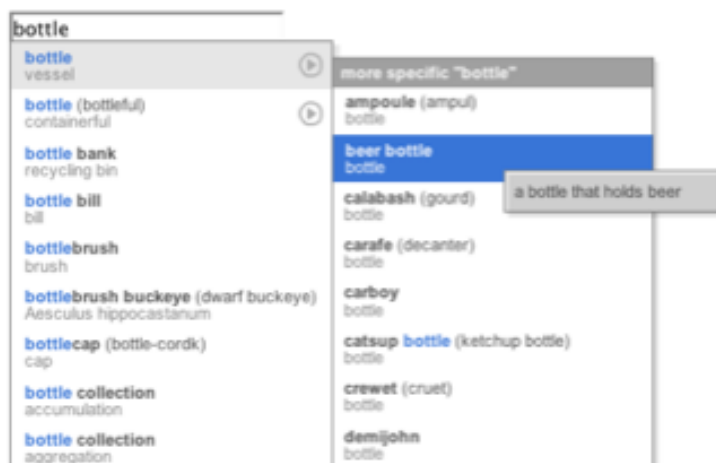
[YSLH03]. Oravassa näkymät muodostuvat käsittehierarkioista. Automaattitäydennys tulostaa syötettä vastaavat ontologian käsitteet näytölle osaksi käsittehierarkiaa. Kuvassa 12 käyttäjä on syöttänyt merkkijonon “str”. Aineistosta on löytynyt siihen täsmäävät käsitteet “stressi” ja “Eroon stressistä”, jotka esitetään osana käsittehierarkiaa.



Kuva 12: Orava-portaalin automaattitäydentävä haku.

Hildebrandin [HvOA<sup>+</sup>07] automaattitäydennyskomponenttissa tavallisen sanastosta täydentämisen lisäksi tulokset ryhmitellään ontologisen tiedon perusteella. Esimerkiksi haettaessa Berliiniä, tulokset esitetään ryhminä maittain. Näin yksi ryhmä muodostuu Saksan sisällä olevista Berliiniin täsmäävistä käsitteistä kuten Berliinin kaupunki ja kaupunginosa Bernau bei Berlin ja toinen ryhmä Etelä-Afrikassa sijaitsevasta Berlin-nimisestä asutusalueesta. Toisena esitystapana Hildebrandin järjestelmässä tuloksille on esittää kaikki täsmäävät käsitteet listattuna ja näyttää niistä käyttäjän pyytäessä WordNet-luokittelun<sup>8</sup> mukaiset alakäsitteet. Kuvassa 13 käyttäjän syötteeseen “bottle” on täydennetty merkkijonon sisältävät käsitteet. Oikeanpuoleisessa listauksessa on näytetty valitun “bottle”-käsitteen alakäsitteet WordNetistä.

<sup>8</sup><http://wordnet.princeton.edu/>



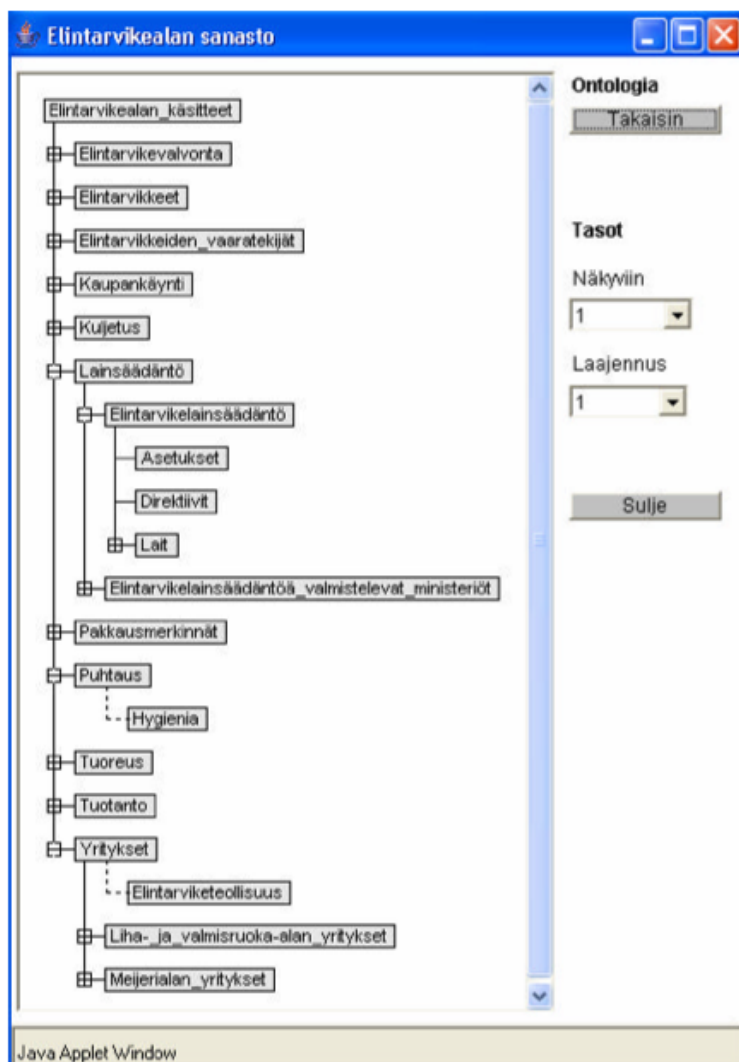
Kuva 13: Hildebrandin autocomplete-komponentti.

### 2.3.2 Puuselaimet

Erityisesti varhaisemmissa semanttisen webin sovelluksissa käsitevalintaan käytettiin *puuselainta* [AJS<sup>+</sup>04]. Puuselaimella tarkoitetaan sovellusta, joka esittää ontologian tai muun hierarkkisen aineiston puuna, jota voi selata laajentamalla solmuja ja päästä näin etenemään hierarkiassa. Kyseessä on samanlainen komponentti, kuin useiden käyttöjärjestelmien graafisissa tiedostoselaimissa. Puuselaimet ovat perinteisin ja yleisin [KHL<sup>+</sup>07] tapa valita käsitteitä semanttisen webin sovelluksissa. Käsitevalitsimena puuselain toimii siten, että käyttäjä etsii haluamansa käsitteen hierarkiasta selaamalla. Joissakin käsitevalitsimissa on manuaalisen selailun lisäksi hakukomponentti, joka avaa hierarkiasta halutun termin kohdan.

Tampereen yliopistossa tehty CIRI, ontologioihin perustuva semanttinen hakuliittymä tiedonhakuun [AJS<sup>+</sup>04], tekee kyselyn dokumenttijoukkoon laajentamalla käyttäjän valitsemia ontologian käsitteitä. Käsitteet valitaan erilliseen käyttöliittymän ikkunaan aukevalla puuvalikolla. Käyttäjä voi laajentaa ontologiapuuta ja hiiren avulla valita haluamansa termit tehtävään tiedonhakuun. Kuvassa 14 esitetään Elin-tarvikealan sanaston hierarkiaa CIRI:ssä.





Kuva 14: CIRI-hakuliittymän käsitevalinta.

Protege<sup>9</sup> [GMF<sup>+</sup>03] on ontologioiden editointiin tarkoitettu sovellus, jolla voi myös visualisoida ontologioita. Protege perustuu avoimeen lähdekoodiin ja sille on mahdollista luoda omia laajennoksia sekä lisäosia. Protege tukee muun muassa OWL- ja RDF-kieliä. Protegen käyttöliittymässä ontologian luokkahierarkia esitetään puuselaimessa. Luokka valitaan editoitavaksi navigoimalla puussa sen kohdalle. Protegessa on myös sanahaku, jolla käsitteet voidaan kohdentaa puusta.

ONKI-ontologiaselaimessa [VTH08] käsitteen valinta toteutetaan ensisijaisesti au-

<sup>9</sup><http://protege.stanford.edu/>

tomaattitäydennyksen avulla (katso luku 2.3.1). Käsitteen valinnan tuloksena esitetään käsittehierarkia käsitepuun juuren asti, sekä mahdollisia ala- ja rinnakkaisluokkia. Käsitteen ontologinen konteksti siis visualisoidaan puuselaimen avulla. Puuselaimin toimii myös käsitevalitsimena, sillä sen kautta voidaan tehdä uusia käsitevalintoja.

### 3 Käsitevalitsimien arviointi

Edellisessä luvussa esiteltiin tapoja visualisoida ontologioita sekä tapoja, joilla käsitteen valinta on toteutettu semanttisen webin sovelluksissa. Tässä luvussa arvioidaan käsitteen valinnan tapoja ja valintajärjestelmiä.

#### 3.1 Automaattitäydentävät valitsimet

Automaattitäydennysvalitsimien etuna on käsitteen haun nopeus. Käyttäjän ei tarvitse kahlata läpi puurakennetta, vaan haluttu käsite palautetaan joidenkin merkien kirjoittamisen jälkeen. Tämä lähestymistapa toimii kuitenkin vain siinä tapauksessa, että käyttäjä tuntee käytettävän sanaston. Tuntemattoman sanaston kohdalla käsitteiden haku muuttuu arvailuksi, sillä käyttäjän täytyy kokeilemalla selvittää, mikä tai mitkä käsitteet sanastosta löytyvät. Erityisesti niiden käsitteiden kohdalla, joita voidaan ilmaista usealla synonyymisellä sanalla, joutuu käyttäjä arvaamaan, mikä käsite löytyy sanastosta. Vaikka käyttäjä löytäisikin etsimänsä käsitteen, ei hän voi olla varma, onko käsite paras mahdollinen sanastosta löytyvä, vai olisiko sille jokin tarkempi tai kuvaavampi termi. Edellisen luvun kuvan 11 tilanteessa, jossa käyttäjä on kirjoittanut automaattitäydennysvalitsimen kenttään “maala”, hän saattaisi valita annotaatiokäsitteeksi “maalaustaide”. Käyttäjältä jää tässä tapauksessa saamatta tieto, että ontologiassa “maalaustaiteella” onkin alakäsitteenä muun

muassa “öljymaalaukset”, joka saattaisikin olla tarkempi termi kuvaamaan annotoitavaa kohdetta.

Toinen ongelma automaattitäydennyksellä toteutetussa käsitevalitsimessa syntyy tilanteessa, jossa käsitteen merkitys ei ole pääteltävissä pelkästään sen nimestä tai ominaisuuksista. Taidealan Iconclass-ontologiassa<sup>10</sup> merkitys muodostuu koko käsitehierarkian perusteella. Esimerkiksi Iconclassin luokan “siivekkäät mielikuvitus-, fantasia- tai satuolennot” yläkäsitteinä ovat muun muassa luokat “hirviöt ja sekasikiöt” sekä “vammat, epämuodostumat, luonnottomuudet; sairaudet”. Näin ollen “siivekäs satuolento” saakin varsin negatiivisen sävyn. Jos käsitehierarkian tietoa ei olisi saatavilla, olisi luokkaa helppo käyttää väärin.

Käsitteen konteksti-informaatio voi myös olla tarpeen monimerkityksisten käsitteiden semantiikan yksikäsitteistämiseksi. Suomalainen paikkaontologia SUO<sup>11</sup> sisältää 491 “Isosaarta”. Nämä ovat instansseja muun muassa luokista “saari”, “metsä” ja “asutusalue”. Jotta tiettyyn Isosaareen voitaisiin viitata, tarvitaan tietoa käsitteen tyypistä.

Automaattitäydentävät käsitevalitsimet jättävät hyödyntämättä käsitteisiin liittyvät ominaisuudet. Ontologioissa on mahdollista määrittää suhteita sellaisten käsitteiden välille, jotka eivät ole samassa hierarkiapolussa keskenään. Ontologisten käsitteiden ominaisuuksilla (property) voidaan tehdä tällaisia linkkejä käsitteiden välille. Esimerkiksi Iconclass-ontologiassa käsitteelle “ihminen ja villi eläin” on määritetty siihen liittyväksi käsitteeksi “metsästys”. Autocompletion-valitsimet eivät hyödynnä tällaisten suhteiden mahdollisuutta opastaa käyttäjää eteenpäin ontologiassa.

Annotoinnissa voidaan käyttää useita linkitettyjä ontologioita. Osa näistä voi toimia vain apuna ja opastamassa monimutkaisemman tai vieraamman ontologian käsitteistöön. Tällaisilla apukäsitteistöillä ei ole tarkoitus annotoida, eikä niiden kä-

---

<sup>10</sup><http://www.iconclass.nl/>

<sup>11</sup><http://www.seco.tkk.fi/ontologies/suo/>

sitteitä siksi voida valita. Iconclass-ontologiaan liittyy erillisiä asiasanoja, jotka eivät ole ontologian luokkia, eikä niitä näin ollen voi valita indeksointitermeiksi. Ne ovat merkitykseltään yleisiä ja niiden tarkoitus on toimia viitteinä ontologian luokkiin. Esimerkiksi asiasanaan “ihminen” liittyy yli kaksikymmentä luokkaa aihealueilta “ihminen yhteiskunnassa”, “ihmiselämä” ja “ulkoavaruuden oliot”. Tavallisella automaattitäydennys-käsitevalitsimella asiasanaa “ihminen” ei voida näyttää eikä hyödyntää, sillä se ei ole ontologian luokka eikä kelvollinen esimerkiksi indeksointiin. Hildebrand [HvOA<sup>+</sup>07] pyrkii parantamaan automaattitäydennys-käsitevalitsimiin liittyviä puutteita näyttämällä järjestelmässään WordNet:in hierarkian alaluokkia hakuun täsmäävien käsitteiden yhteydessä (ks. luku 2.3.1). Sovelluksessa myös ryhmitellään hakutuloksia luokittain. Hildebrandin järjestelmän huono puoli on, ettei sitä voida laajentaa eikä konfiguroida ennalta määräämättömälle ontologialle. Myös käsitteiden visualisointi on kiinnitetty, eikä esimerkiksi paikka-aineiston visualisointi kartalla ole mahdollista.

Tiivistetysti automaattista täydennystä käyttävien valitsimien etuna on nopeus, mutta toisaalta ne tarjoavat suppean näkökulman ontologian sisältöön. Erityisesti entuudestaan tuntemattoman ontologian käyttö autocompletion-valitsimella muodostuu haastavaksi, eikä ontologian sisällöstä pääse muodostumaan käsitystä.

## 3.2 Puuselaimet

Automaattitäydentäviin käsitevalitsimiin verrattuna luokkahierarkian selailun mahdollistavat puuselaimet mahdollistavat ontologiaan tutustumisen ja ovat toimiva ratkaisu käsitevalitsimeksi silloinkin, kun käyttäjä ei tunne ontologian sisältöä. Käsitteen merkityksen arviointi helpottuu, sillä käsitteen sijainti hierarkiassa on näkyvillä. Käsitteistä saattaa olla näkyvillä myös joitakin ominaisuuksia, joiden avulla käyttäjä pääsee navigoimaan ontologiassa tarkempiin käsitteisiin.

Puuselainten heikkous on niiden käytön hitaus. Hierarkiaa voi joutua avaamaan syvällekin ennen kuin haluttuun käsitteeseen päästään. Tilanteissa, joissa käyttäjä ei tunne ontologian rakennetta, voi käsitteen löytyminen hierarkiasta viedä paljonkin aikaa. Ongelmia aiheutuu myös silloin, jos selattavassa ontologiassa on moniperintää, eli käsite voi esiintyä hierarkiassa useissa haaroissa. Näin on esimerkiksi lääketieteen ontologia MeSH:ssä<sup>12</sup>, jossa käsitteen “astma” suorina yläluokkina on “keuhkoputkien sairaudet”, “hengitysteiden yliherkkyys” ja “ahtauttavat keuhkosairaudet”. Osalla näistä yläluokista on myös useampia yläluokkia. Tällaisia suhteita on hankala ilmaista puuselaimella.

### 3.3 Yhteenveto käsitevalitsimista

Edellisen luvun katsaus osoittaa, että ontologioita voidaan visualisoida monella tavalla. Käsitevalitsimiin on toisaalta kiinnitetty tietyntyypinen visualisointitapa, puuselain tai automaattitäydenyksen tuottama listaus. Puuselaimilla voidaan saada lisätietoa käsitteen merkityksestä käsitehierarkian perusteella, mutta käsitteen merkityksen selvittämiseksi muunlainenkin konteksti-informaatio voi olla tarpeen. Paikkaontologian käsitteille esimerkiksi voi olla hyödyllistä esittää kontekstitietona maantieteelliset naapurialueet ja läänit [KHS<sup>+</sup>08] tai historiaontologian tapahtumiluvut ja tapahtumien kestot.

Esitellyt käsitevalitsimet ovat rajoittuneita eikä niillä voida esittää tämäntyyppistä kontekstitietoa. Erityyppisille ontologioille voidaan rakentaa käsitevalitsin tukemaan tietyntyypisen aineiston esittämistä, kuten esimerkiksi karttavisualisointia hyödynnettävässä ONKI-paikassa on tehty [KHS<sup>+</sup>08]. Toisaalta käsitevalitsimen rakentaminen aineistosidonnaisesti rajoittaa mahdollisuuksia sen uudelleenkäyttöön.

Jotta annotaation laatu olisi mahdollisimman hyvä, on tarpeen löytää kuvaavimmat

---

<sup>12</sup><http://www.nlm.nih.gov/mesh/>

termit annotaatioon. Tämän vuoksi on edullista näyttää käsittehierarkia ja muita käsitteeseen liittyviä termejä. Myös hakutuloksista saadaan parhaimmat, kun haku tehdään kuvaavimmalla käsitteellä. Oman hankaluuden muodostavat tilanteet, joissa annotaatioon on käytettävissä useita sanastoja. Tällöin annotoija joutuu etsimään termejä erikseen eri ontologiaselaimista. Prosessia voidaan tehostaa liittämällä käytettävät ontologiat yhteen käsitevalitsimeen.

Jotta käsitevalitsimella voitaisiin auttaa annotoijaa valitsemaan tarkin käsite, tarvitaan käsitevalitsin, joka

- on nopea käyttää
- mahdollistaa usean ontologian samanaikaisen visualisoinnin
- näyttää käsitteen ontologisen kontekstin
- on monikielinen
- on konfiguroitavissa erilaisille aineistoille
- mahdollistaa erilaisia visualisointitapoja

Näiden vaatimusten perusteella suunniteltiin ja toteutettiin käsitevalitsin IRMA [SMKH08].

## 4 IRMA kontekstipohjainen käsitevalitsin

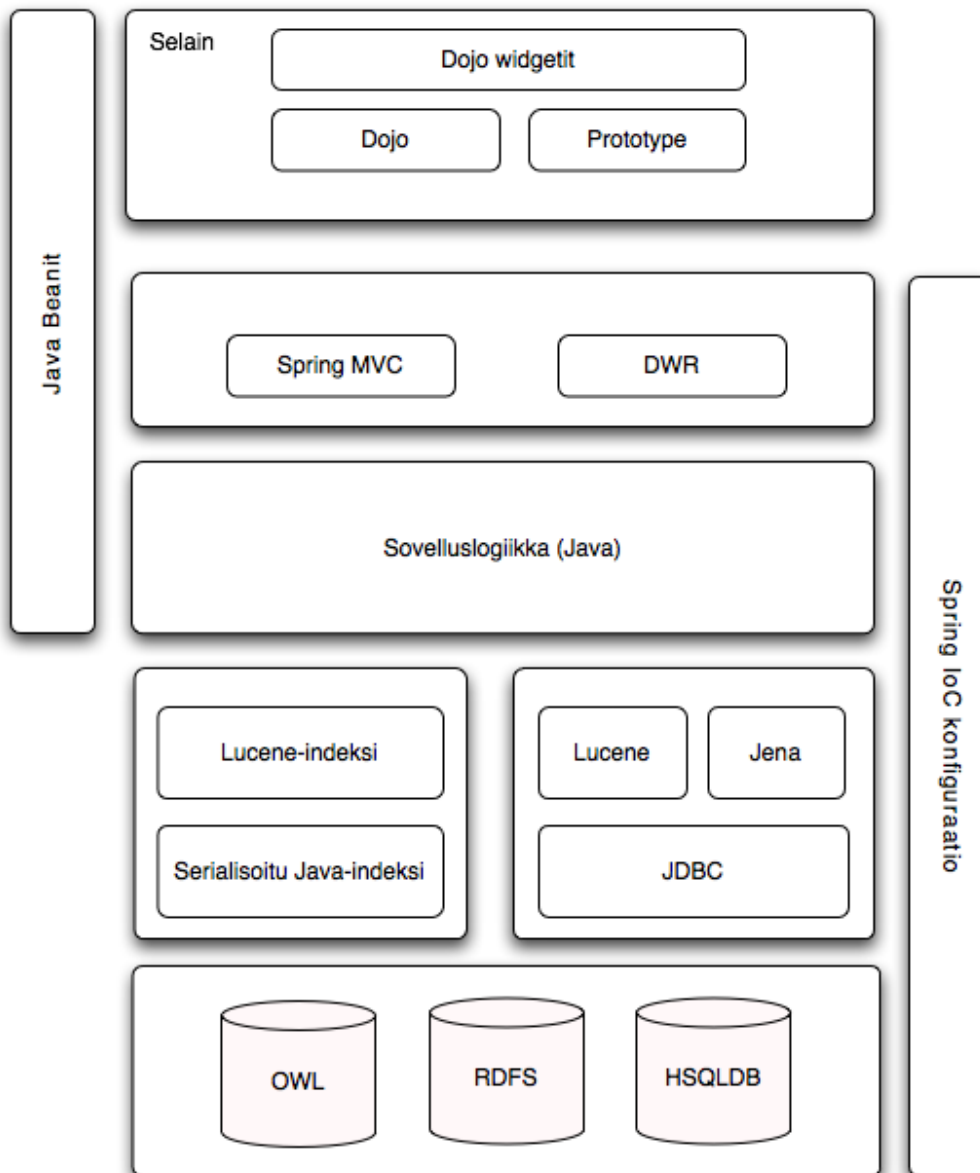
IRMA on FinnONTO-projektissa<sup>13</sup> kehitetty selainpohjainen käsitteenvälintasovellus. Se on suunniteltu mahdollisimman yleiseksi ja moneen käyttökontekstiin muokattavaksi. Monikäyttöisyys tarkoittaa tässä yhteydessä mahdollisuutta käyttää sovelluksessa eri ontologioita sekä niiden yhdistelmiä ja toisaalta mahdollisuutta valita,

---

<sup>13</sup><http://www.seco.tkk.fi/projects/finnonto/>

millaisia kontekstielementtejä käsitteille näytetään. Valmista visualisointiratkaisua ei haluttu kiinnittää, vaan kussakin käyttökontekstissa sovellukseen on mahdollista luoda sille soveltuva visualisointimenetelmä. Järjestelmä perustuu automaattitäydennykseen ja sen tuloksia tarkentavaan hierarkkiseen kontekstinäkymään.

Jotta sovellus olisi mahdollisimman yleinen, päätettiin tehdä siitä modulaarinen. Modulaarisuutta tukee *MVC-suunnittelumalli* (model-view-controller) [KP88], jossa erotetaan toisistaan malli (model), näkymä (view) sekä ohjain (controller). Malli vastaa järjestelmän tietosisällön tallentamisesta ja käsittelystä, näkymä määrittää käyttöliittymän ulkoasun ja ohjain vastaa käyttäjältä tulevat käskyt ja muuttaa mallia ja näkymää niitä vastaavaksi. Koska osiot toteutetaan erillisinä toisistaan, voidaan osia muuttaa tai vaihtaa tarvitsematta muuttaa koko ohjelman toimintaa. Näin esimerkiksi voidaan ontologioille tehdä erilaisia käyttöliittymiä ohjelman toimintalogiikan säilyessä samana. Myös mallia voidaan muuttaa ja säilyttää näkymäntoiminnallaan. Näin ohjelmaa voidaan käyttää eri ontologioilla vain pienin muutoksin. Ohjelma toteutettiin selainpohjaisena Java-sovelluksena, joka on helposti laajennettavissa. Sen ydin on Javalla toteutettu palvelu, jolle rekisteröidään konfiguraation avulla joukko kontekstintuottajia. Nämä ovat ohjelmaluokkia, joista kukin vastaa tietyn tyyppisen konteksti-informaation tuotosta ontologian käsitteille. Ohjain vastaa käyttäjän pyyntöihin tiedustelemalla kontekstintuottajilta, onko näillä vastaustajuuri tämän tyyppiseen pyyntöön. Lopuksi se kokoaa kontekstintuottajien vastaukset ja välittää informaation bean-olioina AJAX:n avulla selaimen käyttöliittymälle. Käyttöliittymässä palvelinpuolen kontekstintuottajia vastaavat JavaScript-moduulit asettelevat ne näytölle. Ohjelman yleinen arkkitehtuuriratkaisu on esitelty kuvassa 15. Ohjelman tekniset ratkaisut käydään seuraavaksi läpi MVC-mallin mukaisessa järjestyksessä.



Kuva 15: Ohjelman yleisarkkitehtuuri.



## 4.1 Malli

### Tietovarastot

Ohjelmaan voidaan ladata mikä tahansa OWL- tai RDF-muotoinen ontologia tai ontologioiden yhdistelmä. Ontologiaa voidaan käyttää joko tekstitiedostosta käsin tai se voidaan ladata tietokantaan. Mahdollisuus tietokannan käyttöön lisättiin ohjelmaan, jotta suuriakin ontologiayhdistelmiä voitaisiin käyttää sujuvasti. Ilman tietokantaa esimerkiksi noin 50 megatavun kokoinen Iconclass-YSO-yhdistelmäontologia on melko raskas käsiteltäväksi muistissa. Tietokantana ohjelmassa on kevyt Java-perustainen HSQLDB<sup>14</sup>. HSQLDB on relaatiotietokanta, joka on yhteensopiva Jenan<sup>15</sup> kanssa. Tämä tarkoittaa sitä, että sinne voidaan tallettaa aineistoa RDF-muodossa. Jenassa itsessään on ohjelmalliset metodit datan tallettamiseksi kantaan. Jena-tietokanta toimii muuten kuten muutkin ohjelmallisesti Javasta käytettävät tietokannat.

### Indeksit

Ohjelmassa käytetään kahdentyyppisiä indeksejä: automaattitäydennyksen haussa Lucene-indeksiä<sup>16</sup> [HG04] ja käsitteiden kontekstitietojen tallettamisessa serialisoituja Java-olioita. Serialisoituja kontekstitietoja käsitellään tarkemmin seuraavassa kappaleessa.

Lucene on Apache-projektin<sup>17</sup> toteuttama tiedonhakumenetelmiin tarkoitettu ohjelmakirjasto, josta on toteutus muun muassa Java-kielellä. Lucene sisältää menetelmiä indeksointiin ja tiedonhakuun. Sen avulla voidaan indeksoida eri formaateissa olevaa tekstiä, mukaan lukien HTML ja PDF. Lucene valittiin indeksointiin sen tehok-

---

<sup>14</sup><http://hsqldb.org/>

<sup>15</sup><http://jena.sourceforge.net/>

<sup>16</sup><http://lucene.apache.org/>

<sup>17</sup><http://www.apache.org/>

kuuden vuoksi [luc]. Ontologian käsitteet nimikkeineen (label) on talletettu Lucene-indeksiin. Kun käyttäjä syöttää automaattitäydennyskenttään merkkijonon, ohjelma tekee kutsun palvelimelle. Palvelinpuolelta palautuu Lucene-indeksin syötettä vastaavat ontologian käsitteet URI-tunnisteineen takaisin käyttöliittymään Java Bean-olioina<sup>18</sup>.

## Kontekstintuottajat

Ohjelman keskusluokalle rekisteröidään konfiguraation avulla palveluita, jotka vastaavat käsitteiden konteksti-informaation tuottamisesta. Näitä kontekstintuottajia luodaan kullekin ontologialle sen mukaan, minkälaista informaatiota sen käsitteille halutaan tarjota. Esimerkiksi lääketieteen ontologiassa (kappale 5.2) on erilliset kontekstintuottajat laajemmille, suppeammille sekä liittyville käsitteille, Iconclass-ontologiassa on indeksit myös kieliversioiden mukaan erikseen sekä englanniksi että suomeksi. Ohjelman suoritusnopeuden parantamiseksi käsitteen kontekstina näytettävä informaatio tuotetaan valmiiksi. Tätä menetelmää kutsutaan tässä yhteydessä indeksoinniksi ja serialisoituja konteksti-tiedostoja indekseiksi. Indeksoinnissa kontekstintuottajat käyvät läpi ontologian käsitteet niille Spring-konfiguraation [Joh02] avulla välitetystä Jena-mallista ja poimivat käsitteen konteksti-informaation tarvittavat ominaisuudet. Indeksit serialisoidaan Java-olioina tiedostoiksi ja ohjelmaa käynnistettäessä luetaan muistiin. Indeksoinnista vastaavat kontekstintuottajat ja indeksointi voidaan suorittaa esimerkiksi aineiston muuttuessa.

Kontekstintuottajien toinen tehtävä on palauttaa käyttöliittymään näitä kontekstintietoja. Ohjelman keskusluokka lähettää käyttäjän syötteen mukaisen pyynnön kaikille sille rekisteröidyille kontekstintuottajille. Kontekstintuottajat ovat Java-luokkia ja toteuttavat yhteisen rajapinnan. Niille on konfiguraatiossa määritelty, minkä tyyppisten käsitteiden kontekstista ne vastaavat. Määrittäminen voidaan tehdä muun muassa

---

<sup>18</sup><http://java.sun.com/javase/technologies/desktop/javabeans/index.jsp>

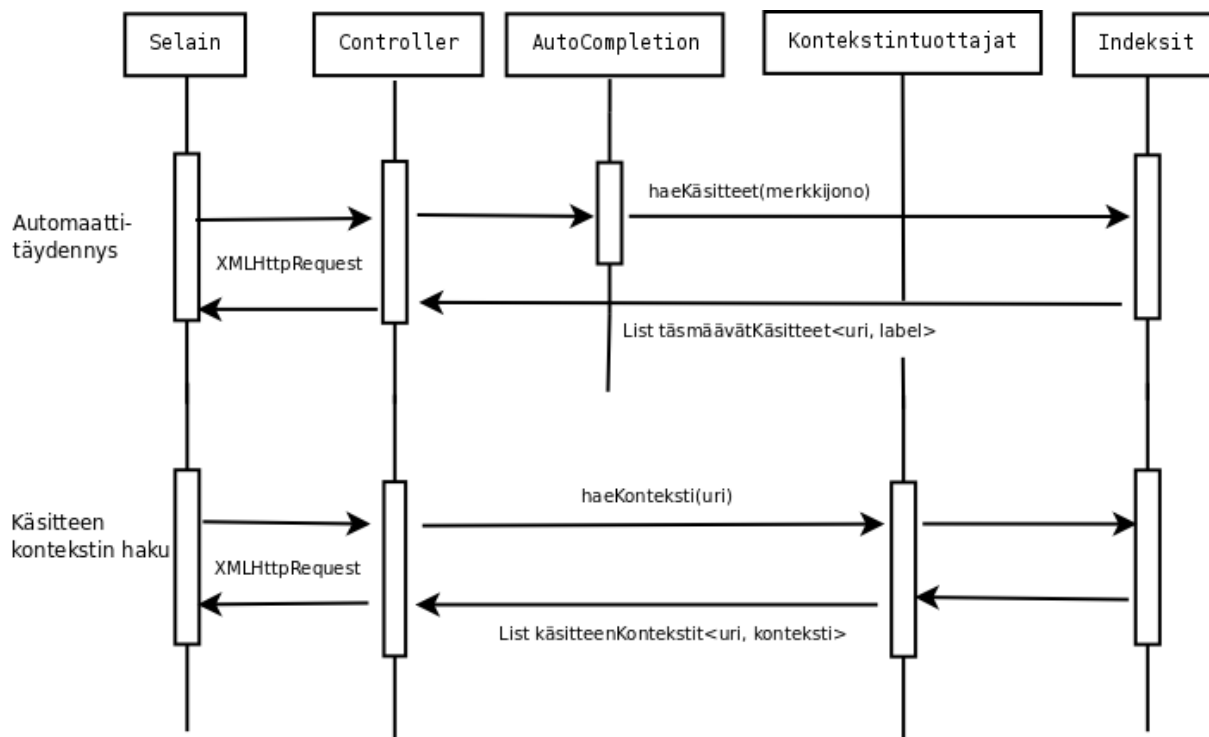
nimiavaruuden tai RDF-tyypin perusteella. Jos kontekstintuottaja vastaa käyttäjän syöttämän käsitteen kontekstista, ja kontekstitietoa on käsittelle olemassa, se palauttaa serialisoidusta indeksista tiedon keskusluokalle eteenpäin prosessoitavaksi.

## 4.2 Ohjain

Ohjelman ydin ja sen suorituksen järjestystä kontrolloiva osa on Javalla toteutettu keskusluokka. Kaikki käyttäjän syötteet ohjautuvat sille ja se välittää tarvittavat kutsut eteenpäin. Keskusluokka itsessään on hyvin yksinkertainen, sen toiminnan ohjaus tehdään Spring-konfiguraatiosta käsin. Keskusluokalle tulee kolmenlaisia pyyntöjä. Kun käyttäjä valitsee käsitteen kontekstin avattavaksi tai vaihtaa ohjelmassa kieltä, välittyy tästä ensimmäisenä tieto keskusluokalle. Keskusluokka välittää tiedon eteenpäin sille rekisteröidyille kontekstintuottajille. Käyttöliittymään palaava tieto kulkee siis myös keskusluokan kautta. Ohjelman toimintakaavio esitetään kuvassa 16. Liikenne käyttöliittymän ja ohjaimen välillä tapahtuu AJAX:in [Gar05] avulla. AJAX ei ole standardoitu itsenäinen teknologia, vaan joukko tekniikoita (HTML, CSS, DOM, JavaScript, XML) yhdessä käytettynä. AJAX mahdollistaa kutsujen lähettämisen käyttöliittymästä palvelimelle siten, ettei koko verkkosivua tarvitse ladata uudelleen. AJAX perustuu XMLHttpRequest-objektiin [vK07], jonka avulla dataa voidaan vaihtaa palvelimen kanssa epäsynkronisesti. Tiedonvälityksessä käytetään apuna DWR-ohjelmakirjastoa<sup>19</sup>. Se mahdollistaa palvelimella ajettavien Java-sovellusten käytön suoraan selaimesta. DWR perustuu siihen, että se luo dynaamisesti JavaScript-tiedostot Java-luokkien pohjalta. JavaScriptin avulla selaimesta voidaan käyttää palvelimen Java-sovelluksia samalla tavoin kuin niitä käytettäisiin suoraan ohjelmassa.

---

<sup>19</sup><http://getahead.org/dwr/>



Kuva 16: Ohjelman suoritus.

## Konfiguraatio

Ohjelman konfigurointi on toteutettu Spring-sovelluskehysellä<sup>20</sup> (framework) [Joh02]. Spring on yleiskäyttöinen Java-sovelluskehys, joka mahdollistaa Inversion of Control (IoC) -suunnittelumallin toteuttamisen. IoC-mallissa perusajatus on kontrollivuon (control flow) kulku toisin päin kuin normaalissa ohjelmansuorituksessa. Ohjelmaan ei siis niinkään määritetä funktiokutsuja, vaan ohjelman tapahtumille rekisteröidään vastaustoimintoja. Erityinen muoto IoC-mallista on Dependency Injection, jossa olio tarvitessaan jotakin palvelua ei pyydä sitä ohjelmallisesti, vaan palvelut tarjotaan oliolle automaattisesti sitä luotaessa. Tarvittavat palvelut määritellään konfiguraatiotiedoston, sovelluskontekstin, avulla. Sovelluskonteksti voidaan esittää yhdellä tai useammalla XML-tiedostolla.

<sup>20</sup><http://www.springframework.org/>

## 4.3 Näkymä

### Käyttöliittymä

Testitarkoitusta varten järjestelmälle tehtiin eräs käyttöliittymä. Olennaista on, että tämän ratkaisun käyttöliittymä on vain yksi vaihtoehto. Toiselle aineistolle sopii ehkä paremmin toisenlainen käyttöliittymä ja juuri tätä varten IRMA on rakennettu modulaariseksi ja toteutettu MVC-arkkitehtuurin mukaan.

Käyttöliittymän elementit on luotu pääasiassa Dojo-JavaScript-kirjaston avulla. Dojo<sup>21</sup> on JavaScript-kirjasto, joka on suunniteltu AJAX-tyyppisiin web-sovelluksiin. Dojo sisältää valmiita käyttöliittymäkomponentteja (widget) liittyen tiedon asemoiintiin ja esittämiseen. Dojo on avoimen lähdekoodin projekti, joten sen valmiita komponentteja voi muuttaa omiin tarpeisiinsa sopivaksi.

Käyttöliittymän ajatuksena on mahdollistaa automaattitäydennyksen palauttamien käsitteiden ontologisen kontekstin tarkasteleminen. Käyttöliittymä muodostuu kahdesta komponentista: automaattitäydentimestä sekä konteksti-informaation näytöstä. Kontekstikomponentit toteuttavat eräänlaisen JavaScript-rajapinnan, muuten niiden toimintaa ei ole kiinnitetty. Komponentti voi esittää vaikka karttakuvia. Kontekstikomponentteja on mahdollista avata hierarkkisesti siten, että tarkastellaan uudessa komponentissa edellisestä valittua käsitettä.

### Käyttöesimerkki

Kuvassa 17 esitetään tilanne, jossa käyttäjä on annotoimassa kuvan 17a maalausta IRMA-selaimen (17b) avulla. Käyttäjä hakee ensin käsitettä ”käärmeet”. Kohdassa 1 käyttäjä on kirjoittanut syötteeksi merkkijonon ”käärmeet”, sanan käärmeet alun. Syötekentän alapuolelle (kohta 2) on täydentynyt syötettä vastaavat käsitteet

---

<sup>21</sup><http://dojotoolkit.org/>

Iconclass- ja YSO-ontologioista sekä Iconclassin asiasanastosta. Seuraavaksi käyttäjä on valinnut listasta käsitteen “25F42: käärmeet” ja käsitteen konteksti-informaatio näytetään automaattitäydentimen listauksen oikealla puolella laatikossa (kohta 3). Konteksti-informaatioon kuuluu tässä tapauksessa käsitteen ontologinen hierarkia ja käsitteeseen liittyvät termit. Kuvan kohta 4. esittää tilannetta, jossa käyttäjä on valinnut tarkasteltavaksi käsitteeseen “25F42: käärmeet” liittyvän termit “23A21: Serpent Ouroboros”. “Ouroboroksen” konteksti-informaatio esitetään uudessa laatikossa ja siihen kuuluu pelkästään käsittehierarkia, ei liittyviä termejä. Halutessaan käyttäjä voisi jatkaa ontologian selailua esimerkiksi valitsemalla “Serpent Ouroboroksen” hierarkiasta käsitteen “23A: ajan allegoriat”, tai palaamalla tarkastelemaan “käärmeiden” hierarkiaa. Olennaista on, että IRMAN avulla käyttäjä voi päästä käsitteisiin, joita ei pelkässä automaattitäydennyslistauksessa oltaisi näytetty.

## 5 IRMA käytännössä

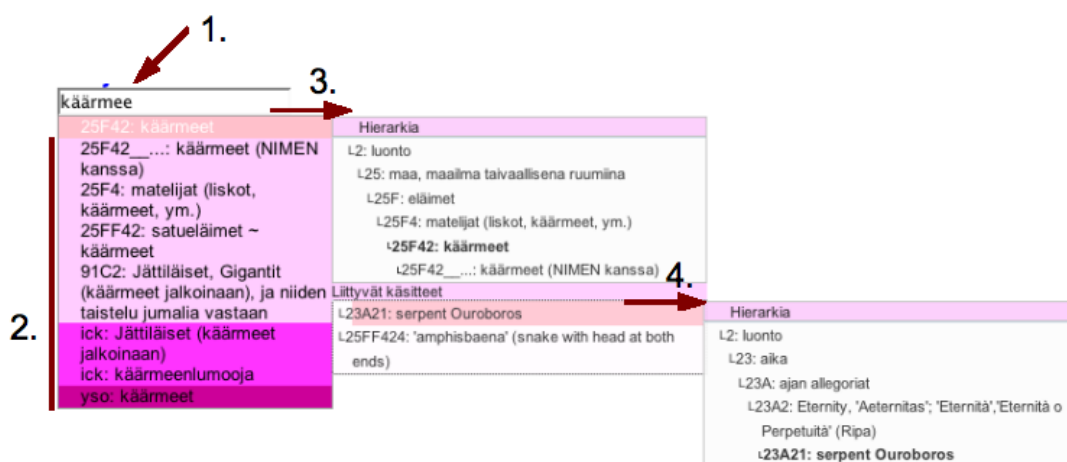
Jotta IRMAN käyttökelpoisuutta ja uusiin aineistoihin mukautuvuutta saataisiin testattua, tehtiin IRMAsta kolme versiota hyvin erilaisilla aineistoilla. Ensimmäinen niistä on valtion taidemuseon annotointitehtäviä varten tehty Iconclass-ontologian ja asiasanat sekä YSO-ontologian yhdistävä versio. Tästä kerrotaan lisää luvussa 5.1. Tällä aineistolla osoitetaan, miten IRMA toimii moniontologiaympäristössä ja millaisia visualisointeja aineistolle voidaan käyttää. Toisessa IRMAN versiossa on lääketieteen ontologia MeSH (Medical Subject Headings). Versio on tehty MediEq-projektia<sup>22</sup> varten ja se sisältää ainoastaan MeSH-ontologian. Tämä versio ohjelmasta näyttää, miten voidaan ratkaista ongelmat moniperintäisen aineiston visualisoinnissa. MeSH-IRMAsta kerrotaan lisää luvussa 5.2. Viimeisessä IRMAN versiossa demonstroidaan ajallisen paikka-aineiston esittämistä kontekstivisualisoinnin

---

<sup>22</sup><http://www.medieq.org/>



(a) Serpent Ouroboros



(b) Serpent Ouroborosin annotointi Iconclassilla Irma-selaimen avulla.

Kuva 17: Ohjelman käyttöesimerkki.

avulla. Tässä esimerkissä IRMA käyttää SAPOa, Suomen ajallista paikkaontologiaa. IRMA-SAPOsta lisää luvussa 5.3.

## 5.1 Kuvataiteen ontologia ICONCLASS

Valtion taidemuseossa tehtävää annotointityötä varten IRMAsta kehitettiin taidealan ontologia Iconclassille oma versio. Valtion taidemuseossa ollaan siirtymässä vapaasta asiasanoituksesta säännöstellyn sanaston käyttöön. Museossa on tutustuttu jo aiemmin Yleiseen suomalaiseen asiasanasto YSAan ja sen ontologiseen versioon YSO. Tarkoituksena on ottaa käyttöön taidealan ikonografisen asiasanaston Iconclassin<sup>23</sup> ontologinen versio, joka on tehty SeCo-tutkimusryhmässä [Haa06]. Iconclass-sanasto on valtion taidemuseon monille annotoijille sisällöltään vieras, toisin kuin Yleinen suomalainen ontologia YSO. Tämän vuoksi annotointia vieraalla ontologialla päätettiin helpottaa yhdistämällä Iconclass-ontologiaan YSO. IRMAN hakukentän tuloksissa näytetään tässä tapauksessa käyttäjän syötteeseen täsmäävät Iconclass-ontologian luokat, asiasanat sekä YSO:n käsitteet. Kuitenkin näistä ainoastaan Iconclassin luokat ovat annotointiin kelpaavia. YSO-käsitteet on yhdistetty OWL:n `equivalentClass`-ominaisuuksilla käsitteitä vastaaviin Iconclassin asiasanoihin ja näin ollen YSO-käsitteistä pääsee navigoimaan ontologiassa eteenpäin aina annotointiin kelpaaviin Iconclass-luokkiin asti.

Iconclass on luokittelujärjestelmä kuvataiteelle ja ikonografialle. Iconclass on hierarkkinen luokittelu, jossa käsitteet merkitään notaatioin. Käsitteiden merkitys muodostuu koko käsitehierarkian perusteella siten, että alempi käsite sisältää aina merkityksessään myös ylempät käsitteet. Esimerkiksi luokka 2 on ”luonto”, luokka 26 ”meteorologiset ilmiöt” ja luokka 26D ”pakkanen”. Iconclass sisältää viittauksia käsitteisiin liittyviin muihin luokkiin, esimerkiksi luokkaan 26D ”pakkanen” liittyy luokat

---

<sup>23</sup><http://www.iconclass.nl>



43C24 “talviurheilu” ja 23D41 “talvi”. Tavallisten luokkien lisäksi Iconclass sisältää avain-luokkia (key). Näillä voidaan tarkentaa edelleen varsinaisia notaatioita. Avaimet merkitään notaation perään +-merkillä ja numerolla. Avaimet on määritetty ontologian ominaisuuksina ja ne liittyvät aina tiettyyn luokkaan. Avainten semantiikkaa on valotettu tarkemmin Heidi Lüscherin pro gradu -työssä [Lü04]. Myös nämä luokat voivat sisältää hierarkiaa. Esimerkiksi äskeisestä talviurheilusta voidaan johtaa merkintä “voittaja, voitto (urheilu, pelit jne.)”, jota merkitään notaatiolla 43C24(+31). Varsinaisen luokittelujärjestelmän lisäksi Iconclassiin kuuluu joukko asiasanoja. Näihin on liitetty viitteet Iconclassin käsitteisiin, ja asiasanat toimivat siis linkkeinä varsinaiseen käsitejärjestelmään. Asiasanoihin kuuluu yleisiä käsitteitä, joita ei useinkaan sellaisinaan käsitejärjestelmästä löydy. Esimerkiksi asiasana “kukko” johtaa muun muassa käsitteeseen 12A63521 “pregnant women holding up a cock and a hen 'Yom Kippur' ”. Näyte Iconclassista RDF-muodossa on kuvassa 18a ja Iconclassiin liittyvistä asiasanoista kuvassa 18b. Kuvassa 18a on katkelma Iconclassin luokista. Luokat on määritetty SKOS-laajennoksena<sup>24</sup>. SKOS on semanttisen webin malli hierarkkisten thesaurusten rakenteen kuvaamiseen [AM06]. Iconclass mallinnettiin SKOS:n laajennoksena, jotta luokat ja avaimet saatiin esitettyä erillisinä tyyppinä. Kuvassa 18a eron näkyy ice:NotationConcept- ja ice:KeyConcept-tyypeistä. Kuvasta näkyy myös ontologiassa käytetyt SKOS-ominaisuudet, kuten skos:broader ja skos:related, joilla kuvataan käsitesuhteita. Kuvassa 18b on Iconclassiin liittyviä asiasanoja, jotka on esitetty tavallisessa SKOS-muodossa skos:Concept-tyyppisinä käsitteinä. Myös asiasanoihin liittyy SKOS-ominaisuuksia.

Kuvassa 19 näkyy tilanne, jossa käyttäjä on kirjoittanut hakukenttään merkkijonon “tork”, lyhenteenä ruotsin kielen sanasta “torka”, kuivuus. Järjestelmän ontologioista ainoastaan YSO sisältää ruotsinkielisiä nimikkeitä, joten tavanomaisella Iconclass-automaattitäydennysjärjestelmällä haku ei olisi tuottanut tuloksia. Valitsemalla lis-

<sup>24</sup><http://www.w3.org/2004/02/skos/>

```

<ice:KeyConcept rdf:about="#key_11_4">
  <skos:prefLabel xml:lang="zxx">11(+4)</skos:prefLabel>
  <skos:definition xml:lang="en">devil(s)</skos:definition>
  <skos:broader rdf:resource="#key_11"/>
</ice:KeyConcept>
<ice:NotationConcept rdf:about="#not_11A1">
  <skos:prefLabel xml:lang="zxx">11 A 1</skos:prefLabel>
  <skos:definition xml:lang="en">God the Creator</skos:definition>
  <skos:broader rdf:resource="#not_11A"/>
  <skos:narrower rdf:resource="#not_11A11"/>
  <skos:related rdf:resource="#not_71A2"/>
</ice:NotationConcept>
<ice:NotationConcept rdf:about="#not_11A2">
  <skos:prefLabel xml:lang="zxx">11 A 2</skos:prefLabel>
  <skos:definition xml:lang="en">Divine Nature</skos:definition>
  <skos:broader rdf:resource="#not_11A"/>
  <skos:narrower rdf:resource="#not_11A21"/>
  <skos:narrower rdf:resource="#not_11A22"/>
  <skos:narrower rdf:resource="#not_11A23"/>
</ice:NotationConcept>
<ice:NotationConcept rdf:about="#not_25D11">
  <skos:prefLabel xml:lang="zxx">25 D 11</skos:prefLabel>
  <skos:definition xml:lang="en">precious and semiprecious stones</skos:definition>
  <skos:broader rdf:resource="#not_25D1"/>
  <skos:narrower rdf:resource="#not_25D11..."/>
<ice:hasKeyConcept rdf:resource="#key_25D11"/> </ice:NotationConcept>
<skos:Concept rdf:ID="evoluutio">
  <skos:prefLabel xml:lang="fi">evoluutio</skos:prefLabel>
  <skos:related rdf:resource="#p2;not_51K4"/>
</skos:Concept>
<skos:Concept rdf:ID="luscinus--e">
  <skos:prefLabel xml:lang="fi">Luscinus (E)</skos:prefLabel>
  <skos:related rdf:resource="#p2;not_98B__FABRICIUS_C."/>
</skos:Concept>
<skos:Concept rdf:ID="palvelus">
  <skos:prefLabel xml:lang="fi">palvelus</skos:prefLabel>
  <skos:related rdf:resource="#p2;not_44E"/>
  <skos:related rdf:resource="#p2;not_45B3"/>
  <skos:related rdf:resource="#p2;not_46C293"/>
  <skos:related rdf:resource="#p2;not_53CC11"/>
</skos:Concept>
<skos:Concept rdf:ID="iisai">
  <skos:prefLabel xml:lang="fi">Iisai</skos:prefLabel>
  <skos:related rdf:resource="#p2;not_71H112"/>
  <skos:related rdf:resource="#p2;not_71H113"/>
  <skos:related rdf:resource="#p2;not_71H114"/>
  <skos:related rdf:resource="#p2;not_71H132"/>
  <skos:related rdf:resource="#p2;not_71H133"/>
</skos:Concept>

```

(a) Iconclassin luokat RDF-muodossa.

(b) Iconclassiin liittyvät asiasanat RDF-muodossa.

Kuva 18: Iconclass RDF-muodossa.

tauksesta käsitteen “torkea”, käyttäjä pääsee etenemään YSON “torkea”:sta siihen liittyvään Iconclass-asiasanaan “kuivuus”. Tähän puolestaan liittyy useita Iconclassin luokkia, joista käyttäjä on kuvassa valinnut luokan “wadi, dry river bend in desert”. Huomattavaa on, että käyttäjä on lopulta valinnut luokan, jonka nimeke on täysin erillinen alkuperäisestä kyselymerkkijonosta “torka”.

Visualisoinnissa eräänlaisena lähtökohtana oli Iconclass-käsitteiden semanttisen merkityksen muodostuminen koko hierarkian perusteella. Iconclass sisältää esimerkiksi termin “kuoleminen”. Tuntematta tarkemmin termin hierarkiaa, sitä voisi helposti käyttää väärin, sillä käsitteen yläluokkana on muun muassa “ihminen”. Termi “kuoleminen” siis tarkoittaa vain ihmisen kuolemaa, vaikkei se kirjoitusasusta ilmenekään. Tästä syystä Iconclassin yhteydessä oli erityisen tärkeää visualisoida koko merkityshierarkia ylimpään luokkaan asti.

Koska Iconclass-tapauksessa IRMAan liitettiin kolme eri sanastoa, päätettiin auto-completion-listaukseen visualisoida eri lähteistä tulevat termit eri värein havainnollisuuden lisäämiseksi. Kuvassa 20 näkyy merkkijonoon “aal” täydentyvät Iconclass-luokat ensimmäisenä, toisena Iconclass-asiasanat merkittynä ick-tunnisteella ja vii-

**Subject:**

|               |                  |  |  |
|---------------|------------------|--|--|
| tork          |                  |  |  |
| yso: torka    | Related Keywords | Related Iconclass Concepts   |  |
| yso: torkning | Lick: kuivuus    | L 25H219: dry river, river bed<br>L 25K162: wadi, dry river bed in desert<br>L 26B9: drought<br>L 34C12: providing water for wild animals during drought<br>L 71E1261: the Israelites come to Rephidim (or Meribah) and complain because there is no drinking-water<br>L 71K2913: after marching for seven days the armies run out of water ~ war against Moab<br>L 71M11: Elijah announces to King Ahab that God will bring a long drought in the land to avenge the apostasy of Israel<br>L 71M45: the end of the drought ~ story of Elijah<br>L 71S113: because the house of the Lord is not rebuilt, the land is afflicted by drought<br>L 71U341: lack of water in Bethuliah; the people complain to the elders | Hierarchy<br>L2: Nature<br>L25: earth, world as celestial body<br>L25K: landscapes in the non-temperate zone, exotic landscapes<br>L25K1: landscapes in tropical and sub-tropical regions<br>L25K16: desert<br><b>L25K162: wadi, dry river bed in desert</b> |

Kuva 19: Iconclass-, YSO- ja Iconclass-asiasanoja sisältävä IRMAN käyttöesimerkki.

meisenä YSOsta peräisin olevat asiasanat yso-alkuliittellä. Kuvan esimerkistä voidaan myös huomata automaattitiedennyksen täsmäävän myös sanoihin, joissa haettu merkkijono ei ole alussa, tässä tapauksessa esimerkiksi termiin “perä-aalto”.

Iconclassin luokkien ja hierarkia visualisoitiin puumaisena. Jos käsitteeseen liittyy avainluokkia, ne visualisoitiin omana puurakenteenaan käsittehierarkian alapuolella. Kuvassa 21 näkyy käsitteen “eläimet” kontekstivisualisointi. Ylimpänä kuvassa on käsitteen hierarkia ja käsitteeseen liittyvät luokat ja näiden alla käsitteeseen liittyvät avaimet. Koska avaimillakin voi olla hierarkkisuutta, visualisoidaan tämä hierarkia vastaavasti kuin tavallinen käsittehierarkia. Kuvassa on laajennettu avaimen “sex and age of animals” hierarkia, jolloin voitaisiin valita siitä esimerkiksi notaatio 25F\_21: “young animal”. Avaimet eivät siis ole ontologisten käsitteiden alaluokkia, kuten tässä tapauksessa luokan “eläimet”, vaan erillinen luokkahierarkia, joka liittyy päänotaatioihin RDF-ominaisuuden avulla. Tässä esimerkissä tulee myös esiin se, ettei Iconclassia ole suomennettu kokonaan, vaan seassa esiintyy englanninkielisiä

|  |
|--|
| aa                                       |
| 46C2231: laituri, aallonmurtaja          |
| 94F6: aaltoileva taistelu (Ilias VIII-X) |
| ick: aallonmurtaja                       |
| ick: aaloe                               |
| ick: aalto                               |
| ick: aalto (aluksen nostattama)          |
| ick: aalto (laivan nostattama)           |
| ick: kreikkalainen aalto                 |
| -ornamentti                              |
| ick: perä-aalto                          |
| yso: aallot                              |
| yso: aaloet                              |

Kuva 20: Iconclass-IRMAN automaattitäydennys-tuloslistaus.

termejä.

|   |  |
|---|--|
| <b>Hierarkia</b>  |  |
| L2: luonto  |  |
| L25: maa, maailma taivaallisena ruumiina                        |  |
| <b>L25F: eläimet</b>  |  |
| L25F1: groups of animals  |  |
| L25F2: nisäkkäät  |  |
| L25F3: linnut   |  |
| L25F4: matelijat (liskot, käärmeet, ym.)                        |  |
| L25F5: sammakkoeläimet  |  |
| L25F6: kalat  |  |
| L25F7: pieneliöt  |  |
| L25F8: sukupuuttoon kuolleet eläimet                            |  |
| L25F9: hirviöt  |  |
| L25FF: satueläimet  |  |
| <b>Liittyvät käsitteet</b>                                      |  |
| L34: ihminen ja eläin   |  |
| <b>Avaimet</b>  |  |
| L25F__0: variant  |  |
| L25F__1: animals used symbolically                              |  |
| L25F__2: sex and age of animals; propagation of animals         |  |
| L25F__3: anatomy of animals                                     |  |
| L25F__4: animal behaviour                                       |  |
| L25F__5: animal(s) in motion; positions, expressions of animals |  |
| L25F__6: disease and death of animal(s)                         |  |
| L25F__7: animals ~ biological investigation                     |  |
| L25F__8: man and animal   |  |
| L25F__9: products of animals                                    |  |
| <b>Hierarkia</b>  |  |
| L   |  |
| <b>L25F__2: sex and age of animals; propagation of animals</b>  |  |
| L25F__21: young animal  |  |
| L25F__22: male animal   |  |
| L25F__23: female animal   |  |
| L25F__24: very old animal                                       |  |

Kuva 21: Iconclass-ontologian käsittehierarkia, liittyvät käsitteet ja avaimet.

## 5.2 Lääketieteen ontologia MeSH

Toinen ilmentymä IRMAsta tehtiin lääketieteen ontologia MeSHiä<sup>25</sup> (Medical Subject Headings) [NPH02] varten. MeSH sisältää englanninkielisen sanaston lisäksi suomenkielisen FinMeSH-sanaston, jossa on yli 20000 käsitettä. Sekä MeSH että FinMeSH on muunnettu SKOS-muotoon SeCo-tutkimusryhmässä. SKOS-muoto mahdollistaa asiasanastojen ja muiden luokittelujen esittämisen semanttisen webin sovellusten edellyttämässä muodossa. Kuvassa 22 esitellään osa MeSH-ontologiaa RDF-muodossa. Kuvassa näkyy, miten MeSH-käsitteet määritellään skos:Concepttyyppeiksi ja miten niihin liittyy SKOS:n ominaisuuksia.

MeSHin ontologinen versio sisältää erittäin paljon moniperintää, jolloin sama termi esiintyy hierarkiapuussa useassa kohtaa eri merkityksissä. Tästä syystä Iconclassin (luku 5.1) yhteydessä käytettyä puumaista visualisointia ei voitu käyttää uudestaan sellaisenaan, sillä Iconclassin puuvisualisointi perustuu tietoon siitä, että luokalla on aina vain yksi yläluokka. Iconclassin lähestymistapaa visualisoinnissa olisi voitu muokata MeSHiä varten siten, että oltaisiin esitetty kaikki hierarkiapuut erikseen. Tämä ratkaisu osoittautui tilaavieväksi ja siihen sisältyi myös paljon redundanttia informaatiota, sillä moniperintää saattaa olla hierarkiapuussa useassa kohtaa. Näin esimerkiksi puun yläosasta saattoi olla useita eri versioita vaikka kiinnostava informaatio sijaitsee puussa alhaalla.

Toisaalta, erotuksena Iconclassin sisältöön, lääketieteen termien merkitys ei MeSHissä muodostu samalla tavoin koko hierarkian perusteella kuten Iconclassissa, joten koko hierarkian esittäminen ei ollut tarpeellista. Ennemminkin on niin, että kaikki termin hierarkiaan liittyvä merkityksen kannalta oleellinen tieto saadaan joko termin ylä- tai alakäsitteestä. Näistä syistä johtuen MeSHin kontekstihierakiassa päätettiin esittää ainoastaan termin läheisimmät ylä- ja alaluokat. Ratkaisu ei sinänsä kuiten-

---

<sup>25</sup><http://www.nlm.nih.gov/mesh/meshhome.html>

```

<rdf:Description rdf:about="http://www.yso.fi/onto/mesh/D008345Q000506">
  <skos:broader rdf:resource="http://www.yso.fi/onto/mesh/D008345"/>
  <rdf:type rdf:resource="http://www.yso.fi/onto/mesh/CompoundConcept"/>
  <mesh3:preferredCombination rdf:resource="http://www.yso.fi/onto/mesh/D020149"/>
</rdf:Description>
<rdf:Description rdf:about="http://www.yso.fi/onto/mesh/D017019">
  <mesh3:recordMaintainer>SYSTEM</mesh3:recordMaintainer>
  <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
  <mesh3:recordAuthorizer>sjn</mesh3:recordAuthorizer>
  <mesh3:recordOriginator>PXP</mesh3:recordOriginator>
  <skos:annotation xml:lang="en">a family of methanogens "found in
any anaerobic environment"
</skos:annotation>
<mesh3:dateEstablished>1992-01-01</mesh3:dateEstablished>
<mesh3:dateRevised>2005-06-30</mesh3:dateRevised>
<skos:prefLabel xml:lang="sv">Methanosarcinaceae</skos:prefLabel>
<skos:prefLabel xml:lang="la-FI">Methanosarcinaceae</skos:prefLabel>
<skos:prefLabel xml:lang="en">Methanosarcinaceae</skos:prefLabel>
<mesh3:activeMeSHYear>2006</mesh3:activeMeSHYear>
<mesh3:activeMeSHYear>2007</mesh3:activeMeSHYear>
<mesh3:historyNote xml:lang="en">92
</mesh3:historyNote>
<mesh3:dateCreated>1991-05-30</mesh3:dateCreated>
<mesh3:publicMeSHNote xml:lang="en">92
</mesh3:publicMeSHNote>
<skos:broader rdf:resource="http://www.yso.fi/onto/mesh/D019640"/>
<skos:scopeNote xml:lang="en">A family of anaerobic METHANOSARCINALES
whose cells are mesophilic or thermophilic and appear as irregular spheroid
bodies or sheathed rods. These methanogens are found in any anaerobic
environment including aquatic sediments, anaerobic sewage digesters and
gastrointestinal tracts. There are four genera: METHANOSARCINA, Methanolobus,
ssMethanothrix, and Methanococcoides.
</skos:scopeNote>
</rdf:Description>

```

Kuva 22: Osa MeSH-ontologiaa.

kaan poista tietoa, sillä ylä- ja alaluokkien hierarkiaa pääsee tutkimaan avaamalla niiden kontekstihierarkian.

Kuvassa 23 on esimerkkinä valittuna termi “maito”, jolla on MeSHissä kolme yläluokkaa: “meijerituotteet”, “eritteet” ja “juomat”. Lisäksi kuvassa näkyy maidon neljä alaluokkaa. Termin “maito” käytön kannalta on olennaista valita oikea merkitys, onko kyse esimerkiksi maidosta meijerituotteena vai äidinmaitona, eritteenä. Epäolennaisempaa on se, että meijerituotteiden yläpuolella hierarkiassa on luokka “ruoka” ja tämän yläpuolella “ruuat ja juomat”. MeSHin moniperinnästä voi saada jonkinlaisen kuvan silläkin perusteella, että “maito” on myös hierarkian kohdassa “ruuat ja juo-

|                           |                        |                   |
|---------------------------|------------------------|-------------------|
| maito                     |                        |                   |
| Eläimen matorauhaset      |                        |                   |
| Maito                     | Hierarkia              |                   |
| Maitohammas               | U Meijerituotteet      | Hierarkia         |
| Maitohampaiden irtoaminen | LEritteet              | LRuoka            |
| Maitohappo                | LJuomat                | U Meijerituotteet |
| Maitohappoasidoosi        | UMaito                 | UJäätelö          |
| Maitohappobakteeri        | LÄidinmaidonkorvikkeet | UMargariini       |
| Maitohappobakteerit       | UMaitoproteiinit       | UMaito            |
| Maitokahviläiskät         | LÄidinmaito            | LVoit             |
| Maitopankit               | UHapanmaitotuotteet    | LJuusto           |
| Maitoproteiinit           | Liittyvät käsitteet    |                   |
| Maitotalous               | LLaktaatio             |                   |
| Maitotauti                |                        |                   |
| Maitovalas                |                        |                   |
| Maitoyliherkkyys          |                        |                   |
| Pagetin matorauhastauti   |                        |                   |

(a) MeSH, maito ja meijerituotteet

|                           |                        |                      |
|---------------------------|------------------------|----------------------|
| maito                     |                        |                      |
| Eläimen matorauhaset      |                        |                      |
| Maito                     | Hierarkia              |                      |
| Maitohammas               | U Meijerituotteet      |                      |
| Maitohampaiden irtoaminen | LEritteet              | Hierarkia            |
| Maitohappo                | LJuomat                | UNesteet ja eritteet |
| Maitohappoasidoosi        | UMaito                 | ERitteet             |
| Maitohappobakteeri        | LÄidinmaidonkorvikkeet | LSmegma              |
| Maitohappobakteerit       | UMaitoproteiinit       | LLapsenkina          |
| Maitokahviläiskät         | LÄidinmaito            | LSienneste           |
| Maitopankit               | UHapanmaitotuotteet    | LYskös               |
| Maitoproteiinit           | Liittyvät käsitteet    | UMaito               |
| Maitotalous               | LLaktaatio             | LEläinmyrkyt         |
| Maitotauti                |                        | LÄidinmaito          |
| Maitovalas                |                        | LSylki               |
| Maitoyliherkkyys          |                        | LSappi               |
| Pagetin matorauhastauti   |                        | UHaimaneste          |
|                           |                        | LLima                |
|                           |                        | LSuolistoeritteet    |
|                           |                        | LTali                |
|                           |                        | UKorvavaha           |
|                           |                        | UHiki                |
|                           |                        | UKyyneleet           |
|                           |                        | LTernimaito          |
|                           |                        | UMahaneste           |

(b) MeSH, maito ja eritteet

Kuva 23: MeSH: hierarkian visualisointi.

mat”, “juomat”, “maito”. Tämä esimerkki entisestään korostaa koko hierarkian tarpeettomuutta tässä yhteydessä. IRMA:n MeSH-visualisointia voidaan verrata MeSHin julkaisijan National Library of Medicine:n verkossa olevaan MeSH-selaimeen<sup>26</sup>. Selain sisältää vain englanninkielisen MeSHin, mutta on muuten vertailukelpoinen IRMAan nähden. MeSH-selaimessa lähtökohtana on ollut esittää kaikki hierarkian haarat erillisinä siten, että hierarkiassa esitetään termin kaikki yläluokat, mutta alaluokista vain lähimmät. Pelkästään “maidon” tapauksessa hierarkia on jo aika tilaavievä, vaikka sillä on vain kolme yläluokkaa. Kuvassa 24 esitetään osa MeSH-selaimen “maito”-hierarkiasta. Tilanne muuttuu hankalampaan suuntaan, kun käsitteen hierarkiassa on moniperintää ylempänäkin. Esimerkiksi termillä “seesamiöljy” on kolme välitöntä yläluokkaa, mutta hierarkiapuita muodostuu viisi. Tätä tilannetta hahmottaa kuva 25, jossa MeSHin termin “seesamiöljy” hierarkia esitetään sekä ONKI-selaimessa että IRMAssa. ONKI-selain (25a) visualisoi hierarkian kokonaan ja siitä voidaan nähdä, että esimerkiksi käsitteiden “lipidit” “rasvat” alle sijoittuu kahteen kohtaan käsite “seesamiöljy”. IRMA-selain (25b) näyttää ainoastaan käsitteen välittömät yläluokat ja näin hierarkia saadaan näyttämään selkeämmältä.

Koska MeSH on käytettävissä sekä suomeksi että englanniksi, on IRMAssa mahdollisuus käyttää kumpaa kieltä tahansa. Kielivalinnan voi tehdä myös kesken selailun, se vaikuttaa vain siten, että kielivalinnan jälkeen avattavat kontekstilaatikot ovat eri kielisiä. Autocompletion-listaukseen sen sijaan tulostetaan oletuksena molempien kieliset termit.

### 5.3 Suomen ajallinen paikkaontologia SAPO

IRMAa testattiin myös paikka-aineistolla. IRMAan kytkettiin Suomen ajallinen paikkaontologia SAPO [KVH08]. SAPO sisältää 667 historiallista paikkaa Suomesta.

<sup>26</sup><http://www.nlm.nih.gov/mesh/MBrowser.html>

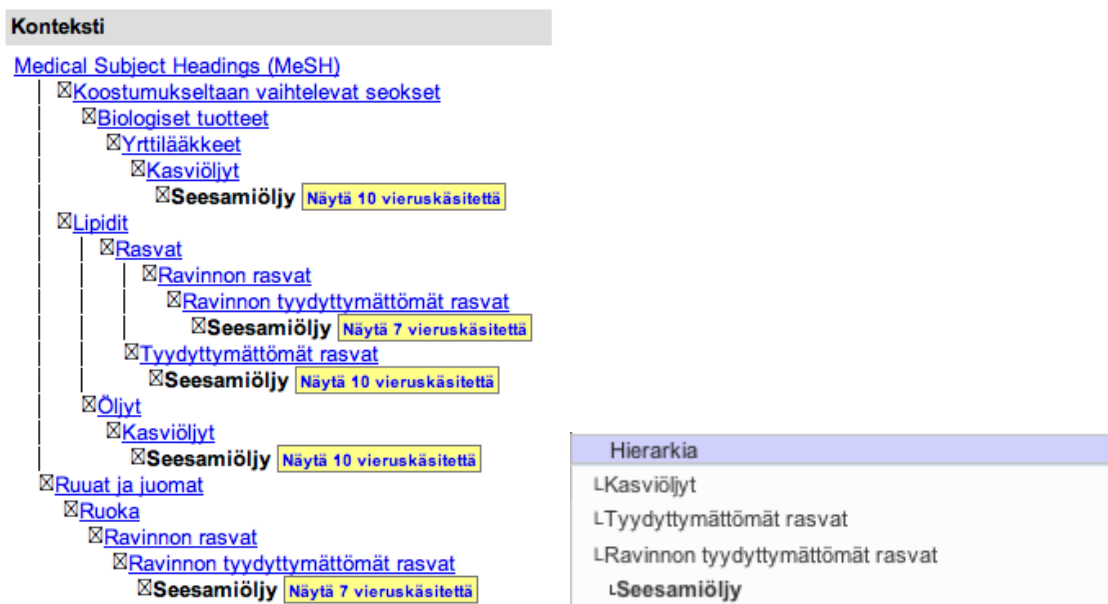




Kuva 24: Osa maito-termin hierarkiasta MeSH-selaimessa.

Nämä ovat entisiä tai nykyisiä kuntia jonakin ajanjaksona (kuten Helsinki vuosina 1946–1965). SAPO sisältää kuntien lisäksi tietoa niiden naapurikunnista kyseisinä aikoina sekä lääneistä, joihin ne ovat kuuluneet. Lisäksi SAPOssa on määritetty kuntien peittävyys historiallisiin alueisiin nähden. Tällä tarkoitetaan kuntien yhteistä maa-alaa esimerkiksi siten, että Helsingin kunnan maantieteelliseen alaan vuodesta 1966 eteenpäin kuuluu Oulunkylän vuosien 1921-1945 maantieteellinen alue. Paikka-aineistoa on 1860-luvulta nykyhetkeen. SAPOn RDF-muodosta on näyte kuvassa 26. Kuvasta nähdään muun muassa SAPOssa käytettyjä ominaisuuksia, sekä kunnille käytetty RDF-tyyppi “kunta”.

Aineisto on perin erilaista verrattuna aiemmin mainittuun Iconclass-ontologiaan, jossa on selvä hierarkia ja linkitys luokkien välillä. SAPOn hierarkia sen sijaan on



(a) MeSHin käsite "seesamiöljy", ONKI-selaimessa

(b) MeSHin käsite "seesamiöljy", IRMA-selaimessa

Kuva 25: MeSHin moniperintä.

matala, käsitehierarkian tasoina ovat kunta, lääni ja Suomi tiettyinä ajanjaksona. Toinen erityinen piirre on SAPOn runsas moniperiytyvyys. Tämä ilmenee siten, että kunta on olemassaolonsa aikana voinut kuulua useampaan lääniin. Esimerkiksi Tuupovaara on vuosien 1913–1945 aikana ollut osana kolmea lääniä. Visualisointi päätettiin moniperinnän vuoksi tehdä erillisinä hierarkioina, sillä kuntien ja läänien yhteyksiä olisi ollut vaikea hahmottaa, jos hierarkian kaikki haarat olisivat olleet samassa puussa.

Kuvassa 27 esitetään paikkahierarkia Tuupovaaran kunnasta vuosina 1913–1945. Vuosina 1917–1944 Tuupovaara oli osa Kuopion lääniä, mutta vuoden 1944 jälkeen se muuttui osaksi Pohjois-Karjalan lääniä. Kontekstihierarkiasta voidaan nähdä myös Tuupovaaralla olleen tuona aikana naapurikunta Korpiselkä, joka sittemmin liitettiin osaksi Neuvostoliittoa ja osaksi Tuupovaaraan. Konteksti-informaatiosta voidaan myös nähdä Tuupovaaran (1913–1945) peittäneen maantieteellisesti osit-

```

<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:suo="http://www.yso.fi/onto/suo/"
  xmlns:sapo="http://www.yso.fi/onto/sapo/"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" >
  <rdf:Description rdf:about="http://www.yso.fi/onto/sapo/Soini">
    <rdfs:label>Soini</rdfs:label>
    <sapo:unionof rdf:resource="http://www.yso.fi/onto/sapo/Soini(1868-)" />
    <rdf:type rdf:resource="http://www.yso.fi/onto/sapo/spaceworm" />
  </rdf:Description>
  <rdf:Description rdf:about="http://www.yso.fi/onto/sapo/Orivesi(1913-1972)">
    <sapo:covers rdf:resource="http://www.yso.fi/onto/sapo/Orivesi(1865-1912)" />
    <sapo:end>1972</sapo:end>
    <rdf:type rdf:resource="http://www.yso.fi/onto/suo/kunta" />
    <sapo:partof rdf:resource="http://www.yso.fi/onto/sapo/Hämeen_lääni(1809-1917)" />
    <sapo:size>555.0</sapo:size>
    <rdfs:label>Orivesi(1913-1972)</rdfs:label>
    <sapo:begin>1913</sapo:begin>
    <sapo:partof rdf:resource="http://www.yso.fi/onto/sapo/Länsi-Suomen_lääni(1944-)" />
    <sapo:neighborOf rdf:resource="http://www.yso.fi/onto/sapo/Juupajoki(1913-)" />
    <sapo:placename>Orivesi</sapo:placename>
    <sapo:partof rdf:resource="http://www.yso.fi/onto/sapo/Hämeen_lääni(1917-1944)" />
    <sapo:coveredBy rdf:resource="http://www.yso.fi/onto/sapo/Orivesi(1865-1912)" />
  </rdf:Description>
  <rdf:Description rdf:about="http://www.yso.fi/onto/sapo/Jakaantumisen33">
    <rdfs:label>Jakaantumisen 33</rdfs:label>
    <sapo:time>1913</sapo:time>
    <sapo:after rdf:resource="http://www.yso.fi/onto/sapo/Haapasaari(1913-1973)" />
    <sapo:after rdf:resource="http://www.yso.fi/onto/sapo/Kymi(1913-1950)" />
    <rdf:type rdf:resource="http://www.yso.fi/onto/sapo/Jakaantumisen" />
  </rdf:Description>
  <rdf:Description rdf:about="http://www.yso.fi/onto/sapo/Nilsinä(1909-1924)">
    <sapo:partof rdf:resource="http://www.yso.fi/onto/sapo/Kuopion_lääni(1809-1917)" />
    <rdf:type rdf:resource="http://www.yso.fi/onto/suo/kunta" />
    <sapo:coveredBy rdf:resource="http://www.yso.fi/onto/sapo/Nilsinä(1907-1908)" />
    <sapo:covers rdf:resource="http://www.yso.fi/onto/sapo/Nilsinä(1907-1908)" />
    <rdfs:label>Nilsinä(1909-1924)</rdfs:label>
    <sapo:partof rdf:resource="http://www.yso.fi/onto/sapo/Kuopion_lääni(1917-1944)" />
    <sapo:placename>Nilsinä</sapo:placename>
    <sapo:neighborOf rdf:resource="http://www.yso.fi/onto/sapo/Varpaisjärvi(1909-)" />
    <sapo:size>0.0</sapo:size>
    <sapo:covers rdf:resource="http://www.yso.fi/onto/sapo/Nilsinä(1869-1906)" />
    <sapo:begin>1909</sapo:begin>
    <sapo:end>1924</sapo:end>
    <sapo:coveredBy rdf:resource="http://www.yso.fi/onto/sapo/Nilsinä(1869-1906)" />
  </rdf:Description>

```

Kuva 26: Näyte SAPOsta RDF-muodossa.

|                                       |
|---------------------------------------|
| <b>Naapurialueet</b>                  |
| L Korpiselkä(1944-1945)               |
| <b>Peittää historiallisia alueita</b> |
| L Joensuu(2005-)                      |
| L Kovero(1902-1912)                   |
| L Tuupovaara(1946-2004)               |
| <b>Aluehierarkia</b>                  |
| L Suomi                               |
| L Suomi(1809-1917)                    |
| L Tuntematon lääni(1809-1917)         |
| L <b>Tuupovaara(1913-1945)</b>        |
| <b>Aluehierarkia</b>                  |
| L Suomi                               |
| L Suomi(1917-1944)                    |
| L Kuopion lääni(1917-1944)            |
| L <b>Tuupovaara(1913-1945)</b>        |
| <b>Aluehierarkia</b>                  |
| L Suomi                               |
| L Suomi(1944-)                        |
| L Pohjois-Karjalan lääni(1944-)       |
| L <b>Tuupovaara(1913-1945)</b>        |

Kuva 27: Ajallisen paikkaontologia SAPOn visualisointi, Tuupovaaran kunta vuosina 1913-1945.

tain aiempaa Koveron kuntaa (1902–1912) sekä että sillä on yhteistä maa-alueetta Joensuun kaupungin kanssa sellaisena kuin Joensuu on vuodesta 2005 eteenpäin.

Kuntahierarkiaa eteenpäin selaamalla voidaan tarkastella esimerkiksi vuosien 1809-1917 Suomen läänijakoa. Kuvassa 28a on valittu Tuupovaaran yläpuolelta hierarkiasta Suomi vuosina 1809-1917. Aukeavasta kontekstilaatikosta voidaan havaita Suomeen tällöin kuuluneen kahdeksan lääniä ja lisäksi yksi tuntematon lääni. Tämä on aineistossa mukana lääniepäselvyyksien vuoksi. Myös lääneihin kuuluvia kuntia voidaan tarkastella valitsemalla haluttu lääni hierarkiasta. Kuvassa 28b on valittu Kuopion lääni vuosina 1809-1917 ja hierarkiassa näkyy siihen kuuluvat kunnat (kuvasta on rajattu osa kunnista pois tilankäytöllisistä syistä). Tähän näkymään oltaisiin päästy myös suoraan kuvan ensimmäisestä hierarkiavisualisoinnista, toisin sanoen Suomen läänijaon tarkastelun olisi voinut jättää väliin.

|  |   |
|--|---|
| tuupo  |   |
| <b>Tuupovaara(1913-1945)</b><br><b>Tuupovaara(1946-2004)</b>   | <b>Naapurialueet</b><br>L Korpiselkä(1944-1945)   |
| <b>Peittää historiallisia alueita</b><br>L Joensuu(2005-)<br>L Kovero(1902-1912)<br>L Tuupovaara(1946-2004)            |   |
| <b>Aluehierarkia</b><br>L Suomi<br><b>L Suomi(1809-1917)</b>   |   |
| L Tuntematon lääni(1809-1917)<br><b>L Tuupovaara(1913-1945)</b>  | <b>Läänit</b><br>L Hämeen lääni(1809-1917)<br>L Kuopion lääni(1809-1917)<br>L Mikkelin lääni(1809-1917)<br>L Oulun lääni(1809-1917)<br>L Tuntematon lääni(1809-1917)<br>L Turun ja Porin lääni(1809-1917)<br>L Uudenmaan lääni(1809-1917)<br>L Vaasan lääni(1809-1917)<br>L Viipurin lääni(1809-1917) |
| <b>Aluehierarkia</b><br>L Suomi<br>L Suomi(1917-1944)<br>L Kuopion lääni(1917-1944)<br><b>L Tuupovaara(1913-1945)</b>  |   |
| <b>Aluehierarkia</b><br>L Suomi<br>L Suomi(1944-)<br>L Pohjois-Karjalan lääni(1944-)<br><b>L Tuupovaara(1913-1945)</b> |   |

(a) SAPO, läänijako

|   |  |
|---|--|
| <b>Naapurialueet</b><br>L Korpiselkä(1944-1945)   |  |
| <b>Peittää historiallisia alueita</b><br>L Joensuu(2005-)<br>L Kovero(1902-1912)<br>L Tuupovaara(1946-2004)   |  |
| <b>Aluehierarkia</b><br>L Suomi<br>L Suomi(1809-1917)<br>L Tuntematon lääni(1809-1917)<br><b>L Tuupovaara(1913-1945)</b>  |  |
| <b>Läänit</b><br>L Hämeen lääni(1809-1917)<br>L Kuopion lääni(1809-1917)  |  |
| <b>Kunnat</b><br>L Eno(1871-)<br>L Hankasalmi(1872-1966)<br>L Iisalmen mlk(1860-1921)<br>L Iisalmi(1860-1969)<br>L Ilomantsi(1875-1901)<br>L Ilomantsi(1902-1943)<br>L Joensuu(1848-1953)<br>L Juuka(1868-)<br>L Kaavi(1875-1920)<br>L Karttula(1874-1925)<br>L Keitele(1879-)<br>L Kesälahti(1873-)<br>L Kiihtelysvaara(1868-1924)<br>L Kitee(1874-1919)<br>L Kiuruvesi(1873-)<br>L Kontiolahti(1873-1921)<br>L Kovero(1902-1912)<br>L Kuopio(1652-1968)<br>L Kuusjärvi(1875-1920)<br>L Lapinlahti(1874-1963)<br>L Liperi(1875-1920)<br>L Maaninka(1874-1924)<br>L Muuruvesi(1907-1925)<br>L Nilsia(1869-1906)<br>L Nilsia(1907-1908)<br>L Nilsia(1909-1924) |  |
| <b>Aluehierarkia</b><br>L Suomi<br>L Suomi(1917-1944)<br>L Kuopion lääni(1917-1944)<br><b>L Tuupovaara(1913-1945)</b>   |  |
| <b>Aluehierarkia</b><br>L Suomi<br>L Suomi(1944-)<br>L Pohjois-Karjalan lääni(1944-)<br><b>L Tuupovaara(1913-1945)</b>  |  |

(b) SAPO, kunnat

Kuva 28: Suomen ajallisen paikkaontologian ominaisuuksia.

SAPOn visualisoinnissa haasteena oli moniperiytyvyys samoin kuin aiemmin mainitussa MeSH:ssä. Mutta toisin kuin MeSH:ssä, SAPOn moniperiytyvyys on systemaattista ja sijoittuu aina tiettyyn kohtaan hierarkiaa, toisin sanoen ainoastaan kunta-lääni-tasolle siten, että kunta on voinut olla osa useaa lääniä. Jos visualisointi oli toteutettu kuten MeSH:ssä, olisi hierarkiasta jäänyt puuttumaan taso, joka esittää Suomea jonakin ajanjaksona, kuten “Suomi (1917–1944)” (vrt. kuva 27). Tämä taso päätettiin kuitenkin ottaa mukaan hierarkiaan, sillä sen kautta pääsee muun muassa selailemaan Suomeen tuona ajankohtana kuuluneita läänejä. Näin ollen SAPOn hierarkia päätettiin visualisoida kokonaisina puina ja esittää kaikki mahdolliset hierarkiat erikseen.

SAPOn aineisto eroaa aiemmin käsitellystä Iconclassista myös siten, että siinä esiintyvät kunnat ja läänit ovat ontologian luokkien instansseja, kun taas Iconclassissa on pelkästään luokkia. Tämä ei kuitenkaan aiheuttanut ongelmia aineiston käsittelyssä, sillä kuntien ja läänien indeksointi oli helposti tehtävissä RDF:type-määreiden perusteella.

SAPOn käyttävää IRMA-selainta hyödynnetään KulttuuriSampo-portaalissa [HRH<sup>+</sup>08]. KulttuuriSammossa haluttiin esittää ajallis-paikallista kuntainformaatiota kartalla, mutta informaation valinnan toteuttaminen osoittautui haasteelliseksi. IRMAN avulla saadaan esitettyä sekä ajallinen kuntahierarkia että kuntien naapurialueet. IRMA-selaimesta voidaan valita haluttu alue, joka sitten näytetään kartalla.

## 6 Arviointi

Toteutettu käsitevalitsin on ollut testikäytössä ja julkaistu Yleinen suomalainen ontologiapalvelu -sivustolla webissä<sup>27</sup>. Ontologiapalvelusivusto on Semanttisen laskennan tutkimusryhmän ylläpitämä palvelu, jossa julkaistaan yleiskäyttöisiä ontologioita.

---

<sup>27</sup><http://www.yso.fi/>

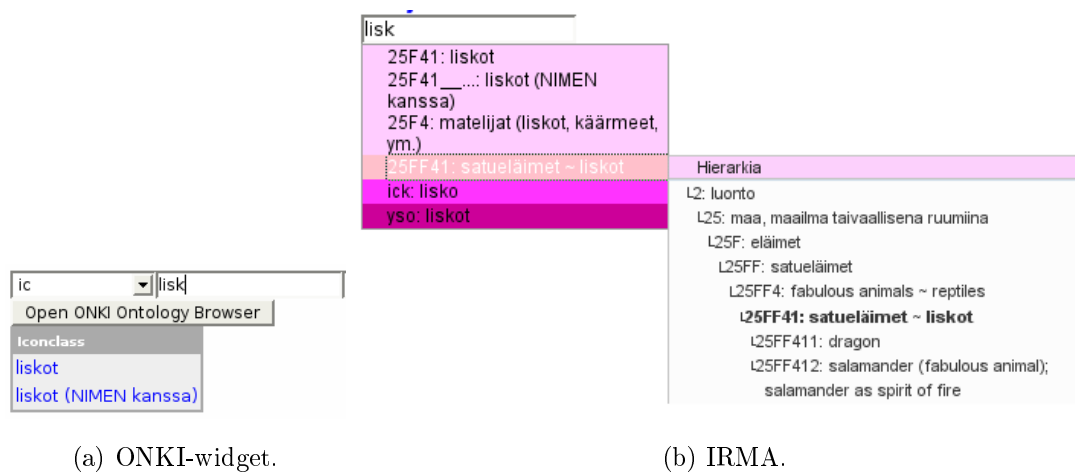
ta julkiseen verkkoon. Palvelun ajatuksena on toimia keskitettynä ontologiapalveluna, josta käyttäjät voivat hankkia palveluita omiin järjestelmiinsä. Ontologiapalveluita tarjotaan kolmessa muodossa: ONKI-ontologiaselaimena [VTH08], ONKI-indeksointikomponenttina [VTH08] sekä IRMA-käsitevalitsimena [SMKH08].

Luvun 2.3 katsauksen perusteella käsitevalitsimista löydettiin useita puutteita. Automaattitäydentävillä valitsimilla parhaan termin löytäminen on vaikeata, käsitteen merkitystä ei saada esitettyä täsmällisesti eikä monimerkityksisiä sanoja erotettua toisistaan, käyttäjää ei ohjailta sanastossa eteenpäin eikä monen ontologian esittämiseen yhdessä ole varauduttu. Puuselaimet taas ovat käytöltään hitaita ja niiden käytössä on ongelmia tilanteissa, joissa käsitteistöä on runsaasti moniperintää tai lähes synonyymiset käsitteet ovat eri puolilla sanastoa.

IRMA-käsitevalitsin mahdollistaa käsitteistön selailun ja sen avulla on mahdollista opastaa käyttäjää eteenpäin ontologiassa valitsemaan tarkempia termejä. Näin helpotetaan täsmällisimmän käsitteen löytymistä ja tällä tavoin parannetaan esimerkiksi annotaation laatua. IRMA mahdollistaa käsitehierarkian visualisoinnin esimerkiksi puuna, kuten se on tässä kirjoituksessa esiteltyissä sovelluksissa toteutettu. Käsitehierarkian lisäksi voidaan näyttää esimerkiksi käsitteeseen liittyviä termejä, mikä osaltaan auttaa käsitteen merkityksen selventämisessä. Monen ontologian aineiston käyttöä IRMAssa esiteltiin luvussa 5.1, jossa IRMAan yhdistettiin ontologiat YSO ja Iconclass, sekä Iconclassiin liittyvä asiasanasto. Nykyisiin ratkaisuihin verrattuna, joissa ontologiaselaimessa näytetään kerrallaan vain yksi ontologia, IRMA tuo parannuksen. Puuselainten ongelma käytön hitaudesta on IRMAssa ratkaistu nopeasti toimivalla automaattitäydentimellä. Myös eri puolilla sanastoa esiintyvät käsitteet voidaan saada palautettua käyttäjälle suoraan automaattitäydennyksen avulla.

Kuvassa 29 on tehty haku Iconclass-ontologiaan merkkijonolla "lisk". Kuva 29a esittää haun tulosta ONKI-indeksointikomponentilla [VTH08] ja kuva 29b IRMA-käsitevalitsimella. Molemmat käyttävät hyväkseen automaattista täydennystä, mutta vain

IRMAN avulla voidaan nähdä käsitteen konteksti. ONKI-komponentti hakee käsitteitä vain niiden nimien alun perusteella toisin kuin IRMA. Myös ainoastaan IRMAssa on mahdollista hyödyntää Iconclassiin liittyviä asiasanoja sekä linkitystä toisesta ontologiasta, tässä tapauksessa YSOsta. ONKI-komponentilla tehdyllä haulla voidaan valita käsitteeksi vain Iconclassin luokat “liskot” tai “liskot nimen kanssa”. IRMA-haulla saadaan tuloksia myös asiasanastosta ja YSOsta ja näistä voidaan navigoida eteenpäin, kuten asiasanasta “lisko” siihen liittyvään luokkaan “Apollo Sauroctonus”, jonka taiteen annotoija heti yhdistää Praxitelesin kuuluisaan veistokseen ajalta 350-340 eaa. Kuvassa 29b on valittu tarkasteluun käsitteen “satueläimet liskot” konteksti, josta voitaisiin valita esimerkiksi käsite “dragon (lohikäärme)”.



Kuva 29: Iconclassin käsitettä 'liskot' haetaan ONKI-widget -komponentilla sekä IRMAlla.

Kuvassa 30 havainnollistetaan, miten ontologian eri osissa olevat lähes samannimiset käsitteet saadaan näytettyä käyttäjälle. Hakusanaksi on kirjoitettu “ihmin”, sanan “ihminen” alku. Palautuvien Iconclass-luokkien lisäksi muita “ihmiseen” liittyviä käsitteitä voidaan navigoida asiasanan “ihminen” kautta.

Kuva 31 esittää tilannetta, jossa on haettu masennukseen liittyviä termejä MeSH-sanastosta. Ensimmäinen kuva on National Library of Medicinen (NLM) MeSH-



| ihmin   |
|---|
| 31: ihminen biologisena olentona  |
| 31A7: sukupuolet (ihminen)  |
| 34: ihminen ja eläin  |
| 34F12: ihminen surmaa eläimen   |
| 34F1: ihminen eläinten uhkaamana  |
| 34F: ihminen ja (villi) eläin   |
| 3: ihminen  |
| 95A__CECROPS: Kekrops: puoliksi ihminen - puoliksi käärme, Attikan ensimmäinen kuningas |
| ick: alkeellinen ihminen  |
| ick: ihminen  |
| ick: kemiallinen ihminen (keskiajan alkemia)  |
| ick: kuoleva ihminen  |
| ick: Pekingin ihminen   |
| ick: pieni ihminen  |
| ick: primitiivinen ihminen  |
| ick: tavallinen ihminen   |
| ick: villi-ihminen  |
| yso: ihminen  |

Kuva 30: Iconclass-ontologiasta löytyvät ihmiseen viittaavat käsitteet.

selaimesta. Siinä hakumerkkijono “depression” (masennus) on täsmätty vastaavaan käsitteeseen. Käsitteestä näytetään sen yläpuolinen hierarkia sekä sen rinnakkaiskäsitteet. Kuvassa 31b on tehty sama haku IRMAlla. Hakuun täsmäävät kaikki merkkijonoon “depress” täsmäävät käsitteet, mukaan lukien “Long-Term Depression” (pitkäaikainen masennus) sekä “Depression, Postpartum” (synnytyksenjälkeinen masennus). Nämä käsitteet liittyvät läheisesti käsitteeseen “masennus”, mutta NLM:n selaimella niitä ei suoraan löydetä. IRMAN avulla saadaan näkyville myös käsitteen “depressio” konteksti-informaatio, johon kuuluu käsite “Antidepressive Agents” (masennuslääkkeet). Tällä helpotetaan annotoijan työtä, sillä “masennuslääkkeitä” varten ei tarvitse tehdä uutta hakua, vaan liittyviä käsitteitä tarjotaan suoraan.

Ohjelman tekniset ratkaisut ovat onnistuneet siltä kannalta, että ohjelmasta saatiin rakennettua yleinen modulaarisuuden ja konfiguroitavuuden avulla. Tavoite saada käyttöliittymä ja visualisoitavat ontologiat helposti vaihdettua saavutettiin. Ohjelman suoritus tapahtuu nopeasti, eikä palvelimen vastaukseen kulu kohtuuttomasti

[Behavior and Behavior Mechanisms \[F01\]](#)

[Behavior \[F01.145\]](#)

[Behavioral Symptoms \[F01.145.126\]](#)

[Affective Symptoms \[F01.145.126.100\]](#)

[Aggression \[F01.145.126.125\] +](#)

[Catatonia \[F01.145.126.156\]](#)

[Child Reactive Disorders \[F01.145.126.159\]](#)

[Coprophagia \[F01.145.126.175\]](#)

[Delusions \[F01.145.126.200\]](#)

[Depersonalization \[F01.145.126.300\]](#)

▶ [Depression \[F01.145.126.350\]](#)

[Encopresis \[F01.145.126.837\]](#)

[Enuresis \[F01.145.126.856\] +](#)

[Hearing Loss, Functional \[F01.145.126.875\]](#)

[Malingering \[F01.145.126.925\]](#)

[Mental Fatigue \[F01.145.126.937\]](#)

[Obsessive Behavior \[F01.145.126.950\]](#)

[Paranoid Behavior \[F01.145.126.962\]](#)

[Schizophrenic Language \[F01.145.126.975\]](#)

[Self-Injurious Behavior \[F01.145.126.980\] +](#)

[Stress, Psychological \[F01.145.126.990\]](#)

(a) Mesh.

|                                    |                         |
|------------------------------------|-------------------------|
| depress                            |                         |
| Appetite                           |                         |
| Depressants                        |                         |
| Central Nervous System Depressants |                         |
| Depression                         | Hierarchy               |
| Depression, Chemical               | ↳ Behavioral Symptoms   |
| Depression, Postpartum             | ↳ Depression            |
| Depressive Disorder                | Related Concepts        |
| Depressive Disorder, Major         | ↳ Antidepressive Agents |
| Long-Term Depression (Physiology)  |                         |
| Myocardial Depressant Factor       |                         |
| Skull Fracture, Depressed          |                         |
| Spreading Cortical Depression      |                         |

(b) IRMA.

Kuva 31: Iconclassin käsitettä 'depression' haetaan National Library of Medicinen selaimella sekä IRMAlla.

aikaa. Mittauksia suoritusnopeudesta ei tehty ja arvio perustuu käyttökokemukseen. Suorituksen jouhevuuteen vaikuttaa osaltaan tiedon asynkroninen vaihto palvelimen kanssa, toisaalta ohjelman indeksit toimivat nopeasti.

IRMA on suunniteltu yhden tai useamman toisiinsa linkitetyn ontologian käsitevalitsimeksi. Sillä voidaan näyttää myös toisistaan erillisiä ontologioita, mutta tällöin ei saavuteta kaikkia etuja. Parhaiten se soveltuu tietyn alueen yhteenlinkitetylle aineistolle. Eräs rajoitus IRMAssa on, ettei sen avulla voi katsoa yhdellä kerralla koko ontologian sisältöä kuten puuselaimilla, sillä IRMAssa on avoinna kerrallaan rajallinen määrä käsitteiden konteksti-näyttöjä.

IRMAN nykyiset käyttöliittymäratkaisut käyttävät kaikki JavaScriptiä sisällön tuottamiseen. Esteettömyyden kannalta tämä on huono ratkaisu, W3C:n<sup>28</sup> julkaiseman verkkosivujen saavutettavuusohjeistuksen mukaan käyttöliittymän pitäisi aina toimia ilman skriptejäkin [acc99].

IRMA ei ole valmis sovellus vaan demonstraatio, jolla pyritään esittämään idea käsitevalitsimesta. Tästä syystä sitä ei ole systemaattisesti testattu ohjelmallisten vikojen löytämiseksi eikä sen toimintaa ole tutkittu suurilla käyttäjämäärillä. Jatkokokehitykseen kuuluukin testaus. Sovellusta olisi myös hyvä testata käyttäjätestein, jotta voitaisiin tehdä tarkempia arvioita sen soveltuvuudesta annotointiin ja jotta nähtäisiin, miten sitä voi parantaa. Tulevaisuudessa IRMA on tarkoitus liittää KulttuuriSampo-portaaliin [HRH<sup>+</sup>08] sen historiallisten paikka-aineistojen hakuliittymäksi. Nykyisin portaalissa haetaan näitä paikkoja automaattitäydentävällä hakukomponentilla. Tarkoitus on että IRMAN avulla hakukyselyjä voidaan muodostaa luontevasti paikkojen kontekstitiedoilla.

---

<sup>28</sup><http://www.w3.org/>

## 7 Yhteenveto

Tässä työssä tutkittiin semanttisen webin sovelluksissa toteutettavaa ontologisten käsitteiden valintaa. Useat tehtävät semanttisessa webissä liittyen sekä hakuun että annotointiin edellyttävät käsitteen valintaa.

Nykyisissä sovelluksissa käsitteen valinta on toteutettu yleensä joko automaattitäydentävällä menetelmällä tai käsitepuuta selaamalla. Näistä menetelmistä voidaan löytää useita ongelmia erilaisissa tilanteissa. Automaattitäydennys ei tarjoa mahdollisuutta tutustua käsitteistöön, joten se on sanastoa tuntemattomalle käyttäjälle ongelmallinen. Siinä ei myöskään ole mahdollista hyödyntää ontologian verkko- maista rakennetta; luokkasuhteet samoin kuin ominaisuuskaaret käsitteiden välillä jäävät huomioimatta. Kärjistetysti voidaan ajatella, ettei automaattitäydennyksellä toteutettu käsitehaku tuo mitään semantiikkaa webiin, sillä automaattitäydennys esimerkiksi sanalistasta tuottaisi samankaltaisen lopputuloksen. Ontologisen tiedon näkymättömyys aiheuttaa ongelmia käsitteiden merkityksen yksiselitteistämässä. Käyttäjän syötteeseen “kurkku” täydentyvästä termistä on olennaista tietää, onko kyseessä ruumiinosa vai vihannes. Tällaista tietoa ei pystytä esittämään pelkän automaattitäydennyksen avulla. Käsitteen valinnassa käytetään myös hierarkkista luokkaselainta. Se osaltaan ratkaisee automaattitäydennyksen ongelmat, mutta on hidaskäyttöinen. Erityisesti jos käsitteistö on rakenteeltaan syvä, voi hierarkiapuuttan selaaminen olla työlästä. Hankaluuksia syntyy myös tilanteessa, jossa käyttäjä ei ole varma käsitteen sijainnista puussa.

Ontologiat eroavat sekä rakenteeltaan että sisällöltään. Tämä on otettava huomioon mietittäessä, millä tavalla ontologiaa voidaan visualisoida edustavasti ja helpottaa käsitteen merkityksen hahmottumista käyttäjälle. Paikka-aineistoa sisältävä ontologia voidaan esimerkiksi visualisoida kartalla ja historiallisia tapahtumia sisältävä aikajanalla. Nykyisissä käsitevalitsimissa on kiinnitetty visualisointitapa eivätkä

ne tue useammanlaisia esitysmuotoja. Näin ollen paikka-aineistolle suunniteltu visualisointikomponentti sopii vain paikka-aineistolle eikä sitä voida hyödyntää muun tyyppisten käsitteiden valinnassa. Työssä esiteltiin erilaisia tapoja visualisoida ontologiaa ja esiteltiin erilaisia visualisointeja käyttäviä sovelluksia.

Automaattitäydennyksen tai hierarkkisen luokkiselaimen avulla tehtävässä käsitteen valinnassa esiityvien ongelmien ratkaisuksi kehitettiin käsitteen valitsin semanttiseen webiin. Sen lähtökohtana on yhdistää automaattitäydennyksen ja hierarkia-puun edut. Käsitevalitsin on toteutettu web-sovelluksena. Tavoitteena oli toteuttaa valitsin, joka on mahdollisimman yleinen ja monelle ontologialle sopiva. Tästä syystä sen rakenteesta tehtiin modulaarinen, jotta sille voidaan helposti tuottaa uudenlaisia visualisointeja. Käsitevalitsinta testattiin kolmella erityyppisellä aineistolla.

Käsitteen valinta on yksi semanttisen webin perusongelma. Jotta semanttinen web voi toimia, täytyy webin aineisto kuvata koneen ymmärtämässä muodossa [BLHL01]. Aineiston kuvailuun puolestaan tarvitaan metatietoa, ja metatiedon tuotannossa pitää valita aineistoa kuvaavia käsitteitä [KK01].

Metatiedon tuottaminen on kriittinen osa semanttisen webin vision toteutumista ja siihen on syytä kiinnittää huomiota. Yritykset ja julkisyhteisöt ontologisoivat sanastojaan sekä Suomessa [Hyv05] että maailmalla [BDBS04] ja ontologioita hyödynetään näiden organisaatioiden tiedonkuvailussa. Jotta laajoja ontologioita voidaan tehokkaasti hyödyntää, pitää siihen luoda edellytykset perustasolta lähtien. Eräs tällainen edellytys on aineiston kuvailuun ja hakuun tarkoitettujen käsitteiden valinnan helpottaminen.

## Kiitokset

Tämä tutkielma on tehty Teknillisen korkeakoulun ja Helsingin yliopiston yhteisessä Semanttisen laskennan tutkimusryhmässä osana FinnONTO-hanketta.

Haluan kiittää professori Eero Hyvöstä työn ohjaamisesta ja hyvistä neuvoista. Lisäksi haluan kiittää työtovereitani Semanttisen laskennan tutkimusryhmässä avusta ja seurasta. Erityiskiitos Eetu Mäkelälle loistokkaista ideoista työhön liittyen.

## Lähteet

- acc99 Web content accessibility guidelines 1.0, W3C, 1999. URL <http://www.w3.org/TR/WAI-WEBCONTENT/>.
- ADR06 Auer, S., Dietzold, S. ja Riechert, T., OntoWiki - A tool for social, semantic collaboration. *International Semantic Web Conference*, Cruz, I. F., Decker, S., Allemang, D., Preist, C., Schwabe, D., Mika, P., Uschold, M. ja Aroyo, L., toimittajat, osa 4273 sarjasta *Lecture Notes in Computer Science*. Springer, 2006, sivut 736–749.
- AH98 Andrews, K. ja Heidegger, H., Information Slices: Visualising and exploring large hierarchies using cascading, semicircular disks. *Proceedings of the Symposium on Information Visualization (InfoVis , '98*, 1998, sivut 9–12.
- AJS<sup>+</sup>04 Airio, E., Järvelin, K., Saatsi, P., Kekäläinen, J. ja Suomela, S., CIRI - an ontology-based query interface for text retrieval. *Proceedings of the 11<sup>th</sup> Finnish Artificial Intelligence Conference STeP 2004, Vantaa, Finland*, Hyvönen, E., Kauppinen, T., Salminen, M., Viljanen, K. ja Ala-Siuru, P., toimittajat, osa 2 sarjasta *Conference Series – No 20*. Finnish Artificial Intelligence Society, 2004, sivut 73–82.
- Ala03 Alani, H., TGVizTab: An ontology visualisation extension for Protégé. *Knowledge Capture (K-Cap'03), Workshop on Visualization Information in Knowledge Engineering*, 2003.
- AM06 Alistair Miles, Thomas Baker, R. S., Best practice recipes for publishing rdf vocabularies. Tekninen raportti, W3C, 2006. URL <http://www.w3.org/TR/2006/WD-swbp-vocab-pub-20060314/>.

- BBP05 Bosca, A., Bonino, D. ja Pellegrino, P., OntoSphere: more than a 3D ontology visualization tool. *Proceedings of the 2nd Italian Semantic Web Workshop*, 2005.
- BDBS04 Baehrecke, E., Dang, N., Babaria, K. ja Shneiderman, B., Visualization and analysis of microarray and gene ontology data with treemaps. *BMC Bioinformatics*, 5,1(2004).
- BLHL01 Berners-Lee, T., Hendler, J. ja Lassila, O., The semantic web. *Scientific American*, 284,5(2001), sivut 34–43.
- BMG00 Bederson, B. B., Meyer, J. ja Good, L., Jazz: an extensible zoomable user interface graphics toolkit in Java. *UIST '00: Proceedings of the 13th annual ACM symposium on User Interface Software and Technology*, New York, NY, USA, 2000, ACM, sivut 171–180.
- CK91 Cousins, S. ja Kahn, M., The visual display of temporal information. *Artificial Intelligence in Medicine*, 3(1991), sivut 341–357.
- dcs08 Dublin core metadata element set, version 1.1. Tekninen raportti, Dublin Core Metadata Initiative, 2008.
- DMD<sup>+</sup>03 Doan, A., Madhavan, J., Dhamankar, R., Domingos, P. ja Halevy, A., Learning to match ontologies on the semantic web. *The VLDB Journal*, 12,4(2003), sivut 303–319.
- ERG02 Eklund, P., Roberts, N. ja Green, S., OntoRama: Browsing RDF ontologies using a hyperbolic-style browser. *First International Symposium on Cyber Worlds*, 2002, sivut 405–411.
- Fen03 Fensel, D., *Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce*. Springer-Verlag New York, 2003.



- Gar05 Garrett, J. J., Ajax: A new approach to web applications, AdaptivePath, 2005. URL <http://www.adaptivepath.com/publications/essays/archives/000385.php>.
- GMF<sup>+</sup>03 Gennari, J. H., Musen, M. A., Ferguson, R. W., Grosso, W. E., Crubézy, M., Eriksson, H., Noy, N. F. ja Tu, S. W., The evolution of Protege an environment for knowledge-based systems development. *Int. J. Hum.-Comput. Stud.*, 58,1(2003), sivut 89–123.
- GMM03 Guha, R., McCool, R. ja Miller, E., Semantic search. *WWW '03: Proceedings of the 12th International Conference on World Wide Web*, New York, NY, USA, 2003, ACM, sivut 700–709.
- Gor88 Gordon, C., An introduction to Iconclass. *Terminology for Museums. Proceedings of an International Conference*. Museum Documentation Association, 1988, sivut 233–244.
- GPS04 Grau, B. C., Parsia, B. ja Sirin, E., Working with multiple ontologies on the semantic web. *International Semantic Web Conference*, McIlraith, S. A., Plexousakis, D. ja van Harmelen, F., toimittajat, osa 3298 sarjasta *Lecture Notes in Computer Science*. Springer, 2004, sivut 620–634.
- Gru93 Gruber, T. R., A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5,2(1993), sivut 199–220.
- Haa06 Haaramo, M., Iconclass-luokittelujärjestelmän ontologisointi ja soveltaminen. Diplomityö, Teknillinen korkeakoulu (TKK), 2006.
- HAK07 Hyvönen, E., Alm, O. ja Kuittinen, H., Using an ontology of historical events in semantic portals for cultural heritage. *Proceedings of the Cul-*

*tural Heritage on the Semantic Web Workshop at the 6th International Semantic Web Conference (ISWC 2007)*, 2007.

- HG04 Hatcher, E. ja Gospodnetic, O., *Lucene in Action*. Manning Publications Co., Greenwich, CT, USA, 2004.
- HM06 Hyvönen, E. ja Mäkelä, E., Semantic autocompletion. *Proceedings of the first Asia Semantic Web Conference (ASWC 2006), Beijing*. Springer-Verlag, New York, 2006.
- Hor Horridge, M., OWLviz. URL <http://protegewiki.stanford.edu/index.php/OWLViz>.
- HRH<sup>+</sup>08 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. *In: K. Robering (ed.): Information Technology for the Virtual Museum. LIT Verlag, Berlin.*, 2008.
- HSVF07 Hyvönen, E., Seppälä, K., Viljanen, K. ja Frosterus, M., Yleinen suomalainen ontologia YSO — kohti suomalaista semanttista webiä. *Tietolinja*, 1(2007).
- HvOA<sup>+</sup>07 Hildebrand, M., van Ossenbruggen, J., Amin, A., Aroyo, L., Wielemaaker, J. ja Hardman, L., The design space of a configurable autocompletion component. Tekninen raportti INS-E0708, CWI, 2007.
- Hyv05 Hyvönen, E., Kohti suomalaista semanttista webiä – suomalaisen semanttisen webin ontologiat (FinnONTO)-hankkeen esittely, Presented at the FinnONTO-symposium, 2005.
- Joh02 Johnson, R., *Expert One-on-One J2EE Design & Development*. Wrox Press Ltd., Birmingham, UK, UK, 2002.

- KC04 Klyne, G. ja Carroll, J. J., Resource Description Framework (RDF): Concepts and abstract syntax. Tekninen raportti, W3C, 2004. URL <http://www.w3.org/TR/rdf-concepts/>.
- Kei02 Keim, D., Information visualization and visual data mining. *Visualization and Computer Graphics, IEEE Transactions on*, 8,1(2002), sivut 1–8.
- KGKN06 Kraines, S., Guo, W., Kemper, B. ja Nakamura, Y., EKOSS: A knowledge-user centered approach to knowledge sharing, discovery, and integration on the semantic web. *The Semantic Web - ISWC 2006*, Cruz, I., Decker, S., Allemang, D., Preist, C., Schwabe, D., Mika, P., Uschold, M. ja Aroyo, L., toimittajat, osa 4273 sarjasta *LNCS*. Springer, 2006, sivut 833–846.
- KH06 Käsälä, T. ja Hyvönen, E., A semantic view-based portal utilizing Learning Object Metadata. *1st Asian Semantic Web Conference (ASWC2006), Semantic Web Applications and Tools Workshop.*, 2006.
- KH07 Kurki, J. ja Hyvönen, E., Relational semantic search: Searching social paths on the semantic web. *Poster Proceedings of the International Semantic Web Conference (ISWC 2007), Busan, Korea*, 2007.
- KHL<sup>+</sup>07 Katifori, A., Halatsis, C., Lepouras, G., Vassilakis, C. ja Giannopoulou, E. G., Ontology visualization methods - a survey. *ACM Comput. Surv.*, 39,4(2007).
- KHS<sup>+</sup>08 Kauppinen, T., Henriksson, R., Sinkkilä, R., Lindroos, R., Väättäinen, J. ja Hyvönen, E., Ontology-based disambiguation of spatiotemporal locations. *1st international workshop on Identity and Reference on the*

*Semantic Web (IRSW2008), 5th European Semantic Web Conference 2008 (ESWC 2008), Tenerife, Spain, 2008.*

- KK01 Kahan, J. ja Koivunen, M.-R., Annotea: an open RDF infrastructure for shared web annotations. *WWW '01: Proceedings of the 10th international conference on World Wide Web*, New York, NY, USA, 2001, ACM, sivut 623–632.
- KP88 Krasner, G. E. ja Pope, S. T., A cookbook for using the model-view controller user interface paradigm in Smalltalk-80. *J. Object Oriented Program.*, 1,3(1988), sivut 26–49.
- KPS<sup>+</sup>06 Kalyanpur, A., Parsia, B., Sirin, E., Cuenca-Grau, B. ja Hendler, J., Swoop: A web ontology editing browser. *Journal of Web Semantics*, 4(2006), sivu 2.
- KVH08 Kauppinen, T., Väätäinen, J. ja Hyvönen, E., Creating and using geospatial ontology time series in a semantic cultural heritage portal. *Proceedings of the 5th European Semantic Web Conference 2008, Tenerife, Spain, 2008.*
- LAWB03 L.Hollink, A.Th.Schreiber, Wielemaker, B. ja B.Wielinga, Semantic annotation of image collections. *In Proceedings of the KCAP'03 Workshop on Knowledge Markup and Semantic Annotation*, Florida, USA, October 2003.
- LPP<sup>+</sup>06 Lee, B., Parr, C. S., Plaisant, C., Bederson, B. B., Veksler, V. D., Gray, W. D. ja Kotfila, C., TreePlus: Interactive exploration of networks with enhanced tree layouts. *IEEE Transactions on Visualization and Computer Graphics*, 12,6(2006), sivut 1414–1426.

- LR96 Lamping, J. ja Rao, R., The hyperbolic browser: A focus+context technique for visualizing large hierarchies. *Journal of Visual Languages & Computing*, 7,1(1996), sivut 33–55.
- luc Apache Lucene - Performance benchmarks. URL <http://lucene.apache.org/java/docs/benchmarks.html>.
- Lü04 Lüscher, H., ICONCLASS - ikonografista luokittelua. Pro gradu, Helsingin yliopisto, Humanistinen tiedekunta, Taiteiden tutkimuksen laitos, 2004.
- NPH02 Nelson, S. J., Powell, T. ja Humphreys, B. L., The unified medical language system (UMLS) project. Teoksessa *Encyclopedia of Library and Information Science*, Kent, A. ja Hall, C. M., toimittajat, Marcel Dekker Inc., 2002, sivut 369–378.
- NSD<sup>+</sup>01 Noy, N. F., Sintek, M., Decker, S., Crubezy, M., Ferguson, R. W. ja Musen, M. A., Creating semantic web contents with Protege-2000. *IEEE Intelligent Systems*, 2,16(2001), sivut 60–71.
- PGB02 Plaisant, C., Grosjean, J. ja Bederson, B. B., SpaceTree: Supporting exploration in large node link tree, design evolution and empirical evaluation. *INFOVIS '02: Proceedings of the IEEE Symposium on Information Visualization (InfoVis '02)*, Washington, DC, USA, 2002, IEEE Computer Society, sivu 57.
- PWG05 Parsia, B., Wang, T. ja Golbeck, J., Visualizing web ontologies with Cropcircles. *Proceedings of the ISWC-2005 Workshop End User Semantic Web Interaction*, 2005.
- RB03 Rivadeneira, W. ja Bederson, B., A study of search result clustering

interfaces: Comparing textual and zoomable user interfaces. Tekninen raportti, University of Maryland HCIL, 2003.

- RMC91 Robertson, G. G., Mackinlay, J. D. ja Card, S. K., ConeTrees: animated 3D visualizations of hierarchical information. *CHI '91: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 1991, ACM, sivut 189–194.
- Sal06 Salminen, M., Kuvien ja videoiden semanttinen sisällönkuvailu. Pro gradu, Helsingin yliopisto, 2006.
- SAvA<sup>+</sup>06 Schreiber, G., Amin, A., van Assem, M., de Boer, V., Hardman, L., Hildebrand, M., Hollink, L., Huang, Z., van Kersen, J., de Niet, M., Omelayenko, B., van Ossenbruggen, J., Siebes, R., Taekema, J., Wielemaker, J. ja Wielinga, B. J., MultimediaN E-Culture Demonstrator. *The Semantic Web - Proceedings of the 5th International Semantic Web Conference 2006*. Springer, 2006, sivut 951–958.
- SDWW01 Schreiber, A. T. G., Dubbeldam, B., Wielemaker, J. ja Wielinga, B., Ontology-based photo annotation. *IEEE Intelligent Systems*, 16,3(2001), sivut 66–74.
- SEA<sup>+</sup>02 Sure, Y., Erdmann, M., Angele, J., Staab, S., Studer, R. ja Wenke, D., OntoEdit: Collaborative ontology engineering for the semantic web. *Proceedings of the First International Semantic Web Conference 2002 (ISWC 2002), June 9-12 2002, Sardinia, Italia*, Horrocks, I. ja Hendler, J., toimittajat, osa 2342 sarjasta *LNCS*. Springer, 2002, sivut 221–235.
- Shn92 Shneiderman, B., Tree visualization with tree-maps: 2-D space-filling approach. *ACM Trans. Graph.*, 11,1(1992), sivut 92–99.

- Shn96 Shneiderman, B., The eyes have it: A task by data type taxonomy for information visualizations. *IEEE Visual Languages*, numero UMCP-CSD CS-TR-3665, College Park, Maryland 20742, USA, 1996, sivut 336–343.
- Sin Sintek, M., OntoViz. URL <http://protege.cim3.net/cgi-bin/wiki.pl?OntoViz>.
- SMKH08 Sinkkilä, R., Mäkelä, E., Kauppinen, T. ja Hyvönen, E., Combining context navigation with semantic autocompletion to solve problems in concept selection. *Proceedings of the Workshop on Semantic Metadata Management and Applications (SEMMA 2008) at the Fifth European Semantic Web Conference ESWC2008*, 2008.
- SMS+01 Storey, M., Musen, M., Silva, J., Best, C., Ernst, N., Fergerson, R. ja Noy, N., Jambalaya: Interactive visualization to enhance ontology authoring and knowledge acquisition in protege, Workshop on Interactive Tools for Knowledge Capture, 2001.
- TH06 Tzitzikas, Y. ja Hainaut, J.-L., On the visualization of large-sized ontologies. *AVI '06: Proceedings of the working conference on Advanced visual interfaces*, New York, NY, USA, 2006, ACM, sivut 99–102.
- Tuf01 Tufte, E. R., *The Visual Display of Quantitative Information*. Graphics Press, 2001.
- vK07 van Kesteren, A., The XMLHttpRequest object. a WD in last call, W3C, 2007. <http://www.w3.org/TR/2007/WD-XMLHttpRequest-20070227/>.
- VTH08 Viljanen, K., Tuominen, J. ja Hyvönen, E., Publishing and using ontologies as mash-up services. *Proceedings of the 4th Workshop on Scrip-*

*ting for the Semantic Web (SFSW2008), 5th European Semantic Web Conference 2008 (ESWC 2008), 2008. Forthcoming.*

- YSLH03 Yee, K.-P., Swearingen, K., Li, K. ja Hearst, M., Faceted metadata for image search and browsing. *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, New York, NY, USA, 2003, ACM, sivut 401–408.