

**TEKNILLINEN KORKEAKOULU**  
**Viestintätekniikka**  
**Diplomityö**

**Palveluiden semanttinen kuvailu ja haku**

Espoossa 19.8.2009

Mikko Salonoja  
msalonoj@cc.hut.fi  
54799D

Tekijä, työn nimi	
Mikko Salonoja	
Palveluiden semanttinen kuvailu ja haku	
Päivämäärä: 19.8.2009	Sivumäärä: 69 + 15
Tiedekunta	Professuuri
Elektroniikan, tietoliikenteen ja automaation tiedekunta	AS-75 Viestintäteknikka
Työn valvoja	
Professori Eero Hyvönen	
<p>Ihminen tarvitsee kuluttajana ja yrittäjänä usein toisten ihmisten ja organisaatioiden tukea palveluiden muodossa. Aiemmin tuntemattomien palveluiden löytäminen voi kuitenkin olla vaikeaa Internetin tiedon määrän lisääntyessä, koska tiedonhakijalle mielekäs sisältö hautautuu helposti epäolennaisen tiedon joukkoon. Tämän ongelman helpottamiseksi semanttisen webin tekniikat tarjoavat uusia mahdollisuuksia.</p> <p>Tässä diplomityössä tutkittiin millaisia palveluiden haun kannalta hyödyllisiä ratkaisuja on aiemmin toteutettu. Tämän jälkeen nämä ratkaisut peilattiin kahden Internetissä sijaitsevan palvelun, PKT-säätiön ylläpitämä Yrityksen palveluhakemisto ja Suomen Asiakastiedon ylläpitämä Aarre-palvelu, parannusehdotuksiksi.</p> <p>Tutkittuja palveluita ehdotettiin tässä diplomityössä parannettavan useilla erilaisilla tavoilla. Näistä osalla ei ollut suoraan mitään tekemistä semanttisen webin tekniikoiden kanssa ja osa taas liittyi semanttiseen webiin hyvin kiinteästi. Erityisesti maantieteellisen tiedon käsittelyssä havaittiin olevan runsaasti kehitettävää molempien tutkittujen palveluiden kohdalla. Myös eroja parannusehdotusten välillä oli havaittavissa. Johtuen Aarre-palvelun suuremmasta tietokannasta ja suuremmasta käyttäjäjoukosta siinä sanojen välisen verkon hyödyntäminen näytti olevan merkityksellisempää kuin Yrityksen palveluhakemiston kohdalla.</p> <p><b>Avainsanat:</b> Semanttinen web, palvelut, palveluluokitukset</p>	

Author, Name of the Thesis

Mikko Salonoja

Semantic description and search of services

Date: August 19<sup>th</sup>, 2009

Number of pages: 69 + 15

Faculty

Professorship

Faculty of Electronics, Communications and Automation

AS-75 Media Technology

Supervisor

Professor Eero Hyvönen

Being a consumer and an entrepreneur a human being often needs support of other people and organizations in the form of services. However, previously unknown services may be difficult to find. This is so because relevant content for a person is easily buried among irrelevant information as the quantity of information in Internet is growing. To ease this problem, semantic web techniques offer new possibilities.

In this master's thesis it was examined what kind of useful solutions have been previously implemented from the service search point of view. Then these found solutions were utilized in suggesting improvements for two Internet services: 1) Yrityksen palveluhakemisto maintained by SME Foundation and 2) Aarre-service maintained by Suomen Asiakastieto.

Examined services were suggested to be improved in many different ways in this master's thesis. Some of these suggestions were not related to the semantic web at all and some were related very closely. Especially there was a lot to be improved in the management of geographical information in both services. In addition, it turned out that there were differences between suggested improvements. Utilizing the network of words seemed to be more important in the Aarre service than in Yrityksen palveluhakemisto. This happened to be so because the Aarre service had a bigger database and more users.

**Keywords:** Semantic Web, services, service classification

## **Alkusanat**

Tämä diplomityö tehtiin osana FinnONTO 2.0 -projektia Semanttisen laskennan tutkimusryhmässä Teknillisessä korkeakoulussa. Haluan kiittää koko projektiorganisaatiota ja erityisesti sen työssä mukana olleita ihmisiä ja organisaatioita (PKT-säätiö, Osuuspankkikeskus ja Suomen Asiakastieto). Semanttisen laskennan tutkimusryhmästä haluan erityisesti kiittää professori Eero Hyvöstä, Osma Suomista, Antti Tania, Jouni Tuomista, Kim Viljasta, Matias Frosterusta, Jussi Kurkea ja Reetta Sinkkilää. Lisäksi erityinen kiitos PKT-säätiön Jarmo Koverolle ja TIEKE:n Timo Simellille positiivisesta vaikutuksesta diplomityöni kokeellisen osuuden edistymiseen.

Espoossa 19.8.2009

Mikko Salonoja

# SISÄLLYSLUETTELO

1 JOHDANTO.....	1
1.1 Aiheen yleiskuvaus.....	1
1.2 Tutkimusongelmat ja niiden rajaus.....	2
1.3 Tutkielman rakenne.....	3
2 SEMANTIikka PALVELUIDEN KUVAILUSSA.....	3
3 ASIASANOITUS JA ONTOLOGIAT.....	4
3.1 Asiasanoitus.....	4
3.2 Ontologiat.....	5
3.2.1 Yleistä.....	5
3.2.2 Yläontologiat.....	5
3.3 Sumeus asiasanoituksessa ja ontologioissa.....	6
4 SEMANTTISEN TIEDON ESITTÄMINEN.....	7
4.1 RDF JA RDFS.....	7
4.1.1 RDF.....	7
4.1.2 RDFS.....	7
4.1.3 RDF:n ja RDFS:n ilmaisuvoima.....	7
4.2 OWL.....	8
4.3 SKOS.....	9
4.4 WSMO.....	10
5 PALVELUIDEN KUVAILUN TYÖVÄLINEITÄ.....	11
5.1 Palveluiden kuvailu.....	11
5.2 Semanttisen tiedon tuottamisessa käytetyt työkalut.....	12
5.2.1 ONKI.....	12
5.2.2 Protégé.....	13
5.2.3 SAHA.....	14
5.3 Sanastot palveluiden kuvailussa.....	14
5.3.1 Käytetyt sanastot palveluiden kuvailussa.....	14
5.3.2 Sanastojen väliset muunnokset.....	16
5.3.3 Sanastojen muokkaus.....	17
6 PALVELUIDEN HAKU.....	19
6.1 Tekstihaku.....	19
6.1.1 Yksinkertaiset hakusovellukset.....	19
6.1.2 Automaattinen kyselyn laajentaminen.....	20
6.1.3 Hakusanojen osilla hakeminen.....	22
6.2 Maantieteellisen sijainnin hyödyntäminen.....	23
6.2.1 Hyödyntämiskohteet.....	23
6.2.2 Ongelmallisten kohteinen visualisointi.....	23
6.2.3 Palvelutiedon maksullisuus.....	25
6.3 Tutkiva tiedonhaku.....	26
6.3.1 Yleistä.....	26
6.3.2 Yhdistävät polut tulosten yhteydessä.....	27
6.3.3 Haun tarkentaminen valmiiden vaihtoehtojen avulla.....	28
6.3.4 Moninäkömahaku.....	29
6.3.5 Hakutermien monisuuntaisten suhteiden visualisointi.....	31
7 AARRE-PALVELUN JA YRITYKSEN PALVELUHAKEMISTON KEHITYS.....	32

7.1	Tavoitteet palveluhakemistojen kehityksessä.....	32
7.2	Lähestymistavat aineistoihin.....	33
7.3	Aineistojen sisältö.....	33
7.4	Aineistojen sisällönkuvailu.....	34
8	AINEISTOJEN JA NIIDEN KUVAILUSSA TARVITTAVIEN SANASTOJEN KAR- TOITUS.....	35
8.1	Yrityksen palveluhakemisto.....	35
8.1.1	Sisällön hahmottaminen hakusovelluksen avulla.....	35
8.1.2	Palveluiden kuvailu hakusovellusta varten.....	36
8.2	Aarre-palvelu.....	38
8.2.1	Toimialaluokitus kuluttajille suunnatun sovelluksen sanastona.....	38
8.2.2	Olemassaolevien sanastojen soveltuvuusarviointi.....	39
8.2.3	Kuluttajaontologian kehitys.....	41
8.2.4	Kuluttajaontologian mielekkyyden testaus.....	42
8.2.5	Kuluttajaontologian käsitteiden sisäinen linkitys.....	44
9	KEHITYSPROJEKTIEEN PROTOTYYPIT.....	46
9.1	Semanttinen web käytännössä.....	46
9.1.1	WSMO.....	46
9.1.2	Exhibit-julkaisukehikon rajoitteet.....	46
9.1.3	Kuluttajapalveluiden kuvailu ja HAKO.....	47
9.1.4	Datan tuonti Exhibit-julkaisukehikosta HAKO:ssa julkaistavaksi.....	48
9.1.5	HAKO toiminnassa.....	49
9.2	Prototyyppien käyttöliittymät.....	51
9.2.1	Yleistä.....	51
9.2.2	Tekstihaku yhdistettynä moninäkömähakuun.....	51
9.2.3	Valmiiden valintojen visualisointi.....	52
9.2.4	Palveluiden sijaintitiedon hyödyntäminen.....	53
9.2.5	Tekstit ja tyylit.....	55
10	YHTEENVETO.....	58

## LIITTEET (8):

1. Tärkeimpiä ONKI:ssa julkaisun määräytyksiä
2. Kehitetyn sovelluksen tægien peilaus KOKO-ontologiaan ja UNSPSC-luokitukseen
3. KOKO-ontologian käsitteitä liitettäväksi yrittäjille suunnattujen palveluiden kuvauksiin
4. UNSPSC:n CSV→OWL-muunnos
5. UNSPSC-luokituksen SKOS-muodon skeema
6. Turtle→OWL-muunnoksen Java-koodi
7. Demosovelluksen datan muuntamisessa käytetty XSL-tiedosto
8. Katuosoitteiden muuntaminen koordinaateiksi

# 1 JOHDANTO

## 1.1 Aiheen yleiskuvaus

Internetin merkitys tiedonhaussa on kasvanut viime vuosina /93/. WWW:n kautta on saatavilla runsaasti tietoa ja sen kautta Internetin käyttäjät yrittävät hakea ratkaisua mitä erilaisimpiin tarpeisiinsa. Toisaalta Internetin käyttö voi olla luonteeltaan myös vähemmän tarkoitushakuista: samoilua samaan tapaan kuin kaupunkien keskustoissa.

Internet voi olla käyttäjälleen tiedonhaun lisäksi myös sosiaalinen kohtaamispaikka tai itseilmaisun media. Näistä esimerkkeinä ovat mm. Facebook /29/, Myspace /86/ ja erilaiset keskustelupalstat kuten Suomi24 /120/. Toisaalta Internetistä voidaan myös hakea tietoa uutisten tai sähkökirjojen muodossa. Näistä molemmat voivat olla käyttäjälleen ilmaisia tai maksullisia riippuen tiedon sisällöstä. Maksullisen sisällön kohdalla puhutaan usein verkkokaupasta, jota käydään myös erilaisilla tavaroilla. Jotkin tavarat, kuten kirjat, sopivat erityisen hyvin mainittuun verkkokauppaan, jos niiden markkinat ovat pirstaleiset ja/tai niiden koko on pieni suhteessa niiden arvoon. Kaikkien tavaroiden ja varsinkin palveluiden ostaminen ilman kasvokontaktia WWW:n kautta ei ole kuitenkaan mielekäästä, mutta Internet pystyy tarjoamaan myös tälle alueelle lisäarvoa käyttäjälleen: helposti löydettäviä yhteystietoja. Tähän haasteeseen ovat oman ratkaisunsa tarjonneet mm. erilaiset hakukoneet, kuten Google /40/, ja puhelinluetteloista sähköiseen muotoon muutetut keltaiset sivut, jotka ovat paperisia edeltäjiään huomattavasti laajempia. Jälkimmäisistä esimerkkinä mainittakoon suomalainen Yritysopas-sivusto /119/.

Internetin tiedon määrän lisääntyessä ongelmaksi sen käytössä nousee tietotulva. Hakiesseen jotain tiettyä tietoa käyttäjä törmää epäolennaiseen informaatioon /63/, jonka osuus hakutuloksista voi nousta usein suuremmaksi kuin olennaisen tiedon. Näin käy johtuen seuraavista syistä: 1) suurin osa hakukoneiden kyselyistä on lyhyitä /65/, 2) valtaosa Internetin navigointi-istunnoista alkaa hakukoneen kautta /91/ ja 3) tietokone ei ymmärrä hakusanojen merkityksiä eikä niiden mahdollisia suhteita, joita voi olla mm. osa-, synonyymisekä ylä- ja alakäsitesuhteet. Vapaissa tekstihauissa hakusanojen valinnassa on siis syytä käyttää harkintaa. Tiedonhakija voi vaikuttaa lopputulokseen myös lisäämällä hakusanoja ja kokeilemalla niiden vaikutusta, joita keinoja käytetään asiantuntijoille tarkoitetuissa järjestelmissä. Tavalliselle ihmiselle nämä keinot eivät kuitenkaan välttämättä sovi kompleksisuutensa vuoksi /25/.

Tietokoneen ymmärtämien semanttisten metatietojen suppeus Internetin sisällöstä haittaa myös tietojen yhdistelyä. Internetin toiminnan kehittämisen kannalta olisi edullista, jos olisi olemassa yhteinen kieli, jolla voitaisiin kuvata tietojen merkityksiä ja niiden liittymistä toisiinsa. Esimerkiksi palveluista tiedonhakija voi haluta tietää pelkkien yhteystietojen lisäksi myös palveluiden käyttökokemuksia. Tähän tarpeeseen on vastattu mm. eat.fi-palvelussa /37/, mutta käyttökokemukset linkittyvät palveluihin vain sovelluksen sisällä. Internetin sisältöjen lisäksi yhteistä kieltä tarvittaisiin myös yritysten ja yleisesti järjestelmien välisessä tiedonvaihdossa. Koska vaihdettavat tietomäärät ovat usein suuria, tietojen

yhteensovitus käsin voi aiheuttaa liiketoiminnalle merkittävän hidasteen ja kustannuksen. Esimerkkinä tietojen yhteensovituksesta käsin on tilanne, jossa yritys A on nimennyt vaihdettavan tuotteen ominaisuuden nimeksi tuotteen\_nimi ja yritys B vastaavasti nimi\_tuote. Ihminen voi ymmärtää kontekstin perusteella, että kyseessä on sama asia, mutta kone taas ei voi tehdä sitä yhtä helposti ja varmasti. Koneellista päättelyä vaikeuttavat erityisesti monimerkitykselliset sanat, joista kangas, luonnontilainen alue tai vaatteissa esiintyvä materiaali, on hyvä esimerkki. Se mihin merkityksiin tällaiset sanat liittyvät, vaihtelee kielten välillä. Esimerkiksi ranskan avocat-sanalla voidaan tarkoittaa avokadoa tai asianajajaa.

Yhdeksi ratkaisuksi edellä mainittuihin ongelmiin on kaavailtu semanttista webiä Internetin nykyisen infrastruktuurin täydentäjäksi /7/. Pelkästä esitysmuodon kuvauksesta semanttinen web eroaa siten, että semanttisen webin tekniikoilla merkitys voidaan liittää sisältöön koneymmärrettävällä tavalla. Palveluiden haussa tästä mahdollisuudesta on hyötyä yritettäessä vastata käyttäjän tarpeita esittäviin kysymyksiin paremmin, nopeammin ja varmemmin. Parannettavaa mm. yleisiin hakukoneisiin nähden riittää, sillä niissä sopivat hakutulokset löytyvät usein vasta monen yrityksen ja erehdyksen kautta /59/.

## 1.2 Tutkimusongelmat ja niiden rajaus

Tässä Semanttisen laskennan tutkimusryhmässä (SeCo) toteutetussa diplomityössä käsitellään palveluiden kuvailua ja hakua yleisesti ja semanttisen webin tekniikoiden avulla. Palveluilla tarkoitetaan tässä reaali maailman palveluita ja niiden kohderyhmä rajataan kuluttajiin ja yrittäjiin.

Kuluttajapalveluiden lähtökohtana on Suomen Asiakastiedon ylläpitämä Aarre-palvelu /117/, jossa palvelut ovat tyypillisesti maksullisia ja palveluntarjoajien kannalta liiketoimintaa. Palveluihin voi kuluttajien kohdalla liittyä jokin fyysinen tavara mm. ravintolan ja ruokakaupan kohdalla, jolloin puhtaan palvelun osuus voi vaihdella. Tässä yhteydessä ei kuitenkaan tehdä jaottelua tuotteisiin ja palveluihin, vaan tuotteiden ymmärretään aina sisältyvän palveluihin.

Yrittäjille suunnattujen palveluiden lähtökohtana on taas PKT-säätiön ylläpitämä Yrityksen palveluhakemisto /97/, jonka sisältö on yhdistelmä julkisen ja yksityisen sektorin palveluita. Hakemistossa esitellyt palvelut liittyvät mm. yritysten väliseen yhteistyöhön ja yritystoiminnan yleiseen kehittämiseen.

Tutkielman keskeiset tutkimuskysymykset ovat:

- Miten Suomen Asiakastiedon ylläpitämän Aarre-palvelun käytettävyyttä voidaan parantaa?
- Miten PKT-säätiön ylläpitämän Yrityksen palveluhakemiston käytettävyyttä voidaan parantaa?
- Miten semanttisen webin tekniikoita voidaan mielekkäästi soveltaa palveluhakusoveluksissa?



- Miten palveluhaun hakujoukkoa tulisi kuvailla, jotta semanttisen webin tekniikoita voitaisiin mielekkäästi soveltaa palveluhaun käyttöliittymässä?
- Miten tulisi muodostaa palveluita kuvailevien käsitteiden joukko, jotta palveluiden kuvailussa ja haussa voitaisiin mielekkäästi soveltaa semanttisen webin tekniikoita?

Ratkaisuksi näihin ongelmiin esitetään seuraavia (molempiin kohderyhmiin liittyen):

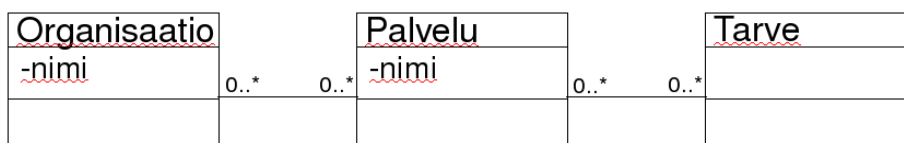
- Palveluiden käyttöön liittyvien kontekstien kuvaaminen. Tämän oletetaan helpottavan joissain tapauksissa ja joillain tavoin annotointia ja tiedonhakua.
- Tarkoitukseen sopivat käyttöliittymäkomponentit. Näihin liittyy mm. valintahaussa valintojen vaikutusten näyttäminen, maantieteellisen tiedon kohdalla karttakuvat sekä erilaiset rajaus-, osoitus- ja navigointimahdollisuudet.

### 1.3 Tutkielman rakenne

Tutkielma alkaa toisessa luvussa lyhyellä esittelyllä semantiikan tuomisesta palveluiden kuvailuun. Kolmannessa luvussa käsitellään yleisesti palveluiden annotointia ja annotoinnissa käytettyjen arvojoukkojen rakenteita. Neljännessä luvussa taas esitellään semanttisen webin menetelmiä, joita voidaan käyttää palveluiden rakenteiden ja niihin liittyvien arvojoukkojen kuvailussa sekä haussa. Viidennessä luvussa esitellään tekniikoita ja aiempia toteutuksia liittyen palveluiden kuvailuun ja kuudennessä luvussa vastaavasti hakuun. Työn kokeellinen osuus, jossa testataan keinoja parantaa kuluttajille ja yrittäjille suunnattujen palveluhakemistojen käytettävyyttä, kuvataan seitsemännessä, kahdeksannessa ja yhdeksännessä luvussa. Lopuksi esitetään yhteenveto koko tutkielmasta.

## 2 SEMANTIikka PALVELUIDEN KUVAILUSSA

Palveluiden kuvauksissa on aina jokin rakenne, jonka mukaan palveluiden sisältö kuvataan. Esimerkiksi kuvassa 1 palvelut on määritelty kokonaisuuksiksi, joilla on nimen lisäksi yhteyksiä organisaatio- ja tarvekokonaisuuksiin.



**Kuva 1.** Esimerkki palvelukuvausten rakenteesta UML-luokkakaaviona

Kuvan 1 kuvaustapa on selkeä, mutta siinä ei ole yksikäsitteisiä merkityksiä liitettyinä tietoon. Semanttisen webin tekniikoiden avulla nämä puutteet voidaan kuitenkin korjata seuraavilta osin:

- Oliot (organisaatio ja palvelu). Esimerkiksi voidaan kuvata, että organisaatiolla tarkoitetaan tässä samaa kuin FOAF:ssa (foaf:Organization-luokka) /12/.
- Olioiden ominaisuudet (organisaatio→nimi jne.). Esimerkiksi voidaan kuvata, että organisaation ja palvelun nimellä tarkoitetaan tässä samaa asiaa, joka on kuvattu myös

Dublin Coressa (dc:title-ominaisuus) /21/ ja JHS 143:ssa (Julkisen hallinnon suositusjärjestelmä - Asiakirjojen kuvailun ja hallinnan metatiedot) /138/.

- Olioiden ominaisuuksien arvot. Esimerkiksi voidaan kuvata, että hiusten leikkaus -palvelu tyydyttää hiustenhoito-tarpeen, missä hiustenhoidolla tarkoitetaan samaa asiaa kuin YSA:ssa (Yleinen suomalainen asiasanasto) /123/. YSA:ssa taas hiustenhoitoon on liitetty merkityksiä (mm. käännöksiä ja assosiativisia asiasanoja), mikä entisestään lisää mainitun hiusten leikkaus -palvelun informaatioisisältöä.

Näistä viimeinen on palveluiden haun kannalta mielenkiintoisinta. Tiedonhaussa käyttäjä ilmaisee hakuehdon usein pelkkänä arvolistana tarkentamatta oliota tai ominaisuutta, johon haku kohdistuu. Näin mm. siksi, koska ominaisuuksien arvoilla kuvataan palveluiden haun kannalta mielenkiintoisin ja rikkain informaatio.

## 3 ASIASANOITUS JA ONTOLOGIAT

### 3.1 Asiasanoitus

Palveluiden sisältöä voidaan kuvata vapailla asiasanoilla tai käyttäen kontrolloitua sanastoa. Jälkimmäinen voi palveluiden kohdalla olla esimerkiksi lista aihepiiriin liittyviä oleellisia sanoja /38/, jotka ovat usein pyritty valitsemaan siten, että ne ovat sanaston sisällä yksilöiviä ja vakioituja. Näin siksi, koska sanat itsessään voivat olla monimerkityksellisiä. Esimerkkinä kontrolloidusta sanastosta mainittakoon Yleinen suomalainen asiasanasto (YSA) /123/, jonka avulla palvelun A tyydyttävä tarve voidaan yksilöidä hiustenhoito-asiasanalla. YSA:n tapauksessa samaan asiasanaan voidaan myös viitata käyttäen sen ONKI-version /123/ URI:a (Uniform Resource Identifier /10/): <http://www.yso.fi/onto/ysa/Y95006>. Tällainen sanaston tunniste viittaa aina johonkin resurssiin, joka voi olla luonteeltaan käsite (esim. mainittu hiustenhoito) tai yksilö. Yksilö on mm. Helsingin rautatieasema, joka voi liittyä mainittuun palveluun esimerkiksi palveluntarjoajan toimipisteen sijainnin kautta.

Sanastoissa palveluiden kuvauksen kannalta on merkityksellistä itse sanojen lisäksi niiden väliset suhteet. Asiasanastoissa näitä on kuvattu tyypillisesti seuraavasti /59, s. 4-5/:

- Ekvivalenssi. Esimerkki: mänty ja honka ovat synonyymeja.
- Hierarkia. Esimerkki: puu on männyn yleistys.
- Assosiativisuus. Esimerkki: veroilmoitus liittyy veroon.

Tietoa voidaan kuvailla myös tägeillä, joiden käyttö voi yhtenäistyä yhteisössä. Näin muodostuneita sanastoja nimitetään folksonomioiksi /85, s. 134-141/, joilla on merkitty mm. YouTuben (suosittu sivusto videoiden jakamiseen) /43/ sisältö.

Kysymykseen siitä kumpia, vapaita tägejä vai kontrolloituja asiasanoja, kannattaa soveluksissa käyttää, ei näytä olevan yksiselitteistä vastausta. Toisaalta kontrolloidulla sanastolla voidaan välttää kirjoitusvirheitä /105/ ja sanojen taivutukseen liittyviä karikoita, mutta toisaalta joidenkin tutkimusten mukaan kontrolloitu sanasto ei tehosta hakuja /57, 87/. Srinivasanin mukaan taas aineistohaku on tehokkainta, kun näitä molempia yhdistetään

/114/. Tämä kuulostaakin loogiselta, sillä tÄgien voidaan ajatella tÄydentÄvÄn kontrolloitua asiasanastoa ennalta-arvaamattomalla tavalla. Toisaalta kontrolloidun, ennalta-mÄÄrÄtyn ja rajoitetun sanaston kÄyttö voi selkeyttÄÄ tiedonkuvailua. TÄstÄ on hyötyÄ mm. moninäky-mÄhakusovelluksissa, joissa kÄyttÄjÄlle tarjotaan joukkopohjaisia valintavaihtoehtoja. Ilmaisuvoimaltaan kehittynyt sanasto voi myöS pitÄÄ sisällÄän paljon informaatiota. TÄtÄ hyödyntÄmällä voidaan parantaa haun kÄytettÄvyyttÄ ainakin joissain sovelluskohteissa riippuen kÄytetyn sanaston sisÄisen informaation mÄÄrÄstÄ ja laadusta.

## 3.2 Ontologiat

### 3.2.1 YleistÄ

Ontologia-sana tulee alunperin filosofiasta, jossa sitÄ kÄytetään puhuttaessa olemassaolon tutkimisesta /7/. Tietotekniikassa ontologialla taas tarkoitetaan ilmaisuvoimaltaan kehittyntÄ asiasanastoa /47/, jonka avulla voidaan kuvata kÄsitteitÄ ja niiden vÄlisiÄ suhteita perinteistÄ asiasanastoa rikkaammin, mistÄ johtuen ontologiaan viittaavasta sisällöstÄ on mahdollista pÄÄtellä enemmÄn. Esimerkiksi seuraavanlaista ontologista tietoa ei ole tyyppillisesti kuvattu asiasanastoissa:

- Muut kuin ekvivalenssi-, hierarkia- ja assosiaatiosuhteet resurssien vÄlillä. Esimerkki: sijaitsee-suhde.
- Formaalien rajoitteiden kuvaukset suhteille. Esimerkki: jokin sijaitsee jossain, missÄ jokin on objekti ja jossain on paikka.

Ontologioilla yleensÄ rajoitutaan kuvaamaan vain jotain tiettyÄ sovellusaluetta, johon halutaan rakentaa tietoteknisiÄ sovelluksia. LisÄksi kuvaaminen tehdÄän yleensÄ vain sillÄ tarkkuudella, kuin se on tarpeellista lähtien tarpeista, joita tietoteknisillÄ sovelluksilla pyritÄän tyydyttÄmÄän /69, s. 2/.

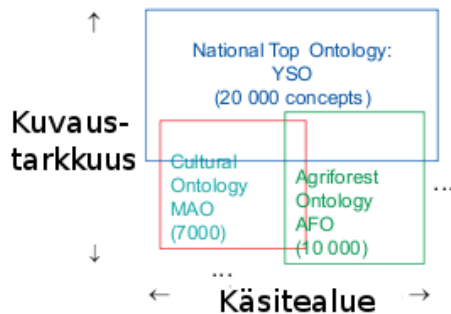
### 3.2.2 Yläontologiat

Olemassaolon ja sen rakenteen hahmottamisessa on tarkkuuseroja, mihin haasteeseen on vastattu kehittÄmällä yläontologioita /69, s. 2/. Näillä tarkoitetaan ontologioita, joiden tarkoitus on kuvata kÄsitteiden ylintÄ tasoa. TÄllÄ tasolla voidaan puhua esimerkiksi autoista yleisesti ja jättÄÄ tarkentavien kÄsitteiden kuvaaminen aihealuekohtaisiin ontologioihin. TÄssÄ aihealueellisesti rajoitettuna ontologiana voitaisiin ymmärtÄÄ esimerkiksi autoontologia, jossa olisi kuvattu erityyppisiÄ autoja osineen, materiaaleineen ja lisÄvarusteineen lähtien auto-kÄsitteestÄ.

Usean ontologian tapauksessa ontologiat voidaan yhdistÄÄ toisiinsa, mitÄ kutsutaan ontologioiden siltaamiseksi /68/. Aiemmin kuvatussa kuvitteellisessa auto-ontologiassa auto-kÄsite voitaisiin ilmaista kuvaavan samaa asiaa kuin yläontologian auto-kÄsite. Ontologiat on mielekÄstÄ sillata myöS niissä tapauksissa, joissa kÄsitteiden kirjoitusasut ovat identtiset, koska osa sanoista on monimerkityksellisiÄ.

Esimerkkeinä kansainvÄlisistÄ yläontologioista mainittakoon seuraavat: 1) OPENCYC /17/. Ontologia on alunperin kehitetty NLP-menetelmien (Natural Language Processing)

tietämysvarastoksi; 2) SUMO (Suggested Upper Merged Ontology) /96/ ja 3) DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering) /35/. Kotimaisista yläontologioista tärkein lienee YSO (Yleinen suomalainen ontologia) /61/, jonka avulla on liitetty yhteen eri alojen käsitteistöjä. Yhdessä eri alojen käsitteistöjen kanssa YSO muodostaa KOKO-ontologian (kuvassa 2).



**Kuva 2.** KOKO-ontologia /61/

Mainittujen yläontologioiden, ulkomaisten ja YSO:n, välillä on havaittavissa merkittäviä eroja mm. käsitteiden määrissä. Esimerkiksi SUMO sisältää hieman yli 10000 käsitettä, kun taas OPENCYC:ssä käsitteitä on tähän verrattuna kertaluokkaa enemmän.

### 3.3 Sumeus asiasanoituksessa ja ontologioissa

Kiinnitettäessä asiasanoja annotoitavaan kohteeseen voi olla mielekästä erilaistaa asiasanojen liittymisen luonnetta. Esimerkiksi kuvailtaessa yksiväristä punaista paitaa ja punaraitaista paitaa liittyy punainen väri niihin eri tavoin. Toisaalta kuvailussa voitaisiin käyttää myös ontologiaa, jos se sisältäisi punainen- ja punaraitainen-käsitteet. Tällöin punaisen värin suhde voitaisiin kuvata ontologian sisällä.

Sumeasta tiedosta puhuttaessa käytetään tiedon kuvailussa painokertoimia, joiden arvot voivat perustua mm. osuuteen kokonaisuudesta /107, s. 10-11, 36-37/. Esimerkiksi jos edellä mainittu punaraitainen paita olisi väriltään punavalkoinen ja sen raidat olisivat yhtä paksuja, voitaisiin merkitä seuraavasti:

- punainen paita: punainen
- punaraitainen paita: punainen (0,5), valkoinen (0,5)

Suhteiden sumeuden kuvauksessa ei ole aina mielekästä mahdollistaa painokertoimen vapaata määrittystä mm. siksi, koska tietoa käsin kuvailevilla ihmisillä on erilaisia taipumuksia käyttää painokertoimia. Näiden suora käyttö voidaan ontologiassa kiertää mm. suhdehierarkian luonnilla seuraavasti: voitaisiin esimerkiksi määritellä liittyy-suhde ja tälle alisuhteet ”liittyy voimakkaasti” ja ”liittyy osittain”. Tällöin merkittäisiin:

- punainen (liittyy voimakkaasti) punainen paita
- punainen (liittyy osittain) punaraitainen paita
- valkoinen (liittyy osittain) punaraitainen paita

## 4 SEMANTTISEN TIEDON ESITTÄMINEN

### 4.1 RDF JA RDFS

#### 4.1.1 RDF

RDF-kielellä (Resource Description Framework) kuvataan mm. WWW:n sisältöjä. Kieltä kehittää W3C (World Wide Web Consortium), joka on Tim Berners-Leen (WWW:n luoja) perustama konsortio. Konsortio määrittelee tämän laajalti käytössä olevan kielen Recommendation-suosituksella. /78/

RDF-kielellä kuvataan väittämiä seuraavasti: Jokaisessa RDF-kielen väittämässä kerrotaan jotain jostain luonnollisen kielen väitelauseen tapaan. Esimerkiksi korva on osa päättä -väitelause on kuvattavissa RDF:llä seuraavasti: korva (subjekti), on osa (predikaatti) ja päättä (objekti). Jokainen RDF-kielen väittäminen on siis subjektin, predikaatin ja objektin muodostama kolmikko. Nämä kaikki osat voivat olla tyypiltään resursseja eli käsitteitä tai käsitteen ilmentymiä. Resurssin lisäksi ominaisuuden (=predikaatin) arvoksi voidaan kiinnittää myös kiinteä arvo, jolla tarkoitetaan tässä tiedon alkeistyyppisiä kuten merkkijonoja ja lukuja. Nyt kuvatussa esimerkissä sekä subjekti että objekti voidaan ymmärtää tyypiltään käsitteiksi. /70/

RDF-väittämien muodostamasta joukosta voidaan hakea tietoa useiden eri kyselykielien avulla, joista W3C:n kehittämä SPARQL (SPARQL Protocol and RDF Query Language) /99/ on käytössä yleisin. Konsortio on standardoinut sen Recommendation-suosituksena. Kielen oppimiskynnys on suhteellisen matala, sillä se muistuttaa SQL:ää, jota käytetään yleisesti tietokantakyselyissä.

#### 4.1.2 RDFS

RDFS eli RDF Schema on skeemakieli, jolla tarkennetaan RDF-väittämien sisältöä. Samoin kuin RDF myös RDFS on W3C:n kehittämä ja suosittama /78/. RDF-väittämien sisältöä voi tarkemmin kuvata RDFS:llä mm. seuraavilla primitiiveillä /13/:

- rdfs:Class ja rdfs:subClassOf kuvaavat resurssit käsitteiksi ja niiden välisen hiarkian.
- rdfs:Literal mahdollistaa tarkentamaan ominaisuuden arvon literaalityyppiseksi.
- rdfs:subPropertyOf mahdollistaa hiarkian kuvaamisen predikaateille.
- rdfs:domain- ja rdfs:range-määrittelyillä on mahdollista rajoittaa RDF-väittämien subjektin ja objektin arvojoukkoja predikaattikohtaisesti (kts. 2.2 Ontologiat).
- rdfs:label ja rdfs:comment mahdollistavat ihmisluettavan tiedon liittämisen resursseille.

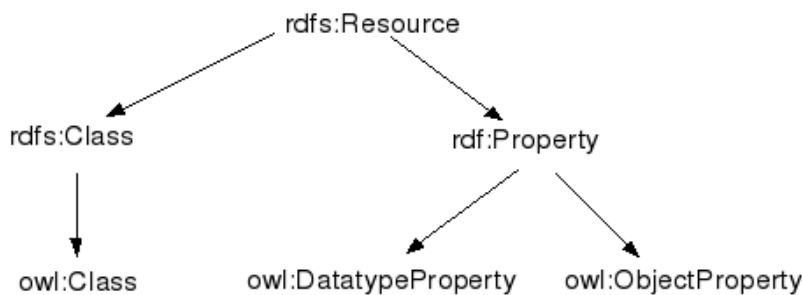
#### 4.1.3 RDF:n ja RDFS:n ilmaisuvoima

RDFS ja RDF yhdessä mahdollistavat verkkomaisen tietorakenteen ja yksinkertaisen olio-perusteisen semantiikan ilmaisemisen formaalisti, mutta tässä yhdistelmässä on ilmaisuvoiman suhteen selkeitä rajoitteita. Näitä ovat mm. seuraavat: /7, s. 111/

- Predikaattikohtaiset rajoitteet voidaan kuvata vain yksinkertaisimmalla mahdollisella tavalla. Mikäli RDF-väitteen subjektina tai objektina on erityyppisiä resursseja, ei tyyppiin perustuvaa rajoitusta voida kuvata.
- Joukko-opin ja Boolean loogisten operaatioiden käyttö käsitteiden ja predikaattien määrittelyssä ei ole mahdollista. Tämä tarkoittaa, että mm. vastakkaisuus-merkitystä ei voida kuvata käsitteelle eikä ominaisuudelle.
- Ominaisuuksien arvoja koskevia lukumäärään perustuvia rajoitteita ei voida kuvata. RDF:n ja RDFS:n yhdistelmällä ei voida mm. ilmaista, että ihmisellä on kaksi kättä.

## 4.2 OWL

Samoin kuin RDF ja RDFS myös OWL (Web Ontology Language) on W3C:n kehittämä suositus /82/. Kieli paikkaa RDF:n ja RDFS:n jättämää aukkoa ilmaisuvoimassa, eli kielet täydentävät toisiaan. Tämä on havaittavissa kuvasta 3, jossa ilmaistaan mm. seuraavaa: 1) on olemassa kahdenlaisia resursseja (rdfs:Resource): luokkia ja ominaisuuksia ja 2) on olemassa tyybiltään kahdenlaisia ominaisuuksia (rdf:Property): owl:DatatypeProperty ja owl:ObjectProperty.



**Kuva 3.** RDF:n, RDFS:n ja OWL:n liittyminen toisiinsa

Laajemman ilmaisuvoiman lisäksi OWL mahdollistaa ontologioiden modulaarisen käytön, sillä viittaaminen muiden luomiin ja muissa ontologioissa sijaitseviin käsitteitteisiin on OWL:ssa mahdollista /112/. OWL-muotoisen tiedon kysely on taas mahdollista mm. OWL-QL:n avulla /31/.

OWL:ssa on otettu huomioon myös laskennallinen hallittavuus /7, s. 112/ rakentamalla siitä kolme ilmaisuvoimaltaan erilaista versiota:

- OWL Full. Koko OWL-kielen sisällään pitävän version etuna suhteessa muihin versioihin on sen laajin ilmaisuvoima ja haittana laskennallinen ratkeamattomuus. Versio pitää sisällään RDF(S):n kokonaisuudessaan.
- OWL DL (OWL Description Logic). Tämän etuna on OWL Full -versioon verrattuna laskennallinen ratkeavuus ja haittana heikompi ilmaisuvoima. Versio on vain osittain yhteensopiva RDF(S):n kanssa.
- OWL Lite. Tämä on ominaisuuksiltaan edelleen karsittu versio OWL DL:stä. Sen etuna on sen yksinkertaisuus ja haittana, luonnollisesti, suppea ilmaisuvoima.

Eri OWL:n versiot ovat tavallaan sisäkkäisiä, sillä jokainen OWL Lite -ontologia on myös oikein muodostettu OWL DL -ontologia ja OWL Full -ontologia. Lisäksi jokainen OWL DL -ontologia on oikein muodostettu OWL Full -ontologia. Yhteensopivuus toiseen suuntaan taas riippuu ontologioiden sisällöstä, eli sitä ei voida taata. /82/

### 4.3 SKOS

SKOS (Simple Knowledge Organization System) on kehikko kontrolloitujen sanastojen määrittämiseen semanttisessa webissä /84/. Toisin kuin OWL-kielessä SKOS:lla pystytään kuvaamaan helposti olemassaolevia epäformaaleita sanastoja ja luokitteluja. Tämän vuoksi SKOS sopii erityisen hyvin palveluluokitusten kuvaamiseen, koska näissä hierarkian merkitys ei ole kovin tiukasti määritetty eikä yhtenäinen. Esimerkiksi UNSPSC:ssä, joka on toimialat ylittävä luokitus tuotteille ja palveluille, kuvataan: 77000000 Ympäristöpalvelut → 77131500 Öljysaaste /48/. Tällä halutaan sanoa, että ympäristöpalvelut on käsitteenä jollain tavalla öljysaasteen laajennus. OWL:n subclassOf-suhteen käyttö olisi tässä harhaanjohtavaa, koska öljysaaste ei ole ympäristöpalvelu.

SKOS on rakennettu RDF:n avulla ja sen ylläpidosta, samoin kuin RDF:n kohdalla, vastaa W3C-organisaatio. SKOS:lla on mahdollista hierarkian lisäksi kuvata mm. 1) käsitteiden nimet sitoen eri kielet yhteen; 2) merkityksiä uudelleenkäyttämällä valmiita RDF-sanastoja, joista esimerkkeinä mainittakoon FOAF (Friend of a Friend) /12/ ja Dublin Core /21/; ja 3) vastaavuuksia käsitteille toisiin (mm. exactMatch-määrittäminen).

Lukuisia sanastoja ja luokituksia on saatavilla SKOS-muodossa. Esimerkiksi van Assem muuntanut SKOS-muotoon Iso-Britanniassa käytetyn IPSV-sanaston (Integrated Public Sector Vocabulary), Alankomaissa audiovisuaalisen materiaalin arkistoinnissa käytetyn GTAA-luokituksen ja lääketieteellisten tekstien luokittelussa käytetyn MeSH-asiasanaston /139/. Osa SKOS-muotoon saatetuista luokituksista on varsin monimutkaisia. Tällainen on mm. MeSH-asiasanasto, jonka SKOS-muotoon muunnoksen yhteydessä SKOS:n eri ominaisuuksia on hyödynnetty erityisen paljon /139/.

Muunnettujen luokitusten määrän ja kirjon perusteella näyttää siis siltä, että SKOS on uskottava vaihtoehto palveluluokitusten kuvaamiseen semanttisessa webissä. Lisäksi luokitusten julkaisu SKOS-muodossa on helppoa mm. SeCo:ssa kehitetyn ONKI-SKOS-julkaisualustan /135/ avulla. Mainitulla alustalla on onnistuneesti julkaistu SKOS-muodossa lukuisia luokituksia ja asiasanastoja, joista tärkeimmät on listattuna taulukossa 1 /123/.

**Taulukko 1.** ONKI-SKOS-julkaisualustalla julkaistuja asiasanastoja ja luokituksia

Nimi	Käsitteitä (kpl)
Yleinen suomalainen asiasanasto YSA	26633
Iconclass	26636
Musiikin asiasanasto MUSA	931
Kaunokki-asiasanasto (fiktiivinen kaunokirjallisuus)	4373
Maailman linnut -ontologia AVIO	11161
Maailman nisäkkäät -ontologia MAMO	6059
Medical Subject Headings MeSH	24355
European Multilingual Thesaurus on Health Promotion HPMULTI	1271
Merenkulkualan asiasanasto MESA	1448
United Nations Standard Products and Services Code UNSPSC	20794
Valtioneuvoston asiasanasto VNAS	6342

#### 4.4 WSMO

W3C:n kehittämät kielet ovat verrattaen yleisessä käytössä. Tästä johtuen niistä löytyy paljon kuvauksia, määrittelyitä, keskustelua ja esimerkkejä Internetistä. Semanttisen tiedon kuvailuun ja käsittelyyn on kuitenkin olemassa muitakin kieliä. Tällainen on mm. WSMO (Web Service Modeling Ontology), joka on DERI:n (Digital Enterprise Research Institute) kehittämä ontologiakieli /30/. Kielen avulla voidaan mallintaa reaali maailman palveluihin liittyvää semantiikkaa ja liittää se web-palveluihin. Näin luodut semanttiset web-palvelut eivät ole itsessään tässä diplomityössä käsiteltyjä palveluita, mutta niitä voidaan hyödyntää apukeinona rakennettaessa sovelluksia reaali maailman palveluiden hakuun.

WSMO:ta hyödynnetään WSMX-sovelluspalvelimen (Web Service Modelling eXecution environment) kautta /23/ ja ilmaistaan WSML-kielen (Web Service Modeling Language) /116/ avulla. Ilmaistut kuvaukset jakautuvat neljään ryhmään:

- Ontologiat (ontology). Näillä voidaan kuvata käsitteitä ja monimutkaisia suhteita niiden välille.
- Web-palvelut (web service). Näillä kuvataan verkossa tarjottavan palvelun toiminnallisuus ja käyttö. Määrittelyksiä näissä ovat:
  - Kyvyt (capability). Näillä kuvataan palvelun toiminta kuten palautettava informaatio sisältö.
  - Rajapinnat (interface). Nämä koostuvat seuraavista:



- Koreografia (choreography). Tämä kertoo palvelun konkreettisen käyttötavan ja sijainnin.
- Orkestraatio (orchestration). Tällä on mahdollista kuvata palveluita, jotka käyttävät toiminnassaan muita palveluita.
- Maalit (goal). Näissä kuvataan käyttäjän tarve jotain ontologiaa käyttäen ja tähän odotetaan vastauksena sopivien palveluiden nimiä.
- Sovittelut (mediation). Näillä kuvataan vastaavuuksia eri resurssien välille. Näin ontologian A avulla määritelty kuvaus web-palvelusta ymmärretään käyttäjän maalin määrittelyssä käyttämän ontologian B mukaisesti. Näin palveluiden tuottajat ja kuluttajat voivat kuvata omaa toimintaansa ja tilaansa itselleen parhaimmin sopivalla käsitteistöllä ja tavalla.

Näiden ryhmien sisällön kuvaaminen WSMML:n avulla on aloittelijalle vaikeaa. Kielen harvinaisuuden vuoksi sen käytöstä löytyy melko vähän esimerkkejä. Lisäksi WSMML:n käyttöä vaikeuttaa myös se, että sitä ei voida kirjoittaa XML:n avulla toisin kuin RDF:n, RDFS:n, SKOS:n ja OWL:n kohdalla, vaan siinä on täysin oma syntaksi. Tästä on esimerkkinä alla osa erään maalin kuvauksesta:

```
wsmmlVariant _"http://www.wsmo.org/wsmml/wsmml-syntax/wsmml-flight"
```

```
namespace { _"http://localhost/XYZgoal#",
            dS _"http://localhost/XYZOntology#",
            dc _"http://purl.org/dc/elements/1.1#" }
```

```
/* XYZ Goal */
```

```
goal _"http://localhost/XYZGoal.wsmml"
```

WSMML-kielellä kuvatut palvelut eivät jää omaksi saarekkeekseen, vaan niitä voidaan käyttää WSMML:stä tarjotun web service -rajapinnan kautta. Lisäksi WSMML-kuvausten kääntäminen OWL-kielelle on mahdollista mm. Innsbruckin yliopistossa kehitetyn kääntäjän avulla /136/.

## 5 PALVELUIDEN KUVAILUN TYÖVÄLINEITÄ

### 5.1 Palveluiden kuvailu

Dokumentteja on kuvailtu perinteisesti aiheeseen perustuen ja toissijaisesti niihin on liitetty myös informatiivisia ja arvioivia yhteenvetoja /51/. Näin on myös palveluiden kohdalla, koska ovathan ne kuvattu usein Internet-sivuina.

Palveluita on mielekästä kuvailla, jos kuvailulla voidaan parantaa tiedonhaun käytettävyyttä enemmän kuin kuvailuun joudutaan näkemään vaivaa. Hyötyä palveluiden semanttisesta kuvailusta voidaan saada mm. annotaatioista päätelystä informaatiosta, sikäli kuin annotaatiot ovat ymmärrettäviä, informaatioisisällöltään hakutarpeen kannalta oleellisia, kattavia ja oikein. Lisäksi saavutettuun hyötyyn vaikuttaa annotaatioihin sisältyvän informaatio-

sisällön suodatusvaikutus. Pääteltyä tietoa voi olla myös termien käännökset, sillä semantiikan hyödyntäminen mahdollistaa monikielisen annotoinnin. Semantiikan avulla voidaan myös joissain tapauksissa nopeuttaa annotointia. Esimerkiksi työkaluja myyvän liikkeen palveluita kuvailtaessa voi olla mielekkäintä listata kaikki työkalut, joista annotoija valitsee tilanteeseen sopivat.

Teknisesti palvelutieto voidaan tuottaa esimerkiksi käyttäen lomakkeistoa, joka sisältää viittauksia kontrolloituun sanastoon, jolla voidaan tarkoittaa mm. asiasanastoa, luokitusta tai ontologiaa. Tämän tarpeen tyydyttämiseksi on kehitetty monenlaisia työkaluja, joista seuraavassa alakappaleessa tarkastellaan ONKI:a, SAHA:a ja Protégé-ohjelmaa. Käytössä nämä kaikki ovat olleet mm. Kulttuurisampo- /60/ ja TerveSuomi-sovelluksia /62/ toteutettaessa.

## 5.2 Semanttisen tiedon tuottamisessa käytetyt työkalut

### 5.2.1 ONKI

Käsitteiden tietovarastoksi ja käyttöä varten on SeCo:ssa kehitetty ratkaisu nimeltä ONKI /140/, joka tarjoaa yksinkertaisen mekanismin käsiteviittauksien hakemiseen minkä tahansa verkkosivun lomakkeen kenttään. Näin ONKI mahdollistaa hajautetun annotoinnin.

ONKI:n käyttöönotto edellyttää kontrolloidun sanaston tai ontologian julkaisua ONKI-palvelimella. Sanastoa julkaistessa syötetietoina annetaan sisältötiedosto OWL- tai SKOS-muodossa (kts. kappale 4) ja joukko sanaston esittämiseen liittyviä määrittämiä (liitteessä 1). Julkaisun lopputulos on nähtävissä kuvassa 4, jossa ONKI-selain on fokusoitu KOKO-ontologian aamutossut-käsitteeseen.

The screenshot shows the ONKI web interface for the KOKO-ontology. The main content area displays the concept 'aamutossut' with its URI: <http://www.yso.fi/onto/koko/p38422>. The interface is divided into several sections:

- Käsitteihaku:** A search bar with the text 'aam' and a dropdown menu showing 'Kaikki tyypit'.
- Konteksti:** A tree view showing the hierarchy of concepts: 'yso-käsitteet (yso)' -> 'pysyvä (yso)' -> 'fyysinen objekti (yso)' -> 'konkreettinen esin objekti (yso)' -> 'pukineet (mao, yso)' -> 'kengät (mao, tao, yso)' -> 'tossut (mao)' -> 'aamutossut (mao)'. A note below says 'Näytä 1 vieruskäsitettä'.
- Ominaisuudet:** A list of properties for 'aamutossut':
  - Asiasana: aamutossut
  - Kuvaus:
  - Tyyppi: mao:MAO-käsite
  - Vastaava käsite: mao:aamutossut\*
  - Yläkäsite: tossut

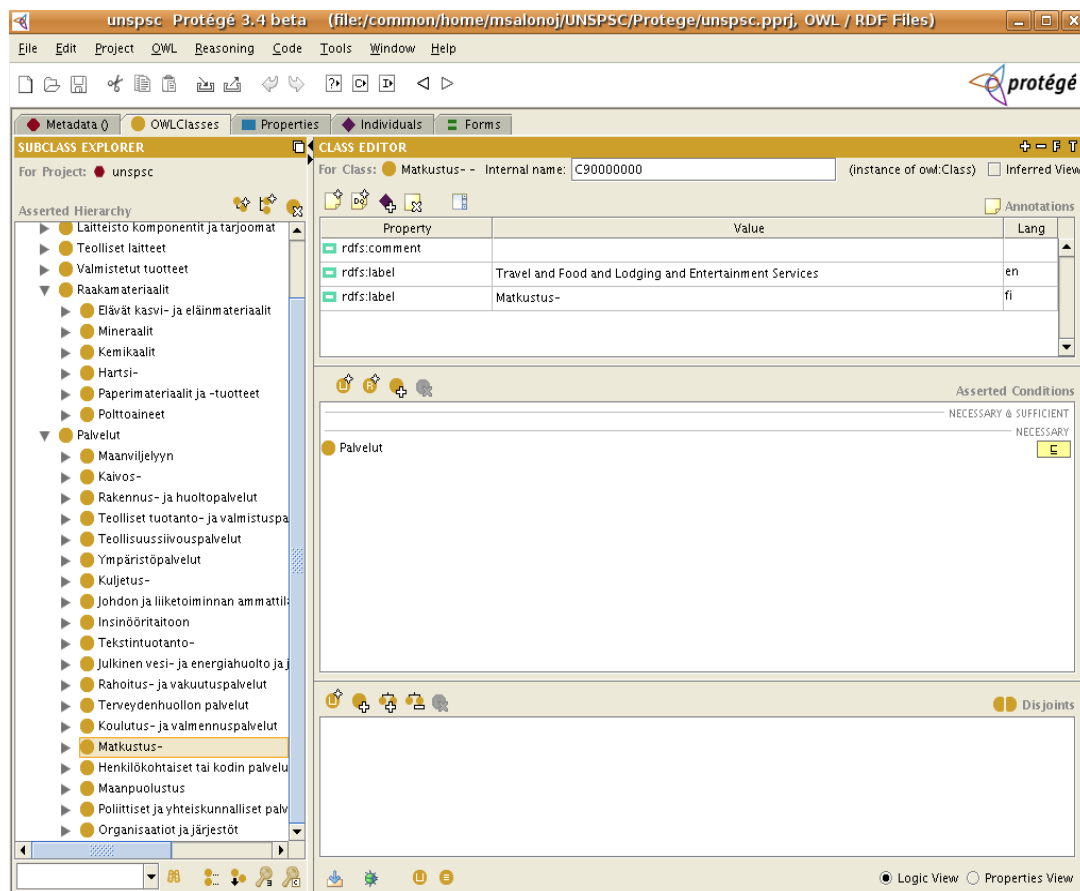
Kuva 4. KOKO-ontologia ONKI-selaimessa /123/

Kuvasta 4 on havaittavissa, että ONKI soveltuu myös sanaston sisällön tarkasteluun hajautetusti. Tämä pätee etenkin suurien (palvelu)luokitusten kohdalla johtuen mm. seuraavista ONKI:n ominaisuuksista:

- Käsitteiden listaus yhdistettynä automaattiseen täydennykseen. Tässä kirjoitusasuun perustuva rajoite voidaan antaa käyttäen skandinaavisia kirjaimia ja jokerimerkkejä. Näitä ominaisuuksia voidaan hyödyntää mm. duplikaattien etsimisessä.
- Hakukielen helppo vaihtaminen. Tämä on erittäin hyödyllinen ominaisuus etenkin niissä tapauksissa, joissa suuri luokitus on vain osittain suomennettu.
- Käsittehierarkian ja muiden suhteiden visualisointi siten, että kaikki käsitteet toimivat linkkeinä.
- Käsitetyyppi-rajoitin.

## 5.2.2 Protégé

Stanfordin yliopiston kehittämän Protégé-ohjelman avulla voidaan luoda mm. RDF(S)- ja OWL-muotoisia ontologioita /115/, mistä johtuen ohjelmalla tuotetut ontologiat voidaan julkaista ONKI-palvelimella. Ohjelmaa on hyödynnetty mm. TerveSuomi-sovelluksen /62/ käsittehierarkioiden kuvaamiseen. Protégé-ohjelman toimintaa on havainnollistettu kuvassa 5, jossa fokus on asetettu matkustamiseen liittyvien palveluiden käsitteeseen. Kuvassa fokuksessa oleva käsite on sijoitettu luokkahierarkiassa palveluiden alakäsitteeksi (SUBCLASS EXPLORER -kohta) ja sille on määritelty nimiä eri kielillä (CLASS EDITOR -kohta).



Kuva 5. Protégé-ohjelma

### 5.2.3 SAHA

Semanttiseen annotointiin on SeCo:ssa kehitetty selainpohjainen SAHA-työkalu /137/, joka luo sivuston lomakkeita RDF-muotoisesta määrittelystä. Näitä lomakkeita käyttäen voidaan SAHA:lla luoda uusia RDF-muotoisia annotaatioita ja muokata vanhoja. Annotatiolla tarkoitetaan tässä yhteydessä lomaketta. Lomakkeen kentät voivat SAHA:ssa olla moniarvoisia ja niissä voidaan viitata myös kontrolloituun sanastoon SAHA:aan integroitua ONKI-toiminnallisuutta hyödyntäen. SAHA:n toimintaa on havainnollistettu kuvassa 6, jossa SAHA:n avulla on annotoitu yrittäjille tarkoitettu palvelu.

The screenshot shows the SAHA Annotatioeditori interface. At the top, there is a header with 'yrittysuomi' and a search bar. Below the header, there is a search bar with the text 'Haku'. To the left, there is a sidebar with 'Annotaatiot' and a list of search results for 'haettava'. The main content area shows the details for the selected annotation, including a table of metadata and a form for editing the name.

**Annotaation tiedot**

Tunniste	http://www.seco.tkk.fi/applications/saha#Instance_ID1228496240214
Luokka	haettava
Otsikko <a href="#">lisää</a>	Agropolis-palvelu (fi)
Kommentti <a href="#">lisää</a>	
Luotu	05.12.2008 18:57
Muokattu	05.12.2008 19:00
Viitattu	<a href="#">TEKEL</a>
Toiminnot	<a href="#">Poista annotaatio</a>

**nimi**

[fi](#) [Agropolis-palvelu](#)

Syötä arvo:

fi

**kuvaus**

Syötä arvo:

Kuva 6. SAHA-työkalu

## 5.3 Sanastot palveluiden kuvailussa

### 5.3.1 Käytetyt sanastot palveluiden kuvailussa

Yritysten toiminnolle ja kulutettuihin tuotteisiin löytyy monia eri luokituksia, jotka kuvaavat palveluihin liittyvää käsitteistöä (kuten toimialoja, tuotteita ja itse palveluita) vaihtelevasti. Keltaiset sivut luokittelevat palveluita omien nimikkeistöjensä mukaan, mutta toisaalta jotkut luokitukset ovat laajemmassa käytössä. Näistä jälkimmäiset ovat houkuttelevia käytettäväksi, koska näihin viitattaessa tieto on helpommin uudelleenkäytettävissä. Uudelleenkäyttöä on pyritty edistämään mm. tilastointiin ja tullaukseen liittyvissä sanastoista, joiden välille on muodostettu muunnosavaimia. Esimerkiksi näin on kuvattu yhteys seuraavien eri tarkoituksiin sanastojen välille /130, 133/:

- SITC (Standard International Trade Classification), joka on YK:n kauppatavaraluokitus. /130/
- MIG (Main Industrial Groupings), joka on virallista statusta vailla oleva suppea luokittelu, jossa toimialat on ryhmitelty niiden tuottamien tavaroiden käyttötarkoituksen mukaan. /133/

- PRODCOM (List of PRODUcts of the European COMmunity), joka on EU:n tasolla laadittu luokitus vajaan 5000:lle hyödykkeelle. /129/
- CPA (Classification of Products by Activities), joka on EU:n jäsenmaiden käyttämä suppea toimialoittainen tuoteluokitus. /130, 133/
- CN (Combined Nomenclature), joka on vienti-ilmoituksissa ja EU:n sisäkaupassa käytetty luokitus noin 15000:lle tavaranimikkeelle. /130, 133/
- NACE (Nomenclature Générale des Activités Economiques dans les Communautés Européennes – Euroopan yhteisön tilastollinen toimialaluokitus). /130, 133/

Muunnoksien lisäksi yhteys sanastojen välille voi syntyä myös tarkennuksien kautta /130/:

- CPA on YK-tasolla määritellyn CPC:n (Central Product Classification) tarkennus.
- CN on Maailman tullijärjestön ylläpitämän HS:n (Harmonized Commodity Description and Coding System) tarkennus.
- NACE on YK-tasolla määritellyn ISIC:n (International Standard Industrial Classification of All Economic Activities) tarkennus. NACE:a taas tarkentaa kansallisella tasolla toimialaluokitus TOL.

Tarkennuksille on tarvetta, koska taloudellisen toiminnan tila ja muutos on erilainen eri alueilla. TOL:n kohdalla tähän haasteeseen on vastattu kehittämällä siitä eri versioita eri vuosille /131/. Suomessa kyseistä luokitusta käytetään rekisteröityjen yritysten toiminnan kuvaamiseen Yritys- ja yhteistietojärjestelmässä (YTJ) /95/, mistä johtuen yhteydet mainittujen sanastojen välille ovat erityisen mielenkiintoisia palveluiden kuvailujen ja haun kannalta. Lisäksi monen yritysrekisteri-palvelun toiminta perustuu vahvasti YTJ:n tietoihin. Näin on havaittavissa mm. Suomen Asiakastiedon Aarre-palvelun kohdalla /117/. Palvelussa hyödynnetään luokituksen hierarkiaa, joka on kuvattuna suoraan käytettyihin koodeihin. Esimerkiksi vuoden 2008 TOL:ssa koodi 60 (Radio- ja televisiotoiminta) on koodin 60202 (Maksulliset TV-kanavat) ylätaso.

Kaupallisesti palveluita on kuvattu lisäksi seuraavien luokitusten avulla:

- COICOP (Classification of Individual Consumption According to Purpose). Luokitus on YK-tasolla laadittu ja se kuvaa yksilöllistä kulutusta käyttötarkoituksen mukaan /132/. Semanttiseen webiin liittyen luokitusta on hyödynnetty mm. Veturi-sovelluksessa /89/.
- UNSPSC (United Nations Standard Products and Services Code). Luokitus kuvaa tuotteita ja palveluita yli toimialojen. Sen määrittelyssä ovat olleet mukana mm. YK ja Dun & Bradstreet -yhtiö, joka tarjoaa kansainvälisiä yritystietopalveluita lukuisissa maissa (myös Suomessa). /44, 48/
- eCI@ss. Luokitus on kehitetty yritysten väliseen yhteistyöhön lähtien Saksan kemianteollisuuden tarpeista, ja siinä kuvataan materiaaleja, tuotteita ja palveluita. Suurteollisuuden näkökulma on nähtävissä luokituksessa mm. käsitteiden jakautumisessa toimialoittain ja hierarkiassa, joka on yritetty optimoida tavaroiden ja palveluiden ostajan kannalta mielekkääksi. Lisäksi on havaittavissa, että merkittävimmät luokituksen

käyttäjäorganisaatiot ovat saksalaisia suuryrityksiä. Semanttisen webin kannalta eCI@ss on mielenkiintoinen, koska se on saatavilla OWL-muodossa /55/. OWL-version on luonut Martin Hepp, joka toimii tutkijana DERI:ssä . /20/

Edellä mainitut luokitukset ovat palveluiden ja tuotteiden kuvailussa hyvin keskeisiä, mistä johtuen seuraavaksi taulukossa 2 esitetään vertailu niiden välillä:

**Taulukko 2. COICOP-, UNSPSC- ja eCI@ss-luokitusten vertailua**

	COICOP	UNSPSC	eCI@ss
Hiearkia	3-tasoinen (pääryhmä, ryhmä ja luokka)	5-tasoinen (segmentti-ryhmä, segmentti, ...)	4-tasoinen (segmentti, pääryhmä, ryhmä ja luokka)
Ominaisuuksia (luokkien lisäksi)	ei	ei	kyllä
Avainsanoja (luokkiin liittyviä)	ei	ei	kyllä
Ilmainen käyttöä	kyllä	kyllä	ei
Nimikkeistön koko (kpl)	n. 1200	n. 21000	yli 30000 (versio 5.1.4)
Käännetty usealle kielelle / suomeksi	kyllä / kyllä	kyllä / kyllä	kyllä / ei

Muista sanastoista mainittakoon NATO:n hankintatarpeista lähtien kehitetty eOTD-terminologia (electronic Open Technical Dictionary), joka identifioi palveluita, tavaroita, paikkoja, yksilöitä ja organisaatioita /56/. Identifioinnin lisäksi terminologialla voidaan kuvata myös ilmentymien ominaisuuksia. Terminologian selailuun ole kuitenkaan tiedossa hyvää työkalua ja sitä ei ilmeisesti ole suomennettu, mikä vaikeuttaa merkittävästi terminologian käyttöä. Lisäksi yrityksille suunnatuissa elektronisissa kaupankäyntiympäristöissä on käytössä erillisiä kontrolloituja sanastoja. Kauppaa käydään mm. RosettaNet- ja ebXML-yhteisöissä, joista ensimmäisessä sanastona käytetään RNTD:tä (RosettaNet Technical Directory) /56, 92/. Tämä sanasto tosin näyttää olevan hyödyllinen koostaan, rakenteestaan ja sisällöstään päätellen vain alkuperäisessä käyttökontekstissään.

### 5.3.2 Sanastojen väliset muunnokset

Palveluilmentymien kuvauksessa käytetty sanasto ei aina sovellu suoraan tiedonhakuun. Tämä havaittiin kehitettäessä mm. Veturi- /89/ ja TerveSuomi-sovelluksia /62/, joissa molemmissa käytettiin tiedonhakuun annotaatioista pääteltyä tietoa. Näistä Veturi-sovelluksessa TOL-koodeihin /130/ perustuvat annotaatiot peilattiin mm. SUMO-ontologiaan /96/ ja COICOP-luokitukseen /132/.

Peilaaminen sanastojen välillä ja/tai annotaatioiden täydentäminen onnistuu riippuen annotaatiokäsitteisiin sisältyvästä informaatiosta, jota on tavallisesti kuvattu formaalilla tavalla hyvin rajallisesti yleisissä ontologioissa. Tästä johtuen peilaussäännöt sanastojen välille joudutaan usein luomaan tyhjästä, kuten tehtiin mm. Veturi-sovelluksen kohdalla. Esimer-

kiksi tomaattien myynti -annotaatiosta voitaisiin ihmisten toimesta tietää ja päätellä mm. seuraavaa:

- Tomaatti on fyysinen kappale: pieni pallo.
- Tomaatti on yleensä väriltään punainen. Myös muun värisiä tomaatteja on olemassa.
- Tomaatteja myydään → niitä voi ostaa.
- Tomaatti on vihannes.
- Vihanneksia syödään → tomaatteja syödään.
- Vihannekset ovat kulutushyödykkeitä → tomaatti ei säily.
- Vihannekset ovat terveellisiä syödä → tomaatit sopivat laihduttajalle.
- Tomaatteja käytetään salaateissa ja salaatteja syödään alkuruokina → alkuruokaa valmistava voi olla kiinnostunut ostamaan tomaatteja.
- Tomaatteja kasvatetaan kasvihuoneissa → tomaatit liittyvät maatalouteen.

Informaatiota esimerkiksi YSO:sta (Yleinen suomalainen ontologia) löytyy kuitenkin vain seuraaville tomaatin ominaisuuksille: vihannes, syötävä ja fyysinen kappale.

### 5.3.3 Sanastojen muokkaus

Palveluiden kuvaukseen voidaan käyttää valmiita sanastoja. Nämä eivät kuitenkaan aina suoraan sovellu käyttöön, mistä johtuen niitä voi olla tarvetta muokata mm. seuraavilta osin:

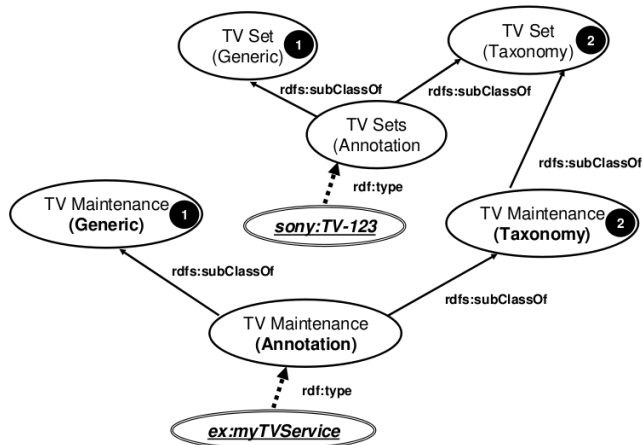
- Hierarkiat. Näitä voidaan haluta muuttaa tai lisätä, koska alkuperäisen käsitteistön hierarkia voi sisältää virheitä, epäloogisia yleistyksiä, tarpeetonta hierarkian syvyyttä tarkasteltaessa rajattua sovellusaluetta tai sitten siitä yksinkertaisesti vain puuttuu joitain käytön kannalta oleellisia näkökulmia.
- Käsitteet. Luokituksista saatetaan haluta poistaa tai siihen saatetaan haluta lisätä uusia käsitteitä. Toisaalta luokituksen nimiä käsitteille saatetaan haluta muuttaa tai niille saatetaan haluta lisätä käännöksiä eri kielille.

Hierarkian tarkentamista helpottamaan on Martin Hepp kehittänyt GenTax-menetelmän /55/ arvioidessaan tuoteluokittelu-järjestelmien sisältöä. Tuoteluokituksen hierarkian merkitys ei vastaa yleisesti rdfs:subClassOf-suhdetta, mistä johtuen tämän suhteen käyttäminen johtaisi virheelliseen päättelyyn. Esimerkiksi UNSPSC-luokituksessa /48/ jälle (koodi: 50202302) on merkitty laajemmaksi käsitteeksi alkoholittomat juomat (koodi: 50202300). Oletetaan, että on olemassa verkkokauppa, jonka varastosta löytyy mm. kaksi erimerkkistä pulloa vettä (koodi: 50202301) ja pussi jääpaloja. Tässä tapauksessa asiakkaalle päätellään hänen hakiessaan alkoholittomia juomia vastaukseksi myös pussi jääpaloja, joka on ylimääräinen.

GenTax-menetelmässä edelläkuvattu ongelma ratkaistaan erottelemalla luokituksen alkuperäisiä nimikkeitä kuvaavista taksonomiakäsitteistä (Taxonomic concept) yleiset käsitteet (Generic concept). Näin siksi, koska Hepp on havainnut, että tuoteluokituksen kokoavat







Kuva 8. GenTax-menetelmä toiminnassa /55/

## 6 PALVELUIDEN HAKU

### 6.1 Tekstihaku

#### 6.1.1 Yksinkertaiset hakusovellukset

Usean palveluiden hakuun käytetyn sovelluksen käyttö perustuu tekstihakuihin, joissa hakujoukkoa voidaan rajoittaa jollain yksinkertaisella tavalla. Esimerkiksi yleisissä hakukoneissa, kuten Googlessa /40/ ja Yahoossa /145/, voidaan hakujoukkoa rajoittaa mm. seuraavasti:

- Sijainti. Esimerkki: sivut Suomessa.
- Tiedostotyyppi. Esimerkki: PDF.
- Aihe. Esimerkki: Shopping (Yahoossa).
- Kieli. Esimerkki: suomenkieliset sivut.
- Sivustot. Esimerkki: www.pkt.fi, www.tekes.fi ja www.te-keskus.fi.

Käytettäessä yleisiä hakukoneita hakujoukkoon sisältyy palveluntarjoajien kotisivuja, joihin hakusanat osuvat vaihtevalta menestyksellä. Lopputulokseen palveluntarjoaja voi kuitenkin vaikuttaa optimoimalla kotisivunsa lähdekoodia tai ostamalla näkyvyyttä valitsemilleen hakusanoille. Kotisivun optimointi voi perustua mm. sivujen linkityksiin, joita mm. Google hyödyntää hakutuloksien järjestämisessä /52/. Lisäksi hakutulokseen voi vaikuttaa myös valittujen hakusanojen keskinäinen läheisyys ja lemmaus, jota käyttämällä voidaan haussa huomioida hakusanojen taivutetut muodot. Ehdoton edellytys palveluiden löydettävyydelle yleisestä Internetistä on kuitenkin niiden julkaisu sinne, mitä läheskään kaikki palveluntarjoajat eivät ole palveluilleen tehneet.

Vapaan tekstin käyttämiseen hakukoneissa liittyy kuitenkin ongelmia johtuen siitä, että käytetyt sanat eivät yksilöi merkityksiä. Tekstihaun avulla saadaankin harvoin löydettyä yli 20 prosenttia hakujoukon yksilöistä, jotka ovat haun suhteen todennäköisesti olennaisia /121/. Käytettäessä NLP-menetelmiä (Natural Language Processing) tekstihaun apuvälineenä monissa järjestelmissä läheskään kaikkea tekstiin sisältyvää informaatiota ei olla

pystytty tunnistamaan /58, 76, 121/. Esimerkiksi Groverin ja Lascaridesin kehittämässä järjestelmässä lääketieteellisistä teksteistä pystyttiin jäsentämään oikein vain noin 30 prosenttia muunneltua täyttä ANLT-kielioppia hyödyntäen /45/.

Samoja tai osittain samoja merkityksiä voidaan kuvailla käyttäen useita eri sanoja kielen sisällä ja niiden välillä. Toisaalta sanojen sama kirjoitusasu ei takaa samaa merkitystä kuten jo johdannossa mainittiin (mm. sanat *avocat* ja *kangas*). Esimerkiksi eri tietosisältöjen kuvailijat voivat kuvailla samaa merkitystä sanoilla *petäjä* tai *honka*. Toisaalta tiedonhakijalle voi olla intuitiivisinta käyttää saman merkityksen kuvaamiseen sanoja *mänty* tai *pine* (englantia) riippuen hänen äidinkielestään. Myöskään kielen sisällä ei ole itsestään selvää, että ihmiset kuvailisivat hakutarvettaan täsmälleen samoilla sanoilla /53/, /67, s. 121/. Eri sanojen huomioiminen on erittäin monimutkaista muotoiltaessa hakua käsin kirjoittamalla. Lisäksi jokainen mainituista toimijoista voi helposti tehdä kirjoitusvirheitä kuvaillessaan merkityksiä. Ongelmista johtuen vapaata tekstiä käyttäessään tiedonhakija saattaa tuntea itsensä epävarmaksi, mikä heikentää käytettävyyttä /52/.

Vapaan tekstin käyttö tiedonhaussa näyttää toimivan tilanteissa, joissa haettava asia on tiedossa ja se on riittävän yksilöivä. Tiedossa oleva asia voi olla esimerkiksi henkilön nimi, yritys tai vaikkapa katuosoite. Sen sijaan käyttäjän aloittaessa tutkimaan jotain aihealuetta vapaata tekstiä käyttäen joutuisi hän kelvollisen hakutuloksen saavuttamiseksi muotoilemaan pitkän ja monimutkaisen hakulauseen. Tämä voi olla kuitenkin mahdotonta johtuen siitä, että alueen sanasto on käyttäjälle ennestään tuntematonta. Toiseksi hakulauseen muotoilulle käsin ei ole välttämättä edes tarvetta. Rajoittamalla sisällön kuvaamisessa käytettyjen vaihtoehtojen määrää tarvitaan lyhyempiä hakulauseita saman asian ilmaisemiseen, ja toisaalta pitkien hakulauseiden muodostamista voidaan nopeuttaa hyödyntämällä sanojen välistä verkkoa semanttisen webin tekniikoilla. Vapaan tekstin käyttö haettaessa palveluita ei siis välttämättä ole kaikissa tilanteissa paras ratkaisu.

### 6.1.2 Automaattinen kyselyn laajentaminen

Kyselyn laajentamisella tarkoitetaan menettelyä, jossa käyttäjän kuvaamaa hakulauseetta muokataan hakutermejä lisäten jonkin periaatteen mukaisesti tavoitteena parantaa haun saantia ja tarkkuutta /102, s. 180/. Lisättyjen hakutermien aineistona voidaan käyttää jotain tekstiaineistoa tai ontologiamuodossa olevaa yleistä tietämystä. Kyselyä taas voidaan laajentaa käsin kirjoittamalla, interaktiivisesti tai automaattisesti. /25/

Automaattisessa kyselyn laajennuksessa hakujärjestelmä vastaa itse hakutermien laajentamisesta, millaista menettelyä on käytetty monissa hakujärjestelmissä. Menettelyä on käytetty käsitteillä annotoidun aineiston haussa mm. museoalaan liittyvän Kantapuu.fi- /134/ ja TerveSuomi-sovelluksen /62/ kohdalla. Lisäksi tekstikuvauksista haettaessa menettelyä on hyödynnetty mm. Guarinon kehittämässä sovelluksessa /50/. Tässä järjestelmässä keltaisten sivujen hakua pyrittiin kehittämään käyttäen apuna Sensus-ontologiaa, joka on laajan WordNet-ontologian johdannainen.

Kaikissa mainituissa automaattista kyselyn laajennusta hyödyntävissä sovelluksissa yleistä tietämystä on hyödynnetty ainakin samaa tarkoittavien sanojen ja/tai tarkentavien käsitteiden suhteen /134/. Samaa tarkoittavia sanoja ovat saman merkityksen erilaiset kirjoitusasut

kielen sisällä ja niiden välillä. Esimerkkinä tästä ovat seuraavat samaa merkitystä kuvaavat kirjoitusasut: mänty, honka ja pine (englantia).

Näyttää siltä, että laajennus samaa tarkoittavien sanojen kohdalla näyttää toimivan kyselyn laajenuksessa tarkentavia käsitteitä paremmin /142/. Erityisesti kielten välisten linkkien hyödyntäminen näyttää ainakin joissain sovelluksissa parantavan niiden käytettävyyttä. Näin on havaittu mm. Kantapuu.fi-sovelluksen kohdalla. Tässä sovelluksessa kyselyn laajennus tehdään automaattisesti alakäsitteisiin (mm. työkalut → kirveet), mistä näyttää olevan hyötyä varsinkin pienten tulosjoukkojen kohdalla. Alakäsitteisiin laajennettaessa hyöty sen sijaan näyttää jäävän kyseenalaiseksi, jos haku kohdistuu aineistoon, joka on jo valmiiksi laajennettu annotointivaiheessa useilla toisiaan tarkentavilla termeillä. Lisäksi jos laajennus on jo sisällytetty alkuperäiseen kyselyyn, tilanne on vastaava: kyselyn laajentaminen tuottaa parhaimpia tuloksia juuri lyhyiden kyselyiden kohdalla /142/.

Kantapuu.fi-hakusovellus ja toteutukset, joissa käytetään suosittuja hakukoneita /18, 90/ osoittavat, että automaattinen kyselyn laajennus ei ole kovin vaikea toteuttaa, koska itse hakujärjestelmää ei välttämättä tarvitse muuttaa. Hakutermien laajentaminen voidaan nimittäin tehdä siten, että muutetaan vain itse hakulausetta, jonka informaatio välittyy Internet-sovelluksissa yleensä URL-osoitteena. Kantapuu.fi-hakusovelluksen käytössä on tosin havaittu, että URL-osoitteiden maksimipituus saattaa rajoittaa laajennettavien termien määrää. Tällöin ongelman lievittämiseksi voisi olla mielekästä käyttää hakulauseessa vain sellaisia termejä, joiden tiedetään esiintyvän hakujoukossa.

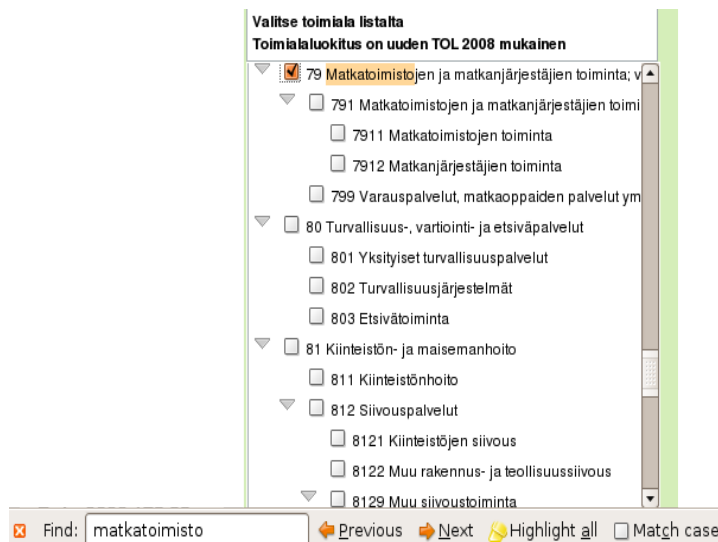
Automaattisen kyselyn laajentamisen tietolähteenä voidaan yleisen tietämyksen lisäksi käyttää myös tekstiaineistoa, joka voi olla kiinnitetty tai jatkuvasti kasvava, joista jälkimmäisen kohdalla puhutaan oppivista järjestelmistä. Tosin ainakaan kiinnitetyn tekstiaineiston kohdalla tutkimustulokset eivät kiistattomasti osoita hakutulosten parantumista /11/. Lisäksi tekstiaineistoa hyödynnetään myös relevanssipalautteeseen perustuvissa sovelluksissa. Näissä käyttäjä osoittaa hakutulosta tarkastellessaan onnistuneita hakutuloksia, joiden sisällön perusteella hakujärjestelmä voi suorittaa uuden haun /8, s. 118/, /106/. Tutkimustulokset tämän keinon käytöstä ovat osoittaneet haun tarkkuuden /102, s. 180/ parantuneen /101, 104/.

Tiedemaailma ei ole päätenyt vielä tähän mennessä selkeään lopputulokseen tutkiessaan automaattisen kyselyn laajennuksen hyödyntämistä tekstikuvauksista haettaessa. Navigli ja Velardi ovat todenneet, että laajentamisesta ei ole merkittävää hyötyä /90/. Toisaalta hyödynnettäessä yhdistettyä sanamuotoon ja yleiseen tietämykseen perustuvaa indeksointia on havaittu, että tulokset paranevat erityisesti järjestetyn tulosjoukon kärjessä /34, 83/. Tuloksien paraneminen juuri tulosjoukon kärjessä on tiedonhaussa erittäin oleellista, sillä käyttäjän ongelman ratkaisuun käyttämä aika on suora järjestelmän käytettävyyden mittari. Käyttäjälle ensimmäiset viisi tulosta ovat todennäköisesti kaikkein merkitsevimpiä ja maksimissaankin hän yleensä käynee tuloksia läpi vain muutamia kymmeniä /65/. Lisäksi myös Guarinon toteutuksen havaittiin parantavan hakutuloksia /50/, mikä on voinut osittain johtua käytetyn ontologian koosta ja laadusta. Tämän lisäksi lopputulokseen voi vaikuttaa myös monimerkitykselliset sanat, joiden merkitykset tulee tapauskohtaisesti päätellä, minkä onnistumisen vaikutuksista lopputulokseen on tosin erilaisia tutkimustulok-

sia /39, 71, 108/. Lisäksi tilanne voi vaihdella aineistojen välillä. Näin mm. Krovetzin ja Croftin mukaan, jotka arvelevat, että monimerkitykselliset sanat eivät muodosta kovin merkittävää ongelmaa sellaisten aineistojen kohdalla, joiden aihepiiri on rajattu /71/. Tämä ehto toteutuu kuluttajapalveluiden ja yrittäjille suunnattujen tukipalveluiden kohdalla.

### 6.1.3 Hakusanojen osilla hakeminen

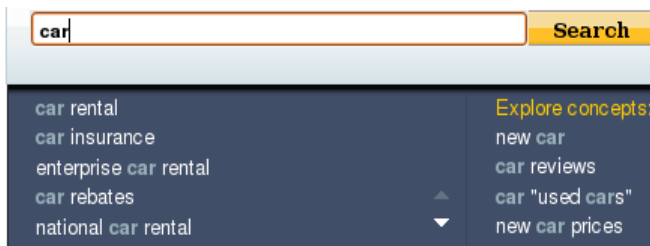
Joissain Internet-selaimissa (mm. Firefox) etsitty merkkijono osoitetaan dynaamisesti tekstistä. Palveluhaussa tätä voidaan hyödyntää minkä tahansa Internetissä sijaitsevan palveluhakemiston kohdalla kuitenkin edellyttäen, että hakemiston sanasto on kuvattu Internet-selaimelle tekstinä eikä esimerkiksi kuvina. Keinon käyttöä on havainnollistettu kuvassa 9, jossa Asiakastiedon Aarre-palvelun toimialaluokitusvalinnoista on osoitettu ensimmäinen valinta, jossa esiintyy matkatoimisto-merkkijono.



**Kuva 9.** Firefoxin etsi-toiminto Aarre-palvelussa /117/

Internet-selaimen tekstihakutoiminnon käyttö on erityisen hyödyllistä, kun palveluhakemiston Internet-sivulla on esitelty runsaasti erilaisia valintavaihtoehtoja. Näiden lukemiseen kun kuluisi kohtuuttomasti aikaa. Asiakastiedon Aarre-palvelun lisäksi näin kävisi mm. Yritysoapas.com-sivustolla. Huolimatta keinon hyödyllisyydestä kaikki Internetin käyttäjät eivät todennäköisesti ymmärrä sitä käyttäen.

Firefoxin etsi-toiminnon tavoin myös jotkut palveluiden haussa käytetyt sovellukset näyttävät hakutuloksia käyttäjän kirjoittaman tekstin tahdissa, minkä on yleisesti havaittu parantavan käytettävyyttä erityisesti tehokkuuden osalta /2/. Hakutuloksiksi sovellukset antavat dokumentteja/palveluita, hakusanoja tai molempia. Näistä hakusanoja annetaan vastaukseksi mm. Yahoo-hakukoneessa (kuvassa 10 hakusanoja, joissa esiintyy car-merkkijono) ja dokumentteja mm. Firefox-selaimen CyberSearch-lisäosan avulla selaimen osoiterivin yhteyteen. Fonectan keltaisilla sivuilla taas tulos on yhdistelmä hakusanoja ja palveluita /32/.



Kuva 10. Yahoos automaattinen täydennys /145/

Aiemmin mainituissa sovelluksissa kirjoitusasulle ei ole mahdollista asettaa monimutkaisia rajoitteita käyttäen jokerimerkkejä, mutta toisin on mm. ONKI:ssa /140/. Lisäksi joissain sovelluksissa hakusanoja voi selata alkukirjaimella /67, s. 67-69/, mikä toiminnallisuus sisältyy mm. TerveSuomi-sovellukseen /62/ ja Yrityksen palveluhakemistoon /97/.

## 6.2 Maantieteellisen sijainnin hyödyntäminen

### 6.2.1 Hyödyntämiskohteet

Monien palveluiden haussa käytettyjen sovelluksien toiminnassa hyödynnetään maantieteellistä sijaintia. Esimerkiksi WikiTravel-sivustolla /3/ palvelutietoa esitellään paikkakohtaisella sivulla aiheittain ryhmiteltynä. Superpages.com-sivustolla /64/, Fonectan keltaisilla sivuilla /32/, eat.fi-palvelussa /37/ ja GoogleMaps-palvelussa /41/ (kuvassa 11) taas haun tuloksena saadut palvelut visualisoidaan kartalle. Mainituista hakupalveluista eat.fi:n ja GoogleMaps:n toiminta perustuu käytännössä täysin karttoihin. Näin on myös OpenStreetMap-sivuston /94/ kohdalla, jonka avulla voidaan hakea maantieteellisiä kohteita ja tarkastella palvelupaikkojen sijainteja niissä.



Kuva 11. GoogleMaps-palvelu /42/

### 6.2.2 Ongelmallisten kohteiden visualisointi

Näkymäksi kuvan 11 Googlen karttaelementissä voidaan valita kartta, satelliitti tai maasto, joiden avulla palvelupaikan sijainti voidaan esittää melko ymmärrettävästi monissa tapauksissa. Jos taas palveluita on keskittynyt paljon pienelle alueelle ja eri korkeuksille maanpinnan suhteen, GoogleMaps-palvelu toimii todennäköisesti heikommin. Esimerkiksi laskettelu- ja ostoskeskusten kohdalla palveluita voi olla vaikeaa erottaa toisistaan, vaikka karttakuvaa tarkennettaisiin pienemmälle alueelle. Ongelma ostoskeskusten kohdalla syntyy siitä, että näkymä rajoittuu rakennusten ulkopinnoille. Kun rakennusten sisäosia ei

visualisoida, palvelupaikkojen sijaintien hahmottaminen ja reittivalintojen suunnittelu vaikeutuu. Lisäksi palveluiden sijaintien visualisoimista haittaa näissä se, että viralliset tiedot palvelupaikkojen sijainneista ovat liian epätarkkoja tiedonhakijan kannalta eivätkä ne useinkaan linkity keskusten näkyviin reittiverkostoihin.

Katukarttaa käyttäen sijaintien visualisoiminen on ostos- ja laskettelukeskusten lisäksi ongelmallista myös joiden kaupunkien keskustoissa. Näin siksi, koska niissä on usein sisäisiä ja maanalaisia käytäviä. Lisäksi suurissa marketeissa tavaroiden etsiminen voi olla joissain tilanteissa hyvin turhauttavaa, vaikka niissä onkin käytössä sisäisiä opaskylttejä ja tavarat on järjestetty usein mielekkäisiin kokonaisuuksiin. Pitkän ostoslistan tai aiemmin tuntemattomien ruoka-aineiden kohdalla karttasovellusten käyttö voisi siis olla mielekästä.

Tiheä rakentaminen voi olla karttakuvassa ongelma myös silloin, kun käyttäjän on vaikeaa hahmottaa kuvasta olennaisia kohteita. Realistisen kuvan näyttäminen kohteesta ei siis välttämättä ole aina mielekästä. Piirrosta, joka tiivistää hyvin oleellisen informaation kohteesta, voi olla helpompi tulkita /1, s. 31-32/. Esimerkiksi piirroksia Sello-ostoskeskuksesta (kuvassa 12) ja Åren laskettelukeskuksesta (kuvassa 13) on todennäköisesti helpompi tulkita kuin vastaavia karttakuvia.



Kuva 12. Piirros Sello-ostoskeskuksen katutasosta /77/

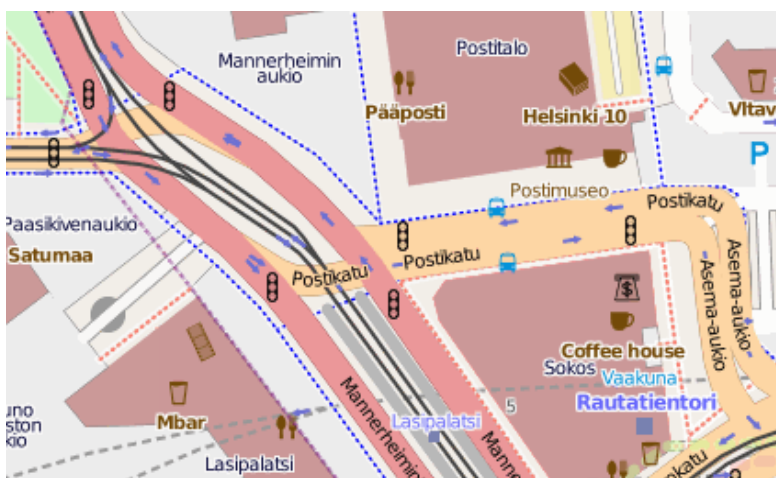


Kuva 13. Piirros Åren laskettelukeskuksesta /143/

Piirroksiin perustuvia ratkaisuja on kehittänyt mm. yritys nimeltä Whatamap.com /143/, jonka kehittämien sovellusten avulla karttapiirroksen voidaan sisällyttää lähes mitä tahansa tietoa palvelupaikoista ja niiden sijainneista. Lopputulosta taas voidaan tutkia Internet-selaimessa tai mobiililaitteessa siten, että palvelupaikat näkyvät karttapiirroksessa samanlaisina lisäinformaatiota sisältävinä pisteinä kuin GoogleMaps-palvelussa (kuvan 11 Parturi-Kampaamo Seija & Helmut).

### 6.2.3 Palvelutiedon maksullisuus

Keltaisissa sivuissa ja erikoissovelluksissa palveluntarjoajien täytyy erikseen maksaa näkyvyydestä niissä, mikä usein heikentää näiden sovellusten kattavuutta. Tätä ongelmaa ei kuitenkaan ole mm. OpenStreetMap-sivustolla /94/, jossa karttojen sisältöä voi muokata kuka tahansa Wikipedian tapaan, mikä tosin voi olla joissain tapauksissa luotettavuuden kannalta ongelmallista. Tässä kartalle voidaan lisätä palvelupaikkoja rajoituksetta ja myös sen karttapohjaa voidaan muokata. Sivuston sisältöä havainnollistaa kuva 14, jossa kartan näkymä on tarkennettu Helsingin keskustaan.



Kuva 14. OpenStreetMap – Helsingin keskusta /94/

## 6.3 Tutkiva tiedonhaku

### 6.3.1 Yleistä

Tiedonhaku voi olla suorien tekstihakujen lisäksi myös tutkivaa (exploratory) /79/. Useimmiten käyttäjä ei tiedä tai muista haettavan kohteen tarkkoja ominaisuuksia, vaan hän muistaa vain suurinpiirtein mihin kohde liittyy. Näin ollen käyttäjä usein samoilee tietoa etsiessään, mitä prosessia tiedonhakua käsittelevä kirjallisuus kutsuu mm. englanninkielisillä termillä orienteering ja browsing /122/.

Tutkiessaan uutta aluetta tai käyttäessään uutta hakusovellusta käyttäjä ei välttämättä aina osaa hahmottaa hakujoukon sisältöä. Tämän vuoksi joihinkin hakusovelluksiin on sisällytetty mahdollisuus hakujoukon esikatseluun, mikä on tutkimuksissa todettu hyödylliseksi ominaisuudeksi /98/. Esikatselu voi tarkoittaa hakutermien listaamista ja ryhmittelyä, mikä on toteutettu mm. Yritysopas.com-sivustolla /119/. Hakutermien yhteyteen on lisäksi joissain sovelluksissa liitetty tieto termien suodatusvaikutuksesta. Esimerkiksi SuperPages-sivustolla (kuvassa 15) hakusivun vasempaan reunaan sijoitetuille valintavaihtoille on ilmaistu niitä vastaavien tulosjoukkojen koot.

The screenshot shows a search interface with a search bar containing 'car' and a location dropdown set to 'MI'. A 'SEARCH' button is visible. Below the search bar, it indicates '39404 Results found for Car, MI'. The results are displayed in a table with columns for 'Category Name' and 'Business Name'. A 'Sort' dropdown is set to 'Relevancy' and a 'Show' dropdown is set to '15 per page'. There is a checkbox for 'Show superguarantee™ only'. The results are categorized into 'Sponsored Results' and 'Best Match'. The 'Sponsored Results' section includes listings for Acura and Honda, each with a star rating, a link to 'Read Reviews', a link to 'Write a Review', and a link to the official website. The 'Best Match' section lists various categories such as 'Auto Dealers (2647)', 'Auto Service & Repair (21268)', 'Antique & Classic Auto Dealers (75)', 'Race Car Dealers (71)', 'Used Cars (2707)', 'Auto Parts Retail (6365)', 'Auto Body Repair & Painting (2946)', 'Auto Customizing, Conversion, & Restoration (2019)', 'Auto Cleaning & Detailing (1761)', 'Brakes Service & Repair (1624)', 'Auto Washing & Polishing (1233)', 'Auto Oil & Lube (1099)', 'Auto Transmissions (875)', 'Cars Service & Repair (518)', and 'Auto Glass Replacement & Repair (195)'. A '20 More' link is also present.

Kuva 15. Superpages.com-sivusto /64/

Hakujoukon esikatselun helpottamiseksi joihinkin sovelluksiin on liitetty myös tagipilvi (tag cloud), jolla kuvataan yleensä hakujoukon tärkeimmät aiheet /67, s. 66-67/. Tätä keinoa tässä tarkoituksessa on hyödynnetty mm. TerveSuomi-sovelluksessa (kuvassa 16).

#### Suurimmat aiheet

aikuiset aikuistyyppin diabetes diabetes ennaltaehkäisy  
 lapsi liikunta miehet naiset nuoret nuoruustyyppin  
 diabetes ohjeet oireet ravitsemus riskitekijät rokotteet rokotus  
 tartuntataudit terveyden edistäminen terveys terveysliikunta  
 tupakointi tupakoitsijat tutkimus vauvat

Kuva 16. TerveSuomi-sovelluksen tagipilvi /62/



Lähtökohtana tutkivassa tiedonhaussa on yleensä jokin konteksti eli enemmän tai vähemmän epämääräinen aihealue, johon käyttäjä haluaa tutustua. Tutustuessaan tähän käyttäjä tekee tyypillisesti useita hakuja, joiden tuloksiin hän voi reagoida monin eri tavoin. Esimerkiksi Evri-hakukoneessa (kuvassa 17) käyttäjä voi reagoida hakujärjestelmän palautteeseen mm. seuraavasti:

- Samoillen yhdistäviä polkuja pitkin (kuvassa 17 - Top connections: Microsoft).
- Tarkentaen hakua (kuvassa 17 - Dig Deeper: Nokia → selling).

**Evri** BETA

Find:  [Share](#)

**Nokia** Company, Electronics Company, and Manufacturing Company

Nokia Corporation ( in Finnish) (, NYSE: NOK, FWB: NOA3) is a Finnish multinational communications corporation, headquartered in Keilaniemi, Espoo, a city neighbouring Finland's capital Helsinki. Nokia is focused on wireless and wired telecommunications, with 112,262 employees in 120 countries, sales in more than 150 countries and global annual revenue of 51.1 billion euros and operating profit of 8.0 billion as of 2007. It is the world's largest manufacturer of ...

Description from Wikipedia [Show more](#)

Top connections <sup>[?]</sup>

See more

About Nokia

**Name:** Nokia Corporation

**Location City:** Espoo

**Location Country:** Finland

**Founded:** Nokia, Finland (1865)

**Stock Symbol:** NOK

**Dig Deeper** Nokia selling

Filter by top connection

- Microsoft
- T-Mobile
- Apple
- Sony Ericsson
- Motorola

Filter by category

- Company
- Electronics Company
- Retailer
- Technology company
- Software Company
- Cell Phone
- Manufacturing Company

Filter by activity

- releasing
- report
- featuring
- missing
- shipping
- losing
- selling

**Nokia's Sales and Profits Plummet**  
...both the high and low ends of the market. In China, which has been the company's largest market, **Nokia** sold 36% fewer devices in the quarter vs. a year earlier as local dealers cleared out excess inventory. **Nokia** has so far refused to be drawn into a...  
55 minutes ago | BusinessWeek Online

**Recession hits mobile phone market**  
...developing markets, where it has provided low-cost phones. Even in China, **Nokia's** largest handset market, the company sold 36 percent fewer devices during the fourth quarter than the same period a year ago. And sales were down 23 percent in the Middle East and...  
23 hours ago | CNET News.com

**Nokia's Sales, Profit and Market Share Drop**  
**Nokia** reported fourth-quarter sales down about 19 percent year on year. The company sold fewer phones than a year earlier, and its market share also dropped, it said Thursday. Sales for the last three months of 2008 totalled EUR12.7 billion (US\$17.9...  
1 day ago | PC World: Latest Technology News

[Show more](#)

**Pictures** <sup>[?]</sup>

**Videos** <sup>[?]</sup>

Kuva 17. Evri /28/

### 6.3.2 Yhdistävät polut tulosten yhteydessä

Evri-hakukoneen (kuvassa 17) lisäksi hakutulosten yhteyteen on kuvattu yhdistäviä polkuja mm. Fonectan keltaisilla sivuilla (kuvassa 18) ja TerveSuomi-sovelluksessa /62/. Palveluita haettaessa ominaisuus on hyödyllinen silloin, kun oikean hakusanan arvaaminen on vaikeampaa kuin jollain tavalla samankaltaisen palvelun löytäminen. Esimerkiksi haettaessa koruliikkeitä käyttäjä saattaa tietää jonkun koruliikkeen nimeltä tai sen osoitteen, mistä johtuen se löytyy nopeasti ja varmasti. Saadun tuloksen yhteydestä taas käyttäjä voi löytää hyödyllisiä hakutermejä, joiden arvaaminen voi olla muuten vaikeaa, koska käyttäjä ei välttämättä tunne hakusovelluksen terminologiaa /53/.

Hae tuotteita/palveluita

Missä?

Koruja ja kelloja

hakusana, toiminta tai yrityksen nimi

maakunta, kunta, postinumero [lisää katuosoite](#)

**HAE**

Tuotteet ja palvelut

Yhteystiedot

**Oulun Koru Oy**

Kirkkokatu 19 90100 OULU

08 3112290

[kotisivu](#) [lähetä matkapuhelimeen](#) [tallenna](#)

Toimialat

Kultasepäntaiteita, Kellosepäntaiteita

Koruja ja kelloja

Kuva 18. Fonectan keltaisten sivujen hakutulos /33/

Kuvasta 18 on havaittavissa, että Fonectan keltaisten sivujen kohdalla tuloksiin liittyvät termit eivät toimi linkkeinä eikä niiden suhteita toisiinsa ole visualisoitu. Toisin on kuitenkin esimerkiksi MuseoSuomi-sovelluksessa /54/, jossa yhdistäviä polkuja pitkin liikkuminen on huomattavasti helpompaa. Tämä on havaittavissa kuvasta 19, jossa osa MuseoSuomi-sovelluksessa tarkasteltavan kohteen ominaisuuksien arvoista (mm. Esine-tyyppi: soittimet > pillit > kukkopillit) toimii linkkeinä uusiin hakuihin.

kukkopilli



**Materiaali:** punasavi  
**Valmistuspaikka:** Suomi  
**Valmistusaika:** 1930-1939  
**Valmistustekniikka:** käsityö  
**Käyttöpaikka:** Suomi, Stensgård, Sepänkylä, Espoo, Suomi  
**Asiasana:** ELÄINLELU  
**Mitat:** Pituus 8cm  
**Museokokoelma:** Museokokoelma  
**Vastuumuseo:** Espoon kaupunginmuseo  
**Asiasanat:** Espoon kaupunginmuseon sanasto  
**Esineen numero:** ECM.3594.58  
 ID: 359458

Muotissa valettu kukko, jolla on pitkä punaiseksi maalattu nokka. Pyrstö on lyhyt, jonka päässä on reikä puhaltamista varten. Pyrstön molemmin puolin on pienet reiät. Kukon rinta on maalattu valkoiseksi ja jalusta vihreäksi. Pintakuviolla matkitaan höyhengeröstä.

**Esinetyyppi:**

- [soittimet](#) (40) > [pillit](#) (1) > [kukkopillit](#) (1)

**Materiaali:**

- [materiaalit](#) (777) > [muut materiaalit](#) (61) > [savi](#) (23)

**Valmistuspaikka:**

- [Eurooppa](#) (254) > [Suomi](#) (2239)

**Valmistusaika:**

- [aikakaudet](#) (302) > [historiallinen aika](#) (3023) > [uusi aika](#) (3013) > [sodat](#) (612) > [II maailmansota](#) (430)
- [vuosiaadat](#) (3012) > [1900-luku](#) (2589) > [1930-1939](#) (466)

**Käyttöpaikka:**

- [Eurooppa](#) (2232) > [Suomi](#) (2227)
- [Eurooppa](#) (2232) > [Suomi](#) (2227) > [Etelä-Suomen lääni](#) (1899) > [Uusimaa-Nyland](#) (670) > [Espoo](#) (513) > [Sepänkylä](#) (97)

**Käyttötilanne:**

- [kansalais- harrastus- ja vapaa-aikatoiminta](#) (210) > [leikit](#) (162) > [leikki](#) (156)
- [valmistustekniikat](#) (1827) > [käsityö](#) (111)

**Kokoelma:**

- [Espoon kaupunginmuseon kokoelmat](#) (1190) > [Museokokoelma](#) (1128)

**Sama käyttöpaikka**

Sepänkylä:

- [nuvin väasa, iakku naisen puvin väasa](#)
- [iakkara kehruuakikara](#)
- [staimen perä staimen peräke](#)
- [palosanki paloampari](#)
- [kansio kiriekansio](#)

Suomi:

- [rukinlapa](#)
- [ruokelina ruokelina damasti](#)
- [kattalina kattalina etupistikiriontaa](#)
- [novtalina pikkulina, kiriotu](#)
- [novtalina ristipistikiriontainen novtalina](#)

**Samaan aiheeseen liittyviä esineitä**

leikkikalut:

- [kuvakirja Juvakirja kangasta](#)
- [leikkiheronen keuhuheronen](#)
- [lehlava leikkilava](#)
- [lehmootorivära leikkimootorivära](#)
- [lehtitraktori leikkitraktori](#)

**Sama materiaali**

savi:

- [kannu koristikannu](#)
- [kynnilinanka](#)
- [maljakko maljakko](#)
- [kannu kannu](#)
- [kuiho kuiho](#)

**Sama valmistusaika**

1930-1939:

- [halvovesipullo parfyymipullo](#)
- [kynnilinapakkaus maljakkokynnilinapakkaus](#)
- [lasmialaus lasmialaus](#)
- [tennismaila tennismaila](#)
- [tennismailan pinnoitusteline](#)

**Sama valmistuspaikka**

Suomi:

- [rukinlapa](#)
- [iakkunen viruu](#)
- [ruokelina ruokelina damasti](#)
- [kattalina kattalina etupistikiriontaa](#)
- [novtalina pikkulina, kiriotu](#)

Kuva 19. MuseoSuomi-sovellus /54/

### 6.3.3 Haun tarkentaminen valmiiden vaihtoehtojen avulla

Hakua voidaan tarkentaa yhdessä ulottuvuudessa ennaltamäärättyjen hierarkioiden tai hakutilanteessa muodostettujen käsiteklusterien avulla. Palveluiden hakuun liittyen ensiksi mainittu mahdollisuus on tarjottu useissa hakukoneissa (mm. Yahoo /145/) ja palveluhakemistoissa (mm. Asiakastiedon Aarre-palvelussa kuvassa 9). Käsiteklusterointia taas hyödynnetään mm. Findex- /74/ ja Grokker-työkaluissa /46/, joissa analysoidaan yleisten hakukoneiden tuottamia tulosjoukkoja, sekä Clusty-hakukoneessa /141/. Ryppäät esitetään näistä Findex-työkalussa listana ja Clusty-hakukoneessa sisennettyinä listoina. Grokker-työkalussa taas käyttäjä voi tarkastella aineistoa erikseen sisennettyjen listojen ja sisennettyjen diagrammien kautta, mitä havainnollistaa kuva 20.

The screenshot shows the Grokker search engine interface. At the top, there's a search bar with the text 'car' and a 'GROK' button. Below the search bar, there are tabs for 'Outline View' and 'Map View', with 'Map View' selected. The main area is a circular visualization of search results, with various colored bubbles representing different search results. Labels for these bubbles include 'Price Quotes', 'Car Dealers', 'Car Reviews', 'Safety Ratings', 'New Cars', 'Car Pictures', 'Car Search', 'Technical', 'Research Cars', 'Compare Prices', and 'Search'. On the right side, there's a list of search results with titles and snippets, including 'Car.com', 'I-CAR Collision Repair Training | worldwide', 'Yahoo! Autos', 'Welcome to I-CAR Online', 'CarDomain.com', and 'Edmunds.com'. The interface also includes a 'Working List' on the left and a 'Search within the map' section with filters for keyword, date, source, and domain.

Kuva 20. Grokker-työkalu

Kuva 20 esittää Yahoo-hakukoneen tulosjoukon ryppäitä haettaessa car-merkkijonolla. Kuvasta on havaittavissa, että ryppäiden koko ilmaistaan Grokker-työkalussa alueiden kokona. Tällöin kaikkein pienimpien ryppäiden tekstikuvausta ei ole mahdollisuutta näyttää suoraan, mutta aluetta osoittamalla teksti voidaan tuoda esiin (kuvassa 20: Safety Ratings). Findex-työkalussa ja Clusty-hakukoneessa taas ryppäiden koko ilmaistaan numeroilla.

Tutkimukset koskien Findex-työkalua osoittavat optimaalisen valmiiden vaihtoehtojen määrän olevan noin 10-20 /73/. Tekstiaineistoa klusteroitaessa näistä osa voi olla käyttäjälle hämmentäviä, ja yleensäkin klusteroinnin hyödyllisyyden on havaittu vaihtuvan eri tulosjoukkojen välillä /74/. Vaikka Jardinen ja van Rijsbergenin esittämä klusterihypoteesi olettaa relevanttien tulosten olevan keskenään samankaltaisia /66/, tämän toteutuminen aina lienee kuitenkin mahdotonta. Yleisesti ottaen Findex-työkalu näyttää olevan kuitenkin parempi ratkaisu perinteiseen tekstihakuun verrattuna mitattaessa nopeutta, tarkkuutta ja mielihyvää /74/. Myös Dennisin tutkimukset koskien Internetin käyttäjien muistikuormaa vahvistavat oletusta siitä, että valmiista haun tarkennusvaihtoehdoista voi olla hyötyä /22/. Näihin turvaudutaan Findex-työkaluun liittyvien tutkimusten mukaan noin joka neljännen haun kohdalla, ja lisähyöty tästä näytetään saatavan erityisesti käytettäessä epämääräisiä hakutermejä, milloin perinteisen tekstihaun tulosten järjestäminen usein epäonnistuu /72/.

#### 6.3.4 Moninäkömähaku

Moninäkömähäussa on ajatuksena tarjota käyttäjälle samanaikaisesti monia toisistaan riippumattomia suuntia lähestyä hakujoukkoa kustakin suunnasta haluamassaan tahdissa. Näitä suuntia kutsutaan tässä käyttöliittymäparadigmassa mm. näkymiksi ja faseteiksi, joiden hierarkiat visualisoidaan tyypillisesti sisennetyillä listoilla. /53, 88/

Esimerkkejä moninäkömahaun sovelluksista on lukuisia ja uusien tekemistä helpottamaan on mm. Lucene-pohjaisessa Solr-hakumoottorissa /128/ tuki moninäkömähäulle. Yksi näistä moninäkömähakua hyödyntävistä sovelluksista on Wine.com /144/, jonka toimintaa havainnollistaa kuva 21. Kuvan esimerkissä ”France”-hakusanalla haettaessa järjestelmä on tunnistanut hakusanan maantieteelliseksi alueeksi, jota voidaan tarkentaa. Lisäksi järjestelmä mahdollistaa haun tarkentamisen hinnan ja tyyppin suhteen.

The screenshot shows the Wine.com website interface. At the top, there is a navigation bar with links like 'wine shop', 'gift center', and 'community'. A search bar is visible with the text 'Enter keyword search advanced search'. Below the search bar, there are filters for 'price' (ranging from \$20 and below to \$80 and above) and 'type' (including Red Wine, White Wine, Rosé Wine, etc.). The search results section is titled 'search results' and shows 'France - Other regions' with 95 results. Two wine products are highlighted: 'Veuve Clicquot Brut Yellow Label' (Champagne, France) and 'Falvey Bourgogne Blanc Chardonnay 2005' (Burgundy, France). Each product listing includes an image, price, and 'add to cart' button.

Kuva 21. Wine.com /144/

Eri sovellusalueista puhuttaessa moninäkömähakua on hyödynnetty mm. seuraavien aineistojen hakua helpottamaan:

- Uutiset. Esimerkki: Guardian-lehden verkkoaineisto /49/.
- Yritysrekisteri. Esimerkki: Asiakastiedon Yritysfilteri-palvelu /118/.
- Verkkokauppa. Esimerkki: eBay Express /19/.
- Ruoka ja juoma. Esimerkkejä: Wine.com /144/ ja Epicurious.com /16/.

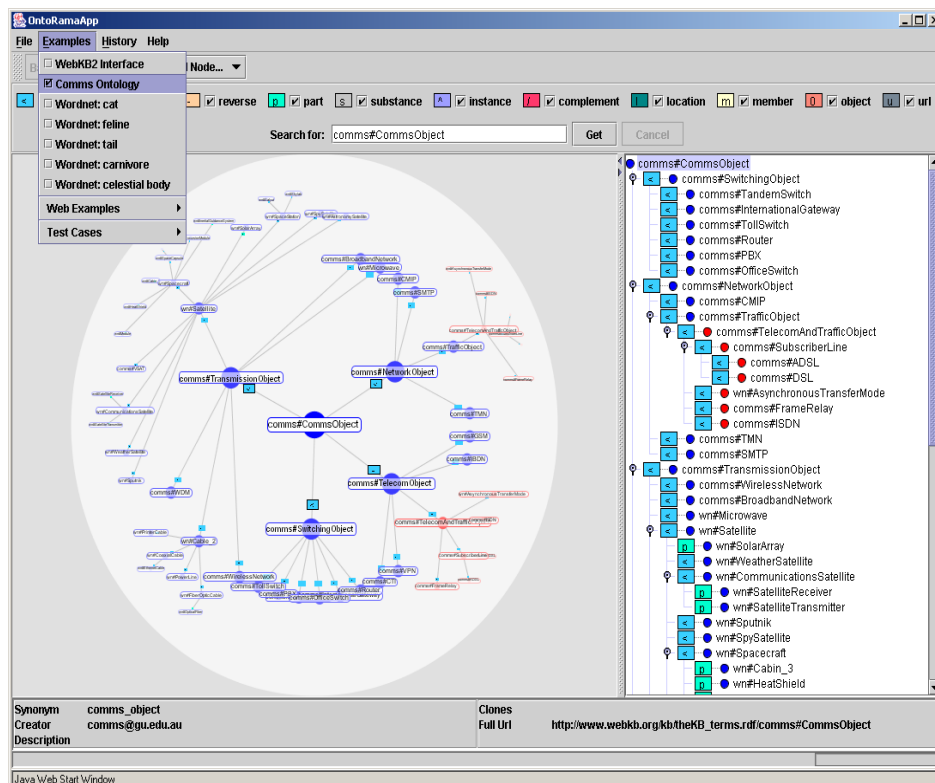
Sovellusalueiden laajan kirjon lisäksi moninäkömähaku näyttää lupaavalta menetelmältä myös käyttäjätutkimusten perusteella. Verrattaessa mm. yhden näkömähakun Yahoo-hakemistoa Flamenco-järjestelmään pohjautuvaan moninäkömähäiseen kuvahakusovellukseen havaittiin, että käyttäjät osasivat hyödyntää käyttöliittymän tarjoamia eri fasetteja ja olivat yleisesti myös paljon tyytyväisempiä moninäkömähakuun perustuvan järjestelmän toimintaan /27, 53/. Tämä voi johtua mm. siitä, että käyttäjät päätyivät umpikujaan erittäin harvoin käyttäessään moninäkömähakua /53/. Myös mSpace-projektin tutkimustulokset klassisen musiikin selailun osalta tukevat tätä olettamusta /109/. Siinä moninäkömähakuun pohjautuva sovellus osoittautui ylivoimaiseksi vertailtaessa yhteen näkömähäkuun ja moninäkömähakuun pohjautuvia sovelluksia keskenään. Selailu tässä ja kuvahakusovelluksessa toteutettiin loppukäyttäjille, mutta näiden perusteella on syytä olettaa, että myös tietovaraston ylläpitäjät voivat hyötyä moninäkömahaun käytöstä. Moninäkömahaun avulla voitai-

siin nimittäin esimerkiksi kartoittaa sitä, kuinka paljon ravintoloita palveluhakemistosta mahdollisesti puuttuu alueittain.

Moninäkömahaun toteutuksen kannalta on keskeistä tarjota tiedonhakijalle mielekkäät lähestymissuunnat tietoon. Se mitä fasetteja kussakin sovelluksessa kannattaa käyttää, riippuu hyvin pitkälle sovellusalueesta ja hakujoukon ominaisuuksista. Yleisiäkin periaatteita fasettien muodostamiseen on kuitenkin kehitelty. Jo ennen tietokoneaikaa 1930-luvulla intialainen S.R. Ranganathan ehdotti, että kirjastoaineistoja voitaisiin luokitella viiden perustavaa laatua olevan fasetin kautta /100/. Nämä viisi ovat: persoonallisuus (teoksen aihe), materia (materiaalit, joita aihe koskee), energia (kuvatut prosessit), avaruus (teoksen tapahtumapaikat) ja aika (teoksen sijoittuminen jollekin aikakaudelle).

### 6.3.5 Hakutermien monisuuntaisten suhteiden visualisointi

Sovelluksissa, joilla voidaan hakea palveluita, ei ole perinteisesti visualisoitu hakutermien monisuuntaisia suhteita. Käsitteiden monisuuntaisia suhteita ontologioihin liittyen on sen sijaan visualisoitu mm. OntoRama- /26/, WSMT- /6/ ja IRMA-työkaluissa /111/. Näistä Java-sovelluksena toimivassa OntoRama-työkalussa ontologian sisältöä visualisoidaan hyperbolisessa puunäkymässä (kuvassa 22). Hieman vastaavalla toimii myös WSML-kielisten ontologioiden kehittämiseen suunniteltu WSMT-työkalu, jossa samassa näkymässä voidaan myös muokata ontologian sisältöä.



Kuva 22. OntoRama-työkalu /27/

SeCo:ssa kehitetyssä IRMA-työkalussa taas käsitteiden monisuuntaiset suhteet visualisoidaan valikkorakenteena. Työkalu on kehitetty ontologiapohjaiseen annotointiin ja sen on tarkoitus helpottaa oikean käsitteen valintaa. IRMA:n toimintaa on havainnollistettu kuvas-

sa 23 matelija-käsitettä haettaessa. Työkalu nopeuttaa kuvan esimerkissä käsitteen hakua tekstin kirjoittamiseen reagoimalla. Lisäksi esimerkissä käsitteen merkitys voi olla helpompi ja nopeampi ymmärtää, koska käsitteen paikka hierarkiassa on visualisoitu.

Subject:	
rept	
L25F4: reptiles	Hierarchy
25FF4: fabulous animals ~ reptiles	L2: Nature
31A454: monsters ~ reptiles	L25: earth, world as celestial body
43C144: tormenting reptiles	L25F: animals
ys: reptiles	L25F4: reptiles
	L25F41: lizards
	L25F42: snakes
	L25F43: crocodiles
	L25F44: tortoises, turtles
	Related Concepts
	L34B31: terrarium

Kuva 23. IRMA-työkalu

## 7 AARRE-PALVELUN JA YRITYKSEN PALVELUHAKEMISTON KEHITYS

### 7.1 Tavoitteet palveluhakemistojen kehityksessä

Tavoitteena Yrityksen palveluhakemiston /97/ ja Aarre-palvelun /117/ kehittämisessä oli maksimoida sovellusten käytettävyys mahdollisimman monen käyttäjän ja tarpeen suhteen, joita molempia on paljon erilaisia. Käytettävyyden mittarina käytettiin standardia ISO 9241-11, joka jakaa käytettävyyden kolmeen osatekijään: tuloksellisuus, tehokkuus ja mielihyvä /110/.

Tuloksellisuuden toteutumisella käytettävyydessä tarkoitetaan mm. sitä, että käyttäjä löytää tarpeeseensa olennaista tietoa. Jotta tämä olisi mahdollista palveluhakemiston kohdalla, tulisi sen sisältää monipuolista ja olennaista sisältöä mahdollisimman paljon. Tiedon olemassaolo ei kuitenkaan riitä tuloksellisuuden toteutumiseen, vaan tiedon tulee olla myös löydettävissä. Ajatuskulut voivat vaihdella eri käyttäjien välillä, ja suuressa palveluhakemistossa yksittäinen tiedonmuru voi helposti hukkaa muiden tiedonmurujen joukkoon. Tiedon löytyminen on siis palveluiden haussa aina enemmän tai vähemmän epävarmaa. Tätä ongelmaa voidaan kuitenkin lievittää liittämällä hakujoukon yksilöihin metatietoa käsitteiden muodossa ja kuvaamalla käsitteiden liittymistä toisiin käsitteisiin mahdollisimman monipuolisella ja mielekkäällä tavalla tiedonhaun kannalta.

Tehokkuudella taas tarkoitetaan hakutehtävän suoritukseen kuluvaan aikaan. Samalla kun parannetaan tuloksellisuutta tarjoamalla useita vaihtoehtoisia navigointireittejä samaan tietoon, parannetaan samalla usein myös tehokkuutta. Lisäksi tehokkuuden lisäys vaikuttaa myös käyttäjän mielihyvän tunteisiin ennaltaehkäisemällä tilanteita, joissa käyttäjä turhautuu.

Palveluhakusovellusten toiminnasta voidaan saada selville käyttäjiltä kysymällä, että he pitävät jotain sovellusta parempana kuin toisia subjektiivisista syistä. Tästä on kysymys puhuttaessa mielihyvän suhteen paremmista sovelluksista. Mielihyvä käytettävyyden osatekijänä on siis hyvin subjektiivinen käsite.

## 7.2 Lähestymistavat aineistoihin

Riippuen tarpeesta ja käyttäjästä strategiaksi tiedonhakuun voi sopia parhaiten suora haku tai selailu, joista selailua tuetaan yleisissä hakukoneissa varsin suppeasti. Suoraa hakua käyttäjä voi haluta käyttää mm. akuuttien helposti muotoiltavien tarpeiden kohdalla. Esimerkkinä näistä ovat seuraavat tilanteet: 1) kuluttaja on nälkäinen kävellessään Helsingin Kampissa klo 13 ja 2) yrittäjä haluaa tietää Espoon Tapiolassa sijaitsevan TE-toimiston puhelinnumeron. Sen sijaan käyttäjä voi haluta selailulla yhdistäviä polkuja pitkin sellaisten tarpeiden kohdalla, jotka ovat vaikeasti muotoiltavissa suoraksi hauksi. Esimerkiksi seuraavat tilanteet ovat tällaisia: 1) kuluttaja miettii asuntonsa seinien ulkonäön kohentamista, 2) kuluttaja kartoittaa vapaa-ajanviettomahdollisuuksiaan ja 3) yrittäjä on aloittamassa vientitoimintaa ulkomaille, mihin hän hakee tukea tiedon ja palveluiden muodossa. Yhteistä näille kaikille selailutilanteille on se, että niihin on useita ratkaisuja, joiden luonnetta käyttäjä ei välttämättä tiedä etukäteen.

## 7.3 Aineistojen sisältö

Yleiset hakukoneet hakevat vastauksia aineistoista, jotka ovat tyypillisesti suuria. Nämä aineistot sisältävät siis paljon sellaista, joka ei ole millään tavalla olennaista kuluttajan ja yrittäjän tarpeiden kannalta. Tämä johtaa siihen, että käytettäessä yleisiä hakukoneita palveluiden hakuun haun kannalta olennainen aineisto-osa toisinaan hukkuu epäolennaisen tiedon joukkoon. Kohdennettujen palveluhakemistojen kohdalla tämä ilmiö on kuitenkin pienempi ongelma. Näin on mm. Yrityksen palveluhakemiston kohdalla, jonka aineisto on käytännössä kokonaan yrittäjälle potentiaalisesti olennaista. Toisin on kuitenkin Aarre-palvelun kohdalla. Osa palvelun tietovaraston tiedosta on erittäin suurella todennäköisyydellä turhaa kuluttajan näkökulmasta. Kovin moni palvelun käyttäjä kuluttajaroolissa nimittäin tuskin on kiinnostunut esimerkiksi kaivosalan palveluista. Näistä palveluista ovat ehkä kiinnostuneita jotkin yritykset, mutta toisaalta Aarre-palvelun potentiaalisia kuluttajakäyttäjiä on niin paljon, että erillisen ylimääräisestä tiedosta siistityn käyttöliittymän ja hakujoukon rakentaminen voisi olla hyödyllistä.

Palveluntarjoajien kotisivujen hakeminen yleisten hakukoneiden kautta ei aina onnistu, koska kaikilla palveluntarjoajilla ei niitä ole. Aarre-palvelun ja Yrityksen palveluhakemiston kautta tietoja näiden palveluntarjoajien toiminnasta on kuitenkin saatavilla, mistä johtuen ne voivat olla tältä osin palveluiden haussa yleisiä hakukoneita hyödyllisempiä käyttää. Pienyritysten suuresta määrästä voidaan päätellä, että näin voisi olla varsinkin kuluttajille suunnattujen palveluiden kohdalla. Toisaalta Yrityksen palveluhakemiston ja Aarre-palvelun tietovarastot näyttävät olevan kaukana täydellisyydestä. Näin on mm. Yrityksen palveluhakemiston kohdalla, josta puuttuu merkittävästi yrittäjille suunnattuja rahoitus- ja koulutuspalveluita. Aarre-palvelun todennäköisiä puutteita sen sijaan ovat palveluntarjoajien puuttuvat ja ylimääräiset toimipisteet. Aarre-palvelun tietovaraston pohjana on nimittäin kaupparekisteri, jonne on voitu ilmoittaa yrityksen konttori, jossa yritys ei palvele asiakkaitaan, ja jättää ilmoittamatta varsinaiset palvelupaikat. Toinen todennäköinen tietovarastosta johtuva ongelma Aarre-palvelussa on franchising-yritysten käsittely. Franchising-yritykset kun palvelevat kuluttajaa tyypillisesti nimellä, joka eroaa

niiden rekisteröidystä nimestä. Aarre-palvelun toiminnassa tämä on todellinen ongelma mm. etsittäessä Espoon Tapiolassa sijaitsevaa Subway-pikaruokaravintolaa, jonka toiminnasta vastaa Eat - Team Oy. Intuitiivisin haku tässä tapauksessa olisi todennäköisesti Subway Tapiola, mutta Aarre-palvelu ei tunne tämän nimestä yritystä.

## 7.4 Aineistojen sisällönkuvailu

Aarre-palvelun ja Yrityksen palveluhakemiston aineistoja on kuvattu jonkin verran kontrolloidulla sanastolla ja vapaalla tekstillä. Vapaat tekstikuvaukset voivat olla hyödyllisiä haettaessa ja tarkasteltaessa yksittäistä palvelua, mutta niiden pohjalta on haasteellista toteuttaa selailussa hyödyllistä moninäkömähakua. Perinteistä tekstihakua voidaan kyllä todistetusti parantaa klusterioimalla tulokset (kts. kappale 6.3.2), mutta mikään kirjoittajan tiedossa olevista klusterointisovelluksista ei pysty visualisoimaan tuloksia aidosti moniulotteisina moninäkömähakusovellusten tapaan (kts. kappale 6.3.4). Lisäksi ryppäiden muodostus ja nimeäminen ei onnistu läheskään kaikissa tapauksissa. Moni palveluiden hakutilanteista on kuitenkin ennakoitavissa, ja käyttäjien kohdistamat paineet hakusovellusten käytettävyydelle ovat niin korkeita, että lisäpanostus tiedon kattavampaan ja yksiselitteisempään kuvailuun on perusteltua. Näin on erityisesti Yrityksen palveluhakemiston kohdalla, sillä nykyisellään siinä haku vapaiden tekstikuvausten pohjalta ei ole useiden tarpeiden kohdalla tuloksellista. Esimerkiksi kansainvälistyvälle yritykselle palvelu antaa tuloksena vain neljä palvelua kansainvälisyys-hakusanaa käytettäessä. Todellisuudessa huomattavasti useampi hakemiston palveluista on suunnattu kansainvälistyvälle yritykselle.

Kontrolloidulla sanastolla kuvailu on monissa tilanteissa työläämpää ja siten kalliimpaa kuin vapaan tekstin käyttö tiedon kuvailussa. Yrityksen palveluhakemisto on kooltaan kuitenkin niin pieni, että sisällön kattavalle kuvailulle kontrolloitua sanastoa käyttäen ei ole todellisia esteitä. Lisäksi vaikka palveluita kuvaillaan vapaalla tekstillä, kuvauksissa ei kerrota palveluiden sisällöstä ja sopivuudesta suhteessa yrittäjään riittävästi. Esimerkiksi Ypäjän Yrittäjät ry kuvaa jäsenpalveluaan melko suppeasti seuraavalla kuvauksella: ”Jäsenpalveluja voivat olla mm. ajankohtaisista asioista tiedottaminen, tilaisuudet ja tapahtumat. ”. Riittämättömien kuvausten kohdalla tekstin koneellinen tulkinta voi muodostua mahdolliseksi tehtäväksi, koska jopa ihmiseltä palveluiden hyödyllisyyden tulkinta voi toisinaan onnistua vain valistuneiden arvauksien avulla.

Hyöty kontrolloidun sanaston kattavasta ja oikeasta käytöstä on siis ilmeisin aineiston selailussa. Kiinnitettyjen rajaustekijöiden suhteet voidaan nimittäin helposti visualisoida käyttöliittymässä, mistä on käsitteiden välillä liikkumisen lisäksi hyötyä moniulotteisessa suodatuksessa. Moninäkömähauksessa rajauksia kun voidaan tehdä helposti useita peräkkäin. Lisäksi peräkkäisten valintojen muodostaman sarjan ei välttämättä aina tarvitse olla ennaltamäärätty, koska rajausvalintojen visualisoiduilla suodatusvaikutuksilla voi olla vaikutusta käyttäjän tekemiin valintoihin. Tilanne on tällainen mm. silloin, kun käyttäjä löytää suuren häntä kiinnostamattoman ryppään, jonka poisto haun tulosjoukosta helpottaa tämän hahmottamista ja läpikäyntiä.



Rajaustekijöistä maantieteellisillä alueilla on suuri merkitys palveluiden kohdalla, koska sijainti rajoittaa useiden palveluiden käyttöä. Suuntaa-antava rajoitus voidaan päätellä palvelupaikkojen sijainnista tai palvelu voidaan määritellä valtakunnalliseksi. Näin on palveluita kuvattu mm. Yrityksen palveluhakemiston kohdalla, mikä on kuitenkin kirjoittajan mielestä epäonnistunut valinta. Näin siksi, koska monien hakemiston palveluiden sopivuus suhteessa käyttäjän sijaintiin voidaan ymmärtää kuntalistana tekemättä kovin suurta virhettä. Vastaavalla tavalla Aarre-palvelussa voidaan myös ymmärtää suuren palvelujoukon sopivuus suhteessa käyttäjän sijaintiin. Molemmissa palveluissa palvelupaikan sijainti on kuvattu yksiselitteisesti, mutta käytettäessä yleistä hakukonetta maantieteellisen termin kirjoitusasun esiintymiselle vapaan tekstin joukossa voi olla monta tulkintaa. Maantieteellisissä nimissä on nimittäin jonkin verran termejä, joiden merkityksiä ei voida päätellä niiden kirjoitusasusta. Toisaalta vaikka termin merkitys onnistuttaisiinkin vapaasta tekstistä eristämään, sen ei voida kuitenkaan varmasti päätellä kuvaavan palvelualueetta. Esimerkiksi Yrityksen palveluhakemistossa Ypäjän kunta tarjoaa yritystontteja ja kuvaa itseään seuraavasti: ”Turkuun matkaa 70 km, Tampereelle 90 km, Helsinkiin 130 km”. Tästä ei voida päätellä, että tarjottu palvelu sopisi Helsingissä toimivalle yritykselle, joka ei aio laajentaa tai siirtää toimintojaan Ypäjälle.

Maantieteellisten alueiden lisäksi aineistoa voidaan rajata myös aiheen perusteella edellyttäen, että aineisto on kuvattu kattavasti käyttäjän kannalta mielekkäitä aiheita käyttäen. Tämä toteutuu molempien aineistojen kohdalla melko heikosti, mistä johtuen Aarre-palvelun aineiston pohjalta kehityksen prototyypin käytettävyyttä ei voi käytännössä nousta sellaisten erikoissovellusten tasolle kuin Wine.com, eat.fi ja eBay Express (kts. kappale 6.3.3). Lisäksi jotta aihetta voitaisiin rajauksessa menestyksekkäästi käyttää, tulisi sovelluksessa käytettyjen sanojen ja niiden välisten suhteiden vastata mahdollisimman pitkälle käyttäjän ajatusmaailmaa. Mahdollisia aiheita on suuressa aineistossa luonnollisesti enemmän kuin pienessä. Näin ollen Aarre-palvelussa mainittu vaatimus on huomattavasti vaikeampi toteuttaa kuin Yrityksen palveluhakemistossa, jossa kontrolloidun sanaston tarkoituksenmukainen koko on paljon pienempi.

## **8 AINEISTOJEN JA NIIDEN KUVAILUSSA TARVITTAVIEN SANASTOJEN KARTOITUS**

### **8.1 Yrityksen palveluhakemisto**

#### **8.1.1 Sisällön hahmottaminen hakusovelluksen avulla**

Yrityksen palveluhakemiston sisällön kartoittamiseksi rakennettiin moninäkömahakäyttöliittymä MIT:n (Massachusetts Institute of Technology) SIMILE-projektissa kehityksen Exhibit-julkaisukehikon pohjalta /81/. Kehitetyn sovelluksen tietovarastona toimi JSON-formaatissa /4/ oleva tekstitiedosto, johon Yrityksen palveluhakemiston aineisto muunnettiin. Muunnoksen lähteenä käytettiin palvelun reaaliaikaisia HTML-sivuja, jotka noudettiin palvelusta Java-kielellä kehitetyn ohjelman avulla. Tämä vaihe oli melko yksinkertainen toteuttaa, koska palveluiden kuvaussivujen WWW-osoitteet oli helppo arvata. Myöskään

JSON-formaatissa olevan tiedoston tuottaminen HTML-sivujen pohjalta ei ollut kovin vaikeaa, koska noudetuilla HTML-sivuilla oli samanlainen rakenne.

Tekstimuotoisen tiedoston luonnin jälkeen tietovarastosta poistettiin sisällöltään toisteinen tieto, jolloin aineiston koko putosi reilusti alle puoleen. Tämä johtui siitä, että monet organisaatiot ja palvelut oli kuvattu palveluhakemistossa alueellisesti monistettuina. Tämän jälkeen aineisto kuvattiin organisaatioina ja niille alisteisina palveluina.

### 8.1.2 Palveluiden kuvailu hakusovellusta varten

Palveluiden alustava skeema määriteltiin hakusovellusta varten koostuvaksi seuraavista kentistä: 1) palvelun nimi, 2) palvelun kotisivu, 3) palvelun organisaation nimi ja 4) palvelun organisaation kotisivu. Näillä kentillä palvelu oli mahdollista yksilöidä ja linkittää käyttökuvaukseen. Hakusovelluksen selailukäytön mahdollistamiseksi skeemaa täydennettiin lisäämällä siihen seuraavat kentät: 1) maantieteellinen sopivuus, 2) palvelun tyyppi, 3) sopivuus suhteessa käyttäjän ominaisuuksiin ja 4) aihealue. Näille kentille annettiin arvoiksi tägejä (tag).

Selailukäytön mahdollistamiseksi lisätyille kentille annettiin tägi-arvoja yrityksen ja erehdyksen kautta tägi kerrallaan. Esimerkiksi havaittiin, että useissa palveluissa oli kysymys rahan lainaamisesta yrittäjille. Tämän havainnon seurauksena Rahan saanti - laina -tägi annettiin kaikkien löydettyjen lainapalveluiden aihealue-kentän arvoksi. Tägien joukko annotoinnin alussa oli siis tyhjä, josta se kasvoi annotointiprosessin edetessä. Annotaatioiden määrän kasvaessa aineiston aiemmin epäselvien palveluiden joukko pieneni. Tämän seurauksena palveluiden sisältöä oli helpompi arvioida, ja tietovarastoon voitiin turvallisesti lisätä uusia palveluita. Näistä tärkeimpinä mainittakoon OP-Pohjolan tuotteet ja sellaiset YritysSuomi-palvelussa esitellyt palvelut, jotka puuttuivat Yrityksen palveluhakemistosta.

Kentistä maantieteellinen sopivuus kuvattiin palveluille vain erittäin karkealla tasolla (valtakunnallinen tai alueellisesti rajoittunut) kuten myös palvelun tyyppi (sähköinen, koulutus, tietopankki tai jokin muu). Aihealueiden ja sopivuuden suhteessa käyttäjään kuvaamiseen taas oli keksittävässä paljon enemmän hakusovelluksen selailukäytön kannalta mielekkäitä tägejä. Tämä on havaittavissa sovelluksen toimintaa esittelevästä kuvasta 24. Sovelluksessa hakua vastaava palvelulistaus on sijoitettu näkymän keskelle, jakauma aihealue-tägien suhteen vasemmalle ja muiden kolmen selailukäyttöön liittyvän kentän tägigen jakauma oikealle.

The screenshot shows the YritysSuomi search interface. At the top left is the YritysSuomi logo. Below it is a search bar with the text 'Hae:'. To the right of the search bar, it says 'LISTA • KARTTA' and '530 Kpl filtered from 620 originally ()'. On the left side, there is a sidebar with 'Aihealueet' (Topic areas) and a list of 126 items, including 'Byrokraatia', 'Vakuutus - pakollinen', 'Kansainvälinen', etc. The main content area displays a list of services, each with a green bar and a title: 'Aloitusrahasato Vera Oy (1)', 'Veran pääomarahaus (Aloitusrahasato Vera Oy)', 'Cursor Oy, Kotkan-Haminan seudun kehittämissyhtiö (2)', 'Riskirahoitus (Cursor Oy, Kotkan-Haminan seudun kehittämissyhtiö)', 'Toimitilat - Kotka-Hamina (Cursor Oy, Kotkan-Haminan seudun kehittämissyhtiö)', 'Finnpartnership - Finnish Business Partnership Programme (1)', 'Kehityksmaaneuvonta (Finnpartnership - Finnish Business Partnership Programme)', 'Finnvera (24)', 'Alustakaus (Finnvera)', 'Finnveratakauk (Finnvera)', 'Investointi- ja käyttöpääomalaina (Finnvera)', 'Investointitakuu (Finnvera)', 'Kansainvälistymislaina (Finnvera)', 'Kansainvälistymistakaus (Finnvera)', 'Kehittämislaina (Finnvera)', 'Luottoriskitakuu (Finnvera)', and 'Ostajaluottotakuu (Finnvera)'. On the right side, there is a filter panel titled 'Yritykseni on (nyt/lähtöleveysuudessa)...' with a list of 486 items and checkboxes for various categories like 'naisyritys', 'vaihe - alkava yritys', 'kansainvälinen', etc. Below that is a 'Palvelun tyyppi' (Service type) filter with 345 items and checkboxes for 'Organisaatio (ei-julkinen)', 'Organisaatio (julkinen)', 'Palvelu - sähköinen', and 'Palvelu - tietopankki'. At the bottom of the filter panel is a 'Maantieteellinen sopivus' (Geographical suitability) filter with 250 items and checkboxes for 'Alueellisesti rajoittunut' and 'Valtakunnallinen'.

**Kuva 24.** Kehitetty sovellus yrittäjien hakutarpeisiin

Kuvailussa käytettyjen tágien määrä jäi vajaaseen 30 kappaleeseen, mikä on huomattavasti vähemmän kuin alkuperäisessä Yrityksen palveluhakemistossa on käytössä. Tämä johtui pääasiassa siitä, että alkuperäisessä Yrityksen palveluhakemistossa oli runsaasti käytössä sellaisia asiasanoja, jotka liittyivät vain yksittäisiin palveluihin tai palveluntarjoajiin. Tällainen oli mm. Ajokortit-asiasana, jolla haettaessa saatiin vastaukseksi vain yksi palvelu. Nämä asiasanat olivat kuitenkin mainittu melko usein palveluiden vapaissa tekstikuvauksissa, minkä seurauksena palvelut ovat todennäköisesti helposti löydettävissä tekstihaun avulla. Yksittäisiä palveluita tai palveluntarjoajia kuvaavilla asiasanoilla annotointi ei siis olisi parantanut käytännössä ollenkaan sovelluksen käytettävyyttä.

Toisin kuin käytettyjen tágien kohdalla alkuperäisen Yrityksen palveluhakemiston asiasanoista oli vaikeaa koostaa ryppäitä. Lisäksi monet asiasanat eivät näyttäneet liittyvän mielekkäällä tavalla käyttäjien tarpeisiin. Kirjoittajalle jäi esimerkiksi epäselväksi, miten YritysSuomi.fi- ja Yrityspalvelut-asiasanat liittyvät yrittäjän ominaisuuksiin tai tarpeisiin. Myös asiasanojen käyttö ei näyttänyt olevan systemaattista. Esimerkiksi Arvonlisäverotus-asiasanalla haettaessa saadaan vastaukseksi seudullisista yrityspalveluista vain Ylivieskan seudun elinkeinopalveluiden tarjoama Seudulliset yrityspalvelut -niminen palvelu.

Toteutuksessa keskityttiin siis määrän sijasta laatuun eli valittuja arvoja annotoitiin palveluille mahdollisimman kattavasti ja lisäksi siten, että arvojen ymmärrettiin olevan poisulkevia. Käyttöliittymässä tällä tarkoitetaan mm. sitä, että jos palvelun A palvelun tyyppi-kentän arvoksi on kuvattu vain tietopankki ja käyttäjä on valinnut kentän suhteen näytettäväksi kaikki koulutus ja muut palvelut, niin palvelu A:n näyttämättä jättäminen käyttöliittymässä on sisällöllisesti oikea päätös. Tätä kautta valintoja vastaavia tulosjoukkoja onnistuttiin pienentämään merkittävästi. Esimerkiksi kuvan 24 tilanteessa käyttäjä on ilmaissut omasta toiminnastaan epäsuorasti seuraavaa: 1) ei harrasta liiketoimintaa millään käyttöliittymässä luetellulla toimialalla, 2) ei ole vakiintuneessa eikä lopettavassa vaihees-

sa, 3) ei ole naisyritys ja 4) ei ole yritysmuodoltaan osuuskunta. Tämän seurauksena varmasti sopimattomia palveluita on onnistuttu poistamaan 91 kpl suhteessa alkuperäiseen hakujoukkoon.

Aihealue-valintoja tehdessä tulosjoukko oli mahdollista saada riittävän pieneksi ajatellen niiden läpikäyntiin tarvittavaa aikaa ja vaivaa. Jos siis käyttäjä haluaisi kartoittaa kaikki palveluhakemiston hänelle mahdollisesti sopivat palvelut, hän voisi tehdä sen parhaassa tapauksessa yhdellä haulilla. Hakua varten käyttäjän tarvitsisi ainoastaan vastata kyllä tai ei pieneen joukkoon helposti ymmärrettäviä kysymyksiä. Tiedonhakuun tarvitsisi siis käyttää vain vähän aikaa ja vaivaa. Kartoituksen tulosjoukosta ei kaikissa tapauksissa löytyisi käyttäjälle sopivia palveluita, mutta tämän tiedon saaminen mahdollisimman nopeasti ja helposti käyttöliittymän avulla on käyttäjän kannalta toivottavaa.

Aihealue- ja sopivuus suhteessa käyttäjään -kenttien kohdalla joidenkin palveluiden kohdalle ei kenttien arvoksi ollut keksittävässä yhtään tägiä (kuvassa 24: {missing this field}). Näiden tapausten kohdalla oltaisiin voitu lisätä jotain muuta -tyyppinen tägi kentän arvoksi, mutta tälle tägille ei keksitty sopivaa nimeä. Lisäämällä tägi jonkun muun tegin kanssa kentän arvoksi oltaisiin voitu ilmaista ero suosittelun ja poissulkevyyden välillä. Kaikki kehitetyn sovelluksen palvelukuvaukset kuitenkin kuvattiin tegin suhteen poissulkevana, koska palveluiden sisällöstä ei ollut riittävästi tietoa saatavilla. Toisaalta pelkkien poissulkevien palvelukuvauksien avulla pystyttiin ilmaisemaan suosittelun ja sopivuuden ero käyttöliittymässä. Kuvan 24 tilanteessa ero on seuraava:

- suosittelu. Tehdessään elinkaari - alkava-, kansainvälinen- ja kansainvälinen (muut maat) -valinnat sopivuus suhteessa käyttäjään -kenttään käyttäjä näkee hänelle erityisesti suositellut palvelut.
- sopivuus. Tehdessään suosittelu-valintojen lisäksi {missing this field} -valinnan käyttäjä näkee kaikki hänelle mahdollisesti sopivat palvelut.

Annotoinnin jälkeen tägit peilattiin KOKO-ontologiaan ja UNSPSC-luokitukseen palvelukuvausten uudelleenkäytettävyyden helpottamiseksi (liitteessä 2). Peilauskuvauksesta on havaittavissa, että melkein kaikille käytetyille arvoille löytyi ainakin osittainen vastine KOKO-ontologiasta, mutta UNSPSC-luokitukseen peilaus taas ei onnistunut kuin joissain tapauksissa. Lisäksi sovelluksen tietovaraston kasvamiseen ja uudelleenkäyttöön varauduttiin myös kartoittamalla KOKO-ontologian käsitteitä, jotka oltaisiin voitu liittää yrittäjille suunnattujen palveluiden kuvauksiin (liitteessä 3).

## 8.2 Aarre-palvelu

### 8.2.1 Toimialaluokitus kuluttajille suunnatun sovelluksen sanastona

Aarre-palvelu käyttää nykyisin sanastonaan toimialaluokitusta (kts. kappale 5.2), joka ei näytä sopivan kuluttajien käyttötarpeisiin kovin hyvin ainakaan sellaisenaan. Toimialaluokitus kun sisältää jonkin verran käsitteitä, joita kuluttaja pitää itselleen erittäin suurella todennäköisyydellä tarpeettomina (mm. kaivosalan palvelut) ja tarpeensa kannalta epäselvinä (mm. muu kustannustoiminta). Toisaalta suhteessa moniin keltaisiin sivuihin merkittäviä määriä kuluttajan kannalta oleellisia käsitteitä puuttuu, mikä haittaa erityisesti suorien

hakujen onnistumista. Esimerkiksi Aarre-palvelu ei sisällä Huvipuistot-nimistä valmista valintaa, joka löytyy Yritysopas.com-palvelusta /119/. Selailukäytössä taas ongelmia syntyy siitä, että toimialaluokitus sisältää paljon termejä, joiden määrän rajaaminen ei onnistu käyttäjältä kovin helposti. Aarre-palvelu nimittäin näyttää toimialaluokituksen täydellisenä ilman informaatiota toimialojen yritysten määristä. Osa toimialoista on mm. aluerajauksen jälkeen todennäköisesti tyhjiä, mutta tätä tietoa ei hyödynnetä käyttöliittymässä. Luokituksen hierarkian juuritasolla valintojen määrä on kohtuullinen, mutta tämä ei auta niissä tapauksissa, joissa käyttäjä ei katso muodostuvien ryhmien olevan hänen tarpeensa kannalta relevantteja. Lisäksi selailukäyttöä haittaa myös se, ettei monia intuitiivisia käsitteiden välisiä suhteita ole kuvattu. Esimerkiksi matkatoimistojen ja hotellien välillä on useiden käyttäjien kannalta vahva käyttöön perustuva assosiaatio, mutta tätä tietoa ei hyödynnetä palvelun käyttöliittymässä.

### 8.2.2 Olemassaolevien sanastojen soveltuvuusarviointi

Toimialaluokituksen soveltuvuusongelmien vuoksi Aarre-palvelun sanastoa yritettiin kehittää hyödyntämällä muiden olemassaolevien sanastojen sisältöä. Ensin tarkasteltiin keltaisissa sivuissa käytettyjä sanastoja, mutta näitä ei päädytty hyödyntämään, koska ne sisälsivät kuluttajan kannalta huomattavia määriä ylimääräisiä käsitteitä. Esimerkiksi kuluttajat eivät ole todennäköisesti kiinnostuneita metallien jalostuksesta. Toisaalta niistä myös puuttui paljon kuluttajan kannalta hyödyllisiä käsitteitä ja käsitteiden välisiä suhteita oli kuvattu niissä niin niukasti, että mielekkään moninäkömahaun toteuttaminen näiden sanastojen avulla näytti käytännössä mahdottomalta. Esimerkiksi hevosiin liittyvien käsitteiden yhteyttä toisiinsa ei ollut kuvattu.

Muista sanastoista COICOP-luokitusta käytettiin Veturi-projektissa toteutussa palveluhakusovelluksessa (kts. kappale 5.2), joten sitä päädyttiin tarkastelemaan seuraavaksi. Luokitus oli saatavilla tekstikuvauksena, joka muunnettiin aluksi ohjelmallisesti OWL-kielelle (kts. kappale 4.2). Luokituksen hierarkiasuhteet kuvattiin RDFS:n subClassOf-suhteella (kts. kappale 4.1), vaikka hierarkiasuhteiden merkitys ei suoraan tätä suhdetta vastannutkaan. Näin kuitenkin tehtiin, sillä tällöin luokitusta oli helppo käsitellä Protégé-ohjelman avulla (kts. kappale 5.1.2). Jotta luokitusta voitiin turvallisesti muokata, OWL-muotoisesta tiedostosta otettiin kopio ja kopion käsitteille kuvattiin ohjelmallisesti yhteydet alkuperäisiin käsitteisiin OWL:n equivalentClass-suhteita käyttäen. Ohjelmointikielenä tämän toteuttamisessa käytettiin Javaa.

Tarkasteltaessa COICOP:n sisältöä havaittiin, että se sisältää paljon hyödyllisiä käsitteitä kuluttajan kannalta. Luokitusta ei kuitenkaan näyttänyt olevan mielekäästä soveltaa sellaisenaan kuluttajille suunnatun palveluhakusovelluksen käyttöliittymässä, koska luokitus sisälsi jonkin verran ylimääräisiä käsitteitä ja käsitteiden nimiä oli tarvetta muokata. Esimerkiksi Henk.koht. hygienian hoitoon tark. sähkökäytt. Laitt. -niminen käsite vaikutti sopimattomalta hakusovelluksen käyttöliittymässä sellaisenaan käytettäväksi, ja Kopiot dokumenteista, lehti-ilmoitukset ja mainokset -käsitteen kirjoittaja olisi jakanut useammaksi käsitteeksi. Lisäksi luokituksen hierarkia ei vaikuttanut kaikelta osin tarkoituksenmukaiselta, sillä paljon intuitiivisesti selviä suhteita puuttui käsitteiden väliltä. Esimerkiksi oluen

ja baarien välistä vahvaa käyttöön perustuvaa assosiaatiota ei ollut kuvattu. Lisäksi osa hierarkiamäärittelyistä näytti suorastaan tarpeettomilta (mm. 02.2.0.2 Sikarit → 02.2.0.2.1 Sikarit → 02.2.0.2.1.1 Sikarit).

Suurimmaksi ongelmaksi COICOP:n käytössä suoraan tiedonhakuovelluksen käyttöliittymässä paljastui sen suppea koko. Näytti nimittäin siltä, että suhteessa keltaisissa sivuissa käytettyihin sanastoihin COICOP:sta puuttui merkittäviä määriä oleellisia käsitteitä. Monet näistä puuttuvista käsitteistä näyttivät kuitenkin löytyvän KOKO:sta (kts. kappale 3.2.1), mistä johtuen se näytti COICOP:ia lupaavammalta kandidaatilta pohjasanastoksi kehitettävään sovellukseen. Toisaalta KOKO:a tarkemmin tarkasteltaessa paljastui, että kuluttajan kannalta ylimääräisten käsitteiden sulkeminen siitä näytti olevan käytännössä mahdotonta. Näin ollen myöskään KOKO ei näyttänyt soveltuvan suoraan hyödynnettäväksi kehitettävän sovelluksen sanastona tai sen osana.

COICOP:n ja KOKO:n jälkeen tarkasteltiin lähemmin UNSPSC-luokitusta (kts. kappale 5.2). Luokitus oli saatavissa Excel-muodossa, josta se muutettiin OWL-muotoon liitteessä 4 esitellyn Java-ohjelman avulla. Samoin kuin COICOP:n kohdalla OWL-muotoon muunnos tehtiin siitä syystä, että sitä voitaisiin tarkastella lähemmin Protégé-ohjelman avulla. Melko pian Protégé-ohjelmalla tarkastelun aloittamisen jälkeen kuitenkin havaittiin, että luokituksen selailu ohjelman avulla ei vaikuttanut mielekkäältä. Erityisesti yksittäisten käsitteiden paikallistaminen oli hidasta, koska ohjelmalla ei ollut mahdollista hakea käsitettä tekstin automaattista täydennystä hyväksikäyttäen. Tämä toiminto oli kuitenkin toteutettu ONKI-selaimen (kts. kappale 5.1.1), jonka avulla luokituksen tarkastelua jatkettiin.

UNSPSC julkaistiin ONKI-palvelimella SKOS-muodossa (kts. kappale 4.3). Julkaistava tiedosto sisälsi yleisiä määrittelyjä (liitteessä 5), joiden yhteyteen lisättiin käsittekohtaiset määrittelyt. Nämä muodostettiin OWL-muotoisista määrittelyistä yksinkertaisilla tekstin korvauksilla, joiden lopputuloksena saatiin seuraavanlaisia tekstilohkoja:

```
<rdf:Description rdf:about="http://www.yso.fi/onto/unspsc/C40101801">
  <skos:broader rdf:resource="http://www.yso.fi/onto/unspsc/C40101800"/>
  <skos:prefLabel xml:lang="fi">Radiaattorit</skos:prefLabel>
  <skos:prefLabel xml:lang="en">Radiators</skos:prefLabel>
  <rdf:type rdf:resource="http://www.w3.org/2008/05/skos#Concept"/>
  <skos:inScheme rdf:resource="http://www.yso.fi/onto/unspsc/conceptscheme"/>
</rdf:Description>
```

Edelläolevasta lohkosta nähdään helposti, että SKOS-muodossa eri käännökset (Radiaattorit ja Radiators) sidottiin käsitteeseen skos:prefLabel-määrittelysillä ja ylätason käsitteeseen (40101800 Lämmityslaitteet ja niiden osat ja tarvikkeet) viitattiin skos:broader-suhdetta käyttäen.

UNSPSC:n tarkastelu ONKI-selaimen avulla (osoitteessa <http://www.yso.fi/onto/unspsc/>) osoittautui erittäin hyödylliseksi. ONKI-selaimen automaattisen täydennyksen avulla luokituksen sisältöä voitiin helposti verrata KOKO-ontologiaan ja muihin sanastoihin. Vertailujen päätteeksi päädyttiin kuitenkin siihen, että UNSPSC ei sovellu suoraan käytettäväksi kehitettävän sovelluksen sanastona tai sen osana. Näin siksi, koska luokitus on mm.

KOKO:on verrattuna suppea tarkasteltaessa kuluttajan kannalta oleellisia käsitteitä. Toisaalta joitakin kuluttajan kannalta oleellisia käsitteitä puuttui KOKO:sta, mutta ne löytyivät UNSPSC:stä. Lisäksi UNSPSC:ssä osa nimikkeistä määritellään KOKO:on verrattuna huomattavasti tarkemmin. Esimerkkinä näistä mainittakoon naisten saappaat -nimike, jolle ei löydy suoraa vastinetta KOKO:sta.

UNSPSC:n jälkeen tarkasteltiin vielä eCI@ss- ja CN-luokituksia, mutta nämä eivät näyttäneet sisältävän juuri mitään uutta keltaisten sivujen sanastoihin, TOL:iin, KOKO:on, COICOP:iin ja UNSPSC:hen nähden. Näin ollen siis mitään olemassaolevaa sanastoa ei ollut mielekästä hyödyntää suoraan. Tämä johti samalla siihen, että GenTax-menetelmää (kts. kappale 5.2.2) ei ollut hyödyllistä käyttää. GenTax-menetelmän käyttö olisi nimittäin johtanut alkuperäisten nimikkeiden määrän kasvamiseen, mikä olisi hankaloittanut merkittävästi nimikkeistön muokkaamista. Toisaalta GenTax-menetelmän hyödyntäminen ei näyttänyt mielekkäältä vaihtoehdolta myöskään siksi, koska alkuperäinen hierarkia nähtiin tarpeellisena säilyttää vain soveltuvin osin kaikkien tarkasteltujen luokituksen kohdalla.

### 8.2.3 Kuluttajaontologian kehitys

Kuluttajille suunnatun palveluhakusovelluksen käyttöliittymässä käytetty käsitteistö päädyttiin luomaan tyhjästä, koska valmista ja tiedonhaun kannalta mielekästä käsitteiden kokoelmaa ei löytynyt siinä käytettäväksi (kts. edellinen alakappale). Tavoitteena oli luoda sellainen käsitteistö, joka sisältäisi sovelluksen kohderyhmän kannalta mahdollisimman paljon mielekkäitä ja mahdollisimman vähän ylimääräisiä käsitteitä. Näin siksi, jotta käyttäjän olisi mahdollista löytää hakemansa suorilla hauilla mahdollisimman nopeasti ja pienellä vaivalla. Tavoitteeseen pääsemiseksi luotiin Protégé-ohjelman (kts. kappale 5.1.2) avulla OWL-muotoinen ontologia (kts. kappale 4.2), johon kerättiin yhteen keltaisten sivujen valmiita hakuvaihtoehtoja. Tämän jälkeen käsiteltävästä joukosta poistettiin ylimääräiset käsitteet, jotta tekstin automaattista täydennystä hyödynnettäessä käyttäjän ei tarvitsisi käyttää turhaan aikaa varmasti sopimattomien vaihtoehtojen tarkasteluun ja ylimääräisten lisärajoitusten muotoiluun.

Hakujoukon vaihettamisen suodattamisen mahdollistamiseksi käsitteet yritettiin ryhmitellä mielekkäiksi kokonaisuuksiksi. Lisäksi käsitteiden välille kuvattiin hierarkia. Lopputuloksena syntyi kooltaan noin kahden sadan käsitteen kokoelma. Käsitteistö jakaantui viiteen näkymään: aika, paikka, toiminta, materia sekä immateria ja fysikaaliset ilmiöt, jotka vastasivat pitkälti Ranganathanin ehdottamia (kts. kappale 6.3.3). Ainoastaan immateria ja fysikaaliset ilmiöt -ryhmälle ei löytynyt selkeää vastinetta.

Syntyneen käsitteistön muokkaaminen ja selailu Protégé-ohjelmalla vaikeutui sitä enemmän mitä suuremmaksi käsitteiden joukko kasvoi. Tämän vuoksi käsitteistö päädyttiin julkaisemaan ONKI-palvelimelle. ONKI-selaimella käsitteiden joukkoa selattaessa havaittiin, että siihen oli kertynyt jonkin verran duplikaatteja ja merkittävä määrä selvästi oleellisia käsitteitä puuttui joukosta. Nämä puutteet korjattiin, minkä jälkeen syntynyttä käsitteistöä verrattiin KOKO-ontologiaan (kts. kappale 3.2.1). Tämän ensijaisena tavoitteena oli kasvattaa kehitetyn käsitteistön kokoa, mikä toteutettiin täydentämällä sitä KOKO:sta yhteisten käsitteiden ympäristöistä. Tavoite onnistui yli odotusten, sillä kehitte-

tyn käsitteistön koko moninkertaistui vajaan tuhanteen käsitteeseen. Runsaasti puuttuvia käsitteitä löytyi erityisesti suppeampien käsitteiden joukosta. Lisäksi uusia käsitteitä löydettiin merkittäviä määriä myös lähi- ja yläkäsitteiden kautta. Useissa tapauksissa yhteisten käsitteiden KOKO:n puolista ympäristöä ei kuitenkaan kopioitu sellaisenaan, koska vain osa KOKO:n käsitteistä nähtiin kuluttajan kannalta tarpeellisina. Lisäksi KOKO:sta suoraa kopiointia rajoitti myös siinä olevat käsitteet, joiden merkitykset näyttivät olevan täysin tai lähes identtisiä. Esimerkiksi seuraavat käsitteet olivat tällaisia: hehkulamput (valo) ja hehkulamput (mao).

Kopioitaessa käsitteitä KOKO:sta havaittiin, että annotoidun tiedon uudelleenkäytön kannalta voisi olla hyödyllistä kuvata käsitteistön yhteydet KOKO:on. Tämä toteutettiin luomalla kehitettyyn käsitteistöön koko-viittaus-niminen ominaisuus, joka sai arvokseen URI-viittauksia KOKO:n vastaaviin käsitteisiin. Ominaisuus kuvattiin kaikille kehitetyn käsitteistön käsitteelle niin tarkasti ja kattavasti kuin oli mahdollista.

KOKO:n jälkeen käsitteistöä täydennettiin aiemmin luotua UNSPSC:n ONKI-versiota hyödyntäen (kts. edellinen alakappale). Lopputuloksena kehitettyyn käsitteistöön lisättiin merkittävä määrä uusia käsitteitä. Lisäyksen yhteydessä yhteys UNSPSC:hen säilytettiin kehitettyyn sanastoon lisätyn UNSPSC-koodi-nimisen ominaisuuden avulla. Ominaisuudelle annettiin arvoiksi UNSPSC:n numerosarjoja (kahdeksan numeroa), jotka olisi helppo muuttaa URI-viittauksiksi yksinkertaisilla tekstinkorvauksilla. Luodun ominaisuuden avulla vastaavuus UNSPSC:hen kuvattiin myös aiemmin lisätyille käsitteille niin tarkasti ja kattavasti kuin mahdollista. Seuraavaksi käsitteistöön haettiin täydennystä vielä selailemalla kuvasanakirjoja (mm. Collins French Visual Dictionary /15/) ja sanoja ryhmitteleviä teoksia (mm. Collins Easy Learning French Words /14/). Tämän seurauksena uusia käsitteitä kehitettyyn käsitteistöön löytyi joitain kymmeniä. Lopputuloksena käsitteiden määrä nousi yli tuhanteen, joista vastaavuus KOKO:n suhteen löytyi noin 90 %:lle käsitteistön käsitteistä. Lisäksi UNSPSC:n suhteen vastaavuus löytyi noin 20 %:lle käsitteistön käsitteistä. Toimialaluokitukseen suhteen täydellistä siltausta kehitetylle käsitteistölle ei ollut mielekästä tehdä. Näin siksi, koska Aarre-palvelun aineiston sisällönkuvailu oli paikoittain liian köyhää ja kuluttajan tarpeiden kannalta epämääräistä (kts. kappale 8.2.1). Alkuperäisissä annotaatioissa oli todennäköisesti paljon hyödyllistä ja paikoittain jopa riittävästi informaatiota, mutta olemassaolevan ja puuttuvan tiedon määrää ei voitu arvioida tämän tutkielman yhteydessä aineiston puutteen vuoksi.

#### 8.2.4 Kuluttajaontologian mielekkyyden testaus

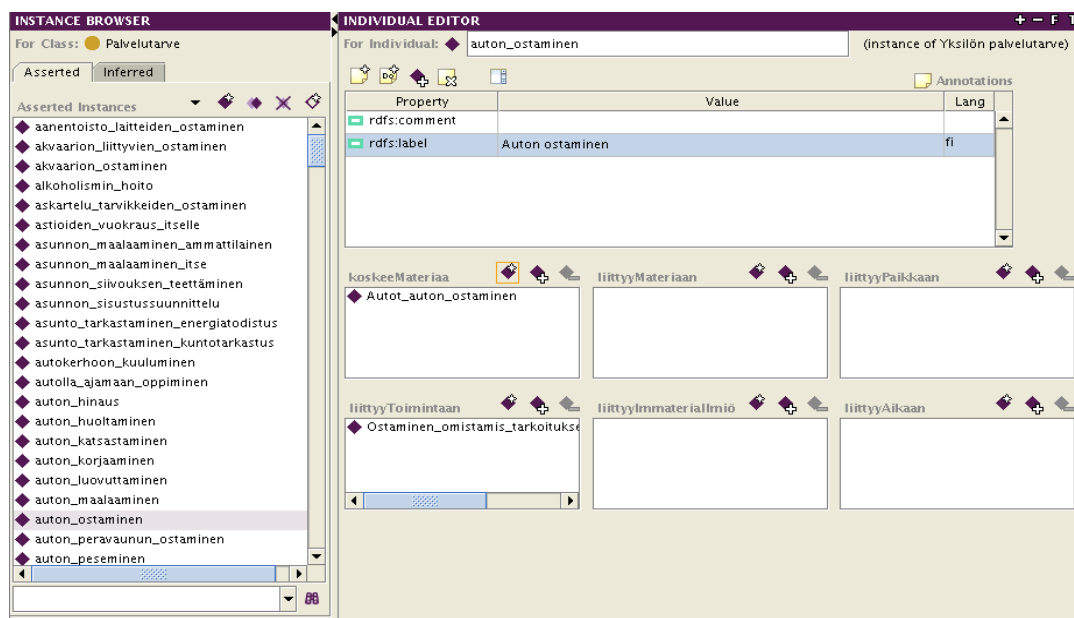
Kehitetyn käsitteistön sisällön mielekkyyttä ajateltiin olevan mahdollista testata epäsuorasti hyödyntämällä sitä tekstin louhinnassa. Tätä suunnitelmaa ei päästy toteuttamaan aineiston puutteen vuoksi tämän tutkielman yhteydessä, mutta Aarre-palvelun kehityksen kannalta tulevaisuudessa tämä testi olisi kuitenkin hyödyllistä tehdä. Näin siksi, koska tekstin louhintaa olisi helppo kokeilla ja toisaalta sen tulokset todennäköisesti kertoisivat jotain kehitetyn käsitteistön mielekkyydestä.

Tekstin louhinta voitaisiin toteuttaa käyttämällä palveluiden vapaita tekstikuvauksia (mm. Internet-sivut) tietolähteenä ja täsmäyttää ne kehitetyn käsitteistön suhteen hyödyntäen



esimerkiksi SeCo:ssa kehitettyä POKA-nimistä automaattisen annotoinnin kehikkoa /125/. Lopputuloksena tämän hyödyntämisestä joukko käsitteitä olisi liitettyä palveluntarjoajiin, mitä kautta käsitteiden jakauma voitaisiin visualisoida sisennettyinä listoina tai Grokker-työkalun tapaan (kuvassa 19) aineiston kokonaiskuvan hahmottamisen helpottamiseksi. Saatu lopputulos olisi todennäköisesti mielenkiintoinen, mutta se ei kuitenkaan välttämättä soveltuisi suoraan kuluttajien katsottavaksi. Osa louhitusta tiedosta olisi todennäköisesti väärää tai puutteellista ja osa taas riittävän oikein, mistä johtuen louhinnan tulos voitaisiin ymmärtää välituloksena. Tämä välitulos taas voitaisiin tarpeen mukaan manuaalisesti tarkistaa ja korjata, mikä olisi todennäköisesti huomattavasti nopeampaa kuin annotointi tyhjästä. Valmista työkalua tarkistamiseen ja korjaamiseen ei ole tiedossa, mutta DYNA-POKA-työkalun /125/ toimintaa muokkaamalla tällainen olisi kuitenkin melko suoraviivaisesti kehitettävissä.

Tekstinlouhinnan lisäksi kehitetyn käsitteistön mielekkyyttä ajateltiin voitavan testata myös yrittämällä peilata joukko todellisia tarpeita sen suhteen. Ideaa kokeiltiin käytännössä Protégé-työkalun (kts. kappale 5.1.2) avulla luomalla kehitettyyn sanastoon erillinen tarve-käsite, jonka ilmentymiksi lisättiin vajaa 300 kappaletta tyhjästä keksittyjä mahdollisia kulutustarpeita (mm. auton ostaminen). Tämän jälkeen lisätylle tarve-käsitteelle määriteltiin ominaisuudet, joiden avulla oli mahdollista kuvata tarpeen liittyminen kehitetyn sanaston näkyymiin. Ominaisuuksien arvoiksi annettiin aiemmin luotujen käsitteiden uusia ilmentymiä, jotka nimettiin seuraavalla syntaksilla: [viitattu käsite]\_[kuluttajatarve]. Tämän vaiheen toteutusta esittävässä kuvassa 25 näin on tehty auton ostamisen tarpeelle, joka on onnistuttu kuvaamaan toiminta- ja materia-näkymien suhteen (kuvan liiittyToimintaan- ja koskeeMateriaa-kohdat).

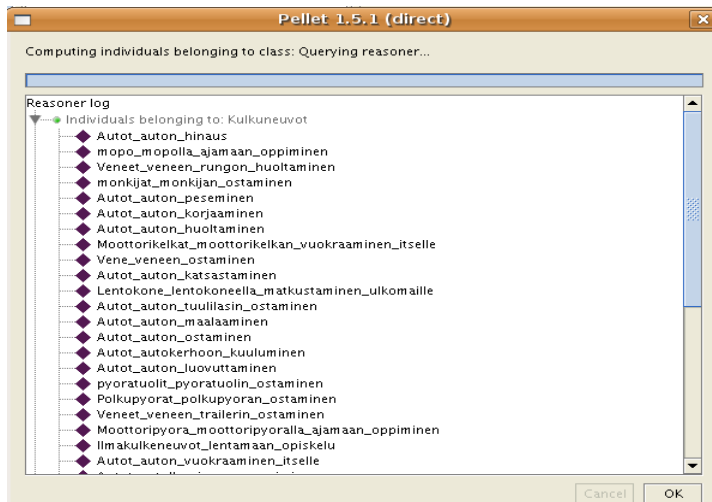


**Kuva 25.** Tarpeiden kuvaus käsitteinä näkyymiin Protégé-työkalussa

Tarpeiden kuvaus näkymien suhteen onnistui kehitetyllä sanastolla melko helposti. Kuvauksen suoraviivaisuudessa havaittiin tosin vaihtelua tarpeiden välillä. Esimerkiksi auton ostaminen onnistuttiin kuvaamaan suoraviivaisesti ja yksikäsitteisesti, kun taas homeen ja kosteuden mittaus asunnossa oli vaikeampi ilmaista kehitetyn käsitteistön avul-

la. Suorien tarve-käsitteiden käyttö kuluttajan hakukäyttöliittymässä on siis mielekästä joidenkin tarpeiden kohdalla.

Lopputulosta tarvekäsitteiden selittämisestä kehityksen käsitteistön avulla päätettiin tutkia Protégé-ohjelman päättelytoiminnallisuuden avulla. Kuvassa 26 tämän vaiheen toteutusta on havainnollistettu päättelystä kulkuneuvoihin liittyviä tarpeita. Kulkuneuvoihin liittyväksi tarpeeksi on päätelty mm. autokerhoon kuulumisen tarve.



Kuva 26. Kulkuneuvot-tarpeiden päättely Protégéssa

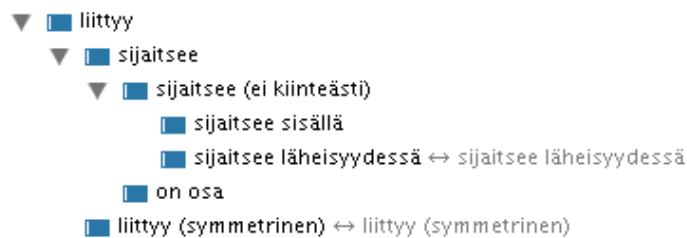
Protégén päättelytoiminnallisuuden käytön lopputuloksena näytti siltä, että sanaston sisältö oli kuluttajan kannalta kohtuullisen mielekäs. Positiiviseen lopputulokseen päädyttiin myös yritettäessä kuvata mielekkäitä suhteita käsitteiden välille (mm. akvaariokalat ja vesi nesteenä) ja kuluttajan kontekstien (mm. ruumiinosat, asunto, kulkuneuvot ja motiivit) suhteen. Mielekkäitä suhteita käsitteiden välille onnistuttiin kuvaamaan melko paljon eli kehitetty käsitteistö sopii luultavasti suunniteltuun käyttötarkoitukseensa melko monissa tapauksissa. Tosin eri ihmiset saattavat kutsua samoja käsitteitä eri nimillä ja ymmärtää niiden merkitykset eri tavoin /1, s. 69-72/ eli sopivuudessa voi olla eroja kuluttajien välillä. Vaihteluiden vaikutusta käytettävyyteen on kuitenkin vaikeaa arvioida ilman perusteellista käytännön testausta. Tätä ei kuitenkaan ollut mahdollista tehdä tämän tutkielman yhteydessä johtuen mm. ajan puutteesta.

## 8.2.5 Kuluttajaontologian käsitteiden sisäinen linkitys

Aineiston selailun mahdollistamiseksi loppukäyttäjän käyttöliittymässä kehitetyn käsitteistön sisäisiä suhteita tarkennettiin. Aiemmin oli kuvattu käsitteiden hierarkia, käsitteiden liittyminen näkyviin ja tarvetermien peilaus käsitteiden suhteen. Tarvetermeistä viitattiin ilmentymiin, mutta nämä viittaukset muutettiin suoriksi käsiteviittauksiksi kehitetyn käsitteistön osien määrän pienentämiseksi. Lisäksi tutkittiin käsitteiden liittymistä toisiinsa muiden kuin hierarkiasuhteiden kautta ja käsitteiden peilausta kuluttajan kontekstien (mm. ruumiinosat, asunto, kulkuneuvot ja motiivit) suhteen.

Käsitteiden välisten suhteiden kuvauksessa hyödynnettiin Aitchisonin tekemien sanatestien tuloksia, joiden mukaan ihmiset muistavat sanoja seuraavasti: 1) samantasoinen assosiaatio

(mm. suola ja pippuri), 2) pareittain esiintyvät sanat (mm. suola ja vesi), 3) yleistyksset (mm. väri ja punainen), 4) synonyymit ja 5) meronymia (mm. jalka ja varvas) /1, s.86, 106/. Yritettäessä kuvata käsitteiden välisiä samantasoisia assosiaatioita havaittiin, että esineiden välille löytyi yllättävän paljon sijaintiin perustuvia yhteyksiä. Näiden havaittiin jakautuvan kahteen osaan: lähekkäinen sijainti (mm. solmio ja solmioneula) ja sisäkkäinen sijainti (mm. huone ja televisio). Lisäksi havaittiin, että meronymian avulla oli melko helposti löydettävissä yhteys monien käsitteiden välille. Näin oli mm. asunnon, auton ja muiden suurempien kokonaisuuksien kohdalla. Synonyymejä taas ei löydetty merkittäviä määriä, mistä johtuen niiden hyödyntäminen jätettiin tulevaisuuteen. Kaikki muut enemmän tai vähemmän epämääräiset mutta hyödylliset yhteydet käsitteiden välillä kuvattiin tarkentamattomalla assosiaatioilla, ja ONKI-palvelimella olevan ontologian sisältö päivitettiin /103/. Sumeuden suhteen painokertoimia annotoinnissa ja suhteiden määrittelyssä ei siis päädytty käyttämään (kts. kappale 3.3), vaan käsitteiden liittyminen toisiinsa kuvattiin hierarkiaan asettuvien diskreettien suhteiden avulla. Protégé-ohjelmassa tämä suhteiden hierarkia kuvattiin sisennettyä listaa käyttäen kuvassa 27 näkyvällä tavalla.



**Kuva 27.** *Suhdehierarkia Protégé-työkalussa*

Esineiden ja tarpeiden peilauksesta kuluttajan kontekstien suhteen olisi monissa tapauksissa merkittävää hyötyä tiedonhaun kannalta. Etenkin ihmiskeho voisi olla intuitiivinen metafora selailukäytössä mm. seuraavissa tapauksissa: 1) päästä särkee → osoita kuvassa päästä ja 2) jalka kipeä → osoita kuvassa jalkaa. Selailu voisi siis olla käyttäjälle erityisen vaivatonta kuvakäyttöliittymien avulla, joiden rakentaminen tosin olisi hitaampaa. Toisaalta keskeisimpien kontekstien/metaforien /67, s. 206-208/ piirtäminen tuskin veisi kovin paljon aikaa.

Päänsäryn ja kipeän jalan tapauksissa metafora toimisi kontekstina käsitteiden välisille poluille. Ongelmallisiin käsitteisiin voitaisiin siis siirtyä helposti ja varmasti löydettävien käsitteiden kautta käyttämättä turhaa aikaa yksittäisen käsitteen kirjoitusasun muotoiluun. Hyötyä kontekstien/metaforien käytöstä olisi lisäksi tarkastellessa ratkaisuja peräkkäisiin tarpeisiin ja yksittäisen tarpeen tyydyttämistä vaihtoehtoisilla tavoilla. Esimerkkejä näistä on listattuna taulukossa 4.

**Taulukko 4.** *Esimerkkejä kontekstien hyödyntämisestä tiedonhaussa*

<b>Tarve</b>	<b>Mikä ontologiassa mahdollistaa helpon siirtymisen käsitteiden välillä?</b>
auton ja erikoisvanteiden ostaminen	autot (on osa)→ renkaat (on osa)→ erikoisvanteet
vapaa-ajan vietto lomakohteessa	vapaa-aikaan liittyy: lento, hotelli, teatteri ja museo
asunnon sisäseinien ulkonäön parantaminen	sisäseinän sisällä sijaitsee: tapetti, taulu, juliste, ikoni ja ryijy

Kontekstien hyödyllisyydestä huolimatta monissa tapauksissa haun muotoilu on kuitenkin niin yksinkertaista, että käyttäjä tuskin haluaa ilmaista hakuhehtoaan sisäänpäinkääntyneen motiivin avulla. Esimerkiksi tietokonetta hankkivalle kuluttajalle tuloksellisen hakutermien keksiminen lienee yksinkertaista.

## 9 KEHITYSPROJEKTtien PROTOTYYPIt

### 9.1 Semanttinen web käytännössä

#### 9.1.1 WSMO

Palveluiden kuvailua kokeiltiin WSMO:n (kts. kappale 4.4) avulla, koska kyseinen kieli vaikutti erittäin ilmaisuvoimaiselta ja kielen avulla näytti olevan mahdollista tuottaa tietoa hajautetusti. Kokeilussa hyödynnettiin WSMT-työkalua /6/ palveluiden kuvailussa ja käyttöönotossa. Ohjeistus WSMO:n hyödyntämiseen oli kuitenkin melko puutteellinen. Näin ollen palveluiden kuvailu ja julkaisu WSMO:ta käyttäen oli erittäin vaikeaa. Myöskään palveluiden kuvauksiin eksyneiden virheiden paikallistaminen ei onnistunut ongelmitta.

Erittäin suuren haasteen WSMO:n hyödyntämiseen aiheutti myös se, että aiheeseen liittyvissä ohjeissa mainitut esimerkkitoteutukset näyttivät kadonneen Internetistä. Toiminnassa olevia web-palveluita onnistuttiin löytämään vain yksi (osoitteesta <http://sws-challenge.org/services/CRMSERVICE?WSDL>), mihin onnistuttiin ottamaan yhteyttä itse tehdystä toteutuksesta. Sen sijaan ulkoiseen itsetehtyyn ja toimivaksi testattuun web-palveluun ei onnistuttu luomaan yhteyttä WSMX:n kautta. Ongelmia ilmeni myös muissa WSMX:n toiminnallisuuksissa, mistä johtuen tehdyt toteutukset eivät toimineet oletetulla tavalla. WSMO ei siis näyttänyt kovin mielekkäältä vaihtoehdolta ratkaista tämän tutkielman tutkimuskysymyksiä.

#### 9.1.2 Exhibit-julkaisukehikon rajoitteet

Kappaleessa 8.1 esitellyn ja Exhibit-julkaisukehikon avulla toteutetun sovelluksen tietovarastoa ajateltiin olevan tarvetta uudistaa tulevaisuudessa, mistä johtuen sovelluksen kehittämisen jälkeen sovelluksen skaalautuvuutta testattiin suuremmille aineistoille. Testaaminen suoritettiin tietokoneella, jossa oli Ubuntu 8.04 käyttöjärjestelmänä, kahden

GB:n käyttömuisti ja Intel Pentium 4 -suoritin. Testitilanteessa tarkkailtiin Internet-selaimen (Firefox) muistin kulutusta ja sovelluksen yleistä toimintaa. Testausta suoritettaessa sovelluksen käyttämän aineiston koko oli 560 palvelua, milloin sovelluksen toiminnassa ei havaittu ongelmia ja Internet-selaimen tarvitsema muisti sovellusta käytettäessä oli 40 MB. Tämän arvioitiin olevan suhteellisen vaatimaton kulutus, sillä lukema asettui melko tarkkaan Google-hakukoneen (24 MB) ja HS.fi-palvelun (65 MB) lukemien puoliväliin. Sen sijaan viisinkertaisella aineistolla joidenkin valintojen purkaminen sovelluksessa kesti erittäin pitkään ja Internet-selaimen tarvitseman muistin määrä yli kaksinkertaistui 87 MB:hen. Tästä voitiin päätellä, että kehitetty sovellus ei skaalaudu suurille aineistoille kovin hyvin. Toisaalta Yrityksen palveluhakemiston kohdalla aineiston merkittävä kasvaminen ei ollut näköpiirissä. Lisäksi skaalautuvuudesta johtuvaa ongelmaa näytti olevan mahdollista siirtää Yrityksen palveluhakemiston kohdalla tulevaisuuteen. Näin voitaisiin tehdä siten, että muodostettaisiin kuntakohtaisia portaaleita, joiden tietovarastot muodostettaisiin automaattisesti eräajona säännöllisin väliajoin suuremmasta tietovarastosta, mikä johtaisi sovelluksen kannalta pienempiin tietovarastoihin. Muodostuvien tietovarastojen luonnissa mukaan otettaisiin kaikki kyseisen kunnan asukkaan kannalta sopivat palvelut riippumatta siitä olisivatko ne kunnan alueen sisä- tai ulkopuolella tarjottavia.

### 9.1.3 Kuluttajapalveluiden kuvailu ja HAKO

Kuluttajille sopivien palveluiden määrä näytti olevan varsinkin suurimpien kaupunkien kohdalla niin suuri, että Exhibit-julkaisukehikon avulla toteutettua sovellusta tuskin voitaisiin tarjota kuluttajien käytettäväksi. Tätä kohderyhmää varten kokeiltiin SeCo:ssa kehitetyn HAKO-hakumoottorin käyttöönottoa, sillä sen käytössä ei ollut tiedossa vastaavia skaalautuvuuteen liittyviä ongelmia kuin Exhibit-julkaisukehikossa.

Toimiakseen HAKO vaati kaksi RDF-XML-muotoista tiedostoa. Toisessa tiedostossa oli data ja toisessa arvojen hierarkia, johon datatiedostosta viitattiin. Arvojen hierarkian sisältävä tiedosto luotiin Protégé-työkalulla (kts. kappale 5.1.2) ja itse data annotoitiin SAHA-työkalua (kts. kappale 5.1.3) käyttäen. Käyttöön SAHA-työkalu saatiin lomakkeiston ja Protégé-työkalulla luodun tiedoston avulla, joista lomakkeisto luotiin kirjoittamalla tekstimuotoinen tiedosto. Tiedoston sisältö kirjoitettiin RDF-XML:ää luettavammalla Turtle-syntaksilla /9/ ja tiedosto muunnettiin SAHA-yhteensopivaan OWL-muotoon Jena /5/ hyödyntävän Java-kielisen muunnosohjelman avulla (liitteessä 6). Lomakkeistossa annotoitavien palveluiden ymmärrettiin kuvaavan tuotetta, tuoteryhmää tai immateriaalista kokonaisuutta, joita voisi liittyä yhteen palveluntarjoajaan useita. Palveluiden skeema määriteltiin taulukon 5 mukaisesti.

**Taulukko 5.** *Kuluttajapalveluiden skeema*

Kentän nimi	Viittaus ulkoiseen merkitykseen	Pakollinen
nimi	foaf:name /12/	kyllä
tuotemerkki	-	ei
tarvenimike	-	ei
materia (koskee)	-	ei
materia (liittyy)	-	ei
toiminta	-	ei
paikka	-	ei
aika	-	ei
immateria/ilmiö	-	ei

Taulukosta 5 havaitaan, että ominaisuuksista vain nimelle on onnistuttu kiinnittämään yleisesti ymmärretty merkitys. Muut skeeman kentät ovat hakusovelluksen toimintaan liittyviä ja ne viittaavat kaikki suljettuihin arvojoukkoihin. Näistä mitään kenttää ei ole määritelty pakolliseksi, koska on suoraan vaikeaa tietää mihin kenttiin haun kannalta mielekäs tieto palveluista sijoittuu.

Suljetuista arvoviittauskentistä tuotemerkki on mukana skeemassa siksi, koska tuotemerkkien käyttö haun rajauksessa olisi todennäköisesti erittäin mielekäs haettaessa monimutkaisten laitteiden ja koneiden (mm. autot, veneet ja pesukoneet) myynti-, korjaus- ja huoltopalveluita. Kenttä viittaa yksinkertaiseen tuotemerkkien listaan, jonka muodostuksessa voitaisiin käyttää hyväksi tuotemerkkirekistereitä. Esimerkkinä kentän tarkoittamasta tuotemerkistä mainittakoon Siemens. Muut arvoviittauskentät viittaavat kehitettyyn kuluttajaontologiaan (kts. kappale 8.2.3).

#### 9.1.4 Datan tuonti Exhibit-julkaisukehikosta HAKO:ssa julkaistavaksi

SAHA:n (kts. kappale 5.1.3) käytön lisäksi etsittiin joustavaa tapaa julkaista Exhibit-sovelluskehikkoon tuotu data HAKO:ssa. Tämä paljastui yksinkertaiseksi ja nopeaksi operaatioksi, sillä Exhibit-sovelluskehikko tarjosi mahdollisuuden tuoda aineisto ulos sovelluksesta RDF-XML-formaatissa. Tässä muodossa data ei kuitenkaan sellaisenaan sopinut HAKO:ssa julkaistavaksi, mistä johtuen datan formaatti muunnettiin SAHA:n tuottamaa RDF-XML-formaattia vastaavaksi XSL-muunnoksella. Käytännössä tämä tehtiin lisäämällä tuotuun XML-tiedostoon xml-stylesheet-määrittys, joka osoitti itse tehtyyn XSL-tiedostoon (liitteessä 7). Tämän jälkeen annettiin Internet-selaimen muuntaa tuotu XML-tiedosto eri muotoon ja kopioitiin selaimen ruudulta teksti tiedostoon. XSL-tiedoston sisältöä katsottaessa on havaittavissa, että muunnos on melko kevyt ja helposti muokattavissa. Näin ollen Exhibit-julkaisukehikon avulla olisi voitu helposti hahmottaa myös Aarrepalvelun aineiston osajoukkojen ominaisuuksia tilaa, jos jotain aineistoa olisi ollut saatavilla.

### 9.1.5 HAKO toiminnassa

SeCo:ssa kehitetty HAKO-hakumoottorin toteutus koostui joukosta määrittämiä ja Java-kielellä tehtyjä luokkia, joiden avulla RDF-muotoista tietoa (kts. kappale 4.1) voitiin tarkastella Internet-selaimella. HAKO:n alkuperäinen toteutus oli suunniteltu toisentyypille tiedolle, mistä johtuen sitä jouduttiin muokkaamaan melko paljon. Lisäksi sovelluksen toiminnan ymmärtäminen edellytti sovelluksen muokkaamista, sillä alkuperäisen HAKO:n rakenne ei ollut täysin intuitiivinen. Välttävasti toimiva versio sovelluksesta onnistuttiin kuitenkin luomaan. Sovelluksessa palveluita oli mahdollista hakea vapaan tekstin, valmiiden valintavaihtoehtojen tai näiden yhdistelmän avulla. Internet-selaimen näkymässä kenttien arvojakaumat sijoitettiin vasempaan reunaan ja tulostaus keskelle (kuva 28).

**Palveluhaku**

**Tulokset (561)**

- o [-] Toiminta-alue 561
  - Alueellisesti rajoittunut 273
  - Valtakunnallinen 288
- o [-] Ratkaisu liittyy 390
  - Autot ja ajoneuvot 11
  - Epäselvä 81
  - [+] Kansainvälinen liiketoiminta 24
  - [+] Koulutus 75
  - Liiketoiminnan aloittaminen 6
  - Maatalous 18
  - Naiset yrittäjinä 10
  - Sukupolvenvaihdos 1
  - Terveys 24
  - Toimitilat/Hautomo 21
  - Verotus 7
  - Yritysysteistyö 142
- o [-] Ratkaisutyyppi 561
  - [-] Finanssituote 43
    - [+] Raha 27
    - [+] Takaus ja takuu 16
  - Kehittämisohjelma 15
  - [+] Organisaatio 496
  - Sähköinen palvelu 7

Results list:

- "AEL"
- "AKTIVA - Etelä-Karjalan aikuisopisto"
- "Adulta Oy"
- "Agropolis Oy"
- "Aikuiskoulutus- ja oppisopimuskeskus AIKUKAM"
- "Aisapari - Härmänmaan ja Järviselän Kehittämisyhdistys ry"
- "Ajoneuvohallintokeskus AKE"
- "Aktiivinen Pohjois-Satakunta ry"
- "Alavieskan kehityskeskus Oy"
- "Alustakausi (Finnvera)"
- "Amiedu"
- "Ammattiopisto Lappia, Aikuiskoulutuspalvelut"
- "Apteekkien Työnantajaliitto"
- "Apteekkitavaratukkukauppiat ry (ATY)"
- "Asiantuntijaosuuskunta Bona Parte"
- "Aurinkosuojaus ry"
- "Autoalan Keskusliitto ry AKL"
- "Autoalan ja Korjaamoiden Liitto ry"
- "Autoliikenteen Työnantajaliitto"

**Kuva 28.** Exhibit-julkaisukehikosta tuotu data HAKO:ssa

Käytännössä yksinkertaistettu HAKO toimi siten, että ensin IndexOntology-luokassa luettiin kertaluonteisena operaationa Jena-malliin /5/ RDF-XML-muotoiset syötetiedostot, joista toinen sisälsi datan ja toinen annotaatioarvojen hierarkian. Tämän lisäksi IndexOntology-luokassa tehtiin joukko konfiguraatiomäärittämiä, jotka välitettiin Jena-mallin kanssa InstanceIndex-luokkaan. InstanceIndex-luokassa Jena-mallin tieto kirjoitettiin määrittysten mukaan tiedostojärjestelmään käyttäen hyväksi CategoryMapper-apuluokkaa, Cache-välimuistiluokkaa sekä Apachen Lucene- ja Commons-kirjastoja /126, 127/.

HAKO:n Internet-selaimessa näytettävä HTML-koodi rakennettiin FreeMarkerin /36/ avulla tekstitiedostossa, johon välitettiin tietoa muuttujiin talletettuina. Muuttujien sisältö taas

muodostettiin InstanceIndexController-luokassa, jonka yhteys tekstitiedostoon määriteltiin Spring-kehikon /113/ avulla. InstanceIndexController-luokassa tiedot käyttäjän valinnoista luettiin sivukutsujen välillä kulkevista parametreista, joita vastaavat tulokset selvitettiin Lucene-kirjastoa käyttäen. Käytännössä kyselyt tehtiin Lucene-kirjaston suuntaan käyttäen erityisiä kyselyobjekteja, joiden muodostus sijoitettiin staattiseksi metodiksi Utils-nimiseen luokkaan. Näin siksi, koska metodia käytettiin hyväksi myös tuloksien jakauman selvittämiseen annotaatioarvojen suhteen, mikä toiminnallisuus taas sijoitettiin InstanceIndexController-luokan käyttämään CategoryMapper-apuluokkaan. Muita kyselyitä tietovarastoon taas ei taas toteutuksessa tarvittu, eli kappaleessa 4 esitellyt semanttisen webin kyselykielet jäivät käyttämättömiksi.

Verrattaessa yksinkertaistetun HAKO:n toimintaa Exhibit-julkaisukehikkoon havaitaan helposti, että Exhibit-julkaisukehikolla oli huomattavasti helpompi kehittää tutkielman tarpeisiin sopiva moninäkömähakuun perustuva sovellus, jos sivuutetaan Exhibit-julkaisukehikon asettamat rajoitteet käytetyn aineiston koolle. Lisäksi kuvan 24 perusteella havaitaan, että yksinkertaistettuun HAKO:on ei ole toteutettu monia Exhibit-julkaisukehikon tarjoamia keskeisiä toiminnallisuuksia, joita ovat mm. 1) näkymien vapaa sijoittelu käyttöliittymässä, 2) näkymien valintojen järjestyksen määrittäminen, 3) monivalinnat näkymisessä, 4) palveluiden ryhmittely palveluntarjoajien mukaan ja 5) palveluiden tietosisällön esittäminen vapaamuotoisena koosteena. Näiden toiminnallisuuksien toteuttaminen HAKO:on olisi käytännössä ollut melko työlästä, mutta HAKO:n yksinkertaistetun rakenteen vuoksi toteutuksessa oltaisiin kuitenkin voitu edetä melko suoraviivaisesti.

Muokkaustarve HAKO:n kehittämisessä olisi käytännössä kohdistunut aiemmin mainittuihin toiminnallisuutta sisältäviin luokkiin, joissa toiminnallisuuksien muodostaman kokonaisuuden kompleksisuus oli kohtuullisella tasolla. Tämä saatiin aikaan tutkimalla luokkien metodikutsujen käyttöä Eclipse-kehittimen /24/ hakutoiminnolla, minkä pohjalta metodien näkyvyys luokan ulkopuolelle minimoitiin ja moni luokkien metodeista muutettiin staattisiksi toiminnallisuuksien vaikutusten eristämiseksi. Myös toiminnallisuuksien vaatima koodirivien määrä onnistuttiin laskemaan kohtuulliselle tasolle (taulukko 6). Verrattaessa yksinkertaistettua HAKO:a kuitenkin Yrityksen palveluhakemiston Exhibit-toteutukseen havaitaan kompleksisuuden suhteen merkittävä ero. Exhibit-toteutuksen toiminnallisuus kun onnistettiin kuvaamaan ainoastaan vajaalla 70:llä itsekirjoitetulla koodirivillä.



**Taulukko 6.** Yksinkertaistetun HAKO:n keskeisimmät osat ja niiden koot

Kokonaisuus	Koodirivejä
IndexOntology-luokka	50
InstanceIndex-luokka	310
CategoryMapper-luokka	120
Cache-luokka	120
search.ftl (HTML-koodin rakennus)	140
IndexInstanceController-luokka	150
UICategoryBuilder-luokka	130
Utils-luokka	50

## 9.2 Prototyyppien käyttöliittymät

### 9.2.1 Yleistä

Hahmoteltaessa parannuksia Aarre-palvelun ja Yrityksen palveluhakemiston käyttöliittymiin havaittiin, että palveluista puuttui monia mielekkäitä toimintoja, jotka ovat käytössä muissa palveluhakusovelluksista (mm. Superpages.com-sivusto /64/, Fonectan keltaiset sivut /32/ ja eat.fi-sivusto /37/). Yrityksen palveluhakemistolle ei löydy varsinaista kilpailevaa sivustoa, mutta Aarre-palvelun kohdalla taas näyttää siltä, että monien kilpailevien sovellusten (mm. Fonectan keltaiset sivut) käytettävyys on monessa suhteessa Aarre-palvelua parempi. Toisaalta kilpailevien palveluhakusovellusten käytettävyyttä ei voida pitää mitenkään täydellisenä, sillä niistäkin puuttuu monia käyttäjän haku helpottavia ominaisuuksia. Näin ollen lähtökohdaksi Aarre-palvelun ja Yrityksen palveluhakemiston käyttöliittymien kehittämiseen otettiin mielekkäiden toiminnallisuuksien yhdistäminen sovelluksiin.

### 9.2.2 Tekstihaku yhdistettynä moninäkömähakuun

Yritettäessä tyydyttää samanaikaisesti mahdollisimman monia selailu- ja suoraikutarpeita moninäkömähakuun yhdistetty tekstihakutoteutus vaikutti mielekkäimmältä lähestymistavalta. Yrityksen palveluhakemiston prototyyppiin (kuvassa 24) tämä saatiin toteutettua melko helposti siten, että sovelluksessa haku tekstillä tehdään automaattisesti pienellä viiveellä käyttäjän kirjoituksesta. Näin ollen hakusanaa ei monissa tapauksissa tarvitse kirjoittaa kokonaisuina, mikä helpottaa ja nopeuttaa tiedonhakuja. Sovelluksessa tekstihaku määritettiin kohdistumaan tuloksien nimien lisäksi myös niiden annotaatioihin. Vapaan tekstihaun lisäksi tekstihakukentän kohdalla kokeiltiin onnistuneesti myös ONKI-valitsimen (kts. kappale 5.1.1) käyttöönottoa. Vaikka kokeilu teknisesti onnistuikin, ONKI-valitsimen käytön hyödyt eivät näyttäneet kovin selviltä tässä yhteydessä johtuen komponentin seuraavista puutteista: 1) automaattinen täydennys kohdistuu aina koko sanastoon riippumatta tulosjoukosta ja 2) hakutermiehdotuksien semanttisia suhteita pitkin ei voi liikkua käyttämättä erillistä selainikkunaa.

### 9.2.3 Valmiiden valintojen visualisointi

Valmiiden valintojen määrä ja hakujoukon koko olivat Exhibit-julkaisukehikon avulla toteutetussa sovelluksessa (kuvassa 24) niin pieniä, ettei erillistä yleistä /67, s. 66-68/ eikä tuloksiin liittyvää asiasanalistausta (kts. kappale 6.3.1) ollut mielekästä toteuttaa. Nämä osiot kun olisivat vieneet käyttöliittymästä vain turhaa tilaa. Toisaalta Aarre-palvelussa tilanne aineiston koon ja asiasanojen määrän suhteen olisi ollut aivan toinen, joten siihen asiasanojen yleinen ja tuloksiin liittyvä listaus olisi todennäköisesti ollut hyödyllistä toteuttaa.

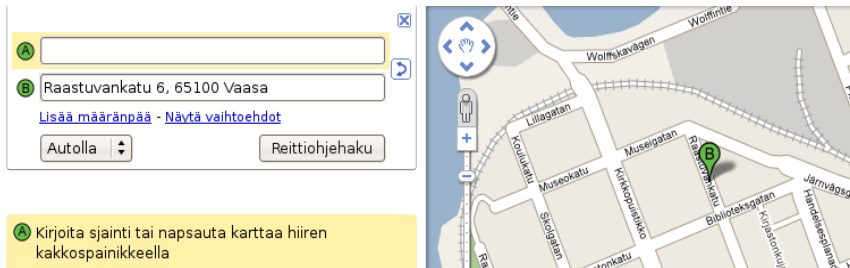
Hicksin lain mukaan valintaprosessiin käytetty aika on logaritmisesti verrannollinen valintavaihtoehtojen määrään /75/. Näin ollen sovelluksien käyttöliittymissä pyrittiin visualisoimaan vain kaikki oleellimmat valinnat aineiston sisällön ja käyttäjän tarpeiden kannalta. Lopputuloksena kuvassa 24 esitetyssä käyttöliittymässä näkyikin valintoja enimmillään yhteensä vain hieman yli 30 kpl, joka tuskin on liikaa. Lisäksi valintojen tekemistä on käyttöliittymässä helpotettu erottamalla valinnat toisistaan tyhjällä tilalla sekä hyödyntämällä valintojen suodatusinformaatiota seuraavasti: 1) epäoleellisten valintojen piilottaminen valintaprosessin edetessä ja 2) valintojen suodatustehon visualisointi. Myös valintojen suorittaminen on tehty mahdollisimman helpoksi määrittämällä teksti valinta-alueeksi. Tämä on parempi ratkaisu kuin Aarre-palvelun pienet valintaruudut, joiden avulla käyttäjän tulee tehdä valintansa kaikissa tapauksissa. Näin siksi, koska osoittaminen hiirellä pienemmälle alueelle vaatii käyttäjältä enemmän keskittymistä ja siten vie myös enemmän aikaa /67, s. 275-276/.

Valintojen kokonaisuuden lisäksi myös näkymäkohtaisesti valintojen määrä on ilmeisesti melko hyvä (kts. kappale 6.3.2) Exhibit-julkaisukehikon avulla toteutetussa sovelluksessa. Aihealueista on ehkä hieman pitkä, mutta käyttäjän katse tuskin harhailee sovellusta käytettäessä kovin paljon, koska näkymissä valinnat on ryhmitelty mielekkäiksi kokonaisuusiksi (mm. yritystoiminnan elinkaari, kansainvälisyys, toimialat ja rahan saanti). Lisäksi valintojen ryhmittely myös nopeuttanee valintojen tekemistä, koska lähekkäisten valintojen välillä käyttäjän käden liike on lyhyt /67, s. 275-276/. Ryhmittely on toteutettu sovelluksessa listana, koska valintojen määrä ja hierarkian esittämisen tarvetta ei käytännössä ollut. Sen sijaan Aarre-palvelun käyttöliittymässä valmiit valintavaihtoehdot kannattaisi esittää esimerkiksi sisennettynä listana avaamatta koko hierarkiapuuta sanaston suuren koon vuoksi.

Kehitetyssä kuluttajaontologiassa on ensisijaisen esittämishierarkian lisäksi myös muita suhteita sanojen välillä, mitä tietoa käyttäjä todennäköisesti haluaisi myös hyödyntää. Käyttäjän kannalta tämä voitaisiin esittää IRMA:a (kuvassa 23) vastaavalla tavalla siten, että valittavan käsitteen ympäristö ontologiassa visualisoitaisiin valittavan käsitteen viereen käyttäjän viedessä hiiren valintavaihtoehdon päälle. Visualisoinnissa jokaisen ontologiassa kuvatun suhteen mukaan voitaisiin piirtää päätasolle oma alipuunsa. Lisäksi myös käsitteiden valintaprosessin tilaa vastaava suodatusteho kannattaisi huomioida valintavaihtoehtojen esittämisessä käyttöliittymässä seuraavasti: 1) tyhjiä tulosjoukkoja palauttavien valintavaihtoehtojen visualisointi muista valintavaihtoehdoista poikkeavalla tavalla ja 2) ei-tyhjiä tulosjoukkoja palauttavien käsitteiden suodatustehon esittäminen numerona.

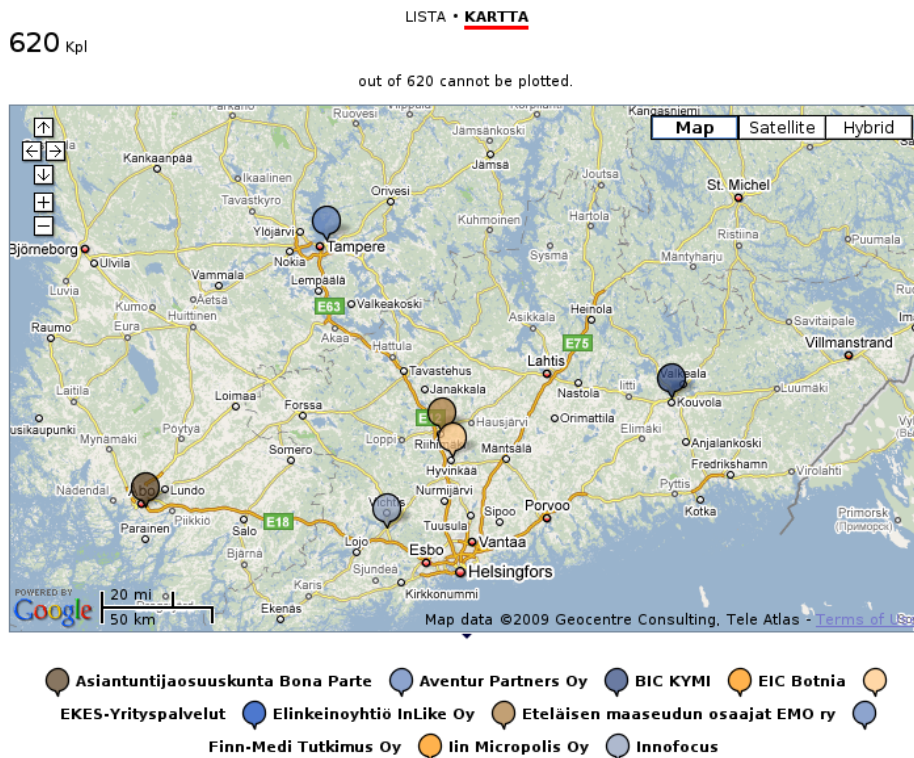
## 9.2.4 Palveluiden sijaintitiedon hyödyntäminen

Monia palveluita käytetään niiden käyttäjien kannalta enemmän tai vähemmän vierassa paikoissa, joten käyttäjät ei välttämättä osaa suunnistaa palvelupaikalle pelkän tekstimuotoisen osoitetiedon avulla. Tilanteen helpottamiseksi Exhibit-julkaisukehikon avulla toteutettuun sovellukseen integroitiin GoogleMaps-reittihaku linkin avulla, joka johdatti määränpään osalta esitetyille reittihaku-sivulle (kuvassa 29). Käytännössä tämä toteutettiin siten, että linkin arvoksi annettiin seuraava: [http://maps.google.fi/maps?daddr=] [URL-koodattu arvo seuraavasta: <katuosoite>, <postinumero> <postitoimipaikka>].



Kuva 29. GoogleMaps-reittihaku /42/

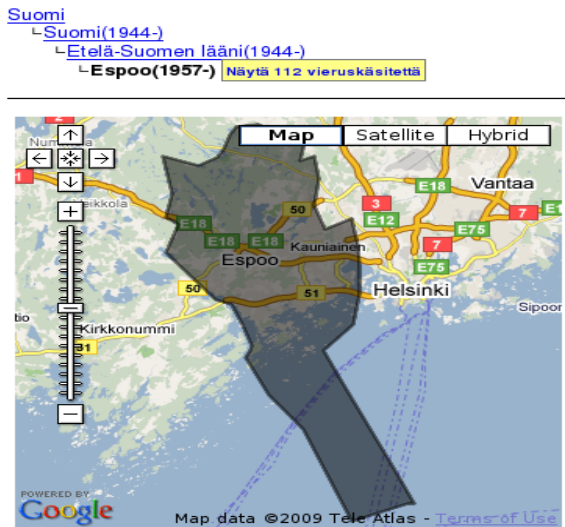
GoogleMaps-reittihaun lisäksi Exhibit-julkaisukehikon avulla toteutettuun sovellukseen integroitiin myös GoogleMaps-karttakomponentti (kuvassa 30), jotta useamman kuin yhden palvelun sijainti voitaisiin hahmottaa yhden karttakuvan avulla. Komponentissa palvelupaikat esitettiin koordinaattien avulla, jotka saatiin palvelupaikkojen osoitteista Googlen tarjoamien Javascript-funktioiden avulla /42/. Funktioiden käyttöä on havainnollistettu liitteessä 8 esitellyssä HTML-sivussa, joka tulostaa koordinaatit palvelupaikan nimen yhteyteen näytöllä.



Kuva 30. GoogleMaps integroituna hakusovellukseen



vaikuttanut todelliselta yrittäjille suunnattuja palveluita tarkastellessa. Aarre-palvelun kohdalla tilanne voisi kuitenkin olla toinen ainakin joissain käyttötapauksissa.



**Kuva 32.** SAPO:n ONKI-käyttöliittymä/123/

Exhibit-julkaisukehikon avulla toteutetussa sovelluksessa maantieteellistä tietoa hyödynnettiin myös tulosjoukon esittämisjärjestyksessä ryhmittelemällä palvelut palveluntarjoajan mukaan. Yrittäjille suunnattujen palveluiden palveluntarjoajat taas listattiin vain aakkosjärjestyksessä, koska niiden määrä tulosjoukoissa jäi melko pieneksi. Sen sijaan Aarre-palvelun tulosjoukon käyttäjä todennäköisesti haluaisi järjestää myös relevanssin mukaan. Käytännössä tämä ei onnistuisi pelkän Boolean hakumallin /33/ avulla, joten tulosjoukon järjestämisessä voisi olla mielekästä hyödyntää jotain tilastollista mallia /80/. Toisaalta tulosjoukko voitaisiin järjestää suoraan myös etäisyyden mukaan. Tätä varten tulisi tietää käyttäjän sijainti, joka voitaisiin kysyä käyttäjältä esimerkiksi karttakuvan avulla ONKI-GEO-sovelluksen (kuvassa 31) tapaan. Toisaalta käyttäjä voisi ilmaista sijaintinsa myös tekstillä vain kunnan tarkkuudella tai tämä tieto voitaisiin päätellä käyttäjän IP-osoitteesta, mikä tosin ei aina onnistu luotettavasti. Tällöin etäisyysmittauksen vertailupisteenä voitaisiin käyttää käyttäjän sijaintikunnan edustavaa pistettä.

### 9.2.5 Tekstit ja tyylit

Exhibit-julkaisukehikon avulla toteutettua palveluhakusovellusta (kts. kappale 8.1) testattiin vapaamuotoisesti muutamalla sovelluksen kohderyhmään kuuluvalla henkilöllä, minkä pohjalta sovellukseen tehtiin monia ulkoasuun kohdistuvia muutoksia. Muutoksien jälkeen kehitetyn sovelluksen käytettävyyttä todennäköisesti nousi, mutta tätä ei onnistuttu kattavasti varmistamaan johtuen käytettävissä olleiden koehenkilöiden pienestä määrästä.

Kuvassa 24 näkyvä käyttöliittymä edustaa kehityksen lopputulosta, johon nähden alkuperäistä käyttöliittymää muokattiin mm. tuloslistauksen osalta. Alkuperäisessä versiossa tuloksia ei listattu erilliseen kehykseen, mikä vaikeutti käyttäjien mielestä valintojen tekemistä tulosten selailun jälkeen. Tuloslistauksen sijoittaminen erilliseen kehykseen ilmeisesti poisti tämän ongelman, mutta muutoksen seurauksena vierityspalkin käyttö todennäköisesti hieman vaikeutui. Tämä ei ole tosin välttämättä minkäänlainen ongelma

kaikille käyttäjille, koska tuloksien selailu onnistuu myös useista hiiristä löytyvän vieritysrullan avulla periaatteessa melko helposti ja nopeasti. Myös tuloksien sivutusta harkittiin alkuperäisen ongelman ratkaisuna, mutta siihen ei päädytty lähinnä kahdesta syystä: 1) Firefox-selaimen etsi-toiminnon käyttö olisi vaikeutunut ja 2) hakujoukon portaaton selailu vaikutti toimivalta keinolta hakujoukon yleiskuvan hahmottamiseen.

Alkuperäisessä versiossa valintojen purku ei ollut käyttäjien mielestä riittävän yksinkertaista, mikä ongelma korjattiin linkittämällä kuvassa 24 näkyvä YritysSuomi-kuva sovelluksen etusivulle. Suhteessa alkuperäiseen versioon muutettiin myös fonttien kokoja ja värejä. Exhibit-julkaisukehikon oletusasetuksilla palveluntarjoajien nimet kirjoitettiin selvästi suuremmalla fontilla, minkä käyttäjät kokivat häiritsevänä. Lisäksi käyttäjiä häiritsevä alkuperäisessä versiossa valintojen edessä näkyvissä numeroissa käytetty harmaa väri. Ongelma korjattiin ylikirjoittamalla Exhibit-julkaisukehikossa määritelty tyyliluokka, joka sääteli kohteiden ulkoasua.

Vapaamuotoisen kokeilun lisäksi sovellukselle tehtiin myös testi ryhmien nimeämiseksi muutamalla henkilöllä. Testiä varten alkuperäisten valintojen nimet muutettiin, minkä jälkeen henkilöitä pyydettiin nimeämään valintoja vastaavat ryhmät ja kuvaamaan ryhmien mielekkyyttä (kuvassa 33: Aihe 1-ryhmä).

The screenshot shows the YritysSuomi application interface. At the top left is the YritysSuomi logo. Below it is a search bar with the text "Hae:" and a search button. To the right of the search bar, it says "16 Kpl filtered from 620 originally ()". Below the search bar is a list of filters under the heading "Aihealueet (tarve)". The filter "16 Aihe 1" is selected. To the right of the filters is a list of services under the heading "OP-Pohjola (15)". The services listed are: Business Eurocard (OP-Pohjola), E-lasku (OP-Pohjola), E-laskupalvelu (OP-Pohjola), Eräsiirtopalvelu (OP-Pohjola), Internet-palvelu (OP-Pohjola), Kassanhallintapalvelu (OP-Pohjola), Kultakortti (OP-Pohjola), Laskujen maksupalvelu (OP-Pohjola), Maksupäätepalvelu (OP-Pohjola), Pankkiyhteysohjelma+Kuitailinki (OP-Pohjola), Suoraveloitus (OP-Pohjola), Ulkomaille lähtevien maksujen hallinta (OP-Pohjola), Ulkomailta saapuvien maksujen hallinta (OP-Pohjola), Verkkomaksupainike (OP-Pohjola), and Viitepalvelu (OP-Pohjola). At the bottom left of the services list, it says "(missing) (1)". On the right side of the interface, there are two panels: "Yritykseni on (nyt/lähitulevaisuudessa)..." with 13 items and "Palvelun tyyppi" with 14 items.

**Kuva 33.** Valintojen nimeämisessä käytetty sovellus

Testin lopputuloksena kaikki ryhmät koettiin mielekkäiksi, mutta muutamille ryhmille oli annettu alkuperäisestä selvästi poikkeavia nimiä. Lisäksi poikkeamaryhmien nimeäminen näytti olevan muita ryhmiä selvästi hitaampaa. Erityisesti alkuperäisessä versiossa ollutta byrokratia-aiheen muodostamaa ryhmää oli vaikeaa nimetä, minkä havainnon pohjalta ryhmä jaettiin kahtia: 1) lomakkeet ja asiakirjat ja 2) lainsäädäntö. Muiden poikkeama-

ryhmien kohdalla taas päädyttiin joiltain osin muokkaamaan niiden käyttöliittymässä näkyviä nimiä.

Erityisen vaikeaa käyttäjille oli nimetä jotain muuta -tyyppistä valintavaihtoehtoa, joka koski aihealue- ja sopivuus suhteessa käyttäjään -kenttiä (kuvassa 33: missing this field). Lisäksi tarkasteltaessa käyttäjien tapaa täyttää omia tietojaan (kuvassa 33: Yritykseni on nyt/lähitulevaisuudessa...) vaikutti siltä, että jotain muuta -tyyppisen valintavaihtoehdon merkitystä ei ymmärretty käytännössä ollenkaan. Tämän vuoksi käyttöliittymään lisättiin väliaikaisena ratkaisuna valinnan käyttöön liittyviä ohjeita, jotka asetettiin aktivoitumaan erillisen ohje-tekstin klikkaamisesta ja hiiren osoittimen viemisestä näkymäotsikoiden päälle.

Yrittäjille suunnatun sovelluksen alkuperäisessä versiossa palveluille ei määritelty erillisiä selitetekstejä. Palveluiden nimien ja niihin kiinnitettyihin linkkien ajateltiin riittävän käyttäjälle, mutta tämä oletus osoittautui vääräksi valintojen nimeämisestä testattaessa. Nimet kertoivat joistain palveluista riittävän paljon, mutta joidenkin palveluiden yhteyteen oli täysin välttämätöntä liittää lyhyt selite. Tällainen palvelu oli mm. ViestinVaihto, jonka kohderyhmänä oli sukupolven vaihdosta suunnittelevat yritykset. Tätä koehenkilöt eivät kuitenkaan osanneet aavistaa pelkän nimen perusteella.

Ryhmien nimeämisen lisäksi koehenkilöitä pyydettiin ratkaisemaan muutamia yrittäjähenkisiä tehtäviä sovelluksen avulla. Tehtävänannot näihin olivat seuraavat:

- 1) Olet aloittamassa yritystoimintaa ja olet kuullut yritystoimintaan liittyvistä pakollisista vakuutuksista. Onko sovelluksesta apua näiden hankkimisessa?
- 2) Olet Helsingissä toimiva pienyritys ja sinulla on rahan puute. Tarvitset rahaa kansainvälistymiseen ja tuotekehitykseen. Mitä vaihtoehtoja ongelman ratkaisemiseksi on löydettävissä sovelluksen avulla?
- 3) Olet Helsingissä toimiva matkailuyrittäjä. Löytyykö sovelluksen avulla sinulle mielenkiintoisia palveluita ja toimijoita mm. yritysysteistyötä ajatellen?
- 4) Olet aloittamassa yhteistyön erään Belgiasta kotoisin olevan yrityksen kanssa. Mitä virallisia tietoja yrityksestä on saatavilla sovelluksen avulla?
- 5) Haluat kehittää yrityksesi toimintaa. Erityisesti yrityksesi myynti ei suju odottamallasi tavalla. Löytyykö sovelluksen kautta ratkaisuehdotuksia ongelmaasi?

Tehtävien suoritusta seurattaessa havaittiin, että käyttäjät löysivät ratkaisut ongelmiin melko helposti valintoja klikkailemalla. Vain neljännen tehtävän ratkaiseminen tuotti pieniä vaikeuksia käyttäjille.

Testitilanteessa korostettiin, että käyttäjä voisi käyttää sovelluksen sisällä mitä tahansa keinoja ongelman ratkaisemiseksi. Jostain syystä kuitenkin vaikutti siltä, että käyttäjät välttelivät tekstihaun käyttöä. Lisäksi testitilanteessa oli havaittavissa, että käyttäjät takertuivat tehtävänannoissa esiintyviin sanoihin. Näin ollen mielekkäämpiä tuloksia todennäköisesti saataisiin todellisilla käyttäjillä ja tarpeilla, minkä vuoksi Yrityksen palveluhakemiston ja YritysSuomi-palvelun käyttölokeja suunniteltiin analysoitavan. Käyttölokeja ei kuitenkaan saatu tarkasteltavaksi, joten suunnitelmaa ei päästy toteuttamaan.

## 10 YHTEENVETO

Tutkielmassa tarkasteltiin PKT-säätiön ylläpitämän Yrityksen palveluhakemiston ja Suomen Asiakastiedon ylläpitämän Aarre-palvelun parantamista yleisesti ja semanttisen webin tekniikoiden avulla. Tarkastelun sivutuotteena kehitettiin Yrityksen palveluhakemistolle vaihtoehtoinen käyttöliittymä, jossa pyrittiin korjaamaan alkuperäisen sovelluksen puutteita käyttäen hyödyksi mm. moninäkömähaku-paradigmaa. Aarre-palvelun kohdalla taas varsinaista loppuunvietyä sovellusta ei kehitetty, koska palveluhakemiston parantamista jouduttiin tarkastelemaan ilman aineistoa.

Ongelmia molempien palveluhakemistojen toiminnassa olivat mm. 1) tietotulva valmiita valintoja käytettäessä, 2) tietotulva tuloksissa, 3) selailun vaikeus, 4) palvelupaikan sijainnin selvittämisen hitaus ja 5) tekstihaun huonot tulokset joidenkin hakujen kohdalla. Näihin ehdotetut ratkaisut on koottu taulukkoon 7.

**Taulukko 7.** *Palveluhakemistojen ongelmat ja niiden ratkaisuehdotukset*

Ongelma	Ratkaisu
tietotulva valmiita valintoja käytettäessä	1) valintavaihtojen järjestäminen mielekkäisiin näkyymiin 2) olennaisten valintavaihtoehtojen tarjoaminen ja tuominen esille 3) käyttäjälle sopimattomien ja hämmentävien valintavaihtoehtojen piilotus
tietotulva tuloksissa	1) palveluiden maantieteellisten käyttörajoitusten määrittely ja hyödyntäminen haussa 2) sisällöllisten duplikaattien poisto hakujoukosta Yrityksen palveluhakemiston kohdalla
selailun vaikeus	1) valintavaihtojen järjestäminen mielekkäisiin näkyymiin 2) toisiinsa liittyvien valintavaihtoehtojen sijoittaminen lähekkäin 3) palveluiden annotointi oikein ja kattavasti 4) annotointi käyttäjän kannalta olennaisten valintavaihtoehtojen avulla 5) käyttäjälle sopimattomien ja hämmentävien valintavaihtoehtojen piilotus
palvelupaikan sijainnin selvittämisen hitaus	1) palveluiden sijaintien jakauman näyttäminen kartalla 2) palvelupaikkojen visualisointi piirroksissa palvelukeskittymien kohdalla 3) esitetyt reittihakulinkkien tarjoaminen tulosten yhteydessä
tekstihaun huonot tulokset	1) mielekkäiden tekstikuvausten liittäminen palveluihin 2) franchising-yritysten kohdalla kuluttajan kannalta havainnollisen hakunimen liittäminen palveluiden kuvauksiin Aarre-palvelussa

Semanttisen webin tekniikoista moninäkömähakua näytti olevan mielekäästä ja mahdollista soveltaa molempien tutkittujen palveluhakemistojen kohdalla. Yrityksen palveluhakemis-



ton kohdalla hakujoukon koon ja pirstaleisuuden vuoksi näytti siltä, että varsinaisesta semanttisesta päättelystä tai ontologian suhteiden hyödyntämisestä selailukäytössä ei näytännyt olevan suuremmalti hyötyä. Aarre-palvelun kohdalla taas tilanne arvioitiin täysin päinvastaiseksi johtuen mm. aineiston suuremmasta koosta.

Aarre-palvelun suuresta koosta johtuen sitä varten kehitettiin kuluttajakäyttöön tarkoitettu ontologia olemassaolevien luokitusten ja ontologioiden pohjalta. Ontologiaan kuvattiin erilaisia yhteyksiä käsitteiden välille palveluiden selailukäytön helpottamiseksi. Yhteyksiä käsitteiden välille löytyi runsaasti mm. seuraavien suhteiden kautta: lähekkäinen sijainti, sijainti jonkun sisällä, osasuhde, yläkäsite ja assosiaatio. Suhteiden visualisointi käyttöliittymässä ehdotettiin toteutettavan SeCo:ssa kehitetyn IRMA-käsitevalitsimen avulla tai tapaan.

Aarre-palvelussa käytössä olleen toimialaluokituksen avulla arvioitiin olevan hankalaa ilmaista hakutarvetta. Ongelman korjaamiseksi muodostettiin uusi käsitteistö, jota ehdotettiin toimialaluokituksen korvaajaksi kuluttajakäyttäjän käyttöliittymässä. Aarre-palvelun toimialaluokituksella kuvatut palvelut pitävät sisällään vaihtelevan määrän kuluttajan kannalta oleellista informaatiota, joten muunnos alkuperäisistä palvelukuvauksista tutkielmassa määriteltyyn skeemaan arvioitiin onnistuvan vaihtelevasti. Joiltain osin muunnokset voisivat onnistua jopa täysin oikein. Tämän merkitystä kokonaisuuden kannalta ei kuitenkaan päästy arvioimaan mm. siitä johtuen, että toimialaluokituksen luokituskohtaisia yritysmääriä ei ollut saatavilla. Todennäköistä kuitenkin olisi, että osa muunnetuista palvelukuvauksista täytyisi annotoida uudelleen korkean käytettävyyden saavuttamiseksi selailukäytössä.

Yrityksen palveluhakemiston aineisto ja aineiston kuvailussa käytetyt käsitteet koettiin ongelmallisena. Asian korjaamiseksi palveluille määriteltiin uusi skeema, jonka mukaan palvelut annotoitiin uudelleen hyödyntämättä alkuperäisiä käsiteviittauksia. Annotoinnissa käytettiin näiden sijaan selailukäytön kannalta mielekkäitä ja tuloksellisia tägejä (tag), jotka peilattiin myöhemmin KOKO-ontologiaan ja UNSPSC-luokitukseen uudelleenkäytettävyyden helpottamiseksi. Tägeillä pyrittiin kuvailemaan palvelut selailukäytön kannalta mahdollisimman mielekkäästi ja kattavasti, minkä vuoksi annotointi suoritettiin tägi kerrallaan. Uusien palvelukuvausten valmistuttua ne tuotiin demosovelluksen tietovarastoon, ja demosovelluksen toimintaa testattiin pienellä joukolla käyttäjiä. Testien perusteella näytti ilmeiseltä, että kehitetty demosovellus toimi selailukäytössä paremmin kuin alkuperäinen Yrityksen palveluhakemisto.

## LÄHDELUETTELO

- /1/ **Aitchison, J.** *Words in the Mind: An Introduction to the Mental Lexicon*. Wiley-Blackwell, 2003. 328 s.
- /2/ **Amin, A., Hildebrand, M., van Ossenbruggen, J., Evers, V. & Hardman, L.** *Organizing Suggestions in Autocompletion Interfaces*. 31th European Conference on IR Research on Advances in Information Retrieval, Toulouse 6-9.4.2009. Springer. s. 521-529
- /3/ **Anon.** *Free Worldwide Travel Guides - Wikitravel*. n.d. (online) [viitattu 16.6.2009]. Saatavilla WWW-muodossa: <URL:<http://www.wikitravel.org>>
- /4/ **Anon.** *Introducing JSON*. n.d. (online) [viitattu 17.6.2009]. Saatavilla WWW-muodossa: <URL:<http://www.json.org>>
- /5/ **Anon.** *Jena – A Semantic Web Framework for Java*. n.d. (online) [viitattu 18.6.2009]. Saatavilla WWW-muodossa: <URL:<http://jena.sourceforge.net>>
- /6/ **Anon.** *The Web Service Modeling Toolkit (WSMT)*. n.d. (online) [viitattu 18.6.2009]. Saatavilla WWW-muodossa: <URL:<http://sourceforge.net/projects/wsmt>>
- /7/ **Antoniou, G. & van Harmelen, F.** *A Semantic Web Primer*. The MIT Press, Cambridge 2004. 238 s.
- /8/ **Baeza-Yates, R. & Ribeiro-Neto, B.** *Modern Information Retrieval*. Addison Wesley, 1999. 544 s.
- /9/ **Beckett, D. & Berners-Lee, T.** *Turtle - Terse RDF Triple Language*. W3C, 2008 (online) [viitattu 18.6.2009]. Saatavilla WWW-muodossa: <URL:<http://www.w3.org/TeamSubmission/turtle/>>
- /10/ **Berners-Lee, T.** *Cool URIs don't Change*. 1998 (online) [viitattu 30.5.2008]. Saatavilla WWW-muodossa: <URL:<http://www.w3.org/Provider/Style/URI>>
- /11/ **Billerbeck, B. & Zobel, J.** *Document Expansion versus Query Expansion for Ad-hoc Retrieval*. Tenth Australasian Document Computing Symposium, Sydney 12.12.2005. Sydney 2005. The University of Sydney. s. 34-41
- /12/ **Brickley, D. & Miller, L.** *FOAF Vocabulary Specification 0.91*. 2007 (online) [viitattu 18.6.2009]. Saatavilla WWW-muodossa: <URL:<http://xmlns.com/foaf/spec/>>
- /13/ **Brickley, D. & Guha, R.** *RDF Vocabulary Description Language 1.0: RDF Schema*. W3C, 2004 (online) [viitattu 18.6.2009]. Saatavilla WWW-muodossa: <URL: <http://www.w3.org/TR/rdf-schema/>>
- /14/ **Collins.** *Collins Easy Learning French Words*. Collins, 2006. 256 s.
- /15/ **Collins.** *Collins French Visual Dictionary*. Collins, 2008. 608 s.
- /16/ **Condé Nast Digital.** *epicurious for people who want to eat*. n.d. (online) [viitattu 17.6.2009]. Saatavilla WWW-muodossa: <URL:<http://www.epicurious.com>>

- /17/ **Cycorp Inc.** *Overview of OpenCyc*. n.d. (online) [viitattu 16.6.2009]. Saatavilla WWW-muodossa: <URL:<http://www.cyc.com/cyc/opencyc/>>
- /18/ **Dey, L., Singh, S., Rai, R. & Gupta, S.** *Ontology Aided Query Expansion for Retrieving Relevant Texts*. Third International Atlantic Web Intelligence Conference, Lodz, 6-9.9.2005. Springer. s. 126-132
- /19/ **eBay Inc.** *eBay Express*. n.d. (online) [viitattu 16.6.2009]. Saatavilla WWW-muodossa: <URL:<http://pages.ebay.com/express>>
- /20/ **eCI@ss e.V.** *eCI@ss, the international standard for the classification of products and services*. n.d. (online) [viitattu 17.6.2009]. Saatavilla WWW-muodossa: <URL:<http://www.eclass-online.com>>
- /21/ **DCMI.** *Dublin Core Metadata Element Set, Version 1.1*. DCMI, 2008 (online) [viitattu 18.6.2009]. Saatavilla WWW-muodossa: <URL:<http://dublincore.org/documents/dces/>>
- /22/ **Dennis, S., McArthur, R., & Bruza, P.** *Searching the World Wide Web Made Easy? The cognitive Load Imposed by Query Refinement Mechanisms*. Third Australian Document Computing Conference, Sydney 20-8.1994. Sydney 1998. University of Sydney. s. 65-71
- /23/ **DERI.** *WSMX - Web Service Modelling eXecution environment*. DERI, 2008 (online) [viitattu 18.6.2009]. Saatavilla WWW-muodossa: <URL:<http://www.wsmx.org>>
- /24/ **Eclipse.org Foundation Inc.** *Eclipse.org Home*. n.d. (online) [viitattu 16.6.2009]. Saatavilla WWW-muodossa: <URL:<http://www.eclipse.org>>
- /25/ **Efthimiadis, E.** *Query Expansion*. Annual Review of Information Systems and Technology 31(1996), s. 121-187.
- /26/ **Eklund, P.W., Roberts, N. & Green, S.P.** *OntoRama: Browsing an RDF Ontology using a Hyperbolic-like Browser*. The First International Symposium on CyberWorlds (CW2002), Tokio 6-8.11.2002. IEEE Press. s. 405-411
- /27/ **English, J., Hearst, M., Sinha, R., Swearingen, K. & Yee, K.-P.** *Flexible search and navigation using faceted metadata*. 2002 (online) [viitattu 18.6.2009]. Saatavilla WWW-muodossa: <URL:<http://flamenco.berkeley.edu/pubs.html>>
- /28/ **Evri Inc.** *Search less, understand more - Evri*. n.d. (online) [viitattu 16.6.2009]. Saatavilla WWW-muodossa: <URL:<http://www.evri.com>>
- /29/ **Facebook Inc.** *Welcome to Facebook | Facebook*. n.d. (online) [viitattu 16.6.2009]. Saatavilla WWW-muodossa: <URL:<http://www.facebook.com>>
- /30/ **Feier, C. & Domingue, J.** *D3.1v0.1 WSMO Primer*. DERI, 2005 (online) [viitattu 18.6.2009]. Saatavilla WWW-muodossa: <URL:<http://www.wsmo.org/TR/d3/d3.1/v0.1/>>
- /31/ **Fikes, R., Hayes, P. & Horrocks, I.** *OWL-QL – A Language for Deductive Query Answering on the Semantic Web*. Journal of Web Semantics 2(2004)1, s. 19-29.

- /32/ **Fonecta Oy.** *Yritykset, palvelut ja yhteystiedot – Tuotteet ja palvelut.* n.d. (online) [viitattu 16.6.2009]. Saatavilla WWW-muodossa: <URL:<http://www.02.fi/keltaiset+sivut>>
- /33/ **Fox, E. & Sharan, S.** *A Comparison of Two Methods for Soft Boolean Operator Interpretation in Information Retrieval.* Virginia Polytechnic Institute & State University, Blacksburg 1986 (online) [viitattu 22.6.2009]. Saatavilla WWW-muodossa: <URL: <http://eprints.cs.vt.edu/archive/00000008/01/TR-86-01.pdf>>
- /34/ **Frosterus, M. & Hyvönen, E.** *Bridging the Search Gap between the Web of Pages and Web of Data by Combining Ontological Document Expansion with Text Search.* Teknillinen Korkeakoulu, Espoo 2009 (online) [viitattu 22.6.2009]. Saatavilla WWW-muodossa: <URL:<http://www.seco.tkk.fi/publications/submitted/frosterus-hyvonen-airoICSD09.pdf>>
- /35/ **Gangemi, A., Guarino, N., Masolo, C., Oltramari, A. & Schneider, L.** *Sweetening Ontologies with DOLCE.* 13th International Conference on Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web, Sigüenz 1-4.10.2002. Springer. s. 166-181
- /36/ **Geer, B. & Bayer, M.** *FreeMarker: Java Template Engine Library - Overview.* n.d. (online) [viitattu 17.6.2009]. Saatavilla WWW-muodossa: <URL:<http://freemarker.org>>
- /37/ **Getnäs Oy.** *Ravintolat Helsinki | Eat.fi | Kaikki ravintolat kartalla.* n.d. (online) [viitattu 16.6.2009]. Saatavilla WWW-muodossa: <URL:<http://www.eat.fi>>
- /38/ **Gilchrist, A.** *Thesauri, taxonomies and ontologies – an etymological note.* Journal of Documentation 59(2003)1, s. 7-18.
- /39/ **Gonzalo, J., Verdejo, F., Chugur, I. & Cigarran, J.** *Indexing with WordNet synsets can improve text retrieval.* COLING/ACL '98 Workshop on Usage of WordNet for NLP, Montreal 16.8.1998. s. 38-44
- /40/ **Google Inc.** *Google.* n.d. (online) [viitattu 16.6.2009]. Saatavilla WWW-muodossa: <URL:<http://www.google.com>>
- /41/ **Google Inc.** *Google Maps.* n.d. (online) [viitattu 16.6.2009]. Saatavilla WWW-muodossa: <URL:<http://maps.google.com>>
- /42/ **Google Inc.** *Google Maps API – Google Code.* n.d. (online) [viitattu 16.6.2009]. Saatavilla WWW-muodossa: <URL:<http://code.google.com/apis/maps>>
- /43/ **Google Inc.** *YouTube – Broadcast yourself.* n.d. (online) [viitattu 16.6.2009]. Saatavilla WWW-muodossa: <URL:<http://www.youtube.com>>
- /44/ **Granada Research.** *Using the UNSPSC. 1998* (online) [viitattu 22.6.2009]. Saatavilla WWW-muodossa: <URL: [http://www.unspsc.org/AdminFolder/Documents/UNSPSC\\_White\\_Paper.doc](http://www.unspsc.org/AdminFolder/Documents/UNSPSC_White_Paper.doc)>

- /45/ **Grover, C & Lascarides, A.** *XML-Based Data Preparation for Robust Deep Parsing*. Annual Meeting of The Association for Computational Linguistics, Toulouse 9.-11.7.2001. Morgan Kaufmann Publishers. s. 252-259
- /46/ **Groxis Inc.** *Grokker*. n.d. (online) [viitattu 17.6.2009]. Saatavilla WWW-muodossa: <URL:<http://live.grokker.com/grokker.html>>
- /47/ **Gruber, T. A.** *Translation Approach to Portable Ontology Specifications*. Knowledge Acquisition 5(1993)2, s. 199-220.
- /48/ **GS1 Finland Oy.** *UNSPSC-koodi*. n.d. (online) [viitattu 16.6.2009]. Saatavilla WWW-muodossa: <URL:[http://www.gs1.fi/muuta\\_toimintaamme\\_liittyvaa/UNSPSC-koodi.html](http://www.gs1.fi/muuta_toimintaamme_liittyvaa/UNSPSC-koodi.html)>
- /49/ **Guardian News & Media Limited.** *guardian.co.uk | Search*. n.d. (online) [viitattu 16.6.2009]. Saatavilla WWW-muodossa: <URL:<http://browse.guardian.co.uk/search>>
- /50/ **Guarino, N., Masolo, C. & Vetere, G.** *OntoSeek: Content-Based Access to the Web*. Intelligent Systems and their Applications, IEEE 14(1999)3, s. 70-80.
- /51/ **Hahn, U. & Mani, I.** *The challenges of automatic summarization*. Computer 33(2000)11, s. 29-36.
- /52/ **Hearst, M.A.** *Next Generation Web Search: Setting Our Sites*. IEEE Data Engineering Bulletin 23(2000)3, s. 38-48.
- /53/ **Hearst, M.A, Elliott, A., English, J., Sinha, R., Swearingen, K. & Lee, K.- P.** *Finding the flow in web site search*. Communications of the ACM 45(2002)9, s. 42-49.
- /54/ **Helsingin yliopisto.** *MuseoSuomi*. n.d. (online) [viitattu 17.6.2009]. Saatavilla WWW-muodossa: <URL:<http://www.museosuomi.fi>>
- /55/ **Hepp, M.** *Products and Services Ontologies: A Methodology for Deriving OWL Ontologies from Industrial Categorization Standards*. International Journal on Semantic Web & Information Systems (IJSWIS) 2(2006)1, s. 72-99.
- /56/ **Hepp, M., Leukel, J. & Schmitz, V.** *A quantitative analysis of product categorization standards: content, coverage, and maintenance of eCI@ss, UNSPSC, eOTD, and the RosettaNet Technical Dictionary*. Knowledge and Information Systems 13(2007)1, s. 77-114.
- /57/ **Hersh, W., Buckley, C., Leone, T. J. & Hickam, D.** *Ohsumed: an interactive retrieval evaluation and new large test collection for research*. 17th annual international ACM SIGIR conference on Research and development in information retrieval, Dublin 3-6.7.1994. Springer. s. 192-201
- /58/ **Hovy, E.H.** *Methodologies for the reliable construction of ontological knowledge*. 13th International Conference on Conceptual Structures, Kassel 17-22.7.2005. Springer. s. 91-106

- /159/ **Hyvönen, E.** *Miksi Asiasanastot eivät riitä vaan tarvitaan ontologioita?* TKK Vies-  
tintäteknikka ja Helsingin Yliopisto, tietojenkäsittelytieteen laitos, 10.10.2005  
[viitattu 16.6.2009]. Saatavilla WWW-muodossa: <URL:  
[http://www.seco.tkk.fi/publications/2005/hyvonen-miksi-asiasanastot-eivat-riita-  
2005.pdf](http://www.seco.tkk.fi/publications/2005/hyvonen-miksi-asiasanastot-eivat-riita-2005.pdf)>
- /160/ **Hyvönen, E., Mäkelä, E., Kauppinen, T., Alm, O., Kurki, J., Ruotsalo, T., Seppä-  
lä, K., Takala, J., Puputti, K., Kuittinen, H., Viljanen, K., Tuominen, J., Palonen,  
T., Frosterus, M., Sinkkilä, R., Paakkari, P., Laitio, J. & Nyberg, K.** *Culture-  
Sampo - Finnish Culture on the Semantic Web 2.0. Thematic Perspectives for the  
End-user.* Museums and the Web 2009, Indianapolis 15-18.4.2009.
- /161/ **Hyvönen, E., Seppälä, K., Viljanen, K., Frosterus, M.** *Yleinen suomalainen onto-  
logia YSO – kohti suomalaista semanttista webiä.* Kansalliskirjasto, 2007 (online)  
[viitattu 22.6.2009]. Saatavilla WWW-muodossa: <URL: [http://urn.fi/URN:NBN:fi-  
fe20071264](http://urn.fi/URN:NBN:fi-fe20071264)>
- /162/ **Hyvönen, E., Viljanen, K. & Suominen, O.** *HealthFinland - Finnish Health Infor-  
mation on the Semantic Web.* 6th International Semantic Web Conference (ISWC  
2007), Busan 11-15.11.2007. Springer. s. 778-791
- /163/ **Hölscher, C. & Strube, G.** *Web Search Behaviour of Internet Experts And Newbies.*  
Computer Networks 33(2000)1-6, s. 337-346.
- /164/ **Idearc Media LLC.** *superpages.com.* n.d. (online) [viitattu 16.6.2009]. Saatavilla  
WWW-muodossa: <URL:<http://www.superpages.com>>
- /165/ **Jansen, B., Spink, A., & Saracevic, T.** *Real Life, Real Users, and Real Needs. A  
Study and Analysis of User Queries on the Web.* Information Processing and Manage-  
ment 36(2000)2, s. 207-227.
- /166/ **Jardine, N. & van Rijsbergen, C.** *The Use of Hierarchic Clustering in Information  
Retrieval.* Information Storage and Retrieval 7(1971)5, s. 217-240.
- /167/ **Kalbach, J.** *Designing Web Navigation.* O'Reilly Media, Sebastopol 2007. 394 s.
- /168/ **Kalfoglou, Y. & Schorlemmer, M.** *Ontology Mapping: State of the Art.* The Know-  
ledge Engineering Review 18(2003)1, s. 1-31.
- /169/ **Kauppinen, T., Ruotsalo, T. & Salminen, M.** *Tietämyksen mallintaminen semantti-  
sessa webissä.* Systemityö (2005)4, s. 32-24.
- /170/ **Klyne, G., Carroll, J. & McBride, B.** *Resource Description Framework (RDF):  
Concepts and Abstract Syntax.* W3C, 2004 (online) [viitattu 18.6.2009]. Saatavilla  
WWW-muodossa: <URL:<http://www.w3.org/TR/rdf-concepts/>>
- /171/ **Krovetz, R. & Croft, B.W.** *Lexical ambiguity and information retrieval.* ACM Tran-  
sactions on Information Systems 10(1992)2, s. 115-141.
- /172/ **Käki, M.** *Findex: Search Result Categories Help Users when Document Ranking  
Fails.* Conference on Human Factors in Computing Systems, Portland 2-7.4.2005.  
ACM. s. 131-140

- /73/ **Käki, M.** *Optimizing the Number of Search Result Categories*. Conference on Human Factors in Computing Systems, Portland 2-7.4.2005. ACM. s. 1517-1520
- /74/ **Käki, M. & Aula, A.** *Findex: improving search result use through automatic filtering categories*. *Interacting with Computers* 17(2005)2, s. 187-206.
- /75/ **Landauer, T.K. & Nachbar, D.W.** *Selection from alphabetic and numeric menu trees using a touch screen: Breadth, depth, and width*. Conference on Human Factors in Computing Systems, San Francisco 14-18.4.1985. ACM. s. 73-78
- /76/ **Leidner, J.L.** *Current issues in software engineering for natural language processing*. Human Language Technology Conference, Edmonton 27.5-1.6.2003. Morristown 2003. Association for Computational Linguistics. s. 45-50
- /77/ **Leppävaaran Liikekeskus Oy.** *Pohjakuvat*. n.d. (online) [viitattu 17.6.2009]. Saatavilla WWW-muodossa: <URL:<http://www.sello.fi/pohjakuvat.php>>
- /78/ **Manola, F., Miller, E. & McBride, B.** *RDF Primer*. W3C, 2004 (online) [viitattu 18.6.2009]. Saatavilla WWW-muodossa: <URL: <http://www.w3.org/TR/rdf-primer/>>
- /79/ **Marchionini, G.**, *Exploratory search: from finding to understanding*. *Communications of the ACM* 49(2006)4, s. 41-46.
- /80/ **Marcus, R.** *The RIAO 94 conference and the status of information retrieval: a personal view*. *ACM SIGIR Forum* 28(1994)2, s. 7-16.
- /81/ **Massachusetts Institute of Technology.** *Exhibit*. n.d. (online) [viitattu 17.6.2009]. Saatavilla WWW-muodossa: <URL:<http://www.simile-widgets.org/exhibit/>>
- /82/ **McGuinness, D. & van Harmelen, F.** *OWL Web Ontology Language Overview*. W3C, 2004 (online) [viitattu 18.6.2008]. Saatavilla WWW-muodossa: <URL: <http://www.w3.org/TR/owl-features/>>
- /83/ **Mihalcea, R. & Moldovan, D.** *Semantic indexing using WordNet senses*. Annual Meeting of the ACL, Hong Kong 1-8.10.2000. s. 35-45
- /84/ **Miles, A., Matthews, B., Wilson, M. & Brickley, D.** *SKOS Core: Simple Knowledge Organisation for the Web*. International Conference on Dublin Core and Metadata Applications (DC-2005), Madrid 12.-15.9.2005. s. 3-10
- /85/ **Morville, P.** *Ambient Findability*. O'Reilly Media, Sebastopol 2005. 188 s.
- /86/ **MySpace Inc.** *MySpace*. n.d. (online) [viitattu 16.6.2009]. Saatavilla WWW-muodossa: <URL:<http://www.myspace.com>>
- /87/ **Muddamalle, M.R.** *Natural language versus controlled vocabulary in information retrieval: a case study in soil mechanics*. *Journal of the American Society for Information Science* 49(1998)10, s. 881-887.
- /88/ **Mäkelä, E., Hyvönen, E. & Sidoroff, T.** *View-based user interfaces for information retrieval on the semantic web*. 4th International Semantic Web Conference (ISWC 2007), Galway 6- 10.11.2005.

- /89/ **Mäkelä, E., Viljanen, K., Lindgren, P., Laukkanen, M. & Hyvönen, E.** *Semantic Yellow Page Service Discovery: The Veturi Portal*. 4th International Semantic Web Conference (ISWC 2005), Galway 6-10.11.2005.
- /90/ **Navigli, R. & Velardi, P.** *An analysis of Ontology-based Query Expansion Strategies*. 14th European Conference on Machine Learning (ECML 2003), Cavtat-Dubrovnik 22.-26.9.2003. Springer. s. 42-49
- /91/ **Nielsen, J.** *When Search Engines Become Answer Engines*. 2004 (online) [viitattu 23.6.2009]. Saatavilla WWW-muodossa: <URL:<http://www.useit.com/alertbox/20040816.html>>
- /92/ **OASIS.** *ebXML - Enabling A Global Electronic Market*. n.d. (online) [viitattu 17.6.2009]. Saatavilla WWW-muodossa: <URL:<http://www.ebxml.org>>
- /93/ **OPA.** *OPA Research: Generational Media Study*. n.d. (online) [viitattu 23.6.2009]. Saatavilla WWW-muodossa: <URL:[http://www.online-publishers.org/media/136\\_W\\_opa\\_generational\\_study\\_sep04.pdf](http://www.online-publishers.org/media/136_W_opa_generational_study_sep04.pdf)>
- /94/ **OpenStreetMap Foundation.** *OpenStreetMap*. n.d. (online) [viitattu 16.6.2009]. Saatavilla WWW-muodossa: <URL:<http://www.openstreetmap.org>>
- /95/ **Patentti- ja rekisterihallitus.** *YTJ - Yritys- ja yhteisötietojärjestelmä — Etusivu*. n.d. (online) [viitattu 16.6.2009]. Saatavilla WWW-muodossa: <URL:<http://www.ytj.fi>>
- /96/ **Pease, A.** *The Suggested Upper Merged Ontology (SUMO)*. n.d. (online) [viitattu 17.6.2009]. Saatavilla WWW-muodossa: <URL:<http://www.ontologyportal.org>>
- /97/ **PKT-säätiö.** *Yrityksen palveluhakemisto*. n.d. (online) [viitattu 16.6.2009]. Saatavilla WWW-muodossa: <URL:<http://www.haeyrityspalvelu.fi>>
- /98/ **Plaisant, C., Shneidermann, B., Doan, K. & Bruns T.** *Interface and data architecture for query preview in networked information systems*. ACM Transactions on Information Systems 17(1999)3, s. 320-341.
- /99/ **Prud'hommeaux, P. & Seaborne, A.** *SPARQL Query Language for RDF*. W3C, 2008 (online) [viitattu 18.6.2009]. Saatavilla WWW-muodossa: <URL:<http://www.w3.org/TR/rdf-sparql-query/>>
- /100/ **Ranganathan, S. R.** *Colon Classification*. Ess Ess Publications, Delhi 2006.
- /101/ **Robertson, S.E. & Jones, S.K.** *Relevance weighting of search terms*. Journal of the American Society for Information Sciences 27(1976)3, s. 129-146.
- /102/ **Rosenfeld, L. & Morville, P.** *Information Architecture for the World Wide Web*. O'Reilly Media, Sebastopol 2002. 461 s.
- /103/ **Salonoja, M.** *Kuluttajan ontologia*. Teknillinen Korkeakoulu, Espoo 2008 (online) [viitattu 17.6.2009]. Saatavilla WWW-muodossa: <URL:[http://www.yso.fi/onto/consumer\\_service/](http://www.yso.fi/onto/consumer_service/)>
- /104/ **Salton G.** *The SMART Retrieval System – Experiments in Automatic Document Processing*. Prentice Hall, Englewood Cliffs 1971.



- /105/**Salton, G. & McGill, M.J.** *Introduction to modern information retrieval*. McGraw-Hill, New York 1986. 400 s.
- /106/**Salton, G. & Buckley, C.** *Improving Retrieval Performance by Relevance Feedback*. Journal of the American Society for Information Science 41(1990)4, s. 288-297.
- /107/**Sanchez, E.** *Fuzzy Logic and the Semantic Web*. Elsevier Science, 2006. 496 s.
- /108/**Sanderson, M.** *Word sense disambiguation and information retrieval*. 17<sup>th</sup> ACM International Conference on Research and Development in Information Retrieval, Dublin 3-6.6.1994. Springer. s. 49-57
- /109/**Schraefel, M., Karam, M. & Zhao, S.** *Listen to the music: Audio preview cues for exploration of online music*. Interact 2003 - Bringing the Bits Together, Zürich 1-5.9.2003. Amsterdam 2003. IOS Press. s. 192-195
- /110/**SFS-EN ISO 9241-11.** *Näyttöpäätteillä tehtävän toimistotyön ergonomiset vaatimukset. Osa 11: Käytettävyyden määrittely ja arviointi*. Suomen Standardoimisliitto SFS 2002. 44 s.
- /111/**Sinkkilä, R., Mäkelä, E., Kauppinen, T. & Hyvönen, E.** *Combining Context Navigation with Semantic Autocompletion to Solve Problems in Concept Selection*. Fifth European Semantic Web Conference (ESWC 2008), Teneriffa 1-5.6.2008. Springer. s. 61-68
- /112/**Smith, M., Welty, C. & McGuinness, D.** *OWL Web Ontology Language Guide*. W3C, 2004 (online) [viitattu 18.6.2009]. Saatavilla WWW-muodossa: <URL: <http://www.w3.org/TR/owl-guide/>>
- /113/**SpringSource Global Inc.** *SpringSource.org*. n.d. (online) [viitattu 17.6.2009]. Saatavilla WWW-muodossa: <URL:<http://www.springsource.org>>
- /114/**Srinivasan, P.** *Optimal document-indexing vocabulary for MEDLINE*. Information Processing & Management 32(1996)5, s. 503-514.
- /115/**Stanford University School of Medicine.** *What is Protégé?* n.d. (online) [viitattu 22.6.2009]. Saatavilla WWW-muodossa: <URL: <http://protege.stanford.edu/overview/>>
- /116/**Steinmetz, N. & Toma, I.** *D16.1v1.0 WSMML Language Reference*. 2008 (online) [viitattu 18.6.2009]. Saatavilla WWW-muodossa: <URL:<http://www.wsmo.org/TR/d16/d16.1/v1.0/>>
- /117/**Suomen Asiakastieto Oy.** *aarre.fi*. n.d. (online) [viitattu 17.6.2009]. Saatavilla WWW-muodossa: <URL:<http://www.aarre.fi>>
- /118/**Suomen Asiakastieto Oy.** *Yritysfiltteri*. n.d. (online) [viitattu 17.6.2009]. Saatavilla WWW-muodossa: <URL:<http://www.aarre.fi/aarre/yritysfiltteri.jsp>>
- /119/**Suomen Yritysoapas Oy.** *Suomen Yritysoapas - Suomalaiset yritykset - Yrityshakupalvelu, jossa on älykäs yrityshaku*. n.d. (online) [viitattu 16.6.2009]. Saatavilla WWW-muodossa: <URL:<http://www.yritysoapas.com>>

- /120/**Suomi24 Oy.** *Suomi24 – Suomen suurin verkkoyhteisö.* n.d. (online) [viitattu 16.6.2009]. Saatavilla WWW-muodossa: <URL:http://www.suomi24.fi>
- /121/**Tam, A.M. & Leung, C.H.C.** *Structured natural-language descriptions for semantic content retrieval of visual materials.* *Journal of the American Society for Information Science* 52(2001)11, s. 930-937.
- /122/**Teevan, J., Alvarado, C., Ackerman, M. S. & Karger, D.R.,** *The perfect search engine is not enough: a study of orienteering behavior in directed search.* Conference on Human Factors in Computing Systems, Wien 24-29.4.2004. ACM. s. 415-422
- /123/**Teknillinen korkeakoulu.** *Kansallinen ontologiapalvelu ONKI.* n.d. (online) [viitattu 16.6.2009]. Saatavilla WWW-muodossa: <URL:http://www.yso.fi>
- /124/**Teknillinen korkeakoulu.** *ONKI-Paikka haku.* n.d. (online) [viitattu 17.6.2009]. Saatavilla WWW-muodossa: <URL:http://demo.seco.tkk.fi/onkipaikka>
- /125/**Teknillinen korkeakoulu.** *Poka - A framework for automatic annotation.* n.d. (online) [viitattu 17.6.2009]. Saatavilla WWW-muodossa: <URL:http://www.seco.tkk.fi/tools/poka/>
- /126/**The Apache Software Foundation.** *Apache Commons.* The Apache Software Foundation, 2009 (online) [viitattu 18.6.2009]. Saatavilla WWW-muodossa: <URL:http://commons.apache.org>
- /127/**The Apache Software Foundation.** *Apache Lucene - Overview.* The Apache Software Foundation, 2009 (online) [viitattu 18.6.2009]. Saatavilla WWW-muodossa: <URL:http://lucene.apache.org/java/docs/index.html>
- /128/**The Apache Software Foundation.** *Welcome to Solr.* The Apache Software Foundation, 2009 (online) [viitattu 18.6.2009]. Saatavilla WWW-muodossa: <URL:http://lucene.apache.org/solr/>
- /129/**Tilastokeskus.** *PRODCOM-luettelo 2008.* n.d. (online) [viitattu 16.6.2009]. Saatavilla WWW-muodossa: <URL:http://www.stat.fi/keruu/tehy/rake\_luettelo.html>
- /130/**Tilastokeskus.** *Toimialaluokitus TOL 2002.* Tilastokeskus, Helsinki 2002. Käsikirjoja 4. 241 s.
- /131/**Tilastokeskus.** *Uusittu toimialaluokitus käyttöön vuonna 2009.* n.d. (online) [viitattu 16.6.2009]. Saatavilla WWW-muodossa: <URL:http://www.stat.fi/til/tol2008.html>
- /132/**Tilastokeskus.** *Yksilöllisen kulutuksenkäyttötarkoituksen mukainen luokitus COICOP.* n.d. (online) [viitattu 16.6.2009]. Saatavilla WWW-muodossa: <URL:http://www.stat.fi/meta/luokitukset/coicop/011-2006/koko\_luokitus.html>
- /133/**Tulli.** *Nimikkeistöt ja luokitukset.* n.d. (online) [viitattu 16.6.2009]. Saatavilla WWW-muodossa: <URL: http://www.tulli.fi/fi/suomen\_tulli/ulkomaankauppatilastot/luokitukset/index.jsp>

- /134/**Tuominen, J.** *Kantapuu: haku*. Teknillinen Korkeakoulu, Espoo 2008 (online) [viitattu 17.6.2009]. Saatavilla WWW-muodossa: <URL:http://www.yso.fi/lusto-demo.html>
- /135/**Tuominen, J., Frosterus, M., Viljanen, K. & Hyvönen, E.** *ONKI SKOS Server for Publishing and Utilizing SKOS Vocabularies and Ontologies as Services*. The 6th Annual European Semantic Web Conference (ESWC2009), Heraklion 31.5. - 4.6.2009. Springer. s. 768-780
- /136/**Universität Innsbruck.** *WSML to OWL Translator*. n.d. (online) [viitattu 17.6.2009]. Saatavilla WWW-muodossa: <URL:http://www.sti-innsbruck.at/results/browse/services/details/?uid=13>
- /137/**Valkeapää, O., Alm, O. & Hyvönen, E.** *An Adaptable Framework for Ontology-based Content Creation on the Semantic Web*. Journal of Universal Computer Science 13(2007)12, s. 1835-1853.
- /138/**Valtioneuvosto.** *JHS 143 kansilehti*. n.d. (online) [viitattu 16.6.2009]. Saatavilla WWW-muodossa: <URL:http://www.jhs-suositukset.fi/suomi/jhs143>
- /139/**van Assem, M., Malaisé, V., Miles, A. & Schreiber, G.** *A Method to Convert Thesauri to SKOS*. 3rd European Semantic Web Conference, Budva 11.-14.6.2006. Springer. s. 95-109
- /140/**Viljanen, K., Tuominen, J. & Hyvönen, E.** *Ontology Libraries for Production Use: The Finnish Ontology Library Service ONKI*. The 6th Annual European Semantic Web Conference (ESWC2009), Heraklion 31.5. - 4.6.2009. Springer. s. 781-795
- /141/**Vivísimo Inc.** *Clusty the clustering search engine*. n.d. (online) [viitattu 16.6.2009]. Saatavilla WWW-muodossa: <URL:http://clusty.com>
- /142/**Voorhees, E.M.** *Query expansion using Lexical-Semantic Relations*. Annual ACM Conference on Research and Development in Information Retrieval, Dublin 3.-6.7.1994. Springer. s. 61-69
- /143/**Whatamap.com Ltd.** *Home*. n.d. (online) [viitattu 16.6.2009]. Saatavilla WWW-muodossa: <URL:http://www.whatamap.com>
- /144/**Wine.com Inc.** *Wine.com - Buy Wine, Wine Clubs, Gift Baskets and more*. n.d. (online) [viitattu 16.6.2009]. Saatavilla WWW-muodossa: <URL:http://www.wine.com>
- /145/**Yahoo! Inc.** *Yahoo!*. n.d. (online) [viitattu 16.6.2009]. Saatavilla WWW-muodossa: <URL:http://www.yahoo.com>

# LIITTEET

## LIITE 1 Tärkeimpiä ONKI:ssa julkaisun määrittäjiä

Kontrolloidun sanaston tai ontologian ONKI:ssa julkaisun yhteydessä on tehtävä muutamia määrittäjiä, jotka vaikuttavat pääasiassa ONKI-selaimen ulkoasuun. Näistä tärkeimpiä ovat:

- Nimi ontologialle.
- Visuaalisen hierarkian kuvaavien ominaisuuksien nimet, joita on tyypillisesti OWL-muodon kohdalla yksi: `rdfs:subClassOf`. Määritettäessä useita ominaisuuksia visualisoimaan hierarkia kaikki määritetyt ominaisuudet ovat päällä eikä tätä voi käytön aikana dynaamisesti muuttaa.
- Resurssin nimen kuvaavan ominaisuuden nimi, joka on tyypillisesti OWL-muodon kohdalla `rdfs:label`.
- Ominaisuudet ja käsitetyypit, jotka halutaan piilottaa ontologiaa selattaessa.

## LIITE 2 Kehitetyn sovelluksen tägien peilaus KOKO-ontologiaan ja UNSPSC-luokitukseen

Kehitettyssä yrittäjille suunnatussa sovelluksessa kenttien arvot kuvattiin tägeinä (tag), joiden vastaavuudet KOKO-ontologian ja UNSPSC-luokituksen suhteen on kuvattu alla olevassa taulukossa.

Tägi / tägin osa	KOKO (nimi: ID)	UNSPSC (nimi:koodi)
Lomakkeet ja asiakirjat	lomakkeet: p36734 asiakirjat: p48398	Lomakkeet tai kysely- kaavakkeet liiketoimin- takäyttöön: 14111806
Lainsäädäntö	lainsäädäntö: p34040	Lainsäädännölliset elimet ja käytännöt: 93101700
Luvat viranomaisilta	luvat: p32857	-
Vakuutus – pakollinen	vakuutus: p4165	Vakuutus ja eläkepalve- lut: 84130000
vapaaehtoinen	-	-
vapaaehtoinen	vapaaehtoinen vakuutus: p13742	-
vientitakuu	vientitakuu: p9432	-
Toimitilat/hautomo	tilapalvelut: p3050 yrityshautomot: p37030	-
Kehittämishjelmat	kehittäminen: p33993	-
Koulutus	koulutus: p36950	Koulutus- ja valmen- nuspalvelut: 86000000
Edunvalvonta ja yritys- yhteistyö	edunvalvonta: p1425 yhteistyö: p32667	-
Laskujen maksaminen - pankkitilit	laskut: p5326	-
Rahan saanti - avustus/tuki	pankkitilit: p10701	-
laina	raha: p8345	-
lainaan takaus	EU-tuet: p24969	Raha-apu: 84101501
pääomasijoitus	lainat: p32357	-
naisyrittäjä	takaus: p17391	-
elinkaari – alkava	pääomasijoittaminen: p10703	-
	naiset: p34608	-
	elinkaari: p35191	-
	aloitus: p31979	-

vakiintunut	toimivuus: p7965	-
vaihtamassa omistajaa	omistajanvaihdos: p7604	-
yrittäjämuodoltaan osuuskunta	osuuskunnat (yrittäjämuodot): p48352	-
kansainvälinen	kansainvälinen kauppa: p6942	-
(Venäjä)	venäläiset: p8327	-
(muut maat)	-	-
toimialalla -	toimialat: p35647	-
autot ja ajoneuvot	autoala: p10969	Moottoriajoneuvot: 25100000
elintarvikkeet ja ruoka	elintarvikeala: p6995	Elintarvike- ja ravintopalvelut: 85150000
käsityöt/tekstiilit/jalkineet	käsityöala: p4595 tekstiiliala: p31426 jalkineiteollisuus: p37036	-
maatalous	maatalousala: p6017	-
matkailu	matkailuala: p3955	Matkailupalvelut: 83121602
terveys ja hyvinvointi	terveydenhuoltoala: p15625	Terveydenhuollon palvelut: 85000000
palvelu - sähköinen	-	-
palvelu - tietopankki	tietopalvelu: p35996	-

### **LIITE 3 KOKO-ontologian käsitteitä liitettäväksi yrittäjille suunnattujen palveluiden kuvauksiin**

Kehitetyn sovelluksen tietovaraston palvelukuvauksia voitaisiin laajentaa uudelleenkäytävyyden helpottamiseksi tässä liitteessä listattujen KOKO-ontologian käsitteiden avulla.

#### **Yksittäisten palveluntarjoajien/palveluiden toimintaa kuvaavia asiasanoja (nimi: ID):**

- tukkukauppa: p35758
- vähittäiskauppa: p1048
- päivittäistavarakauppa: p33295
- logistiikka: p5937
- lihateollisuus: p37298
- laivanvarustus: p17018
- maanmittaus: p34737
- sairaankuljetus: p1151
- kemianteollisuus: p34977
- kumiteollisuus: p17516
- muoviteollisuus: p31298
- kaivostoiminta: p31234
- malminetsintä: p37330
- kivenjalostus: p4154
- ompelu: p30593
- nuohous: p15852
- oppisopimuskoulutus: p14550
- ajokortit: p31754
- starttiraha: p1966
- sähköinen asiointi: p3554
- sähköinen laskutus: p967
- baltit: p3295
- keksinnöt: p35382
- oleskeluluvat: p12191
- vekselit: p14298
- asbesti: p36900
- räjäytystyöt: p13913

**Yksittäisten palveluntarjoajien organisaatiotyyppiä kuvaavia asiasanoja (nimi: ID):**

- työttömyyskassat: p12854
- perheyrietykset: p2862
- sähkösuunnittelijat: p11795
- vartiointiliikkeet: p17849
- kyläkaupat: p15481
- autokoulut: p15536
- apteekit: p33193
- ympäristökeskukset: p3925
- ministeriöt: p31710
- verohallinto: p2796
- verotoimistot: p15939
- lääninhallitukset: p812
- maistraatit: p11000
- työvoima- ja elinkeinokeskukset: p8764
- koulutusorganisaatio: p3224
- kurssikeskukset: p15343
- ammatilliset aikuiskoulutuskeskukset: p10990
- maatalousalan järjestöt: 94101501
- käsi- ja taideteollisuusyhdistykset: p43441
- kauppakamarit: p17003
- yrittäjäjärjestöt: p14276
- etujärjestöt: p34344
- ammattialajärjestöt: p15425
- kaupan järjestöt: p16082
- tuomioistuimet: p7925
- käräjäoikeudet: p11142
- hovioikeudet: p4634
- hallinto-oikeudet: p6965
- tullilaitos: p2371
- tilintarkastajat: p2475



## LIITE 4 UNSPSC:n CSV → OWL -muunnos

UNSPSC-luokitus muunnettiin CSV-muodostosta OWL-muotoon liitteen Java-ohjelman avulla. Syötetiedostossa oli kolme kenttää puolipisteillä toisistaan erotettuna seuraavassa järjestyksessä: koodi, koodia vastaava selite englanniksi ja englanninkielisen selitteen suomenkielinen käännös. Nimet syöte- ja tulostiedostoille määriteltiin ohjelman alussa esitellyissä muuttujissa.

### Java-ohjelma:

```
import java.io.*;
import java.util.StringTokenizer;

public class UnspscCsvToOwl {

    static final String inputFileName="UNSPSC.csv";
    static final String outputFileName="unspsc.owl";
    static final String[][] basicClasses=new String[][]{{
        {null,"unspscRoot","UNSPSC ROOT","UNSPSC ROOT"},
        {"unspscRoot","rawMaterials","Raaka materiaalit","Raw materials"},
        {"unspscRoot","industrialEquipment","Teolliset laitteet","Industrial Equipment"},
        {"unspscRoot","equipmentComponentsAndSupplies",
            "Laitteisto komponentit ja tarjoomat","Equipment Components and Supplies"},
        {"unspscRoot","manufacturedProducts","Valmistetut tuotteet","Manufactured Products"},
        {"unspscRoot","services","Palvelut","Services"}}};

    static final String startText=
        "<rdf:RDF xmlns:protege=\"http://protege.stanford.edu/plugins/owl/protege#\"\\n"+
        "    xmlns:rdf=\"http://www.w3.org/1999/02/22-rdf-syntax-ns#\"\\n"+
        "    xmlns:rdfs=\"http://www.w3.org/2000/01/rdf-schema#\"\\n"+
        "    xmlns=\"http://www.yso.fi/onto/unspsc/\"\\n"+
        "    xmlns:owl=\"http://www.w3.org/2002/07/owl#\"\\n"+
        "    xml:base=\"http://www.yso.fi/onto/unspsc/\">\\n\\n"+
        " <owl:Ontology rdf:about=\"\">\\n"+
        " <protege:defaultLanguage
            rdf:datatype=\"http://www.w3.org/2001/XMLSchema#string\">fi"+
        " </protege:defaultLanguage>\\n"+
        " <rdfs:label rdf:datatype=\"http://www.w3.org/2001/XMLSchema#string\">UNSPSC"+
        " </rdfs:label>\\n"+
        " <owl:imports "+
            " rdf:resource=\"http://protege.stanford.edu/plugins/owl/protege\"/>\\n"+
        " </owl:Ontology>\\n\\n";

    static final String endText="</rdf:RDF>";

    private static void writeBasicClasses(BufferedWriter out) throws Exception{
```

```

for(int i=0;i<basicClasses.length;i++)
    out.write(
        " <owl:Class rdf:about=\"http://www.yso.fi/onto/unspsc/"+
        basicClasses[i][1]+"\">\n"+
        " <rdfs:label xml:lang=\"fi\">"+basicClasses[i][2]+"</rdfs:label>\n"+
        " <rdfs:label xml:lang=\"en\">"+basicClasses[i][3]+"</rdfs:label>\n"+
        (basicClasses[i][0]!=null?
            " <rdfs:subClassOf>\n"+
            " <owl:Class rdf:about=\"http://www.yso.fi/onto/unspsc/"+
            basicClasses[i][0]+"\"/>\n"+
            " </rdfs:subClassOf>\n": "")+
            " </owl:Class>\n\n");
}

private static String getParentId(String code){

    if (code.substring(2,8).equals("000000")){

        int segment=Integer.parseInt(code.substring(0,2));

        if(segment>=10 && segment<=15) return "rawMaterials";
        if(segment>=20 && segment<=27) return "industrialEquipment";
        if(segment>=30 && segment<=41) return "equipmentComponentsAndSupplies";
        if(segment>=42 && segment<=60) return "manufacturedProducts";
        if(segment>=70 && segment<=94) return "services";

    }

    if (code.substring(4,8).equals("0000")) return "C"+code.substring(0,2)+"000000";
    if (code.substring(6,8).equals("00")) return "C"+code.substring(0,4)+"0000";
    return "C"+code.substring(0,6)+"00";

}

public static void main(String[] args) {

    try{
        BufferedReader in = new BufferedReader(new FileReader(inputFileName));
        BufferedWriter out = new BufferedWriter(new FileWriter(outputFileName));

        out.write(startText);
        writeBasicClasses(out);

        String str;
        while ((str = in.readLine()) != null){

            if(str.startsWith("\key\"")) continue;

            StringTokenizer tokenizer = new StringTokenizer(str,";");

```

```

String code=tokenizer.nextToken();
String descriptionEn=tokenizer.nextToken("\\;");
String descriptionFi=tokenizer.nextToken("\\;");
String subClass="http://www.yso.fi/onto/unspsc/"+getParentId(code);

try {
    out.write("<owl:Class rdf:about=\"http://www.yso.fi/onto/unspsc/C"+code+"\">\n"+
        "  <rdfs:label xml:lang=\"fi\">"+descriptionFi+"</rdfs:label>\n"+
        "  <rdfs:label xml:lang=\"en\">"+descriptionEn+"</rdfs:label>\n"+
        "  <rdfs:subClassOf>"+
        "    <owl:Class rdf:about=\""+subClass+"\"/>"+
        "  </rdfs:subClassOf>\n"+
        "</owl:Class>\n");
    } catch (IOException e) {}
}

out.write(endText);
in.close();
out.close();
} catch (Exception e) {}
}
}

```

## LIITE 5 UNSPSC-luokituksen SKOS-muodon skeema

UNSPSC-luokitus julkaistiin SKOS-muodossa ONKI-palvelimelle. SKOS-muodon skeema määriteltiin seuraavasti:

```
<?xml version="1.0"?>
<rdf:RDF <!-- Nimiavaruudet (poistettu) --> >
  <rdf:Description rdf:about="http://www.yso.fi/onto/unspsc/conceptscheme">
    <rdf:type rdf:resource="http://www.w3.org/2008/05/skos#ConceptScheme"/>
    <skos:hasTopConcept rdf:resource="http://www.yso.fi/onto/unspsc/equipmentComponentsAnd-
Supplies"/>
    <!-- Muut skos:hasTopConcept-tagit (poistettu) -->
    <skos:prefLabel xml:lang="fi">UNSPSC</skos:prefLabel>
    <skos:prefLabel xml:lang="en">UNSPSC</skos:prefLabel>
  </rdf:Description>
  <!-- Käsite-lohkot -->
</rdf:RDF>
```

## LIITE 6 Turtle → OWL -muunnoksen Java-koodi

Liitteen Java-ohjelman avulla Turtle-muodossa määritelty lomakkeisto muunnettiin OWL-muotoon. Muunnettavan tiedoston nimi välitettiin ohjelmalle ajon yhteydessä parametrina.

### Java-ohjelma:

```
import java.io.*;
import com.hp.hpl.jena.rdf.model.Model;
import com.hp.hpl.jena.util.FileManager;

public class TurtleToXml {

    public static void main(String[] args) throws IOException {

        String originalFileName=(new File(args[0])).getAbsolutePath();
        String newFileName=originalFileName.replace(".ttl", ".owl");

        Model model=FileManager.get().loadModel(originalFileName, "TURTLE");
        FileOutputStream fos = new FileOutputStream(new File(newFileName));

        model.write(fos);
        fos.flush();
        fos.close();
    }
}
```

## LIITE 7 Demosovelluksen datan muuntamisessa käytetty XSL-tiedosto

Demosovelluksen data muunnettiin HAKO-yhteensopivaan muotoon liitteen XSL-tiedoston avulla.

### XSL-tiedosto:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<xsl:stylesheet version="2.0"
xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:exhibit="http://simile.mit.edu/2006/11/exhibit#"
xmlns:a="http://www.w3.org/2000/01/rdf-schema#"
xmlns:b="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:c="http://simile.mit.edu/2006/11/exhibit#"
xmlns:d="http://www.seco.tkk.fi/u/msalonoj/ys/property#">

<xsl:template match="/">
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:foaf="http://xmlns.com/foaf/0.1/" xmlns:owl="http://www.w3.org/2002/07/owl#"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:es="http://www.seco.tkk.fi/applications/saha/schemas/yrittysuomi#"
xmlns:eserv="http://www.yso.fi/onto/entrepreneur_service/" xmlns:daml="http://www.daml.org/
2001/03/daml+oil#" xmlns:j.0="http://www.seco.tkk.fi/applications/saha#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"&gt;
<xsl:for-each select="rdf:RDF/rdf:Description">
<rdf:Description rdf:about="http://www.seco.tkk.fi/applications/saha#Instance_<xsl:va-
lue-of select="generate-id(exhibit:origin)"/>"&gt;
<rdf:type
rdf:resource="http://www.seco.tkk.fi/applications/saha/schemas/yrittysuomi#Context"/&gt;
<j.0:dateCreated>03.04.2009 12:00</j.0:dateCreated&gt;
<j.0:dateModified>03.04.2009 12:00</j.0:dateModified&gt;
<j.0:localType rdf:resource="http://www.seco.tkk.fi/applications/saha/schemas/yrittysuo-
mi#Context"/&gt;
<es:name xml:lang="fi"><xsl:value-of select="a:label"/></es:name&gt;
<rdfs:label xml:lang="fi"><xsl:value-of select="a:label"/></rdfs:label&gt;
<xsl:apply-templates select="d:area"/>
<xsl:apply-templates select="d:individualType"/>
<xsl:apply-templates select="d:need"/>
<xsl:apply-templates select="d:profile"/>
</rdf:Description&gt;
</xsl:for-each>
</rdf:RDF&gt;
</xsl:template>

<xsl:template match="d:area">
```

```

<es:area rdf:resource="http://www.yso.fi/onto/entrepreneur_service/<xsl:call-template
name="myLocalName"><xsl:with-param name="d:area"/></xsl:call-template"/>
</xsl:template>
<xsl:template match="d:individualType">
<es:individualType rdf:resource="http://www.yso.fi/onto/entrepreneur_service/<xsl:call-
template name="myLocalName"><xsl:with-param name="d:individualType"/></xsl:call-template"/
&gt;
</xsl:template>
<xsl:template match="d:need">
<es:need rdf:resource="http://www.yso.fi/onto/entrepreneur_service/<xsl:call-template
name="myLocalName"><xsl:with-param name="d:need"/></xsl:call-template"/>
</xsl:template>
<xsl:template match="d:profile">
<es:profile rdf:resource="http://www.yso.fi/onto/entrepreneur_service/<xsl:call-template
name="myLocalName"><xsl:with-param name="d:profile"/></xsl:call-template"/>
</xsl:template>

<!-- Tagien muunto -->
<xsl:template name="myLocalName">
  <xsl:choose>
    <xsl:when test=". = 'Lomakkeet ja asiakirjat'">need_forms</xsl:when>
    <xsl:when test=". = 'Lainsäädäntö'">need_law</xsl:when>
    <xsl:when test=". = 'Luvat viranomaisilta'">need_luvat</xsl:when>
    <xsl:when test=". = 'Vakuutus - pakollinen'">need_vakuutus_pakollinen</xsl:when>
    <xsl:when test=". = 'Vakuutus - vapaaehtoinen'">need_vakuutus_vapaaehtoinen</xsl:when>
    <xsl:when test=". = 'Vakuutus - vientitakuu'">need_vientitakuu</xsl:when>
    <xsl:when test=". = 'Toimitilat/hautomo'">need_room</xsl:when>
    <xsl:when test=". = 'Kehittämishjelmat'">need_development</xsl:when>
    <xsl:when test=". = 'Koulutus'">need_education</xsl:when>
    <xsl:when test=". = 'Edunvalvonta ja yritysysteistyö'">need_yritysyhteistyö</xsl:when>
    <xsl:when test=". = 'Laskujen maksaminen'">need_maksaminen</xsl:when>
    <xsl:when test=". = 'Laskujen maksaminen - pankkitilit'">need_account</xsl:when>
    <xsl:when test=". = 'Rahan saanti - avustus/tuki'">need_avustus</xsl:when>
    <xsl:when test=". = 'Rahan saanti - laina'">need_loan</xsl:when>
    <xsl:when test=". = 'Rahan saanti - lainaan takaus'">need_takaus</xsl:when>
    <xsl:when test=". = 'Rahan saanti - pääomasijoitus'">need_paaomasijoitus</xsl:when>

    <xsl:when test=". = 'naisyritys'">profile_women</xsl:when>
    <xsl:when test=". = 'elinkaari - alkava yritys'">profile_aloittavayritys</xsl:when>
    <xsl:when test=". = 'elinkaari - vakiintunut'">profile_toimivayritys</xsl:when>
    <xsl:when test=". = 'elinkaari - vaihtamassa omistajaa'">profile_lopettavayritys</xsl:w-
hen>
    <xsl:when test=". = 'yritysmuodoltaan osuuskunta'">profile_osuuskunta</xsl:when>
    <xsl:when test=". = 'kansainvälinen'">profile_internationalBusiness</xsl:when>
    <xsl:when test=". = 'kansainvälinen (Venäjä)'">profile_russia</xsl:when>

```

```
<xsl:when test=". = 'kansainvälinen (muut maat)'">profile_otherCountries</xsl:when>
<xsl:when test=". = 'toimialalla - autot ja ajoneuvot'">profile_cars</xsl:when>
<xsl:when test=". = 'toimialalla -
käsityöt/tekstiilit/jalkineet'">profile_textiles</xsl:when>
<xsl:when test=". = 'toimialalla - maatalous'">profile_agriculture</xsl:when>
<xsl:when test=". = 'toimialalla - matkailu'">profile_travel</xsl:when>
<xsl:when test=". = 'toimialalla - ruoka'">profile_food</xsl:when>
<xsl:when test=". = 'toimialalla - terveys'">profile_terveys</xsl:when>

<xsl:when test=". = 'Palvelu'">individualType_service</xsl:when>
<xsl:when test=". = 'Palvelu - sähköinen'">individualType_eService</xsl:when>
<xsl:when test=". = 'Palvelu - tietopankki'">individualType_database</xsl:when>

<xsl:when test=". = 'Alueellisesti rajoittunut'">area_regional</xsl:when>
<xsl:when test=". = 'Valtakunnallinen'">area_valtakunnallinen</xsl:when>
<xsl:otherwise><xsl:value-of select="."/;></xsl:otherwise>
</xsl:choose>
</xsl:template>

</xsl:stylesheet>
```



## LIITE 8 Katuosoitteiden muuntaminen koordinaateiksi

Palvelupaikkoja vastaavien koordinaattien selvittämiseksi hyödynnettiin Googlen tarjoamia Javascript-funktioita. Funktioita on käytetty liitteen HTML-sivussa, joka tulostaa koordinaatit palvelupaikan nimen yhteyteen näytöllä. Palvelupaikkojen osoitteet haetaan tiedostosta, jonka nimi on määritelty addressFN-muuttujassa.

### HTML-sivu:

```
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<script type="text/javascript" src="http://maps.google.com/maps?
file=api&v=2&key=ABQIAAAxasFcSKk5bsSA80n92vyxhSdLL1MQViTkRYX3pT6u75Br6WpgRTW0S3M-
cAH4IBrmBGww7NakFFmW1A">
</script>
</head>
<body onload="GUnload()">
<div id="message">Tiedot:<br><br></div>
<script type="text/javascript">
//

// Tiedosto, jossa on kuvattu palvelupaikkojen osoitteet
var addressFN="test_toimipaikat2008.txt";

var delay = 500;
var geocoder;
var lines=[];
var coordPoint;
var address;
var accuracy;

var lineNumber = 0;

function extractCoordinates(){

    if (lines[lineNumber] == "") return;

    var tokens = lines[lineNumber].split(";");
    var kunta = tokens[3];
    var katuosoite = tokens[1];
    var postinumero = tokens[2];

    if (katuosoite.length &lt; 4 ) address = postinumero +"," +kunta;
    else if (katuosoite.substring(0,3) == "PL ") address = postinumero +"," +kunta;
    else address = katuosoite +"," +postinumero + " " +kunta;</pre></div>
```

```

document.getElementById("message").innerHTML +=tokens[0]+'<br>';
document.getElementById("message").innerHTML +=
    "addressRoute": "http://maps.google.fi/maps?daddr='+escape(address)+'",<br>';

geocoder.getLocations(address, function(response) {

if (!response || response.Status.code != 200) coordPoint = "address not found";
else{
    place = response.Placemark[0];
    coordPoint = place.Point.coordinates[1] +"," +place.Point.coordinates[0];
    accuracy = place.AddressDetails.Accuracy;
}

    showResult();
});
}
function showResult() {
    document.getElementById("message").innerHTML +=
        "addressLatLng": "'"+coordPoint+"',<br><br>';
    ++lineNumber;
    if (lineNumber < lines.length) setTimeout("extractCoordinates()", delay);
}

GDownloadUrl(addressFN,
    function(content){
        lines = content.split("\n");
        geocoder = new GClientGeocoder();
        extractCoordinates(); });
//]]>
</script>
</body>
</html>

```