

Semanttiset portaalit

Teemu Sidoroff

Helsinki 7.3.2005

Pro gradu -tutkielma

HELSINGIN YLIOPISTO

Tietojenkäsittelytieteen laitos

Semanttiset portaalit

Tietojenkäsittelytiede

Pro gradu -tutkielma

7.3.2005

63 sivua

Semanttiset portaalit ovat semanttisen webin julkaisukanavia, joissa esitetty tieto on tallennettu konekäsiteltävässä muodossa. Portaalien tietoalkiot on annotoitu RDF-muodossa portaalissa käytettävän portaaliontologian mukaisesti. Portaaliontologia voi koostua useista erilaisista osaontologioista, jotka mahdollistavat tietosisällön kuvaamisen erilaisista näkökulmista.

Semanttisissa portaaleissa esitettävä sisältö voidaan tallentaa RDF-tietämyskantoihin. Sisältöä voidaan hallita ja tuottaa sisällöntuotantotyökaluilla samoin kuten tavanomaisissa portaaleissa, eikä portaalin toteuttamisessa käytettyjen tekniikoiden tarvitse näkyä sisällöntuottajalle. Ongelma semanttisten portaalien käyttöönotossa on olemassa olevien sisällöntuotantojärjestelmien tarjoaman tuen puute ontologiapohjaiselle tiedon kuvailulle.

Semanttisten portaalien tietosisällön rikas metadata mahdollistaa tiedon moninaisen esittämisen. Moniulotteinen metadata tietoalkiosta mahdollistaa näkymäselausparadigman mukaisen käyttöliittymän toteuttamisen. Ontologiapohjainen metadata mahdollistaa myös tietoalkioiden automaattisen päättelysääntöihin perustuvan suosittelem.

Pro gradussa esitellään prototyyppi semanttisesta portaalista joka pohjautuu sisällöltään julkishallinnon Suomi.fi-portaaliin. Portaali on toteutettu OntoViews-järjestelmän pohjalta. Portaalin tietosisältö on annotoitu moniulotteisen portaaliontologian mukaisesti ja näkymäselaukseen perustuva käyttöliittymä mahdollistaa useisiin erilaisiin tiedonhakutilanteisiin mukautuvan tiedonhakukäyttöliittymän.

ACM Computing Classification System (CCS):

H.5.4 [Information Interfaces and Representation]:Hypertext/Hypermedia

H.3.5 [Information Storage and Retrieval]: Online Information Services

Semanttinen web, portaalit, WWW

Kumpulan tiedekirjasto, sarjanumero C-

Sisältö

1	Johdanto	1
2	Semanttinen web ja sen perusteknologiat	3
2.1	RDF	3
2.2	RDFS	6
2.3	Ontologiat ja ontologiakieli OWL	7
3	Semanttiset portaalit	9
3.1	Portaalikäsitteen kehitys: yleisportaalista semanttiseen portaaliin . . .	9
3.2	Yhteisöportaalit	12
3.3	Informaatioportaalit	13
3.4	Esimerkkejä semanttisista portaaleista	14
3.4.1	Open Directory Project	15
3.4.2	OntoWeb	17
3.4.3	Semantic Web Environmental Directory (SWED)	19
4	Semanttisten portaalien tiedonhallinta	22
4.1	Tietämyksen varastointi ja muokkaaminen	22
4.2	RDF-tietovarastot ja tietämuskannat	24
4.3	RDF-kyselykielet	27
5	Semanttiset portaalit tiedonhaussa	29
5.1	Tietämyksen visualisointi	29
5.2	Näkymäselaus	33
5.3	Ontologiaperustainen käsitehaku	36
6	Suomi.fi-julkishallinnon portaali	39
6.1	Portaalin nykyinen versio	39
6.2	Semanttisen webin teknologioiden tuomat mahdollisuudet	41
6.2.1	Portaalin tietojen semanttinen annotointi	42
6.2.2	Näkymäselauksen hyödyntäminen	42
6.2.3	Sanahausta käsitehakuun	43
6.2.4	Semanttinen suosittelu	44
6.2.5	Portaalin ylläpito	45

6.3	Ontologiapohjainen Suomi.fi	46
6.4	Portaalin evaluointi ja jatkokehitys	52
7	Yhteenveto	55
	Lähteet	57

1 Johdanto

WWW:stä on muodostunut valtava tietopankki. Sen nopean kasvun myötä tiedon löytämisestä on tullut haastavaa. Hakukoneiden, kuten Googlen¹ ja AllTheWebin², lisäksi tiedonetsinnässä auttavat portaalit, jotka suodattavat webin valtavaa tietomäärää yhden sivuston alle tai tarjoavat aloitussivun webin samoamiseen.

Web-portaali on laaja käsite, jonka alle on lokeroitu monenlaisia web-sivustoja. Laajemmasta portaalikäsitteestä voidaan hahmottaa kaksi toisiaan leikkaavaa osajoukkoa, joista erityisesti ensimmäiseen tässä tutkielmassa portaaaleja käsiteltäessä keskitytään: informaatioportaalit (*information portals*) ja yhteisöportaalit (*community portals*). Informaatioportaalit toimivat tietämyksen yhteenkerääjinä joko varastoitamalla ja esittämällä tietämystä portaalin sisällä tai tarjoamalla lähtökohdan tiedonhauille esimerkiksi linkkien muodossa. Kanoninen esimerkki tällaisesta portaalista on Yahoo!³, joka tarjoaa muiden palveluidensa ohessa luokitellun hakemiston web-sivuista.

Yhteisöportaalien sisällöstä vastaa portaalialia käyttävä yhteisö, joka tuottaa portaaliiin tietosisältöä omaan käyttöönsä. Yhteisöportaaaleiksi voidaan laskea suuri osa yritysten ja organisaatioiden intraneteistä, joiden sisällöstä vastaa suuri, hajautettu joukko käyttäjiä [So00].

Uuden verkkovision, semanttisen webin (*Semantic Web*) [BLHL01] teknologiat tuovat uusia mahdollisuuksia portaalien sisällönhallintaan, tiedonhakuun ja käyttöliittymäratkaisuihin. Semanttisessa webissä tieto on esitetty koneiden ymmärrettäväsä muodossa ja tiedon sisältö ja esitystapa on täysin erotettuna toisistaan. Tämä mahdollistaa useiden eri näkymien tuottamisen samasta tiedosta sekä useiden tietolähteiden sisältämän tiedon yhteenkeräämisen.

¹<http://www.google.com>

²<http://www.alltheweb.com>

³<http://www.yahoo.com>

Semanttisen, kuten tavallisenkin webin perustana on idea URI-tunnistein (*Uniform Resource Identifier*) esitetyistä resursseista ja niihin liittyvistä metatiedoista. Resursseista ja metatiedoista muodostuu verkko, joka on tarkoitettu koneiden käytettäväksi ja jota ne voivat hyödyntää tiedonhaussa ja tiedon esittämisessä.

Semanttisen webin teknologioita hyväksi käytettäviä portaaleja kutsutaan semanttiseksi portaaleiksi (*semantic portal*) [M⁺03]. Semanttiset portaalit pohjaavat lähtökohtaisesti ontologioihin, jotka kuvaavat formaalisti portaaliin talletettavan tiedon rakenteen [So00]. Portaaleissa ontologioita voidaan käyttää hyväksi tiedon tuottamisessa ja esittämisessä sekä tiedon etsinnän tehostamisessa [UJ99]. Metatiedolla on tärkeä rooli tiedonhaun kannalta, ja metatiedon osuus portaalien tietomäärästä on usein huomattava. Metatietojen avulla portaalien tietoalkioita voidaan luokitella ja etsiä, sekä muodostaa alkioista erilaisia kokonaisuuksia. Kun metatiedot on esitetty standardissa muodossa, voidaan tietämystä siirtää myös eri organisaatioiden välillä helposti, ilman vaivalloista tiedon muunnosta jokaisen eri tietoa hyödyntävän organisaation välillä.

Tässä pro gradu -tutkielmassa selvitetään semanttisen webin tuomia mahdollisuuksia portaalien sisällöntuottamiseen ja tiedon esittämiseen. Luvussa 2 käydään läpi semanttisen webin perusteknologioita, jotka toimivat pohjana myöhemmissä luvuissa käsiteltäville menetelmille. Luvussa 3 keskitytään tarkastelemaan portaalikäsitettä ja määritellään, mitä semanttiset portaalit ovat. Lisäksi luvussa esitellään esimerkkejä erityyppisistä semanttisista portaaleista. Luku 4 keskittyy semanttisten portaalien ylläpitoon ja portaaleihin liittyvään tiedonhallintaan. Luvussa 5 siirrytään käsittelemään semanttisia portaaleja käyttäjän näkökulmasta ja tarkastellaan semanttisten portaalien tuomia mahdollisuuksia tiedonhakuun ja tiedon visualisointiin. Luvussa 6 tarkastellaan Suomi.fi-portaalia ja esitetään kuinka tässä portaalissa voidaan hyödyntää semanttisen webin tuomia mahdollisuuksia.

2 Semanttinen web ja sen perusteknologiat

Alkuperäisen WWW:n kehittäjän, Tim Berners-Leen visioima semanttinen web voidaan nähdä nykyisen WWW:n laajennoksena. Semanttisen webin perustana olevat teknologiat perustuvat WWW:n standardeja julkaisevan ja kehittävän W3C:n (*World Wide Web Consortium*)⁴ suosituksiin.

Berners-Lee esittää kirjassaan *Weaving the Web* valaisevan vertauksen siitä, mikä semanttisen webin tavoite on: “Jos HTML ja web saivat on-line-dokumentit näyttämään yhdeltä valtavalta kirjalta, RDF(S)⁵ ja päättelykielet saavat maailman datan näyttämään yhdeltä valtavalta tietokannalta” [BL99]. Tuo analogia kuvaa hyvin sitä, kuinka semanttisessa webissä pyritään saamaan tieto koneiden ymmärrettävään muotoon, jolloin koneita voidaan käyttää tehokkaammin hyödyksi tiedon etsinnässä ja useista heterogeenisistä lähteistä peräisin olevan tiedon integroinnissa [BLHL01].

Tässä luvussa käsitellään lyhyesti semanttisen webin perustan rakentavien teknologioiden, RDF- ja RDFS-kielten ja ontologioiden perusteet. Semanttisen webin perustana toimii koneiden ymmärtämässä, hyvinmääritellyssä muodossa esitetty RDF-pohjainen tietämys, joka on esitetty ontologioissa formaalisti määritellyjä termejä käyttäen.

2.1 RDF

RDF (*Resource Description Framework*) on tiedonesityskieli, joka on tarkoitettu erityisesti WWW:n resurssien (*resource*) kuvaamiseen [RDF04b]. Resurssi voi olla tavallisen HTML-kielellä kuvatun web-sivun lisäksi myös mikä tahansa entiteetti, joka voidaan identifioida URI-viittauksella (*URI-reference*) [URI98], eli URI:lla (*Uniform Resource Identifier*) tai URI-fragmentilla. Nimensä mukaisesti URI-viittaus identi-

⁴<http://www.w3.org>

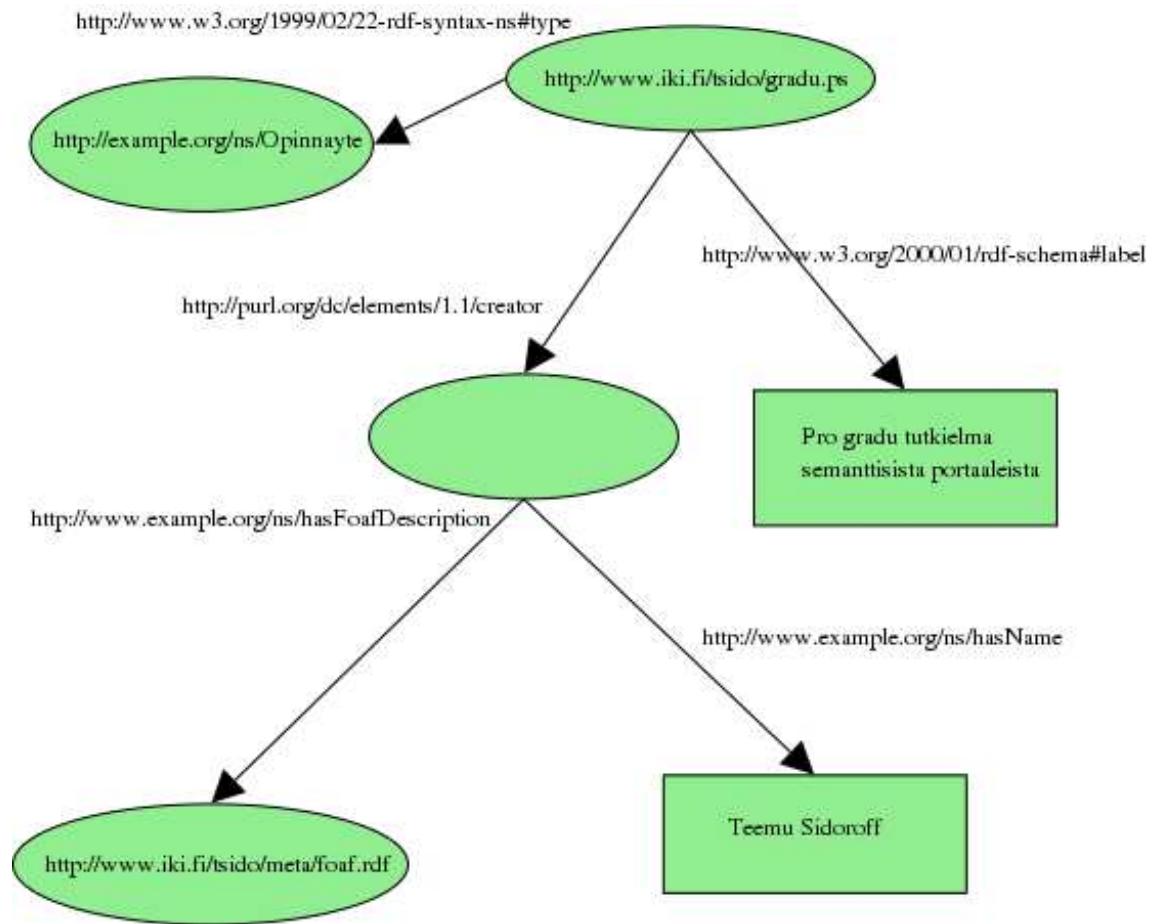
⁵RDF- ja RDFS-kielistä käytetty yhteisnimitys. Ks. 2.1 ja 2.2.

fioi entiteetin verkon resurssina.

Resurssi voi olla mikä tahansa abstrakti tai konkreettinen asia, kuten henkilö, tapahtuma tai vaikkapa joukko useampia entiteettejä. RDF:n avulla näistä resursseista voidaan ilmaista metatietoa, kuten web-sivun kirjoittajan tai julkaisijan tiedot tai vaikkapa johonkin tapahtumaan liittyviä tietoja, kuten tapahtumapaikka tai -aika.

RDF:n perustana on yksinkertainen konsepti, subjekti-predikaatti-objekti -kolmikko [RDF04a]. Kuten yksinkertainen suora suomenkielinen lause, RDF-kielinen lause koostuu kolmikosta, jossa on subjekti, predikaatti ja objekti. Kolmikko kuvaa subjektina olevaan resurssiin liittyvän ominaisuuden ja ominaisuuden arvon. Subjektina on aina jokin URI-viittauksena ilmaistu resurssi. Predikaatti ilmaisee resurssin ominaisuuden ja objekti ominaisuuden arvon. Myös predikaatit ovat URI-viittauksia. Näin ollen ne voivat siis esiintyä myös toisen kolmikon subjektina tai objektina. Tämä mahdollistaa tiedon esittämisen ominaisuuksista ja ominaisuuksille itselleen voidaan näin määrittää omia 'metaominaisuuksia'. Predikaatin, tai ominaisuuden, arvona voi olla jokin toinen resurssi, literaali eli merkkijono, tai nimetön solmu. Jokainen nimetön solmu on yksilöllinen ja se voi käyttäytyä kuten tavallinen resurssi, eli se voi myös esiintyä toisen RDF-lauseen subjektina.

Yksittäinen kolmikko ilmaisee tosiasian resurssista: se kuvaa resurssin jonkin ominaisuuden ja ominaisuuden arvon. Ominaisuus voi riippua pelkästään resurssista itsestään tai se voi kuvata resurssin suhdetta toiseen resurssiin. On siis ilmeistä, että kolmikkoja kertyy yksinkertaisiakin asioita kuvattaessa helposti useita, ja että kolmikoista muodostuu suunnattu verkko. Kuvassa 1 on esitettyä esimerkki RDF-muotoisesta verkosta, joka kuvaa tähän tutkielmaan liittyviä tietoja. Verkossa on mukana nimetön solmu, joka on resurssiin <http://www.iki.fi/tsido/gradu.ps> liittyvä objekti. Tällä nimettömällä solmulla on puolestaan kaksi ominaisuutta, <http://www.example.org/ns/hasFoafDescription>, jonka arvona on resurssi <http://www.iki.fi/tsido/meta/foaf.rdf> ja



Kuva 1: Esimerkki RDF-verkosta, jossa on mukana sekä literaaleja, resursseja että nimettömiä solmuja.

Name, jonka arvona on literaali "Teemu Sidoroff".

Yleisin muoto RDF:n esittämiseen on RDF/XML-syntaksi, jossa nimensä mukaisesti RDF esitetään XML-kuvailukielellä. RDF-muotoista tietämystä voidaan kuitenkin esittää useilla eri syntakseiltaan eroavilla kielillä. RDF/XML-syntaksin lisäksi tällaisia kieliä ovat mm. Notation 3 (N3) ja TRiX [Bol04]. Osa näistä kielistä pohjautuu XML-kieleen, mutta on kuitenkin syytä huomata, ettei RDF ole sidottu XML-esitysmuotoon. Voidaan jopa sanoa, ettei yleisin RDF-kielen esittämisessä käytetty syntaksi, RDF/XML, ole RDF-muotoiselle tietämykselle ihmisen kannalta selkein valinta, sillä RDF-tietämys on verkkomaista, kun taas XML-dokumentit

muodostavat puun. Tämä johtaa siihen, että RDF/XML-syntaksilla esitetty RDF voi on ihmiselle usein vaikeaselkoista. RDF kuitenkin on tarkoitettu erityisesti koneiden hyödynnettäväksi, ja erilaisia XML-kieltä tukevia työkaluja voidaan hyödyntää myös RDF-työkaluja luotaessa ja käytettäessä.

RDF tarjoaa tietämyksen esittämistavan resursseista, joka on helposti laajennettavissa. Jotta erilaista tietämystä voidaan hyödyntää laajamittaisesti, täytyy määrittellä myös tiettyjä yhteisiä ominaisuuksia ja sanastoja.

2.2 RDFS

Jotta RDF-kuvauksia voidaan käyttää eri osapuolten välillä, on kuvauksissa käytetyn termistön oltava tahojen välillä selvillä. Näitä termistöjä voidaan määritellä käyttämällä RDFS-kieltä (*RDF Schema*) [RDF04c]. RDFS määrittelee joukon luokkia ja ominaisuuksia, joita voidaan käyttää kuvaamaan RDF-kuvauksissa esiintyviä resursseja. RDFS on RDF-kielen semanttinen laajennos, joka on itsekin esitetty RDF-muodossa.

RDFS-kielillä voidaan esittää luokkarakenne, jossa voidaan määrittellä luokkien muodostama hierarkia määrittelemällä luokille aliluokkia. Luokkiin voidaan liittää kuuluvaksi ominaisuuksia ja ominaisuuksille voidaan määrittellä sallittu lähtöjoukko (*domain*) ja arvojoukko (*range*). Lähtö- ja arvojoukkoina voi olla tiettyyn luokkaan kuuluvat resurssit, tai arvojoukkojen kohdalla myös literaalit. Myös ominaisuuksille voidaan määrittellä aliominaisuuksia. Näin esimerkiksi ominaisuus **onVeli** voidaan määrittellä ominaisuuden **onSukua** aliominaisuudeksi. Tällöin resurssit, jotka toteuttavat **onVeli** relaation, toteuttavat automaattisesti myös **onSukua** relaation.

RDFS-kielillä esitetystä luokkarakenteesta voidaan nähdä analogia olio-ohjelmoinnin maailmaan, jossa luokkakuvaukset ovat pohjapiirustuksia ja luokkien ilmentymät, eli instanssit, ovat toteutuksia luokista. Samaan tapaan RDFS-kielillä voidaan

määritellä luokkia, joiden jäseniä, eli instansseja, resurssit voivat olla.

Koska RDFS-kieli tarjoaa mahdollisuuden kuvata resurssien luokkia ja luokkien välistä ylä- ja aliluokkasuhteita sekä luokkiin liittyviä ominaisuuksia, voidaan sitä käyttää ontologioiden kuvaamiseen. RDFS ei ole kuitenkaan ilmaisuvoimaltaan yhtä voimakas kuin varsinaiset ontologiakielit, kuten OWL, jota käsitellään seuraavaksi.

2.3 Ontologiat ja ontologiakieli OWL

Tekoälytutkimuksessa ontologialla tarkoitetaan jonkin aihealueen käsitteiden formaalia kuvausta [Gru93]. Ontologia mallintaa tiettyä aihealuetta määrittelemällä aihealueen käsitteistön ja käsitteiden väliset relaatiot. Ontologiat toimivat näin tukirankana aihealueen tiedon käsittelyssä aihealueesta kommunikoidessa.

Asiasanasto, joka listaa luetteloinnissa käytetyn kontrolloidun sanaston ja määrittelee sanaston käsitteiden välisiä suhteita (yleensä ylä- ja alakäsitesuhteet) [Fos97], voidaan nähdä ontologiana. Ontologia voi kuitenkin olla myös monimutkaisempi kokonaisuus, joka tarjoaa mahdollisuuden kuvata luokkiin liittyviä pakollisia ominaisuuksia ja luokkien välisiä sisältyvyysuhteita [McG03].

OWL (*Web Ontology Language*) [OWL04a] on W3C:n suositus ontologiakiielestä, joka on ilmaisuvoimaltaan voimakkaampi kuin RDFS-kieli. OWL tarjoaa ilmaisuvoimansa ansiosta mahdollisuuden esittää monipuolisia ontologioita, joilla voidaan esittää enemmän kuin pelkkiä taksonomisia rakenteita.

Kuten RDFS, OWL on esitetty RDF-muodossa ja se laajentaa RDFS-kielen esiteltäviä mahdollisuuksia luokkien ja ominaisuuksien kuvailuun. Myös RDFS-kielillä voidaan kuvata ontologioita, mutta sen ilmaisuvoimassa on kuitenkin rajoituksia. RDFS-kielen avulla ei voi esimerkiksi ilmaista ominaisuuksien transitiivisuutta tai niiden kardinaalisuusasteita [AvH03]. OWL tarjoaa mahdollisuuden tällaiseen rikkaampaan ominaisuuksien ilmaisemiseen, mikä RDFS-kielestä puuttuu. Kardinaali-

suusasteen ja ominaisuuksien transitiivisuuden lisäksi OWL-kielen avulla voi myös määritellä ominaisuuksien käänteisominaisuuksia.

OWL tarjoaa myös luokkien kuvailuun RDFS-kieltä rikkaamman sanaston. OWL-kielessä luokkia voidaan määritellä muodostamalla yhdisteitä tai leikkauksia toisista luokista tai määrittelemällä luokkaan kuuluviksi sellaisia yksilöitä, joilla on tiettyjä ominaisuuksia [HPSvH03]. Esimerkiksi luokkaan `TKTL-Opiskelija` kuuluvaksi voidaan määritellä ne luokkaan `Opiskelija` kuuluvat yksilöt, joilla on ominaisuuden `opiskeleePääaineena` arvona `Tietojenkäsittelytiede`.

Tällaisten ominaisuuksiensa pohjalta määriteltyjen luokkien johdosta OWL-kielessä se, mihin luokkaan jokin tietty yksilö kuuluu, voidaan päätellä koneellisesti. Kun jokin resurssi on annotoitu OWL-kielisen ontologian mukaisesti, voi semanttisessa webissä toimiva agentti käsitellä resurssia hyödyntämällä ontologioiden tarjoamaa tietämystä resurssin luokasta. Näin esimerkiksi ontologioita hyödyntävä opiskelijaportaalijärjestelmä, joka on käyttäjän rekisteröitymisvastausten perusteella päätellyt tämän olevan luokan `TKTL-Opiskelija` yksilö, voi profiloida portaalijärjestelmän tarjoamaan käyttöliittymässä Matemaattisluonnontieteellisen tiedekunnan opiskelijoille suunnattua käyttöliittymää.

3 Semanttiset portaalit

Tässä luvussa käsitellään portaalikäsitettä ja tarkastellaan sen kehittymistä. Portaaaleiksi miellettyjen web-sivustojen joukko on WWW:n kasvaessa laajentunut sisältämään laajojen yleisportaalien lisäksi myös pienempiä hakemistoja ja yhteisöjen sivuja. Yleisportaalien rinnalle portaaaleiksi on alettu mieltämään yksittäisten aihealueiden linkki- ja tietokokoelmia ja yhteisöjen sivustoja. Luvussa tarkastellaan lisäksi, mitä uutta semanttisen webin teknologiat tuovat portaaaleihin ja mitä semanttiset portaalit ovat. Lopuksi luodaan katsaus olemassa oleviin esimerkkiportaaaleihin, jotka voidaan luokitella semanttisiksi portaaaleiksi.

3.1 Portaalikäsitteen kehitys: yleisportaalista semanttiseen portaaliin

Web-portaali-termi tuli suuren yleisön tietoisuuteen 1990-luvun loppupuolella, kun WWW alkoi kasvaa räjähdysmäisesti. WWW:n kaupallistumisen myötä suurille ihmismassoille kohdistetut yritysten ylläpitämät portaalit on tarkoitettu erityisesti aloitussivuuksi, jotka kokoavat käyttäjälle näkymän verkon tarjontaan [BL99]. Ne kokoavat portaalisivuston alle runsaasti erilaisia hyöty- ja viihdepalveluita säätiedoista sähköpostiin, horoskoopeista uutisiin sekä keskustelupalstoista pörssikursseihin. Usein näiden pääosin mainosrahoitteisten portaalien yhteydessä esiintyy myös verkkokauppoja ja muita portaalin ylläpitäjän tai sen yrityskumppanin tarjoamia kaupallisia palveluita. Tämän tyyppisiä portaaaleja webissä on edelleen runsaasti. Esimerkkejä näistä suurista yleisportaaaleista ovat Excite⁶ ja suomalainen MTV3 Internet⁷. Näistä portaaaleista voidaan käyttää nimitystä yleisportaalit.

2000-luvulla suuren yleisön WWW:hen kohdistaman alkuinnostuksen jälkeen,

⁶<http://www.excite.com>

⁷<http://www.mtv3.fi>

WWW:n arkistumisen myötä ja WWW:n sivutarjonnan kasvaessa portaalikäsité on laajentunut sisältämään näiden yleisportaalien lisäksi myös pienempiä, erilaisille käyttäjäryhmille tarkemmin kohdistettuja sivustoja. Eräs portaalin määritelmä luonnehtii portaaaleja web-sivustoiksi, jotka tarjoavat johonkin aihealueeseen liittyvää informaatiosisältöä [OWL04b]. Määritelmän mukaan portaalit tarjoavat lisäksi kiinnostuneille mahdollisuuden vastaanottaa aihealueeseen liittyviä uutisia, kommunikoida muiden aiheesta kiinnostuneiden kanssa ja löytää linkkejä verkon muihin aihealuetta käsitteleviin resursseihin.

Edellinen määritelmä on kuitenkin liian rajoittava, eikä se täysin kuvasta todellisuutta siitä, minkälaisia sivustoja nykyisin nimitetään portaaaleiksi. Portaaaleiksi kutsutaan usein myös sivustoja, jotka eivät tarjoa kaikkia edellä esitetyistä palveluista, vaan keskittyvät esimerkiksi tarjoamaan pelkästään aihealueen linkkejä tai tarjoamaan puitteet aihealueen yhteisön väliselle kommunikoinnille. Yhteistä näille portaaaleille on kuitenkin määritelmän alkuosa eli se, että portaalit tarjoavat yksittäisen aihealueen informaatiosisältöä.

Seuraavassa keskitytään edellä esitettyjen suurten yleisportaalien sijaan erityisesti tiedonhaussa hyödyksi käytettäviin portaaaleihin, joiden kohderyhmänä on tarkemmin rajattu käyttäjäjoukko, kuten yrityksen työntekijät tai esimerkiksi kunnallispalveluita etsivät henkilöt.

Portaali voi toimia lähtökohtana tiedonhauulle tarjoamalla kokoelman linkkejä tiettyä aihealuetta käsitteleville sivuille tai koota portaalisivuston alaisuuteen aihealueen tietämystä. Tällaisia portaaaleja kutsutaan informaatioportaaaleiksi [RSC04]. Informaatioportaalit helpottavat tiedonhakua WWW:stä toimimalla tiedon tiivistäjinä ja yhteenkerääjinä. Informaatioportaaaleja tarkastellaan lähemmin aliluvussa 3.3.

Informaatioportaaaleja leikkaava toinen portaalityyppi on yhteisöportaali. Yhteisöportaalit ovat sivustoja, joiden sisällöstä ja päivittämisestä vastaa portaalialueen käyt-

tävä yhteisö itse. Portaalin avulla yhteisön jäsen voi julkaista uutta, toisia yhteisön jäseniä kiinnostavaa tietoa ja etsiä portaalista muiden julkaisemaa aiheen tietoa. Yhteisöportaaleja käsitellään aliluvussa 3.2.

O’Murchu et al. esittävät jaottelun, jossa esitetään neljä erilaista portaalityyppiä [OBD04]. Portaalit on jaoteltu julkaisijaorganisaatioittain:

- yritysportaalit (*enterprise portals*) ovat yritysten sisäiseen ja asiakkaiden käyttöön tarkoitettuja portaaleja,
- valtionhallinnon portaalit (*government portals*) tarjoavat julkis palveluita ja tietoa kansalaisille,
- yhteisöportaalit (*community portals*) ja semanttiset yhteisöportaalit (*semantic community portals*) toimivat yhteisön kommunikointialustana ja tiedonvälityskanavana.

Tämä jaottelu on luonteeltaan erilainen kuin edellä esitetty kolmijako yleis-, informaatio- ja yhteisöportaaleihin, jossa portaalit on jaoteltu julkaisijaorganisaation sijaan portaalin käyttötarkoituksen mukaisesti. O’Murchun et al. esittämät portaalit voidaan kuitenkin kuvata edellä esittämääni kolmijakoon. Yritysportaalit voidaan nähdä joko informaatioportaaleina tai yleisportaaleina. Valtionhallinnon portaalit ovat yleensä luonteeltaan informaatioportaaleja. Yhteisöportaalit ja semanttiset yhteisöportaalit ovat nimensä mukaisesti yhteisöportaaleja.

Minkä tyyppinen portaalitahansa voidaan toteuttaa myös semanttisena portaalina. Semanttisella portaalilla tarkoitetaan portaalitahansa, joka hyödyntää semanttisen webin teknologioita ja standardeja. Semanttisessa portaalissa tietosisältö on varastoitu portaaliontologian mukaisesti RDF-muodossa ja portaalitahansan tietosisältö on resursseja, joista on esitetty portaaliontologian mukaista metatietämystä. Portaaliontologia voi koostua useasta erillisestä ontologiasta, joiden avulla portaalitahansan tietosisältöä voidaan

kuvata erilaisista näkökulmista. Ontologinen pohja mahdollistaa tiedon esittämisen loppukäyttäjälle useilla eri tavoilla organisoituna sekä heterogeenisen materiaalin lisäämisen portaalin tietovarastoon.

Kun tieto on esitetty standardissa muodossa, semanttisen webin tapauksessa RDF-muodossa, mahdollistaa se lisäksi tiedon helpon uudelleenkäyttämisen. Avoimen RDF-standardin avulla uudelleenkäyttäminen ei rajoitu ainoastaan organisaation sisälle, vaan myös muut tahot voivat hyödyntää RDF-muodossa julkaistua dataa omissa sovelluksissaan.

3.2 Yhteisöportaalit

Verkossa löytyviä yhteisöjä on useanlaisia, mutta tavallisesti yhteisöä leimaa jokin niiden jäseniä yhdistävä tekijä, jota myös yhteisön yhteisöportaali käsittelee. Yhteisön voi muodostaa esimerkiksi organisaation työntekijät, tai työntekijät ja organisaation sidosryhmät. Tällaisen yhteisön portaali on tavallisesti organisaation intranetissä toimiva yhteistyösivusto. Yhteisö voi olla myös hajautettu. Esimerkki hajautetusta yhteisöstä on jonkin aihealueen tutkijoista koostuva ryhmittymä, joka kerää tuottamiaan tutkimustuloksia yhteisön saataville. Lisäksi yhteisö voi olla virtuaalinen, esimerkiksi jonkin verkkopelin yhteyteen kasvanut ryhmittymä, joka ylläpitää omaa, kyseiseen peliin liittyvää yhteisösivuaan.

Yhteisöportaalilla tarkoitetaan sivustoa, jonka sisältö on suunnattu jonkin yhteisön käyttöön. Yhteisöportaalit tarjoavat kanavan yhteisölle julkaista sitä itseään kiinnostavaa tietämystä [So00]. Yhteisöportaali voi tarjota yhteisön jäsenille samoja asioita kuin yleisportaalitkin, mutta niiden tarjoajana on keskitetyn toimituskunnan sijaan yhteisö itse. Luonnollisesti yhteisöportaalin sisältö on kohdistunut juuri yhteisön tarpeisiin, toisin kuin yleisportaaaleissa, jotka pyrkivät tavoittamaan useita käyttäjätyyppejä.

Pelkkä yhteisön keskustelusivusto ei kuitenkaan muodosta portaalia. Yhteisöportaalien lisämäärittelyksi voitaisiinkin sanoa, että niissä julkaistava tietosisältö on monimuotoista ja voi koostua erityyppisestä tietämyksestä. Lisäksi yhteisöportaali kokoaa yhteisön käytettäväksi erityyppisiä palveluita. Näitä palveluita voivat olla juuri keskustelupalstat ja uutissyötteen. Esimerkin yhteisöportaalista tarjoaa OntoWeb, jota käsitellään osiossa 3.4.2.

3.3 Informaatioportaalit

Yhteisöportaali ja informaatioportaali eivät ole keskenään erillisiä käsitteitä. Myös informaatioportaalien sisältö voi olla portaalia käyttävän yhteisön ylläpitämää tai tuottamaa, jolloin voidaan puhua yhteisöllisestä informaatioportaalista. Esimerkkinä tällaisesta informaatioportaalista käsitellään osiossa 3.4.3 SWED-portaalia.

Informaatioportaalien tunnusomainen piirre on, että ne kokoavat aihealueen tietämystä yhteen verkon pisteeseen ja tarjoavat lähtökohdan tiedonhauulle. Riippumatta siitä, onko informaatioportaaliin tallennettu linkkejä tai muuta tietämystä, voidaan sen sisällön nähdä koostuvan tietoalkioista, jotka on järjestetty jonkinlaisen rakenteen mukaan.

Informaatioportaalit tarjoavat vaihtoehdon tiedonhakuun sanahakuun perustuvilla hakukoneilla. Tämä pätee erityisesti tilanteessa, jossa käyttäjä on hakemassa spesifisen faktatiedon sijaan laajempaa tietokokonaisuutta, jota voi olla hankalaa purkea muutamiksi hakusanoiksi. Tietämystyöntekijöiden parissa tehdyn käyttäjätutkimuksen [SMS02] mukaan tämän tyyppinen laajempi tiedon etsiminen kattaa 35% näiden työntekijöiden verkon käytöstä.

Myös toisessa webin käyttäjien etsintätottumuksia kartoittaneessa tutkimuksessa [TAAK04] havaittiin, että käyttäjät usein löysivät etsimänsä tiedon suunnistamalla, suorittamalla joukon pieniä askelia sen sijaan, että olisivat käyttäneet suoraa

sanahakua silloinkin, kun he tiesivät tarkasti, mitä olivat etsimässä. Etuna tällaisessa suunnistamisessa on se, että käyttäjän ei tarvitse eksplisiittisesti esittää informaatiotarvettaan, vaan hän voi hyödyntää esimerkiksi tuntemaansa verkkosivun rakennetta lähes alitajuisesti siirtyessään hakemaansa informaatioon. Näiden tutkimusten tulosten perusteella informaatioportaaleilla on siis nykyisistä tehokkaista hakukoneista huolimatta tärkeä rooli tiedonhaussa.

Toinen etu informaatioportaaleilla tavallisiin sanahakuun perustuviin hakukoneisiin nähden on se, että portaalin tarjoama tieto on käyttäjälle esiprosessoitua ja luokiteltua. Mikäli portaaliin tallennetuista tietoalkioista on saatavilla riittävästi hyödyllistä metadataa, voi käyttäjä nopeasti arvioida tietoalkion hyödyllisyyden tarpeeseensa. Hyvin toteutetusta informaatioportaalista käyttäjä pystyy löytämään etsimänsä tiedon ilman, että joutuu tarkastamaan jokaisen saadun hakutuloksen sisällön, eli käytännössä lukemaan tai selaamaan tuloksena saadun web-sivun [LRPF04]. Tavallista hakukonetta käyttäessä käyttäjä voi sen sijaan joutua tarkastamaan useita tuloksena saatuja sivuja, koska hakukoneiden palauttamien hakutulosten tarkkuus on usein huono.

Informaatioportaalit puristavat suuren tietomäärän webin yhteen solmuun. Tämä tuo mukanaan haasteita niin käytettävyyden kuin tiedonhallinnankin osa-alueilla. Ongelmaksi nousee tiedon esittäminen siten, että tieto on käyttäjän löydettävissä ja toisaalta se, kuinka tietomassaa voidaan ylläpitää tehokkaasti. Näihin kysymyksiin liittyviä semanttisen webin teknologioiden mahdollistamia ratkaisuja käsitellään luvuissa 4 ja 5.

3.4 Esimerkkejä semanttisista portaaleista

Tässä osiossa esitellään kolme erilaista portaalia, jotka hyödyntävät semanttisen webin teknologioita eri tasoilla. Open Directory Project on webin kattava hakemisto

eli informaatioportaali, joka toimii katalogina laadukkaista sivustoista eri aihealueilta. OntoWeb on tutkijoiden käyttöön tutkimusmielessä luotu yhteisöportaali, joka hyödyntää ontologioita tietämyksen hallinnassa ja esittämisessä. Kolmantena esiteltävä SWED puolestaan on hajautettu hakemisto iso-britannialaisista ympäristöalan organisaatioista.

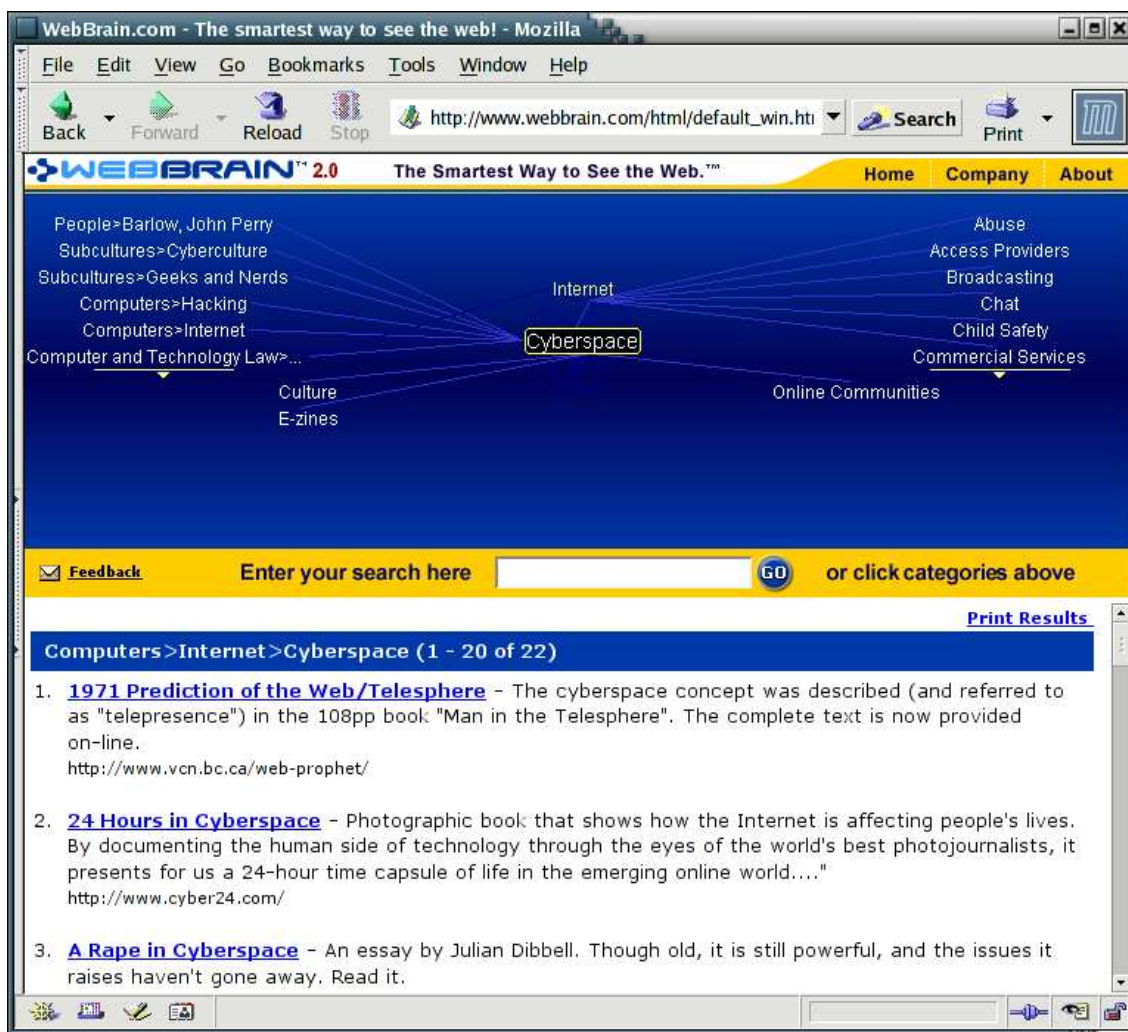
3.4.1 Open Directory Project

Open Directory Project (ODP)⁸ on informaatioportaali, joka toimii koko WWW:n kattavana hakemistona sivustoihin kaikilta aihealueilta. Portaaliin on kerätty linkkejä yli neljään miljoonaan web-sivuun tai -sivustoon. Linkit on jaoteltu n. 590 000 kategoriaan. Kategorioita ylläpitävät kymmenettuhannet vapaaehtoiset toimittajat, jotka lisäävät käyttäjien ehdottamia linkkejä ylläpitämiinsä kategorioihin. Portaalissa ei ole siis mitään keskitettyä hallintoa, vaan sen rakennetta ylläpitää valtava hajautettu yhteisö.

ODP-portaalin linkit on luokiteltu kategorioittain. Kategoriarakenne muodostaa ontologian, jossa jokaisella luokalla voi olla yläluokkia (joista yksi on primäärinen), rinnakkaisluokkia, aliluokkia ja eri kieliversioita. Kategoriat on esitetty käyttäjälle yhdellä web-sivulla, jossa on tarkasteltavissa kategorian alaiset linkit, alikategoriat ja linkitettyt alikategoriat. Linkitettyjä alikategorioita ovat ne alikategoriat, joiden primäärinen yläkategoria ei ole esitetty kategoriana. Jokaisen kategorian yhteydessä esitetään suluissa luku, joka ilmaisee kuinka monta linkkiä kategorian alipuuhun on tallennettu.

Esimerkiksi portaalin WWW:tä käsittelevällä kategoriolla **Computers > Internet > WWW** on englanninkielisen kategorian lisäksi seitsemän eri kieliversiota ja yksi rinnakkaiskategoria (**Computer > Internet On the Web**). Kategorian alaisuuteen ei

⁸<http://www.dmoz.org>



Kuva 2: ODP:n aineiston visualisoiva WebBrain.

ole luokiteltu yhtään linkkiä. Kategoria toimii kuitenkin toissijaisena yläluokkana usealle alikategorialle, mutta sillä ei ole omia suoria aliluokkia. Tämän takia kategorian linkkien määräksi ilmoitetaan hieman harhaanjohtavasti nolla.

Koska portaalin hakemistorakenne on verkkomainen (kategoriolla voi olla useita yläkategorioita), voi navigointi hakemistorakenteessa olla hankalaa. Käyttäjä voi eksyä laskeutuessaan hakemistorakenteessa alikategoriaan, jonka primäärinen yläkategoria ei ole sama, josta kategoriaan ollaan siirtymässä. Lisäksi kategorioiden yhteydessä mainitut linkkien määrät voivat verkkorakenteen takia olla harhaanjohtavia, kuten edellä esitettyssä esimerkissä kävi ilmi.

ODP-portaalista semanttisen portaalin tekee sen tapa kuvata tietosisältöään. Portaalien kategoriat ja linkkikuvaukset on esitetty RDF-muodossa. Koska aineisto on myös vapaasti saatavilla, on ODP:n aineistoa käytetty hyödyksi myös sadoissa muissa tiedonhakuportaaleissa. Esimerkiksi Google on hyödyntänyt ODP-portaalien aineistoa omassa hakemistossaan⁹.

Toinen mielenkiintoinen esimerkki ODP:n aineiston uudelleenkäytöstä on WebBrain-sovellus¹⁰, joka tarjoaa graafisen Java-applettipohjaisen käyttöliittymän hakemistorakenteen selaamiseen (ks. kuva 2). WebBrain-sovelluksessa hakemisto on visualisoitu verkkona, jonka keskuksena oleva solmu on se kategoria, jota käyttäjä kulloinkin tarkastelee. Kategorian linkit on esitetty verkon alapuolella olevassa linkkilistassa samaan tapaan kuin alkuperäisessä ODP-portaalissa. Keskussolmuna olevan kategorian lisäksi verkosta näytetään käyttäjälle kategoriaan suoraan liittyvät ylä- ja alakategoriat ryhmiteltyinä ryppäisiin. WebBrain-sovelluksen navigointiratkaisu on mielenkiintoinen vaihtoehto ODP-portaalien hieman sekavalle rakenteelle.

3.4.2 OntoWeb

OntoWeb-portaali¹¹ [OS01] on semanttisen webin tutkimusyhteisölle ja saman nimisen OntoWeb-projektin jäsenille tarkoitettu yhteisöportaali, jonka avulla yhteisö voi etsiä ja julkaista sitä kiinnostavaa sisältöä.

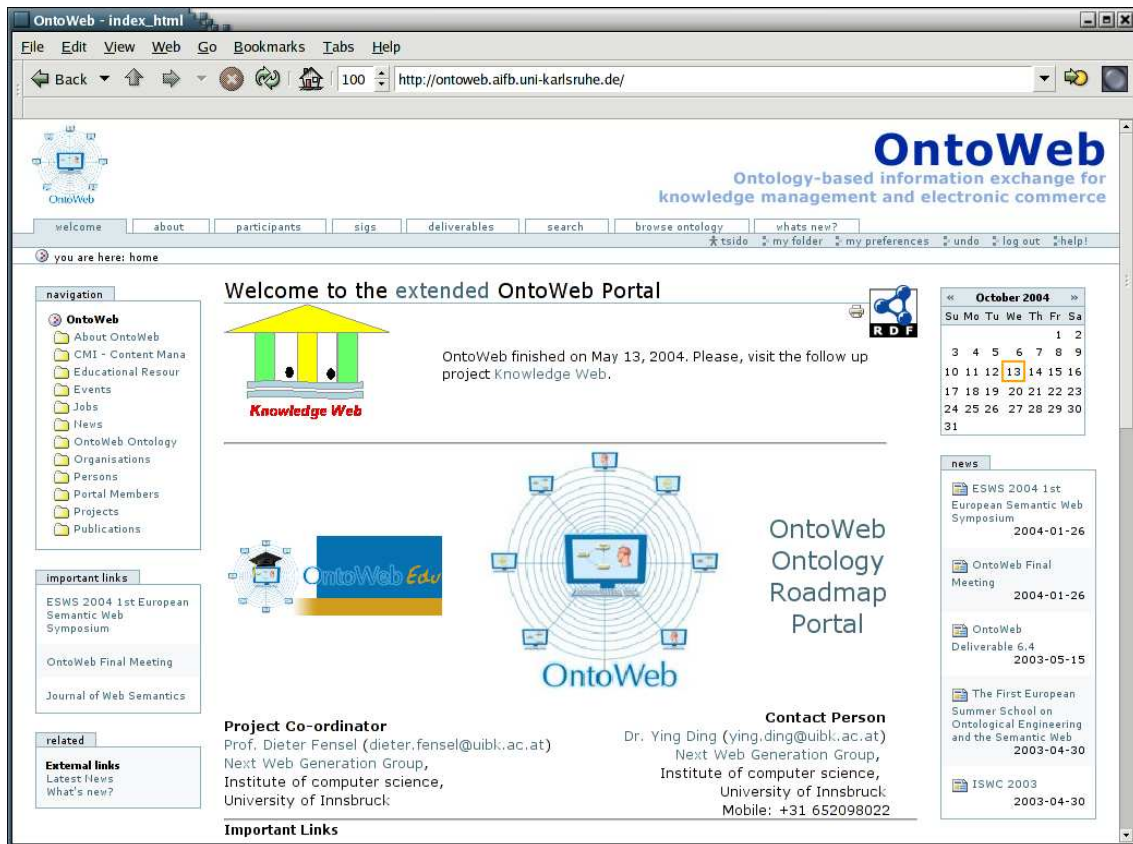
Vuosina 2001–2004 toimineen OntoWeb-projektin tavoitteena oli ontologia-perusteisen tietämyksen hallinta ja eri tiedonhallinta-alan toimijoiden välisen kommunikation tehostaminen, joten on luonnollista, että yhtenä projektin osatavoitteena oli myös luoda yhteisön käyttöön yhteisöportaali nimenomaan semanttisen webin teknologioita hyödyntäen.

⁹<http://directory.google.com>

¹⁰<http://www.webbrain.com>

¹¹<http://www.ontoweb.org>

OntoWeb-portaali tarjoaa yhteisön jäsenille erilaisia resursseja, kuten uutisia, työpaikkailmoituksia, projektiin liittyviä julkaisuja ja raportteja sekä julkaisujen bibliografisia tietoja. Portaali on hyvä esimerkki yhteisöportaalista, jossa portaalin informaatiota käyttävät yhteisön jäsenet itse myös tuottavat portaaliiin sisältöä. Sisältöä ei ole siis tuottamassa mikään keskitetty toimitus, vaan siitä vastaa osajoukko portaaliiin käyttäjistä, jolle on myönnetty oikeus julkaista muiden käyttäjien julkaitavaksi esittämiä dokumentteja.



Kuva 3: Semanttinen yhteisöportaali OntoWeb.

Portaali on käyttöliittymältään (ks. kuva 3) tavallisen yhteisöportaaliiin kaltainen. Sen alla piilee kuitenkin semanttisen webin teknologioita hyödyntävä ydin. Portaali on toteutettu semanttisiiin yhteisöportaaliiin SEAL-kehysrakenteen mukaisesti [M⁺03]. SEAL-kehyksessä portaaliiin tietämys voi koostua heterogeenisistä dokumenttiaineistosta kuten HTML-sivuuista, bibliografiatiedoista ja muista dokumen-

teista. Lisäksi portaalin käyttäjät voivat lisätä portaaliin uutta tietämystä. Uusien dokumenttien lisäämisessä käytetään hyväksi portaalin ontologian mukaisia lomakkeita, joiden käyttö takaa, että lisättävä aineisto on rakenteellisesti eheää.

Ontologia toimii portaalin tukirankana, jonka mukaisesti kaikki portaalissa esitettävä informaatio on annotoitu ja tallennettu portaalin RDF-tietämyskantaan. Tämä mahdollistaa heterogeenisen aineiston yhtenäisen esittämisen niin käyttäjälle kuin koneille, jotka voivat hyödyntää dokumenttien metadatasia. Koska portaalin tietämys on varastoitu RDF-muodossa, OntoWeb-portaalin aineistoa voidaan hyödyntää myös muissa portaaleissa, jotka käyttävät portaalin kanssa samaa ontologiaa tai sen osajoukkoa.

3.4.3 Semantic Web Environmental Directory (SWED)

SWED (Semantic Web Environmental Directory)¹² [RSC04, RSCS04] on yhteisöllinen semanttinen informaatioportaali, joka toimii iso-britannialaisten ympäristöalan organisaatioiden kontaktiluettelona. SWED on suunniteltu semanttisen portaalin prototyypiksi W3C:n SWAD-Europe¹³ semanttisen webin kehityshankkeen yhteydessä.

SWED on luonteeltaan hajautettu portaali, joka koostaa tietämystä eri lähteistä yhdelle sivustolle. Portaalissa esitettyjä organisaatioiden tietoja ei ole tallennettu itse portaaliin, vaan SWED toimii tietojen yhteenkerääjänä, jonka yhtenä komponenttina oleva RDF-etsijä (RDF crawler) käy noutamassa portaaliin rekisteröityjen organisaatioiden tiedot.

Portaalin tarjoama tietämys on esitetty useiden, varsin pienten, ontologioiden avulla. Käytetyt ontologiat kuvaavat organisaatioita ja niiden kiinnostus- ja toiminta-alueita sekä eri organisaatioiden välisiä yhteyksiä. Osa ontologioista on tesaurus-

¹²<http://www.swed.org.uk>

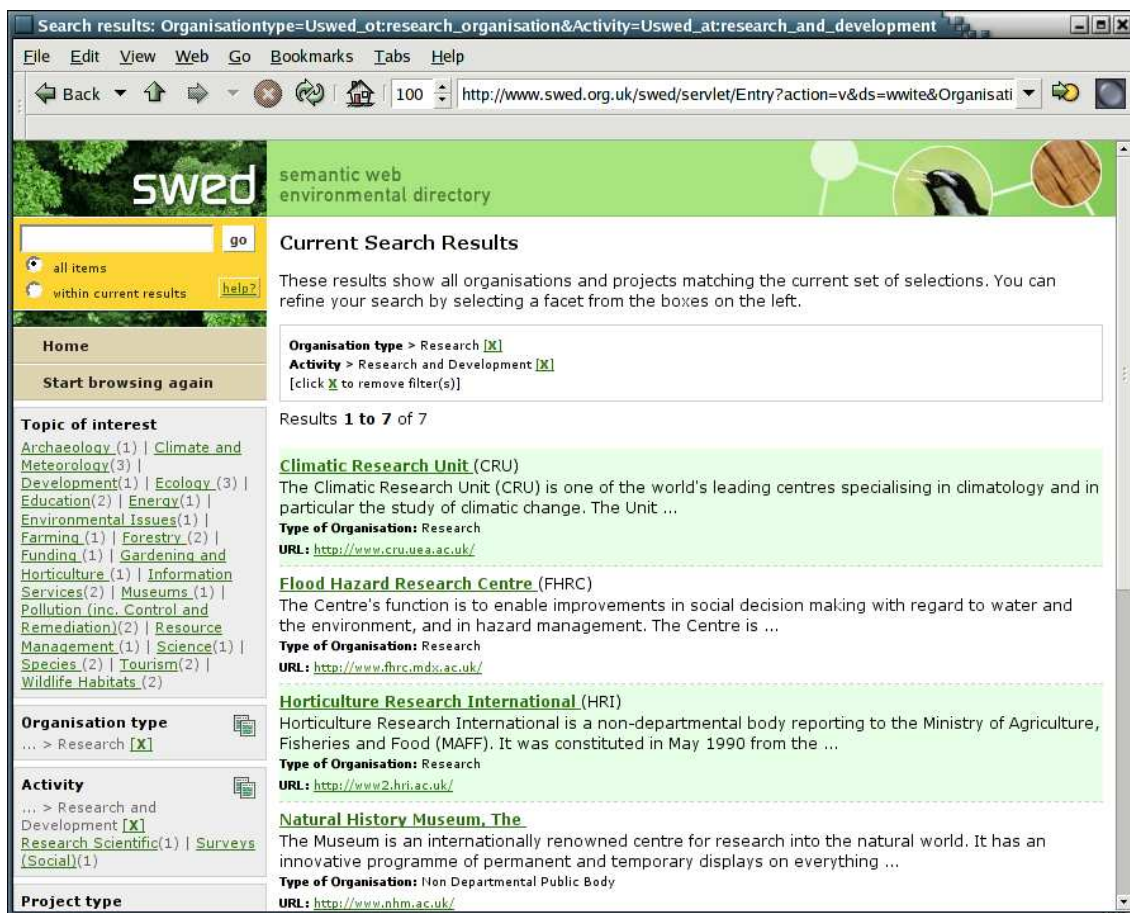
¹³<http://www.w3.org/2001/sw/Europe/>

tyyppisiä sanastoja, jotka määrittelevät käytettyjen termien hierarkiasuhteet, ja osa monimutkaisempia ontologioita, jotka kuvaavat organisaatioiden välisiä kytkentöjä ja organisaatioiden ominaisuuksia. Portaalin tesaurus-tyyppiset ontologiat on määritelty samassa SWAD-Europe -projektissa kehitetyllä RDFS-muodossa esitetyllä SKOS-sanastolla (*Simple Knowledge Organisation System*) [SKO], jossa itse SWED-portaaliakin on kehitetty. Portaalin monimuotoisemmat organisaatioiden suhteita kuvaavat ontologiat on kuvattu OWL-kielellä.

Jokaisen portaalissa olevan organisaation tiedot on tallennettuna portaalin ontologioiden mukaan annotoidussa RDF-muodossa ja tiedot ovat organisaation itse ylläpitämiä. Juuri tämä hajautettu, yhteisön jäsenille jätetty tietojen ylläpito tekee SWED-portaalista yhteisöllisen. SWED tarjoaa portaalin jäseneksi haluaville organisaatioille web-lomakkeen, jonka täyttämällä nämä voivat luoda itselleen tarvittavan RDF-kuvauksen. Kuvauksessa voidaan esittää mm. organisaation toimintalue, tyyppi ja suhteet muihin organisaatioihin tai organisaatioiden osiin. Organisaatio sijoittaa luodun RDF-kuvauksen haluamaansa osoitteeseen ja rekisteröi tämän SWED-portaaliin. Portaalin tarjoama lomake toimii siis helppokäyttöisenä annotaatioeditorina, joka piilottaa käyttäjältä taustalla piilevän vaikeaselkoisen RDF-kuvauksen. Portaalin ylläpitäjän osuus tietojen päivittämisestä on minimaalinen, käyttäjäksi rekisteröitymisen jälkeen organisaatioiden tietojen päivittämisestä vastaavat organisaatiot itse.

Hajautettu tiedonhallinta edesauttaa portaalissa esitettävien tietojen ajantasaisuutta. Koska tiedot ovat organisaatioidensa itsensä hallittavissa, voivat organisaatiot itse tehdä muutoksia tietoihinsa ilman mainittavaa viivettä tietojen päivittämiseen portaalissa. Organisaatio voi myös hyödyntää julkaisemaansa dataa muissa yhteyksissä, esimerkiksi muissa portaaleissa. Kun data on käytössä useissa yhteyksissä, toimii se myös organisaatiolle kannustimena pitää se ajantasaisena.

Portaalin käyttöliittymä on toteutettu näkymäselauksen (*faceted browsing*) [H⁺02]



Kuva 4: SWED-portaalin käyttöliittymä.

periaatteella (ks. aliluku 5.2). Portaalissa mukana olevia yhteisöjä ja järjestöjä voi etsiä selaamalla portaalia siinä käytettyjen ontologioiden muodostamien ulottuvuuk-
sien mukaan (ks. kuva 4). Näitä ulottuvuuksia ovat mm. organisaation tyyppi, toi-
mintakenttä ja maantieteellinen toiminta-alue. Näkymäselauksen periaatteen mukai-
sesti sivustoa voi selata rajaamalla samanaikaisesti attribuutteja eri ulottuvuuksista.
SWED demonstroi kiinnostavalla tavalla semanttisten portaalien tarjoamia mahdol-
lisuuksia. Siinä käytetty hajautettu tiedon varastointi mahdollistaa helpon tietä-
myksen uudelleen käyttämisen ja tietoalkioista esitetty rikas metadata mahdollistaa
portaalin käyttöliittymässä käytetyn näkymäselauksen hyödyntämisen.

4 Semanttisten portaalien tiedonhallinta

Tässä luvussa käsitellään semanttisiin portaaleihin liittyvää tiedonhallintaa. Tarkasteltavana on semanttisten portaalien ylläpito ja se, kuinka RDF-tietämys voidaan tallentaa ja kuinka kerran tallennettua tietämystä voidaan noutaa tietovarastosta erilaisten kyselykielten avulla.

Semanttisissa portaaleissa esitetty tietämys on aina jonkin ontologian mukaista. Valmiissa, tuotantokäytössä olevassa portaalissa ontologiat ovat kiinnitettyjä, ja muutoksia tapahtuu ainoastaan ontologioiden mukaisesti annotoituihin tietoalkioihin. Portaalin tietoalkioiden ylläpitämisestä vastaa tavallisesti portaalin sisällöntuottajat tai toimittajat.

Yleinen malli laajojen web-sivustojen tiedonhallinnassa on tallentaa portaalin sisältö tietokantaan ja generoida sivut tietokannassa olevan datan perusteella. Datan ylläpitoa voidaan hallita julkaisujärjestelmän avulla. Samaa tiedon julkaisumallia voidaan hyödyntää myös semanttisissa portaaleissa, jolloin portaalin sisällöntuottajan ei tarvitse tuntea pohjalla olevaa RDF-tietomallia.

Tiedonhallintaa sisällöntuottajan näkökulmasta käsitellään osiossa 4.1. Luvun loppuosissa käsitellään tiedonhallintaa dataa lähempänä olevalla tasolla. Osiossa 4.2 tarkastellaan erilaisia tapoja tallentaa RDF-muotoista dataa ja lopuksi osiossa 4.3 tarkastellaan, kuinka tallennetusta tietomassasta voidaan hakea kiinnostavaa tietoa.

4.1 Tietämyksen varastointi ja muokkaaminen

Portaalin sisällöntuottajille portaalissa esitettävän tiedon tietomallin ei tarvitse näkyä lainkaan, eikä sen näkyminen ole yleensä edes mielekäästä. Riittää, että sisällöntuottajille on tarjolla helppo tapa päivittää portaalin sisältöä. Portaalin ylläpidossa voidaan siis käyttää julkaisujärjestelmää (*web publishing system*) tai sisällönhallin-

tajärjestelmää (*content management system (CMS)*), joka tallentaa tietoalkiot portaalissa käytettyjen ontologioiden mukaisesti. Tällaisen sisällönhallintajärjestelmän keskeisenä osana voidaan käyttää annotaatioeditoria, joka tarjoaa käyttäjälleen yksinkertaisen tavan lisätä portaaliin uutta ja editoida jo olemassa olevaa tietämystä ilman, että käyttäjän tarvitsee seikkaperäisesti tuntea järjestelmän perustana olevia ontologioita. Annotaatioeditori voi ohjata käyttäjän työtä tarjoten tälle käyttöliittymän, jonka avulla annotoidaan tietoalkio ja joka tarkistaa, että tuotettu annotointi todella on portaalin ontologioiden mukainen.

Ontologiapohjaisen sisällönhallintajärjestelmän toteuttaminen ja tehokas käyttö edellyttää, että portaalin käyttämät ontologiat ovat stabiileja. Mikäli portaalin perustana käytettyjen ontologioiden rakenne muuttuu usein, on sisällönhallintajärjestelmän toteuttaminen huomattavasti vaativampaa. Lisäksi rakenteiden muuttaminen hankaloittaa myös järjestelmän käyttäjän työtä.

Yksinkertaisimmillaan annotaatioeditorina voidaan käyttää ontologiaeditoria, kuten Protégéa [N⁺01]. Ontologiaeditorit on kuitenkin tarkoitettu pääasiassa nimensä mukaisesti ontologioiden editointiin, eivät ainoastaan ontologioiden mukaisten instanssien (esim. portaalin tietoalkioiden) annotointiin ja soveltuvatkin paremmin portaalin ontologioiden kehitysvaiheessa hyödynnettäviksi työkaluiksi. Portaalin ylläpitoon Protégén toiminnallisuudesta ei tarvita kuin murto-osaa, ja sen käyttö voi tämän takia olla ylläpitovaiheessa hankalaa. Uusien tietoalkioiden annotointi esimerkiksi Protégélla voi olla työlästä, jos tietoalkioihin liittyy runsaasti olio-ominaisuuksia (*object properties*), joiden arvoiksi täytyy valita tai luoda Protégén käyttöliittymän avulla uusia instansseja. Protégén tehokas käyttö edellyttää myös RDF-tietomallin perusteiden tuntemista, jota ei portaalin sisällöntuottajalta voida edellyttää.

Käytettävyyden kannalta parempi ratkaisu onkin käyttää ontologiaeditorista erillistä annotaatioeditoria. Annotaatioeditori voi olla käyttöliittymältään hyvinkin yksinkertainen, esimerkiksi web-pohjainen työkalu, joka tarjoaa käyttäjälle joukon lo-

makkeita instanssien kuvailemiseen. Tällainen editori on käytössä mm. edellä käsitellyissä ODP- ja SWED-portaaleissa. Myös IWebS-projektissa [IWE] on kehitetty helppokäyttöinen web-pohjainen annotaatioeditori AnnE [Api05], jota voidaan käyttää sisällöntuotannon pohjana.

Kuvassa 5 on esitetty SWED-portaalissa käytetty ratkaisu annotointien tekemiseen. Editori on tarkoitettu ei-teknisten ihmisten käytettäväksi, eikä sen käyttäminen vaadi portaalissa hyödynnettyjen ontologioiden tuntemista. Editori koostuu joukosta HTML-lomakkeita, jotka portaalijärjestelmä generoi konfiguraatitiedostojen pohjalta. Lomakkeen avulla kuvataan portaaliiin lisättävän uuden tietoalkion ominaisuudet joko tekstikenttään, tai valitsemalla ominaisuuden arvoksi jokin lomakkeen tarjoamista vaihtoehdoista.

Käyttäjää voi editoida antamia tietoja myös myöhemmin, käyttäen samaa annotaatioeditoria. SWED-portaalin annotaatioeditori on pystytty pitämään helppokäyttöisenä, koska uuden tietoalkion luonti ei tuota luotavan tietoalkion lisäksi portaaliiin uusia instansseja, joille tulisi kuvata myös omat ominaisuudet. Tämä johtuu siitä, että portaalin tietoalkioiden metadata on pidetty varsin yksinkertaisena, ja on sidottu tiukasti portaalin ontologioihin. Tilanne kuitenkin monimutkaistuu, jos annotoitavaan tietoalkioon halutaan luoda ominaisuuden arvoksi toinen resurssi, jota ei löydy portaalin tietovarastosta. Tilanne voi syntyä esimerkiksi silloin, jos halutaan liittää tietoalkioon tekijäksi henkilö, joka puolestaan kuvataan resurssina. Tällöin editorin on tarjottava mahdollisuus molempien luotavien instanssien annotointiin.

4.2 RDF-tietovarastot ja tietämuskannat

Semanttisen portaalin tietovarastoksi on valittavissa useita erilaisia ratkaisuja. Yksinkertaisimmillaan tietämys voi olla tallennettuna tiedostojärjestelmässä tiedostoon tai tiedostoihin. Toinen vaihtoehto on hyödyntää tietokantoja. Monipuolisin ratkaisu

for Bristol)

e-mail **e-mail** This should be the main e-mail address of the project. **Please be aware** that (as with any web pages) it is possible people who produce spam e-mail may take the addresses from SWED to add to spam lists. We do not recommend that you add personal e-mail addresses to SWED information.

Description* This field provides space for you to describe your project

Description **Description** Enter a description of the project, up to 200 words. This text should tell users about your project in sufficient detail for them to understand its nature and goals e.g. purpose(s), role(s) or responsibilities, types of activity, involvement with other organisations/projects... Please **do not** use HTML or other formatting.

Classification As you will have seen from the SWED directory web site, the categories below are used to help users find or explore the information that they are looking for.

Click on the to expand the list. To add a category click on the term - you can add as many terms from each category as appropriate. To find out more about what the term means click on the [?] next to the term.

NOTE: in some cases the same term appears in different 'branches' of the lists (which are often often called trees) below. This is to make it easier to find those appropriate to your project - you only need to add a term once.

Type of Project

- Development (Economic and Social) [?]
- Economic/Social Development Project [?]
- Educational Project [?]
- Evaluation Project [?]
- Funding Project (Grant Giving) [?]
- Fundraising Project [?]
- Initiative Implementation Project [?]
- Networking Project [?]
- Pilot Project [?]
- Policy Development Project [?]
- Regeneration Project [?]
- Research and Development Project [?]
 - Development Project [?]
 - Product Development Project [?]
 - Service Development Project [?]
 - Research Project [?]

Type of Project Choose the term or terms that describe the kind of project that your project is - choose the terms that are as specific as possible

Kuva 5: SWED-portaalin tarjoama annotaatioeditori uusien tietoalkioiden lisäämiseen.

on hyödyntää tietämyskantaa, josta on portaalissa tallennettujen RDF-kolmikoiden lisäksi saatavilla myös ne RDF-kolmikot, joita ei ole eksplisiittisesti aineistossa ilmaistu, mutta jotka voidaan siitä päätellä. Eri ratkaisujen käyttökelpoisuus riippuu siitä, minkätyyppistä sisältöä portaaliin on odotettavissa ja siitä, kuinka muuttumatonta sisältö on.

RDF-pohjaisen tiedon tallentaminen suoraan palvelimen levyjärjestelmälle on yksinkertaisin ratkaisu tietovarastoksi. Ratkaisu on käyttökelpoinen mikäli portaalin

sisältö generoidaan tietyin väliajoin joukoksi staattisia sivuja tai koko portaalin tietämystä säilytetään palvelimen käyttömuistissa.

Tietokannat tarjoavat tiedon varastointiin joustavamman ja luotettavamman ratkaisun kuin pelkkään levyjärjestelmään perustuvat mallit. Mikäli data on luonteeltaan muuttuvaa ja portaalin sisällön halutaan olevan nopeasti tai automaattisesti päivitettävissä, on tietokantapohjainen ratkaisu mielekäs. RDF-mallin mukaista dataa voidaan tallentaa tavalliseen relaatiotietokantaan useiden erilaisten tietokantaskeemojen mukaan, joiden soveltuvuus riippuu tiedon käyttötavoista [Mel00]. Yksinkertaisimmillaan tietokantaskeemana voidaan käyttää yhtä taulua, jonka riveiksi tallennetaan jokainen tietämuskannan RDF-lause. Monimutkaisimmissa ratkaisuisa tietokantaskeema voidaan luoda käytetyn RDF-skeeman luokkarakenteen ja siinä määriteltyjen ominaisuuksien perusteella.

RDF-tietämyksen tallentaminen ja erityisesti noutaminen suoraan relaatiotietokannoista voi olla vaivalloista, koska tiedonhaussa on otettava huomioon RDFS-kielen määrittelemä semantiikka, kuten luokkahierarkiat. OWL-ontologioiden mukaisessa tietämyksessä transitiivisten ja käänteisten ominaisuuksien ja muiden OWL-kielen ominaisuuksien huomiointi tietokantakyselyissä vaikeuttaa edelleen toimivien kyselyjen muodostamista. Ratkaisuna on joko tallentaa tämä pääteltävissä oleva tieto suoraan tietokantaan, tai suorittaa päättely ajonaikaisesti.

Suoran tietokantoihin tiedon tallentamisen sijasta on järkevää käyttää erilaisten RDF-hallintajärjestelmien tietokantoihin tarjoamia rajapintoja. Tällaiset hallintajärjestelmät tarjoavat pääsyn RDF-tietämyksen tarkastelemiseen ja yleensä myös kyselykielen, jolla voidaan noutaa tietämuskannasta halutut ehdot täyttävät resurssit tai RDF-verkot. Yksi tällainen järjestelmä on Sesame [BKvH02]. Sesame on avoimen lähdekoodin RDF-tietokantajärjestelmä, joka tarjoaa käyttöön RDF-tietämuskannan ja kyselykielen tietämuskannan tietojen noutamiseen.

Sesamessa tieto voidaan tallentaa monenlaiseen alla olevaan tietovarastoon; RDF-tietämys voidaan esim. tallentaa tietokantaan, tiedostojärjestelmään tai säilyttää muistissa. Pääpaino on kuitenkin tietokantaratkaisuissa. Sesame muodostaa käytetystä tietokannanhallintajärjestelmästä riippuen erilaisen tietokantaskeeman, johon tallennetaan käytetty RDFS-ontologia ja sen mukainen RDF-tietämys. Sesamea tietovarastonaan käyttävälle sovellukselle tallennusratkaisulla ei ole suoranaista merkitystä, sillä Sesamen SAIL-rajapinta tarjoaa käytetystä tallennusratkaisusta riippumattoman yhtenäisen näkymän tallennettuun RDF-dataan.

4.3 RDF-kyselykielet

Edellä käsitelty Sesame tarjoaa mahdollisuuden noutaa tietovarastosta RDF-tietämystä RDF-kyselykielen avulla. Nimensä mukaisesti RDF-kyselykielten avulla tietovarastoon voi kohdistaa kyselyjä, joilla voidaan saada käsiteltäväksi haluttu osajoukko koko tietovaraston tietämyksestä.

RDF-tietämykseen voidaan kohdistaa kyselyitä tietomallin eri tasoille [BKvH02]. Yksinkertaisimmillaan tietomallia voidaan tarkastella joukkona yksittäisiä RDF-kolmikoita tai esimerkiksi RDF/XML-serialisoinnin tuottamalla syntaksitasolla XML-työkaluja hyödyntäen. Aidosti RDFS-kieltä tukevan kyselykielen on kuitenkin toimittava semanttisella tasolla ja tuettava RDFS-kielen tuomaa `rdfs:type` ja `rdfs:subClass` kaltaisten ominaisuuksien tuomaa semantiikkaa. RDF-kyselykielellä voidaan hakea tietämuskannasta helposti esimerkiksi kaikki resurssit, jotka ovat opinnäytteitä, ts. ne resurssit, joilla on ominaisuus (`rdf:type`), jonka arvona on `Opinnayte`. Lisäksi, kyselykielestä riippuen, voidaan hakea näistä resursseista vaikkapa vain ne, joilla on ruotsinkielinen otsikko, ts. ominaisuus `rdfs:label`, jonka arvona on ruotsinkielisellä kielikoodilla varustettu literaali.

Sesamessa käytetty RDF-kyselykieli on nimeltään *SeRQL* (*Sesame RDF Query Lan-*

guage) ja sitä kehitetään osana Sesamea. Kuten useat muutkin RDF-kyselykielet, muistuttaa SeRQL syntaksiltaan relaatiotietokannoissa käytettyä SQL-kieltä, mutta sen kyselyt kohdistuvat taulujen sijaan tietovaraston RDF-verkkoon.

Esityksiä ja toteutuksia RDF-kyselykielistä on esitetty useita ja ne eroavat niin ilmaisuvoimansa kuin syntaksinsa puolesta toisistaan [HBEV04]. Tällä hetkellä lähimpänä standardin asemaan pääsemistä on SPARQL [SPA04], joka on kehitteillä W3C:n suositukseksi RDF-kyselykielestä. Kuten SeRQL, perustuu myös SPARQL-kielissä tiedonhaku kyselyissä esitettyjen verkkohahmojen vertaamiseen RDF-tietovaraston muodostavaan verkkoon. Kehitteillä olevan SPARQL-kielen lisäksi muista kyselykielistä vahvoilla on SeRQL ja RQL. Molempien kielten vahvoina puolina on niiden ilmaisuvoima ja sisäinen RDFS-tuki [HBEV04].

On syytä huomata, etteivät nämä RDF-kyselykielet huomioi OWL-kielen rikkaampaa semantiikkaa. Mikäli käytettävä tietämys on ilmaistu OWL-kielillä, ei pelkkä SeRQL-kielinen kysely välttämättä palauta oikeaa vastausta. OWL-kielisen tietämyksen käsittely vaatii kuitenkin laajamittaisempaa päättelyä kuin RDFS-kielen käsittely, eikä edellä käsiteltyjä kyselykieliä ole triviaalia laajentaa kattamaan myös OWL:n tuomaa semantiikkaa. OWL-kielille on olemassa erilaisia kyselykieliä, kuten OWL-QL [FHH04], ja niiden toimintatapa on erilainen kuin pelkälle RDF:lle tarkoitettujen kielten.

5 Semanttiset portaalit tiedonhaussa

Käyttäjälle portaalit on hyödyllinen vain, jos hän pystyy hyödyntämään portaalitiedontarpeensa täyttämiseksi. Kuten web-ympäristössä aina, on semanttisten portaalien kohdalla muistettava, että käyttäjiä on erilaisia. Semanttisten portaalien kohdalla myös tietokoneohjelmistot on syytä muistaa erillisenä käyttäjäryhmänä. Ohjelmistoja eivät ole ainoastaan eri hakukoneiden verkkoryömijät (*web crawlers*) jotka tallentavat kokonaisia web-sivuja hakukoneiden indeksointiohjelmistojen käytettäväksi. Semanttisessa webissä toimivat myös älykkäät ohjelmistoagentit [Hen01], jotka hyödyntävät portaalin tarjoamaa RDF-muotoista tietämystä omilla tehtävissään. Ohjelmistojen lisäksi myös ihmiskäyttäjiä on erilaisia ja erilaiset päätelaitteet esittävät tiedon erilaisessa muodossa. Ohjelmistoagenteille portaalien tulisi tarjota suora pääsy RDF-muotoiseen tietoon, mikäli portaalin sisältämä tietämys halutaan tarjota kaikkien semanttisessa webissä toimivien agenttien käyttöön. Ihmiskäyttäjille tieto tulisi visualisoida siten, että se on ymmärrettävässä muodossa ja lisäksi mahdollisimman monien käyttäjien saavutettavissa.

Tässä luvussa käsitellään semanttisia portaalitiedon käyttäjän näkökulmasta. RDF-muotoisen tietämyksen visualisointia käsitellään osiossa 5.1. Osiossa 5.2 käsitellään näkymäseläus-paradigmaa, joka on esiintynyt käyttöliittymäratkaisuna useissa eri web-portaaleissa, myös semanttisissa portaaleissa. Osiossa 5.3 käsitellään suoraa tiedonhakua ja sitä, mitä hyötyä ontologiat tuovat tavalliseen avainsanahakuun verrattuna.

5.1 Tietämyksen visualisointi

Semanttisen portaalitiedon tietämys on tallennettuna RDF-muodossa. Portaalitiedon käyttäjälle tämä tietämys on kuitenkin esitettävä muodossa, joka on ymmärrettävä ja selkeä. Mikä tärkeämpää, tietämystä esittäessä on lisäksi päätettävä, mikä osa sii-

tä käyttäjälle ensinkään esitetään, koska portaalin tietämuskannassa voi tietämystä olla yksittäisistäkin resursseista runsaasti, eikä kaikki tieto välttämättä ole tarkoitettu loppukäyttäjälle.

Erilaisia graafisia visualisointiratkaisuja ontologioille ja RDF-muotoiselle datalle on tutkittu ja esitetty paljon [GC02]. Suurin osa ratkaisuista on kuitenkin tutkimusasteella olevia prototyyppisiä, eikä niiden soveltaminen web-maailman portaalien rakennuspalasiksi ole suoraviivaista. Lisäksi visualisoinnissa täytyy muistaa, että pelkkä tietämyksen esittäminen havainnollisessa muodossa ei yksin riitä vaan tietämyksessä on pystyttävä myös navigoimaan.

Visualisointiratkaisujen yhteydessä on muistettava myös saavutettavuus (*accessibility*) [WAI99]. Saavutettavuus on tärkeää suurille yleisöille tarkoitetuissa portaalissa, sillä käyttäjäkunta ja portaalin selaamiseen käytettävät päätelaitteet ovat heterogeenisiä. Saavutettavuuden merkitys kasvaa jatkuvasti, sillä erilaisten päätelaitteiden ja sitä myötä erityyppisten web-selainten kirjo on mobiililaitteiden lisääntymisen myötä kasvanut rajusti. Lisäksi väestön ikääntyminen johtaa väistämättä myös heikkonäköisten joukon kasvuun. Saavutettavuus ei liity siis ainoastaan marginaaliseen käyttäjäjoukkoon, vaan se koskettaa myös suurta yleisöä.

W3C:n suosituksen asemassa oleva SVG-kieli (*Scalable Vector Graphics*) [SVG03] voisi olla yksi tapa visualisoida mielivaltaisia RDF-verkkoja. SVG-kielillä voidaan esittää grafiikkaa ja interaktiivisia animaatioita web-selaimissa. Esimerkki SVG-kieltä hyödyntävästä sovelluksesta on Foafnaut¹⁴, joka visualisoi FOAF-ontologian [FOA04] mukaisia FOAF-dokumentteja ja dokumenttien välisiä yhteyksiä. FOAF (*Friend of a Friend*) on suosittu ihmisten yhteystietoja ja kontakteja kuvaamaan tarkoitettu OWL-kielillä toteutettu ontologia. Foafnautin avulla voidaan visualisoida näiden FOAF-dokumenttien kuvaamia ihmisten välisiä verkostoja. Nykyiset selaimet eivät kuitenkaan vielä tue SVG-kieltä laajalti, vaan sen esittäminen vaatii se-

¹⁴<http://www.foafnaut.org>

lainlaajennoksia. Tämä tekee SVG-kielestä kömpelön ratkaisun tiedon esittämiseen suurille yleisöille, mutta tilanne voi muuttua muutaman seuraavan vuoden aikana.

Portaalien yhteydessä luonnollisin tiedon esitysmuoto on saavutettavuuden kannalta perinteinen HTML, jota voidaan esittää kaikilla selaimilla useissa erilaisissa ympäristöissä. Mielivaltaista RDF-dataa voidaan esittää HTML-sivuina vaihtelevin tuloksin. BrownSauce [Ste03] on järjestelmä, joka muodostaa RDF-datasta HTML-muotoisen esityksen, jossa tarkasteltavana on kerrallaan yksittäinen resurssi. Järjestelmä esittää tarkasteltavana olevaan resurssiin liittyvät ominaisuudet ja niiden arvot avain-arvo pareina HTML-taulukossa. Järjestelmän avulla voidaan myös samoilla resurssien muodostamassa verkossa ja siirtyä tarkastelemaan mitä tahansa HTTP-yhteydellä saavutettavaa resurssia. Toinen RDF-dataa HTML-muodossa esittävä järjestelmä on RDF Model Browser¹⁵, joka esittää verkosta tai levyltä ladattavan RDF(S)-mallin laajennettavana puurakenteena, josta voidaan suoraan selata mallin resursseja ja niiden ominaisuuksia.

Koska saman merkityksen omaavaa RDF-dataa voidaan esittää usealla tavalla, on selkeän ja yleispätevän ihmiselle sopivan esityksen muodostaminen erittäin hankalaa. Portaalien kohdalla ei ole edes tarkoituksen mukaista esittää mielivaltaista RDF-verkkoa, vaan voidaan keskittyä esittämään portaalissa käytettyjen ontologioiden mukaista dataa. Lisäksi portaalissa ei ole yleensä tarpeen esittää resursseista kaikkea ontologiankaan mukaista tietämystä, vaan osa metatiedoista voi olla kiinnostavaa ainoastaan ylläpitäjälle. Tällaista metatietoa voi olla esimerkiksi resurssista tehdyn annotaation versionumero tai tiedot annotoinnin tekijästä.

Haystack-järjestelmä visualisoi mielivaltaista RDF-muotoista tietämystä ja esittää sitä jäsennetyssä muodossa käyttäjälle [QHK03]. Haystack-järjestelmä yhdistää tietämystä käyttäjän sähköpostikontakteista ja muista käyttäjän tietokoneella ja verkossa olevista lähteistä. Järjestelmän kehittäjät argumentoivat, että RDF-datan mie-

¹⁵<http://visualrdf.sourceforge.net>

lekkääseen esittämiseen tarvitaan ontologiaa, joka kuvaa kuinka jonkin alueen tietämystä tulisi esittää [QK04]. Haystackissa tätä visualisointiin käytettävää ontologiaa kutsutaan nimellä VOWL (*View Ontology Web Language*). VOWL on RDF-muotoinen kuvaus siitä, kuinka tietyn tyyppiset resurssit tulisi esittää ja mitä luokkiin liittyviä ominaisuuksia käyttäjälle esitetään. VOWL:n avulla voidaan myös esittää virtuaalisia ominaisuuksia, joita ei alkuperäisessä RDF-datassa ole määritelty, mutta jotka voidaan siitä päätellä.

Jonkinlainen säännöstö on tiedon visualisoinnissa aina tarpeen. Se, tarvitaanko formaalia visualisoinnin kuvailevaa ontologiaa on kuitenkin kyseenalaista. Vastaus riippuu visualisointijärjestelmässä käytetyistä toteutusratkaisuista. Jos datajoukko on yhtenäinen, voidaan selvittää yksinkertaisemmalla säännöstöllä. Tämä on mahdollista erityisesti informaatioportaaleissa, joiden sisältönä olevien tietoalkioiden eri tyyppitunnetaan. Visualisoinnin pohjana voidaan tällöin käyttää itse datajoukon kuvaavaa ontologiaa ja joukkoa sääntöjä, jotka kuvaavat mitä ominaisuuksia erityyppisistä tietoalkioista esitetään ja missä muodossa esittäminen tapahtuu.

W3C-konsortion, MIT-yliopiston ja Hewlett Packardin yhteishankkeen, Similen¹⁶, osana kehitetylle Longwell RDF-selaimelle määritellään visualisointisäännöt RDF-muodossa. Longwell tarjoaa käyttöliittymäksi näkymäpohjaisen käyttöliittymän (ks. aliluku 5.2) ja visualisointisäännöissä määritellään, mitkä datajoukon osat muodostavat minkäkin näkymän, ja kuinka yksittäiseen resurssiin liittyviä ominaisuuksia visualisoidaan.

Semanttisissa portaaleissa tieto on esitettävä käyttäjälle kuten normaaleilla web-sivuilla, HTML-muodossa. Portaalien sisältö on usein riittävän yhtenäistä, jotta muunnos RDF-esityksestä voidaan tehdä joukolla sääntöjä, jotka kuvaavat, minkä tyyppiset resurssit esitetään ja kuinka niiden ominaisuudet näytetään käyttäjille.

¹⁶<http://simile.mit.edu/longwell/>

5.2 Näkymäselaus

Suuret aineistot on tiedonhaun mahdollistamiseksi yleensä luokiteltu jonkin ulottuvuuden mukaan. Esimerkiksi useissa kirjastoissa kirjat jaotellaan hierarkkiseen Deweyn desimaalijärjestelmään perustuvan luokittelun mukaan aihealueisiin, jotka puolestaan jaetaan aliaihealueisiin. Näin esimerkiksi yksittäinen Platosta, antiikin filosofista, kertova kirja voidaan luokitella kuulumaan aihealueen filosofia aliaihealueeseen historia eli filosofian historiaan. Toisaalta kirja voitaisiin luokitella myös aihealueen historia aliaihealueeseen henkilöhistoria.

Digitaalisen informaation maailmassa tiedon sijainnilla ei ole kuitenkaan samanlaisia fyysisiä rajoituksia kuin kirjastoissa. Nyt edellinen Platoa käsittelevä kirja voidaan siis sijoittaa sekä filosofian historiaan että henkilöhistorian aliaihealueisiin. Tietoalkio voidaan luokitella useiden kategorioiden mukaan, ja sama tietoalkio voidaan löytää erilaisia polkuja pitkin. Esimerkiksi digitaalisessa kirjastossa kirjoja voitaisiin luokitella aihealueiden lisäksi julkaisuvuoden ja kirjailijan kansallisuuden mukaan ja hyödyntää tiedonhaussa näitä eri ulottuvuuksia samanaikaisesti.

Näkymäselauksessa aineisto on luokiteltu yhden ulottuvuuden sijaan useiden ortogonaalisten, toisistaan riippumattomien, ulottuvuuksien mukaan. Usean ulottuvuuden mukainen luokittelu mahdollistaa aineiston selaamisen eri näkökulmien perusteella ja antaa käyttäjälle useampia mahdollisuuksia löytää yksittäinen tietoalkio [H⁺02]. Edellisessä digitaalisessa kirjastoesimerkissä näkymäselaukseen perustuvan käyttöliittymän näkymät olisivat julkaisuvuosi, aihepiiri ja kirjailijan kansalaisuus. Näkymät muodostavat kategoriat voivat lisäksi olla hierarkkisia [Y⁺03]. Digitaalisen kirjastomme julkaisuvuodet voivat muodostaa hierarkkisen jaottelun, jossa ylimmällä tasolla vuodet esitetään vuosisadan tarkkuudella, toisella tasolla vuosikymmenittäin ja alimmalla tasolla jopa yksittäisen vuoden tarkkuudella.

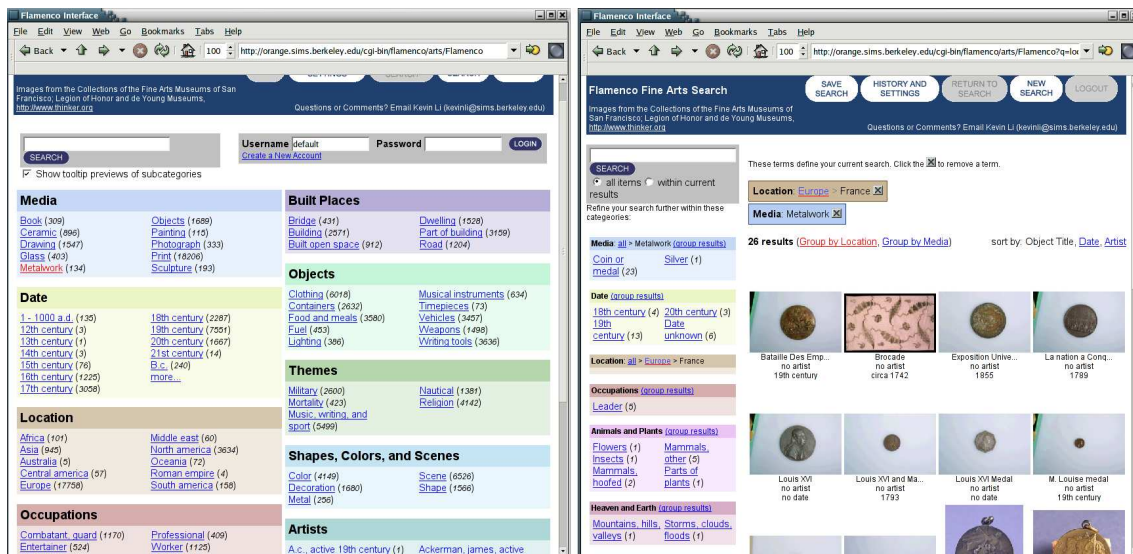
Näkymäselausta on hyödynnetty erilaisissa muodoissa useilla web-sivustoilla. Esi-

merkki näkymäseläusta hyödyntävästä web-portaalista on osiossa 3.4.3 esitelty SWED-portaali. Portaalin käyttäjä voi esimerkiksi valita tarkasteltavaksi ne tietoalkiot, tässä tapauksessa organisaatiot, jotka toimivat koko maailmanlaajuisesti ja joiden toimintakenttänä on eläinten suojeleminen. Järjestelmä muodostaa tällaisen kyselyn tulosityoukoksi leikkauksen kaikista portaalin tietoalkioista, jotka on annotoitu maantieteellisen ontologian mukaan toimimaan koko maailman laajuudella ja tietoalkioista, jotka liittyvät toiminta-ontologian eläintensuojelu-luokkaan.

Näkymäseläus vaatii aineiston, josta on saatavilla riittävästi metadataa [H⁺02]. SWED-portaalissa näkymäseläuksen toteuttamiseen on hyödynnetty ontologioiden mukaisesti annotoitua aineistoa. On kuitenkin syytä huomata, ettei näkymäseläus ole sidoksissa mihinkään tiettyyn tiedonesitysmalliin. Näkymäseläuksessa eri näkyminä esitettävät kategoriat on kuitenkin usein suoraviivaista muodostaa ontologioiden perusteella ja näin ontologioiden mukaisesti annotoidun aineiston tapauksessa näkymäseläus on helppoa toteuttaa.

SWED-portaalin näkymäseläuksen toteutus perustuu Berkeleyyn yliopiston Flamenco-projektin [FLA] työhön. Flamenco-projekti on soveltanut näkymäseläusta ja kehittänyt siihen perustuvan järjestelmän tiedonhaun apuvälineeksi suuria aineistoja sisältäville web-sivustoille. Järjestelmää kehitettäessä lähtökohtana on ollut erityisesti käytettävyys [H⁺02]. Seuraavassa tarkastellaan projektin esittämää ratkaisumallia suuren, yhtenäisen aineiston tiedonesittämisjärjestelmäksi.

Flamenco-projektin Lee et al. vertaavat suureen tietomassaan tutustumista kolmivaiheiseen shakkipeliin, johon kuuluu pelin avausvaihe, keskivaihe ja loppuvaihe [Y⁺03]. Erotuksena shakkipeliin analogiassa on se, että vaiheita ei tarvitse käydä läpi järjestyksessä, vaan niiden välillä voidaan siirtyä mielivaltaisesti. Avausvaiheessa järjestelmä tarjoaa käyttäjälle yleisnäkömän aineistosta, jossa on esitettyä ylimmän tason kategoriat jokaisesta järjestelmän eri näkymästä. Tässä vaiheessa käyttäjä saa yleiskuvan järjestelmän tietosisällöstä ja aihealueista, joiden kannalta aineistoa



Kuva 6: Flamencon käyttöliittymä selauksen aloitusvaiheessa (vasemmalla) ja keskivaiheessa (oikealla).

voi tarkastella (ks. kuva 6). Valitessaan yhden esitetyistä kategorioista lähempään tarkasteluun käyttäjä siirtyy pelin keskivaiheeseen.

Keskivaiheessa käyttäjälle esitetään ruudun oikealla puoliskolla hakutulos, eli kaikki valittuun kategoriaan kuuluvat tietoalkiot (ks. kuva 6). Ruudun vasemmalla laidalla esitetään eri näkymistä ne kategoriat, jotka liittämällä nykyiseen hakulausekkeeseen voidaan rajata hakutulosta, tuottamatta kuitenkaan täysin tyhjää tulosjoukkoa. Eri kategorioita valittaessa muodostetaan siis valittujen kategorioiden leikkaukseen kuuluvat tietoalkiot. Keskivaiheessa hakutulosta voi myös laajentaa poistamalla hakulausekkeesta esitettyjä kategorioita. Kun käyttäjä valitsee tarkasteltavaksi yhden tietoalkion, siirrytään käyttöliittymässä analogiassa esitettyyn loppupeliin. Loppupelissä esitetään yksittäinen tietoalkio ja siihen liittyvät metatiedot.

Flamenco-järjestelmän puutteeksi voidaan nähdä se, että yhdestä näkymästä voidaan valita ainoastaan yksi kategoria tulosjoukon muodostamiseen. Laajemman valinnanvapauden tarjoaisi mahdollisuus valita tulosjoukkoon useita eri kategorioita samasta näkymästä. Luonnollinen tapa olisi muodostaa yhdiste saman näkymän ka-

tegorioista ja sen jälkeen muodostaa eri näkymien valituista osajoukoista leikkaus. Tällainen ratkaisu on tarjolla esimerkiksi Relation Browser++-järjestelmässä [Z⁺04], joka tarjoaa näkymäselaukseen perustuvan käyttöliittymän suurten datajoukkojen tarkasteluun.

Näkymäselausta voidaan hyödyntää semanttisissa informaatioportaaleissa tiedonhaun apuvälineenä helposti, mikäli portaalin tietoalkiot muodostavat riittävän yhtenäisen kokonaisuuden. Näkymäselauksen toteuttamiseen tarvittava metadata saadaan tietoalkioiden rikkaasta metadatasta. Lisäksi tämä metadata on annotoitu ontologioiden mukaan, joiden rakenteesta voidaan muodostaa eri näkymät muodostavat kategoriat.

5.3 Ontologiaperustainen käsitehaku

Tässä osiossa tarkastellaan, kuinka RDFS- ja OWL-kielisten ontologioiden avulla sanahaku voidaan laajentaa käsitehauksi, jossa pelkkien avainsanojen lisäksi etsitään myös ontologian mukaisia käsitteitä ja niihin liittyviä tietoalkioita.

Portaalin käyttöliittymässä on hyvä tarjota käyttäjälle myös oikoreittejä tiedonhakuun. Tämä on suotavaa erityisesti, kun käyttäjällä on selkeä kuva siitä, mitä tietoa hän on hakemassa eikä hän halua navigoida kohteeseensa pitkien linkkiketjujen kautta. Yksinkertaisen käyttöliittymän aineistoon tarjoava sanahaku täydentää hyvin selaamista ja tarjoaa selaamista tehokkaamman tiedonhakatavan silloin, kun käyttäjän tiedontarve on tarkasti selvillä [Mar95].

Suoraviivainen tapa hyödyntää ontologioita tiedonhaussa on tarjota käyttäjälle mahdollisuus valita joukko ontologian käsitteitä ja palauttaa tuloksena ne tietoalkiot, jotka ovat jonkin valitun käsitteen ilmentymä, tai liittyvät jonkin ominaisuuden kautta johonkin valituista käsitteistä. Tampereen yliopistossa kehitetty CIRI-järjestelmä on tällaiseen selattavaan ontologiaan perustuva tiedonhakupohjainen järjestelmä [A⁺04]. Jär-

jestelmä tarjoaa puunäkymän ontologiaan, josta käyttäjä voi poimia käsitteitä, joiden perusteella järjestelmä muodostaa automaattisesti hakulausekkeen. Käyttäjä voi myös tehdä käsin omia lisäyksiä ja korjauksia hakulausekkeeseen, ennen kuin sitä täsmäytetään dokumenttiaineistoon.

Ontologia voidaan piilottaa käyttäjältä tarjoamalla tälle tavallisen sanahaun kaltainen yksinkertainen, pelkästä tekstikentästä koostuva käyttöliittymä. Erityisesti portaaliympäristössä tämä on perusteltua sillä portaalin muu käyttöliittymä tarjoaa selailuun perustuvan käyttöliittymän, esimerkiksi edellä esitellyn näkymäselauksen muodossa. Tällöin on turha rasittaa käyttäjää tarjoamalla tälle vielä erillinen sanasto, josta hän voi poimia hakulausekkeeseen termejä. Tällainen ratkaisu on käytössä MuseoSuomi-portaalissa¹⁷, joka kokoaa yhteen suomalaisten museoiden kokoelmia semanttiseen portaaliin.

MuseoSuomi-portaalissa käytössä oleva sanahaku kohdistuu portaalin tietoalkioiden lisäksi myös portaalin ontologioihin. Haun avulla voidaan etsiä jokin portaalin käsite ja siirtyä tätä kautta tarkastelemaan portaalissa käytössä olevan näkymäselauksen perustuvan käyttöliittymän avulla käsitteeseen liittyviä tietoalkioita [Saa04]. Käsitehaun avulla käyttäjä voi MuseoSuomessa myös kohdistaa hakunsa yksittäiseen homonymiin valitsemalla hakusanan tuottamista kategorioista sen, joka vastaa käyttäjän kiinnostuksen kohdetta.

Ontologiapohjaisessa käsitehaussa kyselyä voidaan laajentaa automaattisesti. Automaattisessa kyselyn laajentamisessa hakutulokseen voidaan tuoda instansseja, jotka eivät suoraan ole yhteydessä hakulausekkeessa määriteltyihin käsitteisiin, mutta jotka ovat yhteydessä esimerkiksi hakukäsitteiden alakäsitteisiin tai yläkäsitteisiin. Sovelluskohtaisesti laajentamisessa voidaan myös hyödyntää erilaisia assosiativisia suhteita, kuten osa-kokonaisuussuhdetta. Hakujen toteuttamisessa voidaan hyödyntää OWL-kielen tarjoamia transitiivisiä ominaisuuksia. Esimerkiksi haettaes-

¹⁷<http://museosuomi.cs.helsinki.fi>

sa sukupuu-tyyppisestä aineistosta henkilön jälkeläisiä kyselyä voidaan laajentaa automaattisesti koskemaan myös jälkeläisten jälkeläisiä. Myös FOAF-tyyppisestä aineistosta jonkin henkilön tuttavien haettaessa voidaan hakua laajentaa asteittain koskemaan myös tuttavien tuttavien seuraamalla kohteiden välisiä `foaf:knows`-ominaisuuksia¹⁸ solmusta solmuun.

¹⁸Ominaisuutta `foaf:knows` ei ole FOAF-ontologiassa määritelty transitiiviseksi, joten OWL-kielen ominaisuuksia tukevat työkalut eivät osaa tehdä laajennosta automaattisesti.

6 Suomi.fi-julkishallinnon portaali

Suomi.fi on valtiovarainministeriössä toimivan valtioneuvoston tietohallintoyksikön ylläpitämä julkishallinnon portaali, joka tarjoaa kansalaisille lähtökohdan eri virastojen ja julkishallinnon organisaatioiden sivuilta löytyvän informaation etsimiseen. Tässä luvussa tutkitaan semanttisen webin tuomia mahdollisuuksia portaalin kehittämisessä käytettävyyden ja ylläpitämisen kannalta.

Aluksi osiossa 6.1 tarkastellaan portaalin nykytilaa. Osiossa 6.2 siirrytään tutkimaan, mitä hyötyä semanttisen webin teknologioista portaalille on. Osiossa 6.3 esitellään demonstraatiojärjestelmä, semanttinen portaali, johon nykyisen Suomi.fi-portaalin sisältöä on siirretty. Demonstraatioportaalista ja sen jatkokehitysmahdollisuuksia arvioidaan lopuksi osiossa 6.4.

6.1 Portaalin nykyinen versio

Suomi.fi-portaali kerää julkishallinnon eri virastojen ja organisaatioiden sivuilla olevaa tietämystä yhteen pisteeseen. Portaali tarjoaa linkkejä eri organisaatioiden sivuilla olevaan tietoon luokiteltuina kolmeentoista eri aihealueeseen, jotka on ryhmitelty pääasiassa elämäntilanteiden mukaan. Aihealueita ovat esimerkiksi asuminen ja perhe, terveys ja ravinto, sekä ympäristö ja luonto. Aihealueet tarjoavat lyhyen aihealuetta kuvaavan ingressitekstin sekä joukon linkkejä. Esimerkki aihealueen yläsivusta on esitetty kuvassa 7. Jokainen aihealue on lisäksi pilkottu alikategorioihin, jotka sisältävät aihealueen johonkin osa-alueeseen rajattuja linkkejä. Portaalin linkkikokoelma on varsin laaja. Sen nykyisessä versiossa linkkejä on noin 3000, joista noin puolet sisältyvät suomenkieliseen versioon ja loput ruotsin- ja englanninkielisiin portaalisivuihin.

Rakenteeltaan portaali on puumainen. Aihealueet alikategorioineen muodostavat

The screenshot shows the Suomi.fi website interface. At the top, there's a navigation bar with 'Palaute | Sivukartta | Tietoa sivustosta'. Below that is a search bar with 'Hae' and 'Ohjeet' buttons. The main content area is titled 'Terveys ja ravinto' and includes several sub-sections: 'Terveystieteiden tutkimuskeskus', 'Seksuaalielämä', 'Päihitteet', and 'Ravinto ja elintarvikkeet'. There are also links for 'Terveystieteiden tutkimuskeskus', 'Kansallinen terveysprojekti', and 'Terveys 2015 - kansanterveysohjelma'. A sidebar on the right contains 'Muuta aiheeseen liittyvää', 'Lomakkeet', 'Keskustelut', and 'Aihealueen ajankohtaista'. The footer shows the page was updated on 30.09.2004 and is managed by the Finnish Government.

Kuva 7: Suomi.fi-portaalin tietosivu.

puun ja jokainen linkki kuuluu vain yhteen aihealueeseen. Tämä tarkoittaa, että jokainen linkki on löydettävissä ainoastaan yhden selaamispolun päästä. Puurakenteen rajoitteita on pyritty kiertämään lisäämällä käyttöliittymän sivustaan palsta, jossa esitetään linkkejä muihin aihealueeseen liittyviin portaalin aihealuesivuihin. Koska nämä linkit ovat erillään käyttöliittymän muusta rakenteesta, eivät ne kuitenkaan eksytä käyttäjiä, toisin kuin osiossa 3.4.1 esitellyn ODP-portaalin verkko-omainen navigointirakenne.

Aihealueiden sivuille liittyy kuvausten lisäksi asiasanoitusta. Asiasanoja liittyy myös

sivuilla olevaan tärkeimpään tietosisältöön, eli muihin palveluihin osoittaviin linkkeihin. Asiasanojen lisäksi linkkeihin liittyy tieto niiden päässä olevan sivun omistajaorganisaatiosta sekä lyhyt kuvaus linkitetyn sivun tai sivuston sisällöstä.

Asiasanoitusta voidaan käyttää hyödyksi etsimällä linkkejä ja aihealueita asiasanahaun avulla. Asiasanahaku on portaalin toinen tiedonhakutapa hakemiston selaimisen lisäksi. Asiasanahaku täsmäyttää hakukyselyn merkkijonoa linkkien ja aihealueiden otsikoiden, kuvausten ja asiasanojen kanssa. Lisäksi asiasanahaku kohdistuu myös Suomi.fi-portaalin sisärsivustoihin¹⁹ sekä kunta- ja virkamieshakemistoon.

Portaalia ylläpidetään julkaisujärjestelmän avulla. Ylläpitäjä voi järjestelmän avulla hallita aihealueiden sisältöä, asiasanoitusta ja hajautettujen ylläpitäjien oikeuksia. Hajautetut ylläpitäjät voivat ehdottaa portaalin toimitukselle lisättäväksi uusia linkkejä ja muokata omistamiaan linkkejä selainpohjaisen käyttöliittymän avulla. Hajautetut ylläpitäjät eivät siis pääse lisäämään uusia linkkejä suoraan, vaan linkit lisätään portaaliin vasta toimituksen hyväksynnän jälkeen.

Suomi.fi-portaalin nykyinen versio on siis varsin perinteinen keskitetysti hallittu informaatioportaali, jossa vastuu aineiston ylläpidosta on keskitetyllä toimitusjoukolla. Portaalin sisältö koostuu asiasanoin ja kuvaustekstein annotoiduista linkeistä, jotka on sijoitettu yksiulotteiseen hierarkiaan.

6.2 Semanttisen webin teknologioiden tuomat mahdollisuudet

Tässä osiossa tarkastellaan lähemmin, mitä mahdollisuuksia luvuissa 4 ja 5 esitetyt semanttisten portaalien ratkaisumallit tarjoavat Suomi.fi-portaalille. Asiaa lähestytään sekä käyttäjän että portaalin ylläpitäjän kannalta. Käyttäjälle tärkeää on

¹⁹Verkkopalveluopas <http://www.asiointiopas.fi>, elektroninen lomakepalvelu <http://www.lomake.fi> ja keskustelufoorumi <http://www.otakantaa.fi>

tiedon vaivaton löytäminen ja ylläpitäjälle taas vaivaton tiedonhallinta.

6.2.1 Portaalin tietojen semanttinen annotointi

Suomi.fi-portaalissa esitetyistä tietoalkioista tallennettua metadataa hyödynnetään tällä hetkellä lähinnä tiedonhaussa asiasanojen muodossa. Jokaiseen tietoalkioon liittyy joukko asiasanoja, jotka huomioidaan portaalin sanahaun täsmäytyksessä. Asiasanojen lisäksi toinen käyttäjälle esitetty metatieto linkitetystä sivusta on sen julkaisijaorganisaatio. Nämä tiedot on tallennettu portaaliin tietovarastoon tavallisin merkkijonoina. Tämän takia esimerkiksi sama julkaisuorganisaatio on voitu tallentaa eri tavoin ilmaistuna. Esimerkiksi opetusministeriö on voitu esittää portaalin tietovarastossa muodossa `OPM` tai `Opetusministeriö`.

Ontologisen rakenteen mukaisesti annotoidulla metadatatalla voidaan portaalin tietoalkioihin liitettävä metadata saada yhtenäiseksi ja tietosisällöltään rikkaammaksi. Kun tieto sivun julkaisuorganisaatiosta esitetään pelkän merkkijonon sijaan käsitteenä, tässä tapauksessa esimerkiksi luokan `organisaatio` instanssina, voidaan kaikki saman organisaation julkaisemat sivut hakea portaalin tietovarastosta helposti. Kun itse julkaisijaorganisaatiokin on esitetty käsitteenä, voidaan myös siihen liittää metatietoa. Tämän metatiedon avulla portaalin tarjoamaa tiedonhakua voidaan laajentaa koskemaan yksittäisten portaalissa esitettyjen sivujen lisäksi tietyn tyyppiä palveluita tarjoavia organisaatioita. Julkaisijaorganisaatioiden lisäksi myös muu sivuihin liittyvä metadata voidaan esittää käsitteinä. Nämä käsitteet muodostavat portaalin tietämyskantana toimivan semanttisen verkon.

6.2.2 Näkymäselauksen hyödyntäminen

Käyttäjän kannalta portaalin nykyisessä versiossa oikean tiedon löytäminen voi olla ongelmallista. Hakemiston puurakenne rajoittaa linkit yksittäiseen lehteen, johon

on olemassa ainoastaan yksi hakupolku. Mitä syvemmällä puussa linkki on, sen hankalammaksi sen löytäminen tällöin tulee, koska käyttäjän on osattava jokaisella hierarkiatasolla valita oikea alikategoria tai löydettävä haettava linkki kyseiseltä hierarkiatasolta. Tämä jokaisen tason linkkien selaaminen aiheuttaa tiedonhaussa myös lisäkustannuksen ylimääräisenä selaamistarpeena. Lisäkustannus puolestaan voi johtaa käyttäjän huomiokyvyn herpaantumiseen, jolloin todennäköisyys etsityn linkin havaitsemiseen heikkenee dramaattisesti [Mar95].

Yksi tapa helpottaa tiedon löytämistä selaamalla on hyödyntää osiossa 5.2 käsitellyä näkymäselausta. Näkymäselauksen avulla haettu tieto voidaan löytää useiden reittien kautta, eikä tieto ole sidottu yhteen hakupolkuun. Näkymäselauksen avulla käyttäjä voi lisäksi rajata tietoalkioiden tulosjoukkoa asteittain, kunnes se on riittävän pieni helposti selattavaksi. Kun portaalissa esitettyjen linkkien metadata on annotoitu yhteisen ontologian mukaan, näkymäselaus voidaan ottaa käyttöön ilman merkittävää lisäkustannusta, koska tätä ontologiaa voidaan hyödyntää näkymäselauksessa käytettävien kategorioiden muodostamisessa.

6.2.3 Sanahausta käsitteeseen

Ontologioiden avulla voidaan kehittää myös portaalin sanahakua. Jos asiasanojen sijaan portaalin tietoalkiot on annotoitu mm. aihealueiden ja elämäntilanteiden mukaan, voidaan haku kohdistaa myös näihin ontologian käsitteisiin ja palauttaa hakutuloksena ne tietoalkiot, jotka liittyvät esimerkiksi talvirenkaisiin tai verotukseen. Ontologioiden mahdollistamalla käsitteellä erilaisia homonyymisiä käsitteitä voidaan erottaa toisistaan. Klassisena esimerkkinä voidaan käyttää hakusanaa **Nokia**. Käsitteehaku voi palauttaa tuloksena kaksi erilaista käsitettä, Nokia-nimisen yrityksen ja Nokia-nimisen kaupungin, tai Suomi.fi-portaalin ympäristössä näihin kahteen käsitteeseen liittyvät tietoalkiot ryhmiteltyinä.

Suomi.fi-portaalin nykyisen version sanahaku toimii tällä hetkellä moniportaisena. Se täsmäyttää tuloksia mm. portaalin sivuilla olevaan tekstiin ja asiasanoitukseen, linkkeihin, ja virkamieshakemiston henkilöihin. Tässä mielessä sanahaku toimii jo eräänlaisena käsitehakuna. Linkkien lisäksi se täsmäyttää hakutulosta myös portaalin sivuihin eli siihen kontekstiin, jossa linkit sijaitsevat.

6.2.4 Semanttinen suosittelu

Kolmas portaalista tiedon hakemista ja löytämistä helpottava ominaisuus on suosittelu. Suosittelemalla tarkoitetaan käyttäjälle esitettäviä tietolähteitä, jotka on jollain mielenkiintoisella tavalla yhteydessä hänen tällä hetkellä tarkastelemaansa tietoon [Bur00].

Kun käyttäjä on tarkastelemassa yksittäistä linkkiä, voidaan hänelle esittää samalla esimerkiksi saman organisaation ylläpitämät tai saman sivuston muut portaaliiin tallennetut sivut. Tämän tyyppistä suosittelua on olemassa jo portaalin nykyisessä versiossa. Kategorioiden yhteydessä on usein viittaus johonkin toiseen kategoriaan tai aihealueen Lomake.fi-portaalin lomakkeisiin. Tämä yksinkertainen suosittelumeکانismi voidaan kuitenkin korvata siirtymällä edellä esitetyllä tavalla puumaisesta luokittelusta moniulotteisempaan tietoalkioiden luokitteluun.

Ontologioiden avulla suosittelusta voidaan tehdä laajempaa ja suosittelulinkkien muodostaminen voidaan automatisoida. Suosittelemalla voidaan hyödyntää joko ontologioissa suoraan esitettyjä yhteyksiä tai rakentaa käytettyjen ontologioiden perusteella ontologioista erillisiä päättelysääntöjä, jotka liittävät tarkasteltuun tietoaalkioon sääntöjen perusteella liittyviä muita tietoalkioita [MHSV04]. Sääntöihin perustuvassa semanttisessa suosittelussa voidaan esimerkiksi esittää sääntö, jonka perusteella portaalii esittää käyttäjälle, joka tarkastelee linkkiä lakitekstiin, myös portaalii muut kyseisen lakitekstin aiheeseen liittyvät linkit.

Portaalin ontologiassa voidaan myös suoraan määritellä käsitteiden välille semanttisia yhteyksiä, joita portaalijärjestelmä voi hyödyntää linkkien suosittelussa. Erilaiset kohderyhmät voidaan yhdistää näitä lähellä oleviin aihealueisiin ja jopa organisaatioihin. Näitä yhteyksiä hyödyntämällä käyttäjälle, joka valitsee näkymäselauksessa esimerkiksi *opiskelija*-kategorian, järjestelmä voi suositella muiden aihealueiden kategorioita, jotka portaalin ontologian suunnittelija on katsonut opiskelijoita kiinnostavaksi, ja linkittänyt käsitteeseen *opiskelija* semanttisella yhteydellä.

Suosittelumekanismit muodostavat erilaisia semanttisia linkkejä portaalin tietoalkioiden välille. Nämä suosittelujärjestelmän automaattisesti muodostamat linkit tekevät portaalin sisäisestä linkkirakenteesta tiheämmän ja luovat tietoalkioiden välille oikopolkuja, jotka mahdollistavat portaalin aineiston assosiatiivisen selailun. Assosiatiivisessa selailussa mielenkiintoinen tietoalkio voidaan löytää seuraamalla kahden toisistaan ensisilmäyksellä rippumattoman tietoalkion välistä linkkiä.

6.2.5 Portaalin ylläpito

Kun portaalin tietämys on tallennettu avoimen ontologian mukaisesti, voidaan myös sisällöntuotantovastuuta hajauttaa helpommin useiden organisaatioiden kesken. Portaaliin lisättävät tietoalkiot eli sivut on annotoitava portaalin ontologian mukaisesti. Sivusta on mm. kuvattava, mihin aihealueisiin ja kohderyhmiin se liittyy. Lisäksi voidaan rajata sivun sisältämän tiedon maantieteellistä kattavuutta esimerkiksi kunta- tai maatasolla.

Ylläpitäjät voivat annotoida portaaliin uusia tietoalkioita annotaatioeditorin avulla. Editori voidaan toteuttaa web-sovelluksena, joka tarjoaa käyttäjälle lomakepohjaisen käyttöliittymän uuden sivun annotointiin. Tällainen ratkaisu on käytössä esimerkiksi osiossa 3.4.3 käsitellyssä SWED-portaalissa. Geneerisempi web-pohjainen editor on ollut myös kehitteillä IWebS-projektissa [Api05]. Editorin avulla käyttä-

jä voi annotoida sivun ainoastaan käyttämällä ontologian mukaisia termejä. Näin voidaan kontrolloida sitä, että sivujen annotointi tapahtuu ontologian mukaisesti.

Portaalin ylläpitoprosessi voidaan toteuttaa joko käyttämällä samanlaista keskitettyä kontrollia kuin nykyisessä julkaisujärjestelmässä, jossa kaikki eri hajautettujen ylläpitäjien sivuehdotukset tarkastetaan keskitetyssä toimituksessa ennen portaalin tietämuskantaan lisäämistä. Mikäli hajautettujen ylläpitäjien käyttämä annotaatioeditori on tarpeeksi helppokäyttöinen ja mahdollistaa laadukkaiden sivujen annotoinnin, voidaan tämä toimitusvaihe kuitenkin jättää pois. Kolmas vaihtoehto on avata järjestelmään lisättävien annotointien tekeminen kaikille, kuten ODP-projektissa (ks. 3.4.1). Tällöin annotoinnit tarkastava toimitus on kuitenkin välttämättömyys.

6.3 Ontologiapohjainen Suomi.fi

Nykyisen Suomi.fi-portaalin aineiston pohjalta tämän pro gradu-tutkimuksen yhteydessä toteutettiin semanttinen informaatioportaali. Portaaliin on otettu tietoalkioiksi osajoukko nykyisen Suomi.fi-portaalin linkeistä. Toteutettu prototyyppiportaali ei ole sisällöltään yhtä laaja kuin alkuperäinen Suomi.fi, vaan siihen on kerätty n. 160 tietoalkioita. Tietoalkioina prototyypissä on alkuperäisen portaalin tapaan linkkejä erilaisille sivustoille ja yksittäisille web-sivuille. Lisäksi erilaiset linkitettyihin sivustoihin liittyvät organisaatiot muodostavat osan tietoalkioista.

Portaali on toteutettu Semantic Computing -tutkimusryhmän kehittämää OntoViews-järjestelmää [MHSV04] hyödyntäen. Portaalissa käytetään OntoViews-järjestelmän yhteydessä Ontodella-suositelijaa, jonka avulla portaalin tietoalkioiden väliset HTML-linkit luodaan automaattisesti järjestelmälle esitettyjen SWI-Prolog-kielisten [SWP] sääntöjen perusteella.

Portaalin rakenne pohjautuu joukolle ontologioita, jotka on kehitetty mallintamaan

portaalissa esitettäviin sivustoihin liittyvää metadataa. Ontologian eri osa-alueet kuvaavat erilaisia tietoalkioihin liittyviä tietoja:

- Aihealueita, joihin sivustot ovat jakautuneet. Aihealueina on käytetty samaa jaottelua kuin alkuperäisessä portaalissa, jotta niihin liittyviä olemassa olevia kuvaustekstejä on voitu hyödyntää portaalin käyttöliittymässä. Jatkokokehityksessä luontevampi vaihtoehto olisi käyttää JUHTA-työryhmän JHS-145 suositusta [JHS04], jolloin aihealueiden jaottelu ei olisi päällekkäinen muiden portaalin ontologian osa-alueiden kanssa.
- Käyttäjryhmiä, joita sivuilla olevat tiedot lähinnä koskevat, ja joita silmällä pitäen tietoalkiona oleva sivu tai sivusto on tehty.
- Elämäntilanteita, jotka koskettavat portaalin käyttäjryhmiä.
- Maantieteellisiä alueita. Ilmoittamalla portaalin tietoalkioiden sisältämän tiedon maantieteellinen kattavuus portaalissa voidaan esittää yksittäisien kuntien ja kaupunkien palveluita.
- Tietoalkioina olevien sivustojen tarjoamat palvelut ja sivustojen tyypit. Tietoalkioina olevat resurssit voivat tarjota esim. lomakkeita, yhteystietoja viranomaisiin, elektronisia julkishallinnon palveluita tai yleistä informaatiota jostain aiheesta.
- Sivustojen julkaisijoina toimivien organisaatioiden organisaatiotyyppejä. Sivustoihin liittyvät organisaatiot on esitetty tietoalkioina ja erityyppiset organisaatiot on luokiteltu niiden toimintamuodon perusteella.

Näistä eri osaontologioista projisoidaan myös portaalin näkymäselaukseen (ks. osio 5.2) perustuvan käyttöliittymän eri näkymät. Kuvassa 8 on esitetty portaalin selauskäyttöliittymä selauksen keskivaiheessa. Näkymäselaukseen käyttöliittymässä voidaan va-

The screenshot shows the SW-Suomi.fi Semantic Information Portal interface. At the top, there is a navigation bar with the logo of the Helsinki Institute for Information Technology and the text 'SW-Suomi.fi - A Semantic Information Portal'. Below this, there is a search bar and a list of search filters. The filters include 'Käsitteet', 'Aihealue', 'Elämäntilanne', 'Organisaatiohakemisto', and 'Alueellinen kattavuus'. The search results are displayed in a grid format, with two results for 'Nuorten asuminen' visible. Each result includes a title, a brief description, and a link to 'Siirry sivustolle'.

Kuva 8: Semanttisen Suomi.fi portaalin näkymäselauksikäyttöliittymä selauksen keskivaiheessa, jossa käyttäjä on valinnut kaksi kategoriata hakuehdoiksi.

lita eri näkymistä kategorioita, joiden perusteella hakutuloksena näytetään kategorioiden kuuluvien tietoalkioiden leikkaus. Lisäksi kategorioihin voi liittyä omaa tietosisältöä. Tietosisältö koostuu kategorian HTML-muotoisesta kuvaustekstistä. Tietosisältö näytetään oletusarvoisesti aihealue-kategoriasta, mutta myös muiden luokitusten kategorioihin voidaan liittää omat kuvaukset.

Selaukskäyttöliittymän lisäksi portaalin tietosivuja on mahdollista hakea käsittehaun avulla. Käsittehakuna portaalissa on käytetty OntoViewsin Ontogator-hakumootorin tarjoamaa hakupalvelua. Haku täsmäyttää kyselyssä esiintyvää merkkijonoa kategorioiden ja tietoalkioiden `rdf:label` ominaisuuksien arvoihin.

Kuvassa 9 on esitetty portaalin tietosivu. Jokainen portaalin tietosivu visualisoi

The screenshot shows a web browser window displaying the SW-Suomi.fi Semantic Information Portal. The page title is "SW-Suomi.fi - A Semantic Information Portal". The browser address bar shows the URL: http://localhost:8081/item?&l=fi&m=0&g=c%2501%2501&c=%2500%2500&c=%2501%2501&sb=http%3A%2F%2Fv... The page content includes a navigation menu with "Uusi haku", "Takaisin hakusivulle", "Ohjeet", and "Tietoa sivustosta". The main content area displays search results for "Nuoret (2)" and "Nuorten asuminen pääkaupunkiseudulla (> Nuorten asuminen)".

Aihealue:

- [Asuminen ja perhe](#) (33) > [Asuminen](#) (7) > [Asunnon hankinta](#) (3)

Kohderyhmä:

- [Nuoret](#) (10)
- [Yksityshenkilöt](#) (62) > [Nuoret](#) (10)

Elämäntilanne:

- [Asunnon hankinta](#) (8)

Alueellinen kattavuus:

- [Suomi](#) (69) > [Etelä-Suomen lääni](#) (4) > [Helsinki](#) (4)

Nuorten asuminen pääkaupunkiseudulla

Thumbnail not available

Nuorille suunnattua tietoa asumisesta ja asunnon hankinnasta pääkaupunkiseudulla. Tietoa vuokra-asumiseen liittyvien prosessien lisäksi myös muista erilaisista asumismuodoista, kuten asumisoikeusasumisesta.

URL: http://www.kompassi.info/page.asp?_item_id=317

Kuuluu sivustoon:

Sivusto:

- [Nuorisotiedotuskeskus Kompassi](#)

Linkejä samalta aihealueelta:

Asunnon hankinta:

- [Nuorten asuminen](#)
- [Suomen opiskelija-asunnot Oy](#)

Linkejä samaan elämäntilanteeseen:

Asunnon hankinta:

- [Lomake.fi](#)
- [Nuorten asuminen](#)
- [Suomen opiskelija-asunnot Oy](#)
- [Asumisterveys ja turvallisuus](#)
- [Asuminen.fi](#)

Linkejä samalle kohderyhmälle:

Nuoret:

- [Kuljettajaopetus](#)
- [ALLISON - tietosivut nuorille](#)
- [Nuorisotiedotuskeskus Kompassi](#)
- [Maailmalle.net](#)
- [Mopokortti](#)
- [Nuorten asuminen](#)
- [Nuoris on tietosivut opiskelusta](#)
- [Opintotukijärjestelmä](#)
- [Työelämästä nuorille](#)

Kuva 9: Semanttisen Suomi.fi portaalin tietosivu jossa esitetään yhden tietoalkion tiedot.

portaaliontologian mukaisesti annotoidun tietoalkion. Tietosivulla esitetään loppukäyttäjälle tarkoitettuja ominaisuuksia tietoalkiosta sekä ne portaaliontologian luokat, joihin kyseinen tietoalkio on liitetty. Lisäksi sivulla on esitetty linkkejä toisiin tietoalkioihin, jotka liittyvät tietoalkioon jonkin suosittelusäännön kautta.

Eri näkymien perustana olevien luokittelujen lisäksi portaaliontologiassa luokitellaan myös portaalin tietoalkiot erilaisiin luokkiin sivujen sisällön perusteella. Erilaiset sivutyypit on esitetty taulukossa 1. Sivutyyppi tarjoaa osin rinnakkaisen tavan luokitella tietoalkion palvelutyyppinä sivustotyyppiontologian kanssa. Määrittelemällä tietoalkion sivutyypin eksplisiittisesti voidaan ilmaista tietoalkion luonne.

Luokan nimi	Kuvaus
WebPage	Web-sivu, kaikkien sivujen yläluokka.
InformationPage	Yksittäinen sivukokonaisuus, joka kertoo jostain palvelusta.
Forum	Keskustelusivusto.
Form	Elektroninen lomake.
BenefitPage	Tukia ja etuisuuksia käsittelevät sivut.
Legislation	Lakeja ja asetuksia käsittelevät sivut.
ServicePage	Elektroninen palvelu. Tarjoaa konkreettisen verkon välityksellä toimivan palvelun. Vaativat usein autentikoinnin esim. pankkiavaimilla.
Site	Sivuston etusivu.
InternalPage	Suomi.fi-portaalin sisäinen sivu.

Taulukko 1: Portaaliontologian WebPage-luokka ja sen aliluokat.

Sivutyypeistä erikoisin on `InternalPage`, jonka avulla voidaan luoda portaalin sisäisiä tietosivuja. Sisäisten sivujen avulla voidaan luoda esimerkiksi kokonaisuuksia, jotka muodostuvat yhdistelmistä portaalin ulkopuolisia sivuja kuvaavia tietoalkioita ja portaalin omaa sisältöä sisältäviä tietoalkioita. Kokonaisuudet muodostetaan ketjuttamalla tietoalkioita listoiksi, jotka esitetään portaalin käyttöliittymässä kuvassa 10 esitetyllä tavalla. Kuvassa 10 esiintyvässä sisäisessä sivussa on esitetty `InternalPage`-tyyppinen tietoalkio, joka kokoaa yhteen yhden avioliittoa käsittelevän loogisen kokonaisuuden. Kuvassa esitetty sivu sisältää portaalin omaa aihetta käsittelevää esittelytekstiä sekä joukon linkkejä, jotka on tässä tapauksessa ryhmitelty pieniksi loogisiksi avioliiton liittyvää laillista aspektia ja hääseremoniaa käsitteleviksi kokonaisuuksiksi.

`InternalPage`-sivutyypin lisäksi portaaliin on mahdollista luoda myös täysin staattisia sivukokonaisuuksia muodostamalla näitä `OntoViews`-järjestelmän pohjalla toimivan `Cocoon`-julkaisualustan [ACP] avulla. Staattisten sivujen liittäminen varsinaisen portaalin sisällön muodostaviin tietoalkio-sivuihin ei ole kuitenkaan suoravi-



Kuva 10: Useita tietoalkioita yhdistävän IntenalPage tietoalkion visualisointi.

vaista, koska niitä ei ole luotu portaalin ontologioiden pohjalta. Cocoonin hyödyntäminen mahdollistaa kuitenkin sen, että laajempaan portaalikokonaisuuteen voidaan liittää yhtenä osana myös ontologiaperustaista RDF-materiaalia.

Portaalissa tietoalkioiden yhteydessä esitetyt linkit toisiin tietoalkioihin (kuvassa 9 oikealla) on toteutettu Ontodella-järjestelmän avulla. Linkit on muodostettu esittämällä sääntöjä, jotka määrittelevät minkälaisen tietoalkioiden välille muodostetaan suosittelulinkki. Sääntöjen avulla voidaan yhdistää kaksi tietoalkiota toisiinsa ominaisuudella, jota ei eksplisiittisesti ole portaali-ontologiassa määritetty. Käyttö-

liittymässä OntoViews linkittää eri sääntöihin sopivien tietoalkioiden muodostamat tietosivut toisiinsa. Esimerkiksi kuvassa 9 olevalta tietosivulta on linkki mm. tietosivulle *Nuorisotiedotuskeskus Kompassi*, koska tietosivulla esitetty tietoalkio ja linkin päässä oleva tietoalkio sopivat Ontodellassa määriteltyyn sääntöön. Nämä oikopolkuja tietoalkioiden välille luovat suosittelusäännöt ovat luonteeltaan hyvin samankaltaisia kuin luvussa 5.1 käsitellyn Haystack-järjestelmän virtuaaliset ominaisuudet, joiden avulla Haystackissa voidaan yhdistää toisiinsa resursseja, joiden välille ei ole esitetty eksplisiittistä ominaisuutta.

6.4 Portaalin evaluointi ja jatkokehitys

Edellä esitelty semanttinen portaalit on prototyyppi, joka ei sellaisenaan ole tuotantoympäristöön riittävä. Tässä osiossa tarkastellaan portaalin osa-alueita, jotka vaativat jatkokehitystä ja esitetään ehdotuksia kehityksen mahdollisista suunnista. Lisäksi osiossa arvioidaan toteutetun käyttöliittymän toimivuutta ja sen rajoituksia. Prototyyppiportaalissa näkymäselauksessa hyödynnettävät näkymät eivät ole täysin ortogonaaliset. Ongelma on ilmeinen elämäntilanteiden ja kohderyhmien muodostamissa näkymissä, jotka ovat useilta osin toistensa kaltaiset. Nämä näkymät voitaisiinkin yhdistää yhdeksi näkymäksi. Portaaliontologiaa voidaan joko muuttaa siten, että joko elämäntilanneontologian instanssit yhdistetään kohderyhmiin, tai kohderyhmäontologian instanssit elämäntilanneontologiaan. Toinen vaihtoehto on säilyttää portaaliontologia ennallaan ja muuttaa sen sijaan Ontodellan projisointisääntöjä, jotka luovat portaalin käyttöliittymän. Toinen näkymäselaukseen liittyvä ongelma on portaalin aineiston pienuus. Näkymäselauksen edut eivät tule näkyviin pienellä aineistolla, joka rajautuu jo yhden näkymän kategorioita valitsemalla helposti hahmotettavaksi hakutulokseksi.

Julkishallinnon portaalit on kaksikielisessä maassa oltava monikielinen. Prototyyppi

on kuitenkin toteutettu ainoastaan yksikielisenä. Käytetty teknologia mahdollistaa kuitenkin pienellä lisätyöllä myös monikielisen portaalin toteuttamisen. Portaaliontologiassa jokaiseen resurssiin voidaan liittää useita arvoja `rdfs:label` ominaisuuden arvoksi, joille jokaiselle voidaan määritellä kielityypitetty literaaliarvo. Nämä literaalit toimivat eri kieliversioiden perustana. Ontologian laajentaminen monikieliseksi on kuitenkin varsin työlästä, eikä eri kieliversioiden käyttöönotto OntoViews-järjestelmän kaikissa komponenteissa portaalin kehittämishetkellä ollut helposti järjestettävissä, joten prototyyppi toteutettiin tässä vaiheessa ainoastaan yksikielisenä.

Toinen julkishallinnon portaaleille tärkeä ominaisuus, saavutettavuus, toteutuu portaalissa suurelta osin W3C:n suositusten mukaisesti [WAI99]. Portaalin sivut skaalautuvat ja sivujen selaaminen tekstipohjaisella selaimella on sujuvaa. Julkishallinnon JUHTA-työryhmän sivustojen suunnitteluohjeen [JHS00] peräänkuuluttama löydettävyyys kärsii hieman siitä, etteivät portaalin dynaamisesti generoidut tietosivut indeksoidu hakukoneisiin. Tämä on tosin mahdollista korjata portaalin myöhemmissä versioissa.

Prototyyppinä toteutetussa portaalissa ei olla keskitytty portaalin hallinnassa käytettävien työkalujen kehittämiseen, vaan portaalin tietämystä on käsitelty Protégé-ontologiaeditorilla. Samaisella editorilla luotiin myös portaalin ontologiat ja kansoitettiin ne portaalin tietoalkioilla. Kuten osiossa 4.1 mainittiin, on pelkkä Protégé-editoriin pohjautuva portaalin tiedonhallinta tuotantokäytössä liian kömpelö ratkaisu, joten portaalin sisällönhallintajärjestelmän suunnittelu on tarpeen. Komponenttina sisällönhallintajärjestelmässä voidaan käyttää esim. AnnE-editoria, jota voidaan laajentaa mahdollistamaan uusien tietoalkioiden lisäämisen lisäksi myös tietoalkioiden poistamisen, muuttamisen, sekä portaaliontologioihin liittyvien annotointien editoinnin. Toinen vaihtoehto on hyödyntää itse Protégé-editoria, ja karsia siitä pois sellaista toiminnallisuutta, jota ylläpitovaiheessa ei enää tarvita, kuten ontologioiden luokkarakenteen muuttamisen.

Portaaliprototyyppiä toteuttaessa perustyyppistä poikkeavien tietoalkioiden, käytännössä portaalin sisäisten tietosivujen luominen ja ylläpitäminen oli joustavan sisällönhallintajärjestelmän puuttumisen takia erittäin työlästä. Sisäiset tietosivut luotiin kehitysvaiheessa muun sisällön tavoin Protégé-editorilla, jonka hyödyntämisessä ilmeni käytettävyyso ongelmia, jotka hankaloittavat esimerkiksi sisäisen tietosivun muodostavien tietoalkioiden järjestyksen muokkaamista²⁰. Sisäisten tietosivujen avulla voidaan julkaista portaalin kiinnostavin sisältö, joten näiden sivujen luonnin on oltava myös helppoa. Näiden tietosivujen luontiin AnnE tai Protégé-editori eivät ole riittäviä, vaan vaativat paremmin tehtävään räätälöidyn järjestelmän.

Toinen ylläpitoon liittyvä aspekti, johon prototyyppiä toteutettaessa ei kiinnitetty huomiota on portaalin hajautettu ylläpito. Portaalin ylläpito olisi mahdollista hajauttaa usean toimijan välille tarjoamalla web-pohjainen rajapinta, jonka kautta portaaliin voidaan lisätä uusia tietoalkioita. Hajautukseen liittyy myös se, että portaalin tietämuskantaan voidaan tarjota pääsy myös muille organisaatioille, kuten seutuportaalien ylläpitäjille, jotka voisivat näin hyödyntää Suomi.fi-portaalissa varastoitua tietämystä.

Kolmas hallinnointiin liittyvä seikka, johon portaalin ja koko OntoViews-järjestelmän, kehittämisessä kannattaa panostaa, on dynaamisempi uuden sisällön integrointi portaaliin. Tällä hetkellä portaalin ulkoasua voidaan XSLT-tekniikan avulla muuttaa ilman että portaaliohjelmisto tulisi käynnistää uudelleen, mutta portaalin ei voi tuoda uutta sisältöä ilman, että se käynnistetään uudelleen. Uudelleen käynnistys pitää sisällään portaaliontologian projisoimisen ja kategorioiden lataamisen uudelleen OntoViews-järjestelmään. Isoilla aineistoilla uudelleenkäynnistys voi kestää useita tunteja.

²⁰Osa näistä käytettävyyso ongelmista on korjautunut editorin uudemissa versioissa.

7 Yhteenveto

Tässä pro gradu -tutkielmassa on tarkasteltu semanttisia portaaleja ja niihin liittyviä teknologioita. Semanttiset portaalit voivat toimia semanttisen webin sisältämän tietämyksen yhteenkerääjinä, jotka tarjoavat yhtenäistetyn näkymän jonkin alueen tietämykseen.

Semanttisen webin teknologioiden tuoma hyöty portaaleille perustuu yhtenäiseen, koneymmärrettävään tapaan käsitellä ja esittää portaaleiden tietoalkioita. Semanttisessa webissä data on tarkoitettu myös koneiden ymmärrettävään muotoon. Tällöin datan esitystapana on RDF ja tätä dataa on usein syytä myös visualisoida käyttäjille. Tämä visualisointi voidaan tehdä semanttisissa portaaleissa.

RDF-muotoista tietämystä voidaan siirtää helposti eri organisaatioiden välillä. Kun tietämys on esitetty yhteisten ontologioiden mukaisesti, voidaan samaa tietämystä hyödyntää eri organisaatioissa ilman, että jokaisen eri organisaation välillä on tietoa siirrettäessä muokattava myös tiedon esitysmuoto skeemasta toiseen. Näin samaa tietoa voidaan käyttää uudelleen erilaisissa yhteyksissä niiltä osin, joiden katsotaan soveltuvan kulloiseenkin käyttötarkoitukseen.

RDF-tietämyskannat tarjoavat hyvän alustan semanttisten portaalien tiedonvarastointiin. Tietämyskannoista voidaan esittää portaalin käyttöliittymässä kiinnostavat osat tietoalkioiden sisällöstä sekä projisoida tietoalkioiden annotointiin käytetyistä ontologioista erilaisia kategorioita, joiden mukaan tietoalkiot voidaan portaalissa esittää. Mitä rikkaampaa metadataa tietoalkioista on käytössä, sitä useampia erilaisia tapoja voidaan hyödyntää muodostettaessa portaalin hakukäyttöliittymää. Hakukäyttöliittymänä voidaan hyödyntää näkymäseläystä ja ontologioihin pohjautuvaa käsitehakua, jolla voidaan korvata tavallinen, usein ainoastaan kuvausteksteihin täsmäyttävä tekstihaku, joka useissa portaaleissa on käytössä. Lisäksi tietämyskantojen yhteydessä hakukäyttöliittymässä voidaan hyödyntää RDF-kyselykieliä, jol-

loin voidaan tarjota käyttäjälle mahdollisuus formuloida kompleksisia hakuja portaalin sisältöön.

Haaste semanttisten portaalien luomisessa on portaalien sisällönhallintajärjestelmien toteuttaminen siten, että ne tuottavat portaaliontologioiden mukaisesti annotoitua tietämystä portaaliin. Tietämyksen luomiseen voidaan käyttää ontologia- tai annotaatioeditoria. Ontologiaeditorit tarjoavat mahdollisuuden muokata portaalin rakenteen muodostavia ontologioita, ja annotaatioeditorilla voidaan luoda portaaliin sisältöä, annotoimalla portaalin tietoalkiot. Olemassa olevat ontologia- ja annotaatioeditorit eivät kuitenkaan vastaa täysin nykyisten portaalien helppokäyttöisiä sisällönhallintajärjestelmiä, eivätkä pysty näitä portaalien hallintaan erikoistuneita järjestelmiä täysin korvaamaan. Perinteisillä sisällönhallintajärjestelmillä ulkoasun ja sisällönmuokkaus voidaan tehdä samoilla työkaluilla, mutta ne eivät useinkaan tarjoa riittäviä mahdollisuuksia sisällön metadatan käsittelyyn.

Pro gradu -työn yhteydessä toteutettu semanttinen portaaliprototyyppi laajentaa nykyisen Suomi.fi-portaalin toiminnallisuutta tarjoamalla tiedon etsimiseen monipuolisemman käyttöliittymän näkymäselauksen ja semanttisen linkityksen myötä. Portaalin tietosisältö on esitetty portaaliontologian mukaisesti. Ontologian muodostaa tietoalkioihin liittyvä metadata, joista projisoidaan portaalin hakukäyttöliittymässä käytetyn näkymäselauksen kategoriat. Kategorioita hyödynnetään lisäksi portaalin käsittehaussa. Itse portaalin tietosisältö, tietoalkioina olevat sivunkuvaukset, on annotoitu portaaliontologian mukaisesti.

Demonstraatioportaalin toteuttamisen yhteydessä ilmeni, että ontologiat tarjoavat toimivan tavan esittää portaaleissa esitettävää tietoa. Semanttisen webin teknologiat mahdollistavat semanttisten portaalien hyödyntämisen, mutta tehokkaaseen portaalien hallintaan vaaditaan nykyisiä ontologiaeditoreita tiiviimmin sisällöntuotantoketjuun integroitu sisällönhallintajärjestelmä, joka tarjoaa käyttäjälle helppokäyttöisen käyttöliittymän portaaliontologian editointiin ja tietoalkioiden muokkaamiseen.

Lähteet

- A⁺04 Airio, E. et al., CIRI - An Ontology-based Query Interface for Text Retrieval. *Proceedings of the 11th Finnish AI Conference, Web Intelligence, vol 2*. Finnish AI Society, syyskuu 2004.
- ACP The Apache Cocoon Project. <http://cocoon.apache.org/>. [2.3.2005]
- Api05 Apiola, M., Ontologiapohjainen annotointi, Pro gradu. Tekninen raportti C-2004-95, Tietojenkäsittelytieteen laitos, Helsingin yliopisto, 2005.
- AvH03 Antoniou, G. ja van Harmelen, F., ”web ontology language: Owl”. Teoksessa *Handbook on Ontologies in Information Systems*, Staab, S. ja Studer, R., toimittajat, Springer-Verlag, 2003, sivut 67–92.
- BKvH02 Broekstra, J., Kampman, A. ja van Harmelen, F., Sesame: A Generic Architecture for Storing and Querying RDF and RDF Schema. *Proceedings of the First International Semantic Web Conference (ISWC2002)*, Sardinia, Italia, heinäkuu 2002, Springer Verlag.
- BL99 Berners-Lee, T., *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor*. Harper San Francisco, 1999.
- BLHL01 Berners-Lee, T., Hendler, J. ja Lassila, O., The Semantic Web. *Scientific American*, 284,5(2001), sivut 34–43.
- Bol04 Bolzer, O. M., *A Brief History of RDF Serialization Formats*, 2004. <http://www.fakeroot.net/sw/rdf-formats/>. [15.1.2005]

- Bur00 Burke, R., Knowledge-based recommender systems. Teoksessa *Encyclopedia of Library and Information Science Vol. 69*, Marcel Dekker, 2000, sivut 180–201.
- FHH04 Fikes, R., Hayes, P. ja Horrocks, I., OWL-QL - A Language for Deductive Query Answering on the Semantic Web. *Journal of Web Semantics*, 2,1(2004), sivut 19–29.
- FLA Flamenco Project. <http://bailando.sims.berkeley.edu/flamenco.html>. [2.3.2005]
- FOA04 D. Brickley ja L. Miller, *FOAF Vocabulary Specification*, syyskuu 2004. <http://xmlns.com/foaf/0.1/>. [2.3.2005]
- Fos97 Foskett, D., Thesaurus. Teoksessa *Readings in Information Retrieval*, K. Sparck Jones, P. W., toimittaja, Morgan Kauffman, 1997, sivut 111–134.
- GC02 Geroimenko, V. ja Chen, C., toimittajat, *Visualizing the Semantic Web*. Springer-Verlag New York, Inc., 2002.
- Gru93 Gruber, T. R., A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, 5,2(1993), sivut 199–220.
- H+02 Hearst, M. et al., Finding the Flow in Web Site Search. *Communications of The ACM*, 45,9(2002), sivut 42–49.
- HBEV04 Haase, P., Broekstra, J., Eberhart, A. ja Volz, R., Comparison of RDF Query Languages. *Proceedings of the Third International Semantic Web Conference (ISWC2004)*, Hiroshima, Japani, marraskuu 2004, Springer Verlag.

- Hen01 Hendler, J., Agents and the Semantic Web. *IEEE Intelligent Systems*, 16,2(2001), sivut 30–37.
- HPSvH03 Horrocks, I., Patel-Schneider, P. F. ja van Harmelen, F., From SHIQ and RDF to OWL: The Making of a Web Ontology Language. *Journal of Web Semantics*, 1,1(2003), sivut 7–26.
- IWE Intelligent Web Services (IWebS) Project. <http://cs.helsinki.fi/group/iwebs/>. [2.3.2005]
- JHS00 Julkishallinnon www-sivuston suunnittelun ohjeet. Tekninen raportti JHS 129, Julkisen hallinnon tietohallinnon neuvottelukunta JUHTA, 2000.
- JHS04 Palvelutietojen ryhmittely ja osoitteet asiointia varten monta toimialaa kattavissa julkisen sektorin portaaleissa. Tekninen raportti JHS 145, Julkisen hallinnon tietohallinnon neuvottelukunta JUHTA, 2004.
- LRPF04 Lara, R., Roman, D., Polleres, A. ja Fensel, D., An Evaluation of Semantic Web Portals. *IADIS Applied Computing International Conference*, Lissabon, Portugali, maaliskuu 2004.
- M⁺03 Maedche, A. et al., Semantic portal - The SEAL approach. Teoksessa *In Spinning the Semantic Web*, Fensel, D. et al., toimittajat, MIT Press, 2003, sivut 317–359.
- Mar95 Marchionini, G., *Information Seeking in Electronic Environments*. Cambridge University Press, 1995.
- McG03 McGuinness, D. L., Ontologies Come of Age. Teoksessa *In Spinning the Semantic Web*, Fensel, D. et al., toimittajat, MIT Press, 2003, sivut 171–194.

- Mel00 Melnik, S., *Storing RDF in a relational database*, 2000. <http://www-db.stanford.edu/~melnik/rdf/db.html>. [2.3.2005]
- MHSV04 Mäkelä, E., Hyvönen, E., Saarela, S. ja Viljanen, K., OntoViews - A Tool for Creating Semantic Web Portals. *Proceedings of the Third International Semantic Web Conference (ISWC2004)*, Hiroshima, Japani, marraskuu 2004, Springer Verlag.
- N⁺01 Noy, N. et al., Creating Semantic Web Contents with Protege-2000. *IEEE Intelligent Systems*, 16,2(2001), sivut 60–71.
- OBD04 O’Murchu, I., Breslin, J. ja Decker, S., Online Social and Business Networking Communities. *Proceedings of the ECAI 2004 Workshop on Application of Semantic Web Technologies to Web Communities*, Valencia, Espanja, elokuu 2004.
- OS01 Oberle, D. ja Syns, P., The Knowledge Portal ”OntoWeb”. Teoksessa *Handbook on Ontologies*, Staab, S. ja Studer, R., toimittajat, Springer-Verlag, 2001, sivut 499–517.
- OWL04a W3C Recommendation, *OWL Web Ontology Language Overview*, helmikuu 2004. <http://www.w3.org/TR/owl-features/>. [2.3.2005]
- OWL04b W3C Recommendation, *OWL Web Ontology Language Use Cases and Requirements*, helmikuu 2004. <http://www.w3.org/TR/webont-req/>. [2.3.2005]
- QHK03 Quan, D., Huynh, D. ja Karger, D. R., Haystack: A platform for authoring end user semantic web applications. *Proceedings of the Second International Semantic Web Conference (ISWC2003)*, Sanibel Island, Florida, USA, lokakuu 2003, Springer Verlag.

- QK04 Quan, D. ja Karger, D. R., How to Make a Semantic Web Browser. *Proceedings of the 13th International World Wide Web Conference*, New York, NY, USA, toukokuu 2004, ACM Press.
- RDF04a *Resource Description Framework (RDF): Concepts and Abstract Syntax*, helmikuu 2004. <http://www.w3.org/TR/rdf-concepts/>. [2.3.2005]
- RDF04b W3C Recommendation, *RDF Primer*, helmikuu 2004. <http://www.w3.org/TR/rdf-primer/>. [2.3.2005]
- RDF04c W3C Recommendation, *RDF Vocabulary Description Language 1.0: RDF Schema*, helmikuu 2004. <http://www.w3.org/TR/rdf-schema/>. [2.3.2005]
- RSC04 Reynolds, D., Shabajee, P. ja Cayzer, S., Semantic Information Portals. *Proceedings of the 13th International World Wide Web Conference on Alternate track papers & posters*, New York, NY, USA, toukokuu 2004, ACM Press.
- RSCS04 Reynolds, D., Shabajee, P., Cayzer, S. ja Steer, D., SWAD-Europe Deliverable 12.1.7: Semantic Portals Demonstrator - Lessons Learnt, 2004. http://www.w3.org/2001/sw/Europe/reports/demo_2_report/. [2.3.2005]
- Saa04 Saarela, S., Näkymäpohjainen RDF-haku, Pro gradu. Tekninen raportti C-2004-48, Tietojenkäsittelytieteen laitos, Helsingin yliopisto, 2004.
- SKO SKOS Project. <http://www.w3.org/2004/02/skos/>. [2.3.2005]
- SMS02 Sellen, A., Murphy, R. ja Shaw, K. L., How Knowledge Workers Use the Web. *Proceedings of the SIGCHI conference on Human factors in computing systems, CHI Letters 4(1)*, Minneapolis, Minnesota, USA, huhtikuu 2002, ACM.

- So00 Staab, S. et al., Semantic Community Web Portals. *Proceedings of the 9th International World Wide Web Conference*, Amsterdam, Hollanti, toukokuu 2000, Elsevier.
- SPA04 W3C Working Draft, *SPARQL Query Language for RDF*, lokakuu 2004. <http://www.w3.org/TR/rdf-sparql-query/>. [2.3.2005]
- Ste03 Steer, D., Brown Sauce: an introduction. Tekninen raportti HPL-2003-10, Hewlett-Packard Laboratories, 2003.
- SVG03 W3C Recommendation, *Scalable Vector Graphics (SVG) 1.1 Specification*, tammikuu 2003. <http://www.w3.org/TR/SVG/>. [2.3.2005]
- SWP SWI-Prolog. <http://www.swi-prolog.org/>. [2.3.2005]
- TAAK04 Teevan, J., Alvarado, C., Ackerman, M. S. ja Karger, D. R., The Perfect Search Engine is not Enough: A Study of Orienteering Behavior in Directed Search. *Proceedings of the 2004 conference on Human factors in computing systems*, Wien, Itävalta, 2004, ACM Press.
- UJ99 Uschold, M. ja Jasper, R., A Framework for Understanding and Classifying Ontology Applications. *Proceedings of the ICAJ-99 Workshop on Ontologies and Problem Solving Methods*, Tukholma, Ruotsi, elokuu 1999.
- URI98 *Uniform Resource Identifiers (URI): Generic Syntax*, elokuu 1998. <http://www.ietf.org/rfc/rfc2396.txt>. [2.3.2005]
- WAI99 W3C Recommendation, *Web Content Accessibility Guidelines 1.0*, toukokuu 1999. <http://www.w3.org/TR/WAI-WEBCONTENT/>. [2.3.2005]

- Y⁺03 Yee, K.-P. et al., Faceted Metadata for Image Search and Browsing. *Proceedings of the conference on Human factors in computing systems*. ACM Press, huhtikuu 2003.
- Z⁺04 Zhang, J. et al., Relation Browser++: an information exploration and searching tool. *Proceedings of the The National Conference on Digital Government Research*, toukokuu 2004.