

hyväksymispäivä arvosana

arvostelija

## **Toimijaontologiat ja niiden käyttö semanttisessa webissä**

Jussi Kurki

Helsinki 1.3.2011

HELSINGIN YLIOPISTO  
Tietojenkäsittelytieteen laitos

Tiedekunta — Fakultet — Faculty		Laitos — Institution — Department	
Matemaattis-luonnontieteellinen tiedekunta		Tietojenkäsittelytieteen laitos	
Tekijä — Författare — Author			
Jussi Kurki			
Työn nimi — Arbetets titel — Title			
Toimijaontologiat ja niiden käyttö semanttisessa webissä			
Oppiaine — Läroämne — Subject			
Tietojenkäsittelytiede			
Työn laji — Arbetets art — Level		Aika — Datum — Month and year	
Pro gradu -tutkielma		1.3.2011	
		Sivumäärä — Sidoantal — Number of pages	
		55 sivua + 14 liitesivua	
Tiivistelmä — Referat — Abstract			
<p>Henkilöt, paikat, ajat, teokset ja muut niin kutsutut nimetyt entiteetit ovat keskeisessä osassa tiedon eristämisessä ja haussa. Nimettyjen entiteettien avulla voidaan vastata kysymyksiin, jotka ovat muotoa kuka, mitä, missä, milloin. Lisäksi entiteetit toimivat esimerkiksi kirjastojärjestelmissä keskeisinä haku- ja liitäntäpisteinä. Tyypillinen tarve on löytää johonkin tekijän kirjoittamat kaikki teokset tai jonkin teoksen kaikki eri versiot. Tällöin on tärkeää kyetä erottelmaan samannimiset tekijät ja teokset.</p> <p>Tässä tutkielmassa keskitytään henkilö- ja organisaatiotiedon mallintamiseen ja hyödyntämiseen. Väitämme semanttisen webin tuovan lisäarvoa kirjastoissa ja muissa organisaatioissa kerätyille auktoriteettitiedoille (esimerkiksi kirjailijarekisterit) paremman monikielisuuden, linkittyvyyden, siirrettävyyden ja uudelleenkäytettävyyden kautta. Keskeisinä ratkaisukeinoina esitetään semanttisen webin avoimia standardeja datan esitysmuodoksi, jaettuja sanastoja tiedon kuvailuun sekä keskitettyjä kaikille avoimia entiteettirekistereitä tiedon hakuun. Entiteettirekistereissä toimijatiedot julkaistaan WWW:ssä sekä helppokäyttöisten käyttöliittymien että erilaisten avoimien sovellusrajapintojen kautta.</p> <p>Kirjastomaailmassa entiteettien hallintaan käytetään tyypillisesti ns. auktoriteettitietokantaa, jossa jokaisesta entiteetistä tehdään tietue, joka noudattaa tiettyä tarkasti määriteltyä muotoa ja joka tunnustetaan kontrolloidulla nimen muodolla. Semanttiseen webiin pohjautuvassa mallissa entiteetti tunnustetaan yleismaailmallisella muuttumattomalla tunnisteella (URI) ja data koodataan laajennettavalla graafimuotoisella esityksellä (RDF). Edellä mainittuihin tekniikoihin pohjautuen olemme määritelleet olemassa oleviin sanastoihin perustuvan toimijaontologian. Lisäksi olemme toteuttaneet ”ONKI People” -ontologiapalvelun, jota voidaan hyödyntää keskitettynä entiteettirekisterinä sekä moninäkömähakuliittymän että sovellusrajapintojen kautta. Lisäksi esittelemme tapaustutkimuksena toimijaontologian hyödyntämistä kulttuuriportaalissa.</p> <p>ACM Computing Classification System (CCS):  H.3.5 [Information Storage and Retrieval]: Online Information Services  H.5.4 [Information Interfaces and Presentation]: Hypertext/Hypermedia  I.2.4 [Artificial Intelligence]: Knowledge Representation Formalisms and Methods</p>			
Avainsanat — Nyckelord — Keywords			
semanttinen web, ontologiat, ontologiapalvelut, moninäkömähaku			
Säilytyspaikka — Förvaringsställe — Where deposited			
Muita tietoja — övriga uppgifter — Additional information			

# Sisältö

<b>1</b>	<b>Johdanto</b>	<b>1</b>
<b>2</b>	<b>Toimijan idea</b>	<b>3</b>
2.1	Erlaisia toimijoita . . . . .	3
2.1.1	Henkilö . . . . .	4
2.1.2	Ryhmä . . . . .	4
2.2	Identiteetti . . . . .	4
2.2.1	Henkilörekisterit . . . . .	5
2.2.2	Verkkoidentiteetit . . . . .	6
2.2.3	Yksityisyys ja tietosuoja . . . . .	6
2.3	Roolit . . . . .	8
2.4	Sosiaaliset verkostot . . . . .	8
<b>3</b>	<b>Toimijatietojen mallintaminen</b>	<b>10</b>
3.1	Vaatimukset . . . . .	10
3.1.1	Tunniste . . . . .	11
3.1.2	Nimen esittäminen . . . . .	12
3.1.3	Rooliluokitukset . . . . .	13
3.1.4	Kansallisuus ja paikkatiedot . . . . .	15
3.1.5	Aikaan liittyvät tiedot . . . . .	17
3.2	Olemassa olevia malleja . . . . .	18
3.2.1	MARC . . . . .	18
3.2.2	ULAN . . . . .	18
3.2.3	FOAF . . . . .	22
3.2.4	REL . . . . .	22
3.2.5	BIO . . . . .	24
3.2.6	Muita sanastoja . . . . .	24
3.3	Toimijaontologia . . . . .	25

3.3.1	Toimijaontologian URI-tunnisteet . . . . .	26
3.3.2	Nimikentät . . . . .	27
3.3.3	Rooli ja kansallisuus . . . . .	27
<b>4</b>	<b>Toimijoiden hakeminen - ONKI People</b>	<b>28</b>
4.1	Moninäkömahaku . . . . .	28
4.2	Sosiaalisen verkoston visualisointi . . . . .	32
4.3	Sovellusrajapinnat . . . . .	32
4.4	Tekninen toteutus . . . . .	33
4.5	Arviointi . . . . .	34
<b>5</b>	<b>Tapaustutkimus: Kulttuurisampo</b>	<b>38</b>
5.1	Toimijatietojen näyttäminen . . . . .	38
5.2	Sosiaalisten polkujen hakeminen . . . . .	39
5.2.1	Idea . . . . .	41
5.2.2	Tekninen toteutus . . . . .	41
5.2.3	Tulokset . . . . .	42
5.2.4	Arviointi . . . . .	43
<b>6</b>	<b>Jatkotutkimus</b>	<b>43</b>
6.1	Käyttäjän mallinnus . . . . .	44
6.2	Hajautettu palvelu . . . . .	45
<b>7</b>	<b>Yhteenveto</b>	<b>46</b>
	<b>Lähteet</b>	<b>49</b>
	<b>Liitteet</b>	
	<b>1 Toimijaontologian skeema</b>	
	<b>2 Rooli- ja kansallisuusontologiat</b>	

# 1 Johdanto

Henkilöt, paikat ja ajat (ns. nimetyt entiteetit) toimivat keskeisinä tiedon haku- ja liitännäispisteinä. Kirjastojärjestelmissä tyypillinen tarve on löytää jonkin tekijän kirjoittamat kaikki teokset tai jonkin teoksen kaikki eri versiot. Perinteisenä lähtökohta on tehdä toimijoista rekisteri, ns. auktoriteettitietokanta, johon listataan toimijan nimi ja muita keskeisiä ominaisuuksia haun ja yksilöinnin helpottamiseksi [Til04, Tay06].

Auktoriteettijärjestelmän avulla voidaan

1. löytää teos, jonka tekijä, nimike tai aihe on tunnettu (finding objective),
2. luetella kaikki teokset, jotka liittyvät johonkin tekijään (collocating objective).

Monissa kirjastoissa käytetään Yhdysvaltain kongressin MARC-tietokantaformaattia<sup>1</sup>. MARC pohjaa 70-luvulla suunniteltuun tiedonesitysmalliin ja sitä vaivaavat lukuisat ongelmat<sup>2</sup>. Esimerkiksi eri maissa on käytössä erilaisia versioita MARC:sta, kansallisuustietoja ei koodata yhdenmukaisesti ja translitterointiskeemat vaihtelevat maakohtaisesti. Lisäksi tietoja on hyvin vaikea jakaa kirjastojen välillä. Käytännössä jokainen kirjasto ylläpitää omaa auktoriteettikantaansa.

Semanttinen web<sup>3</sup> tarjoaa tehokkaita työkaluja entiteettien yksilöintiin ja yhdistämiseen [VA08, BPSV09]. Keskeisinä rakennuspalikoina toimivat universaalit muuttumattomat URI-tunnisteet (Uniform Resource Identifier<sup>4</sup>) ja yksinkertainen graafimuotoinen tietomalli RDF (Resource Description Framework<sup>5</sup>).

RDF-tietomallissa data kuvataan kolmikkoina. RDF-kolmikko koostuu subjektista, predikaatista ja objektista. Kolmikön jäsenet ovat joko URI-resursseja, anonyymejä resursseja tai literaaleja.

Toisiinsa viittaavista kolmikoista muodostuu verkko, josta voidaan koneellisesti päätellä tietoa. Kuvassa 1 on havainnollistettu henkilötiedoista muodostuvaa graafia.

Koska URI-tunnisteet ovat yleismaailmallisia ja pysyviä, periaatteessa kuka tahansa voi viitata esimerkiksi minun RDF-mallissa määriteltyihin resursseihin. Teoreettinen idea ei kuitenkaan toteudu ilman tehokkaita tapoja hakea ja jakaa URI-resursseja.

---

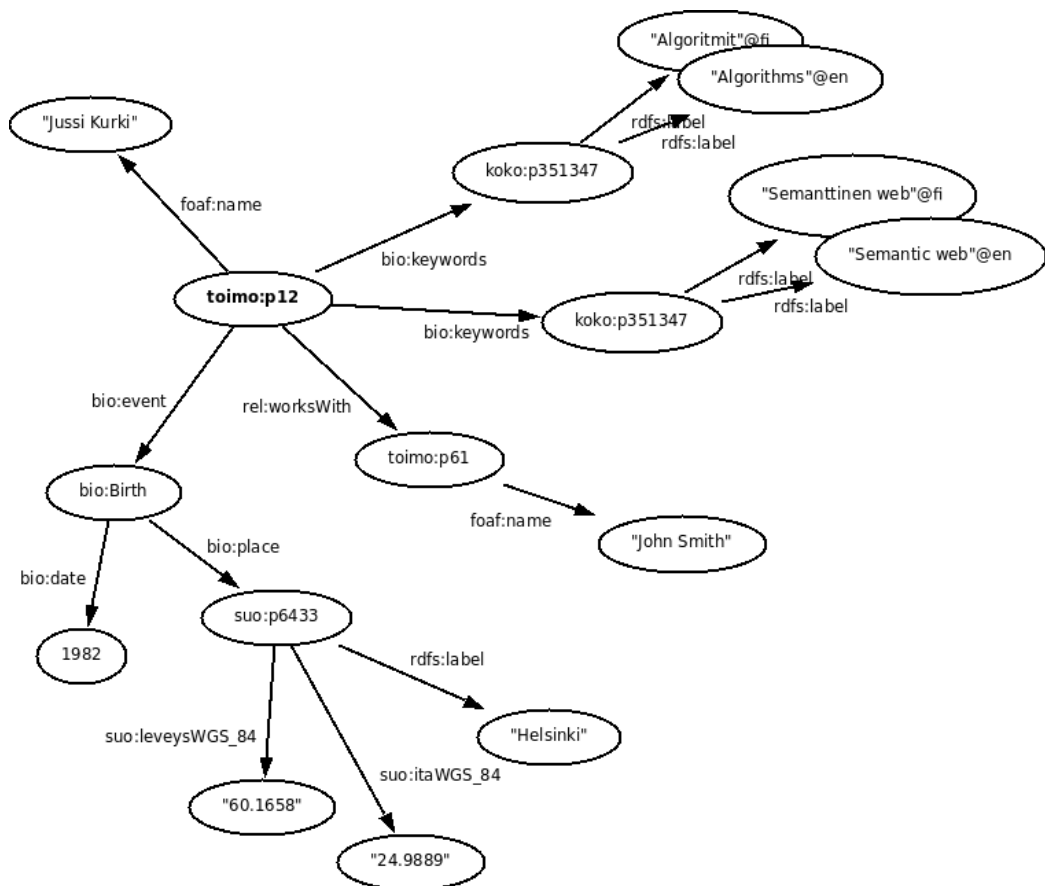
<sup>1</sup><http://www.loc.gov/marc/>

<sup>2</sup>[http://www.loc.gov/catdir/bibcontrol/tillet\\_paper.html](http://www.loc.gov/catdir/bibcontrol/tillet_paper.html)

<sup>3</sup><http://www.w3.org/2001/sw/>

<sup>4</sup><http://www.w3.org/Addressing/>

<sup>5</sup><http://www.w3.org/RDF/>



Kuva 1: Resurssien kytkeytyminen henkilöön semanttisessa webissä.

Tässä tutkielmassa esitellään keskeisinä tuloksina semanttisen webin tekniikoihin pohjautuva tietomalli toimijoiden (henkilöt, ryhmät ja organisaatiot) kuvaamiseen sekä mallin päälle rakennettuja sovelluksia niin yksittäisten toimijoiden hakemiseen kuin sosiaalisten verkostojen visualisointiin. Tavoitteena on hamotella yleinen infrastruktuuri toimijatietojen hallintaan. Keskeisenä teesinä on, että semanttisen webin keinoin toteutettu järjestelmä tukee paremmin monikielisyyttä, linkittyvyyttä, siirrettävyyttä ja uudelleenkäytettävyyttä verrattuna perinteisiin kirjailija- ja teosrekistereihin. Tutkielman painopiste on toimijatiedon esittämisessä ja hallinnassa. Käyttäjämalleja – esimerkiksi toimijatietoihin perustuvaa päättelyä, automaattista profilointia ja yhteisöllistä suosittelua – sivutaan jatkotutkimusosuuksissa.

Tutkielma rakentuu seuraavasti. Luvussa 2 määritellään toimijan käsite ja esitellään keskeiset toimijaa kuvaavat ominaisuudet. Luvussa 3 käydään läpi vaatimuksia ja olemassa olevia malleja toimijatietojen kuvaamiseen. Lisäksi määritellään toimijaontologia henkilöiden ja ryhmien ominaisuuksien ja niiden arvojoukkojen kuvaamiseen. Luvussa 4 esitellään ”ONKI People” -palvelu toimijoiden hakemiseen ja sosiaalisten suhteiden visualisointiin. Luvussa 5 tutkitaan toimijatiedon yhdistämistä ja soveltamista Kulttuurisampo-portaalissa. Luvussa 6 pohditaan joitakin jatkotutkimusmahdollisuuksia. Luvussa 7 tehdään lyhyt yhteenveto tutkielman aihepiiristä.

## **2 Toimijan idea**

Toimijalla tarkoitetaan esimerkiksi henkilöä, yritystä tai organisaatiota. Toimija nimensä mukaisesti toimii ja muuttaa asioiden tilaa jollain mahdollisesti päämäärähakuisella tavalla.

Tämä luku esittelee toimijatietojen mallintamisen lähtökohtia. Luvussa määritellään vapaasti kuvaillen toimijan käsite ja toimijan määrittelevät tärkeimmät ominaisuudet. Tarkoituksena on perustella ja motivoida seuraavissa luvuissa tehtyjä teknisiä valintoja.

### **2.1 Erilaisia toimijoita**

Henkilö muodostaa pienimmän yksittäisen komponentin toimijoiden joukossa. Oikeastaan toimijat voidaan jakaa kahteen luokkaan: henkilöihin ja niiden muodostamiin erilaisiin ryhmiin. Joissain yhteyksissä toimijoina voidaan nähdä myös älykkäät tietokoneohjelmat. Tässä tutkielmassa kuitenkin oletetaan, että toimintaa ohjaa jol-

lain tavalla ihminen. Oletus ei sinänsä ole välttämätön, mutta se selkeyttää jakoa tekoälyn älykkäiden agenttien tutkimuksen [WJ95] ja toimijatiedon mallintamisen [Che76] välillä.

### 2.1.1 Henkilö

Henkilö on yksi ihminen, jolla on oma identiteetti. Juuri muuta pysyvää ei henkilössä olekaan. Nimet vaihtuvat, ikää karttuu ja asuinpaikka muuttuu. Ihmisestä voi kirjoittaa elämäkertoja, sijoittaa sukupuuhun, luetella lempiartisteja, -ruokia tai -värejä. Itsensä voi nähdä suomalaisena, opiskelijana, ihmisenä, maapallon asukkaana tai iloisena pykniikkona. Silti mikään persoonallisuusluokittelu tai faktalista ei riitä kuvaamaan ihmistä tyhjentävästi.

### 2.1.2 Ryhmä

Ryhmä on joukko yksilöitä, joita yhdistää jokin tekijä. Esimerkkejä ryhmistä ovat yritys, organisaatio, seura, yhdistys, oppilaitos, museo ja säätiö. Joissakin tapauksissa ryhmällä itsellään ei ole suurempaa merkitystä vaan ryhmä saa merkityksen jäsenistään. Toisissa tapauksissa, esimerkiksi monikansallisen suuryrityksen kohdalla, yksittäiset jäsenet eivät välttämättä ole merkittäviä. On vaikeaa löytää mitään yksittäistä ominaisuutta, joka olisi kaikille yhteinen. Ryhmä voi muodostua ja hajota nopeasti. Ryhmä voi olla ilman ainuttakaan jäsentä tai siihen saattaa kuulua tuhansia yksilöitä. Ryhmät voivat koostua myös toisista ryhmistä.

## 2.2 Identiteetti

Pysyvä identiteetti mahdollistaa luotettavan informaation kulun eri toimijoiden välillä. Passit, ajokortit, kanta-asiakaskortit ja loputtomat henkilörekisterit luovat tunnistajien verkoston, jossa ihmiset ja heidän tekemiset voidaan rekisteröidä ja yksilöidä. Kun asiakas näyttää kassalla luottokorttia, hänet tunnistetaan yhteiskunnan tietyn kansalaiseksi, jolla on pankkihistoria ja luottokelpoisuus. Kanta-asiakaskortit tallentavat ostohistoriaa ja kaupat räätälöivät tarjoukset vastaamaan ostotottumuksia, linja-autoyritykset keräävät matkustajatietoja, matkapuhelinoperaattorit puhelutietoja ja verkkopalvelut sivupyynnölokeja.



### 2.2.1 Henkilörekisterit

Identiteettitietoja voidaan hallita ja ylläpitää henkilörekisterin avulla. Suomen henkilötietolaissa henkilörekisterillä tarkoitetaan ”käyttötarkoituksensa vuoksi yhteensuoravista merkinnöistä muodostuvaa henkilötietoja sisältävää tietojoukkoa, jota käsitellään osin tai kokonaan automaattisen tietojenkäsittelyn avulla taikka joka on järjestetty kortistoksi, luetteloksi tai muulla näihin verrattavalla tavalla siten, että tiettyä henkilöä koskevat tiedot voidaan löytää helposti ja kohtuuttomitta kustannuksitta”<sup>6</sup>.

Henkilörekisterin suunnittelussa kenttien valinta on keskeisin kysymys. Kun rekisteri on suunniteltu pitää Suomessa tietosuojavaltuutetulle tehdä virallinen rekisteriseloste, joka sisältää tarkan kuvauksen rekisterin sisältämistä tiedoista ja sen tarkoituksesta<sup>7</sup>.

Väestörekisteriseloste<sup>8</sup> listaa mm. seuraavia tietoja tallennettaviksi: henkilötunnus, nimet, osoite, perhesuhdetiedot, kansalaisuustiedot, kuolinaika ja henkilön ilmoittamat tiedot ammatista. Selosteen mukaan tiedot on tarkoitettu käytettäväksi tuomioistuimenmenettelyssä, hallinnollisessa päätöksenteossa ja tilastojen laatimisessa. Lisäksi mainitaan mielipide- ja markkinatutkimukset sekä suoramarkkinointi ja osoitepalvelut.

Henkilö- ja asiakasrekisterit jakavat mielipiteitä. Ne mahdollistavat entistä parempia yksilöityjä palveluita. Toisaalta ne sisältävät mahdollisuuden tietovuotoon ja ihmisten häiritsemiseen. Hyvin toteutettuna etuja on kuitenkin molemmille osapuolille. Mitä vaivattomammin ja varmemmin voi vakuuttaa toisen omasta identiteetistä sitä helpompi on maksaa verkossa laskuja, matkustaa ulkomaille, selailla omia sairauskertomuksia ja uusia myöhästyneitä kirjastolainoja.

Henkilötietojen sovelluskohtaisesta tallentamisesta seuraa, että jokaisen palveluntarjoajan täytyy ylläpitää omaa rekisteriä, palkata omat tietoturva-asiantuntijat ja vakuutella käyttäjät siitä, että palvelua on turvallista käyttää. Käyttäjän kannalta tämä johtaa siihen, että hänen pitää rekisteröityä kymmeniin eri palveluihin, syöttää aina samat tiedot ja keksiä taas yksi uusi hyvä salasana ja vielä muistaakin se. Käyttäjän pitää olla alati valppaana tunnistamaan huijausta yrittävät web-sivut ja sähköpostiviestit. Oikeastaan sähköpostiin ei voi luottaa ja huijaussivuja voi pystyttää lähes kuka tahansa. Käyttäjää profiloidaan usein myös automaattisesti, jolloin

---

<sup>6</sup><http://www.finlex.fi/fi/laki/ajantasa/1999/19990523>

<sup>7</sup><http://www.tietosuoja.fi/>

<sup>8</sup><http://vaestorekisterikeskus.fi>

yksityisyyden rajat hämärtyvät entisestään.

Salausmenetelmät on pitkälle kehitettyjä ja teoria luottamuksellisuudesta, todenuksesta, eheydestä jne. on laajasti pohdittua. Verkossa tapahtuvat huijaukset ovat usein yksinkertaisia, mutta silti ne toimivat. Sähköposti käyttää protokollaa, jota ei ole suunniteltu sen nykyiseen suurusluokkaan. Rekisteröitymistä vaativia palveluita ilmestyy kuin sieniä sateella. Lukuisat eri sivustot kertovat tietoa rock-tähdistä, suurista tiedemiehistä ja kansallisankareista. Viimeisin suuri sisällöntuottaja on käyttäjä itse. Kaikkialla on henkilöitä, tietoa ja tiedonsiirtoa. Käyttäjät ovat syystäkin huolestuneita yksityisyydestään.

### 2.2.2 Verkkoidentiteetit

Verkossa henkilöllä voi olla useita identiteettejä. Identiteettien keveys ja verkon kattavuus on luonut lähes jokaiselle alakulttuurille omat fooruminsa, joissa käyttäjät voivat esiintyä oman nimensä sijaan nimimerkin turvin [WC05].

Verkkossa identiteettiä voi vaihtaa helposti ja samalla henkilöllä voi useita eri persoonia eri palstoilla. Palstoilla ei usein pyritä täydelliseen anonymiteettiin, sillä sanan painoarvo laskee, jos kuka vain voi sanoa mitä vain. Esimerkiksi keskustelupalstoilla on usein tärkeää, että kunkin jäsenen identiteetti pysyy samana, joilloin henkilölle voi ajan mukaan muodostua maine järkevänä ja vakavastiotettavana keskustelijana. Pysyvyys ja maine luovat omassa ympäristössään toimivan aseman. Virtuaaliset henkilöt muodostavat ryhmiä, jossa jokaisella on paikkansa. Käyttäjä voi kuitenkin sujuvasti vaihtaa rooliaan ja olla jokaisessa ympäristössä toisista ympäristöistä riippumaton täysipainoinen jäsen.

Ryhmien merkitys käyttäjäkeskeisissä palveluissa on suuri. Tiettyyn ryhmään kuulumisella viestitetään omia mielipiteitä. Ryhmien avulla paikataan usein järjestelmän ilmaisuvoiman rajoitteita. Omassa käyttäjäprofiilissaan ei voi järkevästi ilmaista kuin hyvin rajallisen määrän asioita. Jos vapaassa kuvailussa mainitsee pitäväänsä psykedeelisestä konemusiikista, hukkuu tieto loputtomaan määrään muuta tietoa. Jos liittyy ryhmään ”psytrance”, tulee tieto automaattisesti kaikkien saataville. Ryhmät muodostavatkin tietynlaisen yhteisen sanaston.

### 2.2.3 Yksityisyys ja tietosuojaja

Cranorin ja kumppanien mukaan [CRA99] yli 80% Internetin käyttäjistä on huolestuneita yksityisyyden suojasta. Cranorin haastattelemat henkilöt olivat vähintään-

kin käyttäneet WWW-sivuja ja sähköpostia. Myönteisimmin suhtauduttiin tietojen luovuttamiseen räätälöityjen palveluiden vastineeksi. Negatiivisimmin suhtauduttiin käyttäjätietojen perusteella kohdistettuun sivurajat ylittävään mainostamiseen.

Broder [Bro00] pohtii käyttäjien yksityisyyden ja palveluntarjoajien markkinoinnin välisten etujen ristiriitaa teknisestä näkökulmasta. Kysymykset ovat saaneet yhä suurempaa huomiota tiedonlouhintamenetelmien yleistyessä ja kehittyessä (esittely aihepiiristä löytyy esim. [SCDT00]). Keskeisin tiedonkeräysmenetelmä on HTTP-palvelimen sivupyyntölokin louhiminen. Pyynnöt sisältävät tiedon niin haetusta sivusta kuin aiemmasta katsotusta sivusta (referer). Pyynnössä näkyy sivun pyytäneen koneen osoite, josta voidaan joissain tapauksissa päätellä yrityksen tai jopa yksittäisen henkilön identiteetti. Huolta aiheuttaa myös uuden IPv6-protokollan konekohtaiset yksilölliset osoitteet, joka tekee yksittäisen käyttäjän jäljittämisen nykyistä helpommaksi [ZDLW10].

OECD julkaisi 80-luvulla ohjeistuksen<sup>9</sup> yksityisyyden suojasta ja henkilötietojen siirtämisestä. Tähän perustuen EU:ssa on työstetty tarkempaa direktiiviä 95/46/EC<sup>10</sup>. Henkilötietojen keräykseen liittyy esimerkiksi seuraavia periaatteita:

- Kohteelle kerrotaan tietojen keräämisestä
- Tietoja tulee kerätä vain ilmoitettuun tarkoitukseen
- Kohde on antanut suostumuksensa tietojen keräämiseen
- Kerätyt tiedot tulee pitää ulkopuolisilta salassa
- Kohteelle tulee ilmoittaa kuka tietoja kerää
- Kohteella tulee olla pääsy hänestä kerättyihin tietoihin

Automaattisen louhinnan kohdalla näiden periaatteiden noudattaminen on selvästi lähes mahdotonta. Jos tiedot kerätään esimerkiksi monimutkaiseen matemaattiseen malliin, kukaan ei voi konkreettisesti sanoa minkälaisia tietoja järjestelmä sisältää. Periaatteet eivät myöskään ota kantaa tilanteeseen, jossa käyttäjää ei voida eikä haluta yksilöidä.

---

<sup>9</sup>Guidelines on the Protection of Privacy and Transborder Flows of Personal Data – [http://www.oecd.org/document/18/0,3343,en\\_2649\\_34255\\_1815186\\_1\\_1\\_1\\_1,00.html](http://www.oecd.org/document/18/0,3343,en_2649_34255_1815186_1_1_1_1,00.html)

<sup>10</sup>Data Protection Directive – [http://www.cdt.org/privacy/eudirective/EU\\_Directive\\_.html](http://www.cdt.org/privacy/eudirective/EU_Directive_.html)

## 2.3 Roolit

Rooli on identiteetin lisäksi toinen keskeinen elementti toimijaa määriteltäessä. Roolien avulla voidaan kuvata toimijan luonnetta ja määritellä toimijan suhteita muihin toimijoihin. Masolo [MVB<sup>+</sup>04] kuvaa rooleja käsitteinä, joita tietyt entiteetit voivat esittää ollessaan vuorovaikutuksessa toisten entiteettien kanssa.

Masolo huomauttaa, että termiä rooli käytetään useissa eri merkityksissä. Esimerkiksi tietämyksen esittämisessä ja kuvauslogiikassa roolilla voidaan tarkoittaa mitä tahansa binääristä relaatiota. Loebe [Loe03] puolestaan jakaa roolit relelaatioihin liittyviin rooleihin, prosesseihin liittyviin rooleihin ja sosiaalisiin rooleihin.

Rooli kuulostaa väliaikaiselta, mutta voi olla osa toimijaa koko toimijan elinkaaren. Loebe mainitsee esimerkkinä, että muusikko on muusikko nukkuessaankin. Roolista tulee harkitun esittämisen vaikutelma, mutta toisaalta jokainen tilanne asettaa toimijan johonkin roolin. Näkökulmasta riippuen toimijan voi sanoa olevan luke-mattomissa eri rooleissa. Esimerkiksi kaupassa ihminen voi olla asiakkaan roolissa. Samoin se on tietyssä ajassa ja paikassa olevan henkilön roolissa. Toisaalta henkilö on ehkä katsovan tai kävelevän ihmisen roolissa.

Roolit sopivat hyvin toimijan mallintamiseen, sillä niiden avulla voidaan rajata tietojoukko kulloinkin kiinnostavaan tilanteeseen. Esimerkiksi toimijan ollessa roolissa ”asiakas”, saa toimija uudet kentät ”pankkitili” ja ”asiakasnumero”. Ja vaikka rooliin ei liittyisi mitään erityisiä ominaisuuksia, jo pelkkä roolissa oleminen antaa runsaasti informaatiota. Henkilö roolissa ”isä” kaipaa todennäköisesti erilaisia palveluita kuin henkilö roolissa ”yritysjohtaja”, vaikka kyse olisi samasta henkilöstä.

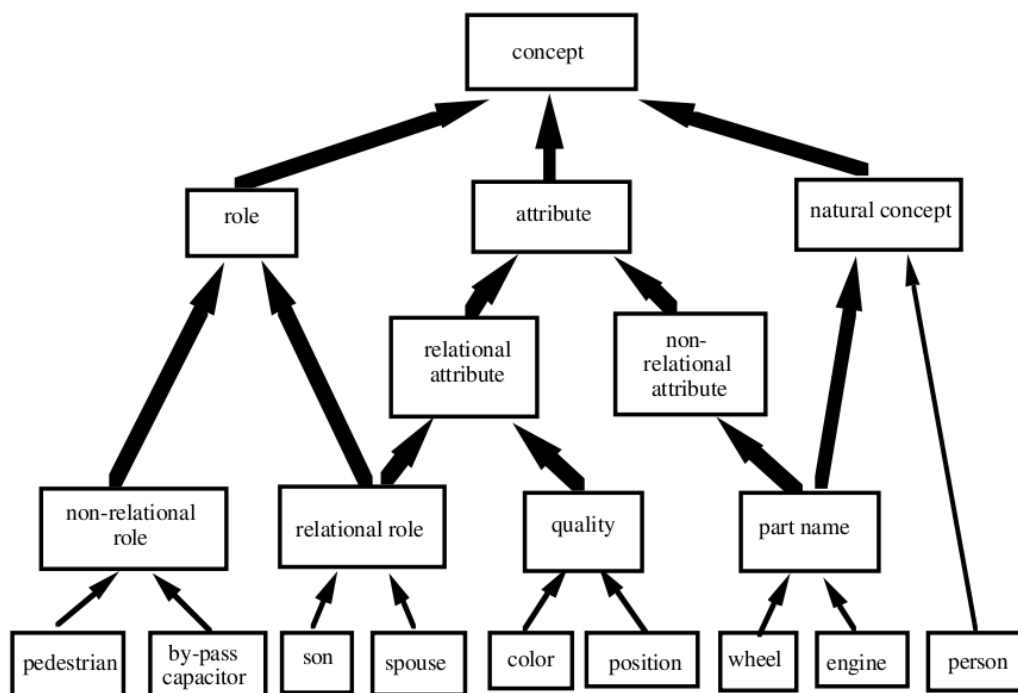
Ryhmään kuuluminen on merkitykseltään usein samankaltaista kuin roolin esittäminen. Esimerkiksi tietojärjestelmän oikeuksia voidaan hallita niin käyttäjäryhmien kuin roolien avulla. Esimerkiksi UNIX:ssa ryhmät ovat tiedostojen oikeuksien pohjana. Vastaavasti tietylle roolille voidaan määrittää tiettyjä oikeuksia. Keskeisenä erottavana tekijänä on, että roolin esittäminen on luonteeltaan paikallista ja ryhmään kuuluminen yleismaailmallista.

## 2.4 Sosiaaliset verkostot

Roolit kuvaavat toimijaa vain tietyllä hetkellä – niiden avulla on vaikea mallintaa toiminnan jatkuvuutta. Roolit eivät myöskään kuvaa toimijan suhteita toisiin toimijoihin. Henkilö roolissa ”isä” kertoo itse toimijasta jotain, mutta tieto kenen isästä

on kyse jää toissijaiseksi.

Guarino määrittelee kaksipaikkaisen roolin käsitteen [Gua92], joka on edellisessä kappaleessa kuvatun roolin kaltainen, mutta lisäksi ilmaisee relaatiota. Guarinon malli on esitetty kuvassa 2. Guarino jakaa käsitteet rooleiksi, attribuuteiksi ja luonnollisiksi käsitteiksi. Roolit ovat jaoteltu kahteen alityyppiin, joista toisella voidaan ilmaista myös relaatioita, esimerkiksi sukulaisuussuhteita.



Kuva 2: Guarinon attribuutti-luokittelu [Gua92].

Friend of a Friend -projektissa<sup>11</sup> tavoitteena on mallintaa sosiaalisten kontaktien verkostoa. Siinä suhteita määriteltäessä on lähdetty mahdollisimman yleisestä suhteesta, toisen henkilön tuntemisesta. Verkostot rakentuvat tämän määrittelyn varaan, sillä mallissa ei ole oleellista esimerkiksi kuinka hyvä ystävä on kyseessä. Monissa yhteyksissä tuttujen luokittelu ystäviin, tuttuihin jne. olisikin lähinnä kiusallista. Tärkeintä on, että henkilöiden välillä on kontakti, jonka kautta voi kommunikoida. Verkostojen kautta tieto ja vaikutteet leviävät nopeasti. Krauz ja kumppanit [KSS97] huomioivat, että asiantuntijat eivät usein pysty siirtämään taitojaan ja tietojaan vain kirjallisesti. Toisaalta asiantuntijat eivät aina haluakaan tuoda tietojaan julkisesti saatavaksi. Tällöin tiedonhakijan ongelmana on oikean asiantuntijan löytäminen. Krauz ehdottaa Internetin olemassa olevia sosiaalisia verkostoja ratkaisuksi.

<sup>11</sup>Friend of a Friend - <http://www.foaf-project.org/>

Hän listaa esim. kotisivujen linkit, keskustelupalstojen arkistot ja artikkelien kirjoittajat kiinnostaviksi louhittaviksi resursseiksi.

Myös ”PageRank” -tyyppistä relevanttien sivujen poimimista on tehty sosiaalisten verkostojen avulla [PMB<sup>+</sup>07]. Ideana on hyödyntää vertaisverkkojen paikallisia keskittymiä ja tietoja yhdistelemällä muodostaa yleiskuva verkosta. Keskittymät tunnistetaan yksinkertaisesti solmuun tulevien kaarten perusteella – solmu on kiinnostava, jos siihen tulee runsaasti kaaria.

Sosiologi Mark Granovetter havaitsi jo 70-luvulla, että ”tunteminen” vaikuttaa myös informaation perillemenoon [Gra73]. Kun tietää keneltä viestit ja määräykset tulevat, niihin luottaa ja niiden mukaan toimii. Ihmiset luottavat usein kontakteihinsa enemmän kuin virallisiin auktoriteetteihin. Granovetter havaitsi myös, että merkittävä osa työpaikkatiedosta kulkee sosiaalisten verkostojen kautta. Erityisesti ”heikot” linkit, kuten tutut ja vanhat koulu- ja työtoverit, toimivat keskeisinä yhdistävinä polkuina ihmisryhmien välillä. Juuri tätä näkökulmaa yritetään hyödyntää webin sosiaalisissa palveluissa.

### 3 Toimijatietojen mallintaminen

Tässä luvussa määritellään yleisiä vaatimuksia, joita toimijatietojen mallintamisessa tulee ottaa huomioon. Vaatimusten jälkeen esitellään olemassa olevia XML- ja RDF-pohjaisia sanastoja toimijatietojen esittämiseen. Lisäksi luvussa kuvataan toimijaontologiassa tehtyjä ratkaisuja. Luvun loppuun pohditaan koneellisesti muodostettuja käyttäjämalleja.

Toimijatietojen mallintamisen vaatimuksia on pohdittu [AS99]. Vastaan tulevat ongelmat esiintyvät myös muualla semanttissa webissä ja kirjastoalalla, kun tarvitaan pysyviä ja yhteentoimivia entiteettejä ja niiden luokitteluja [Fra03]. Myös tarve yhä suurempien tietomassojen automaattiseen integrointiin aiheuttaa kasvavaa huolta [FPS97]. Tietokantaskeemojen eroavaisuudet ja tietojen koodauksen epätäsmällisyydet tekevät yhdistelystä vaikeaa ja virhealtista. Tarvetta paremmille malleille niin tiedon esittämiseen kuin entiteettien yksilöintiin näyttäisi siis olevan.

#### 3.1 Vaatimukset

Tunniste ja nimi ovat keskeisimmät rakennuspalikat toimijatietoja kuvatessa. Näiden lisäksi vaatimuksissa pohditaan muita tietoja lähinnä monipuolisempien haku-

toiminnallisuuden mahdollistajina.

### 3.1.1 Tunniste

Tunniste voi olla paikallinen ja vain tietyssä kontekstissa yksilöivä. Tietokannan juokseva tunnistenumero on esimerkki paikallisesta tunnistenumerosta. Vastaavasti ystäväpiirissä tuttuja voidaan kutsua pelkillä etunimillä. Ratkaisu on toimiva, jos järjestelmän ei tarvitse kytkeytyä isompiin kokonaisuuksiin. Laajemmassa mittakaavassa henkilö voidaan yksilöidä esimerkiksi sosiaaliturvatunnuksella.

Yleismaailmallinen tunniste ratkaisisi toimijoiden yksilöintiongelmat, mutta on todennäköisesti mahdoton toteuttaa. Tuskin voi luetella ja numeroida jokaista maapallon ihmistä, ryhmää ja organisaatiota. Jo kaikkien kirjailijoiden tai taiteilijoiden luetteleminen yksikäsitteisesti on vaikeaa, ellei mahdotonta. Lisäksi tunnisteiden hallinta on ongelmallista, jos vain joku tietty auktoriteetti päättää uusista tunnisteista ja valvoo ”tunnisteavaruutta”.

Eräs tapa hoitaa ongelma on luopua kokonaan tunnisteista. Tunnistaminen tapahtuu silloin vain toimijan tietojen perusteella. Päätetään, että tietyt kentät yhdessä muodostavat identiteetin antavan ”klusterin” [FPS00]. Ongelmana on epävarmuus, sillä vaikka tiedot täsmäisivät ei voida olla täysin varmoja, että kyseessä on sama henkilö tai ryhmä. Lisäksi tietojen muuttaminen muuttaa myös tunnistavaa klusteria, vaikka toimijan identiteetin ei pitäisi periaatteessa muuttua tietojen muuttuessa. Jos kuitenkin tarvitaan täsmäystä, klusterointi on tehokas keino.

URI (Uniform Resource Identifier) on semanttisen webin malli tunnisteiden esittämiseen. URI:t ovat yksilöllisiä ja yleismaailmallisia, mutta toisin kuin jonkin tietyn auktoriteetin määrittelemät universaalit tunnisteet, URI:t mahdollistavat joustavamman käytön semanttisen webin semantiikan kautta. On esimerkiksi helppo määrittellä kaksi eri URI:n osoittamaa kohdetta samoiksi. Toisessa järjestelmässä voi olla henkilö <http://helsinki.fi/matti> ja toisessa <http://tkk.fi/matti>. Jos todetaan, että kyse on samasta henkilöstä, voidaan henkilötietojen osaksi liittää ”sama kuin”-tyyppinen kenttä, joka osoittaa toiseen URI:in. Jos lisäksi URI:t ovat todellisia web-soitteita (URL, Universal Resource Locator), voidaan osoitteessa tarjota lisätietoa, joka helpottaa yksilöimistä. URI:en käsittelyä helpottamaan voidaan rakentaa esimerkiksi ONKI-palvelin tyyppinen ratkaisu [KVH05, HVTS08].

### 3.1.2 Nimen esittäminen

Nimi on tärkein ominaisuus toimijaa tunnistettaessa ja yksilöitäessä. Nimi on harvoin pysyvä tai yksikäsitteinen. Eri ihmisillä on samoja nimiä, sukunimi voi vaihtua esimerkiksi avioliitossa ja nimiä voidaan translitteroida useilla eri tavoilla. Esimerkiksi picasson lukuisia nimiä on listattu taulukossa 1<sup>12</sup>. Nimeen perustuvien henkilörekistereiden hallintaa on pohdittu erityisesti kirjastoalalla, jossa auktoriteettitietokannat ovat olleet pitkään tutkimuksen alla [Zha01]. Esimerkiksi tiedonsaannin kannalta on tärkeää, että löydetään kaikki tietyn henkilön tuotokset erilaisista nimistä huolimatta.

Picasso, Pablo

Picasso, Pablo Diego José Francisco de Paula Juan Nepomuceno

Crispín Crispiniano de la Santissima Trinidad Ruiz Blasco

Picasso, Pablo Ruiz

Picasso, Pablo Ruiz y

Ruiz Picasso, Pablo

Ruiz y Picasso, Pablo

Ruiz, Pablo

Ruys Picasso, Pablo

Ruys, Pablo

Taulukko 1: Pablo Picasson nimen erilaisia kirjoitusasuja.

Nimen ja entiteetin välisiä suhteita on pohdittu mm. FRANAR-ryhmässä<sup>13</sup>. FRANAR-ryhmän mallissa nimistä, taiteilijanimistä, toimijoista ja persoonista muodostetaan omat entiteettinsä, jotka linkityksillä yhdistetään järkeviksi kokonaisuuksiksi (taulukko 2). Taulukon esimerkin kirjailija Barbara Mertz on julkaissut teoksia useilla taiteilijanimillä.

Verkkopalveluissa identiteetti kiteytyy usein käyttäjätunnukseen tai nimeen – joskaan ei välttämättä henkilön oikeaan nimeen. Tämä on johtanut mitä monipuolisimpiin lempinimiin. Tyypillinen ongelma rekisteröidytessä uuteen palveluun on keksiä jokin nimi tai käyttäjätunnus, joka ei vielä ole käytössä. Monissa uudemmissa palveluissa käytetäänkin esimerkiksi sähköpostiosoitetta yksilöivänä tunnuksena.

Esimerkiksi sosiaalinen musiikkiportaali Last.fm<sup>14</sup> käyttää nimeä tunnisteena (kuva

<sup>12</sup><http://www.getty.edu/research/tools/vocabularies/ulan/>

<sup>13</sup><http://www.ifla.org/VII/d4/wg-franar.htm>

<sup>14</sup><http://last.fm/>



Authorized heading

Mertz, Barbara

Information note/see also references:

Barbara Mertz also writes under the pseudonyms Barbara  
 Michaels and Elizabeth Peters.

For works written under those pseudonyms, search also under:

>> Michaels, Barbara, 1927-

>> Peters, Elizabeth

See also reference tracings:

<< Michaels, Barbara, 1927-

<< Peters, Elizabeth

Taulukko 2: Franar-ryhmän ehdotus auktoriteettimallista.

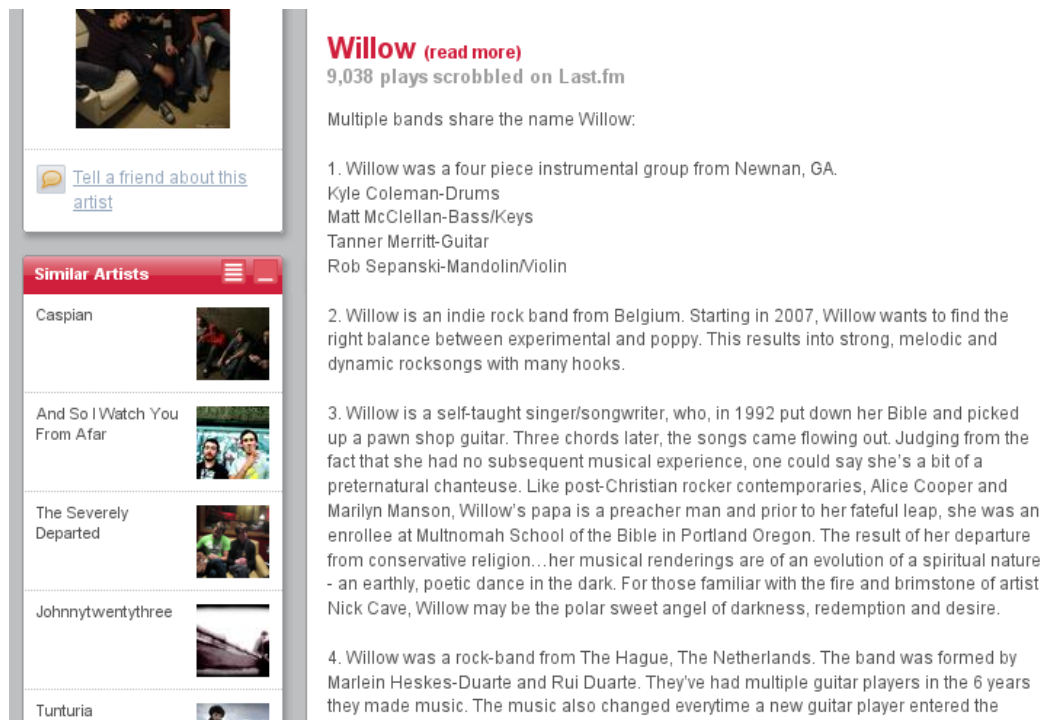
3). Ainoastaan käyttäjien syöttämät vapaat tekstikuvaukset kertovat päällekkäisyyksistä. Palvelussa on kehittyneet käyttäjämallit ja monimutkaiset yhteisölliset suositelukoneet. Kuitenkin identiteettien sotkeutuminen aiheuttaa sekaannusta, kun useampi artisti yhdistyy yhden nimen ja identiteetin alle. Suositukseen, esittelyteksteihin, musiikkipätkiin ja tageihin ei voi luottaa, kun ei tiedä mitä artistia ne koskevat.

Vastaavasti palvelussa tulee samalle artistille useita identiteettejä, kun nimen kirjoitusasu vaihtelee. Erilaisten translitterointien takia palvelusta löytyy esimerkiksi venäläisiä artisteja useaan kertaan. Jälleen kerran vaikutukset ulottuvat niin suositeluun kuin tiedonhakuun. Nimien täsmäystä ja hallintaa on pohdittu laajasti muissa tutkimuksissa [GMA07, BS92]. Täsmäyksessä voidaan hyödyntää esimerkiksi merkkijonojen editointietäisyyttä mittaavia algoritmeja. Getty-säätiön ULAN-tietokannassa auktorisoitu nimi<sup>15</sup> koostuu lukuisista ominaisuuksista, jotka on listattu taulukossa 3.

### 3.1.3 Rooliluokitukset

Rooleja voi käyttää osana tietojärjestelmää ja niistä voidaan rakentaa hierarkkinen luokittelu. Voidaan määritellä sosiaaliseen asemaan liittyviä rooleja, kuten isänä, siskona, ystävänä, työtoverina tai puolisona oleminen. Rooleja voidaan liittää ammattiin. On lääkäreitä, taiteilijoita ja leipureita. Jaottelua voi tarkentaa jakamalla taiteilijat edelleen esimerkiksi kuvataiteilijoihin, valokuvaajiin, kuvanveistäjiin. Näin

<sup>15</sup>[http://www.getty.edu/research/conducting\\_research/vocabularies/ulan/ulan\\_xml\\_dd06.pdf](http://www.getty.edu/research/conducting_research/vocabularies/ulan/ulan_xml_dd06.pdf)



**Willow** (read more)  
9,038 plays scrobbled on Last.fm

Multiple bands share the name Willow:

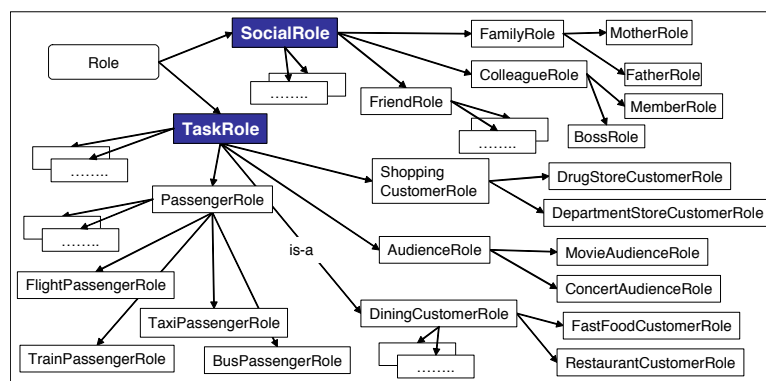
1. Willow was a four piece instrumental group from Newnan, GA.  
Kyle Coleman-Drums  
Matt McClellan-Bass/Keys  
Tanner Merritt-Guitar  
Rob Sepanski-Mandolin/Violin
2. Willow is an indie rock band from Belgium. Starting in 2007, Willow wants to find the right balance between experimental and poppy. This results into strong, melodic and dynamic rock songs with many hooks.
3. Willow is a self-taught singer/songwriter, who, in 1992 put down her Bible and picked up a pawn shop guitar. Three chords later, the songs came flowing out. Judging from the fact that she had no subsequent musical experience, one could say she's a bit of a preternatural chanteuse. Like post-Christian rocker contemporaries, Alice Cooper and Marilyn Manson, Willow's papa is a preacher man and prior to her fateful leap, she was an enrollee at Multnomah School of the Bible in Portland Oregon. The result of her departure from conservative religion...her musical renderings are of an evolution of a spiritual nature - an earthly, poetic dance in the dark. For those familiar with the fire and brimstone of artist Nick Cave, Willow may be the polar sweet angel of darkness, redemption and desire.
4. Willow was a rock-band from The Hague, The Netherlands. The band was formed by Marlein Heskes-Duarte and Rui Duarte. They've had multiple guitar players in the 6 years they made music. The music also changed everytime a new guitar player entered the

Kuva 3: Last.fm palvelussa nimen käyttäminen tunnisteena aiheuttaa ongelmia. Vasta sivun leipätekstistä selviää, että Willow-nimisiä artisteja onkin itse asiassa neljä. Käyttäjälle jää epäselväksi koskevatko ”Similar Artists” -suositukset amerikkalaisita itseoppinutta laulajaa vai indie-rock-yhtyettä belgiasta vai jotain muuta neljästä vaihtoehdosta.

Term ID	Numeerinen tunniste, joka yksilöi nimen tietokannassa
Term Text	Nimi.
AACR2 Flag	Ilmaisee onko nimi Kongressin kirjaston käyttämässä muodossa.
Display Name	Ilmaisee onko nimi tarkoitettu ensisijaisesti näytettäväksi.
Display Order	Nimille on määritelty järjestys, jossa ensimmäisenä listataan tyypillisimmin käytetyt nimet.
Languages	Ilmaisee nimen kielen.
Other Flags	Ilmaisee esim. onko nimi virallinen.
Term Contributors	Instituutiot tai projektit jotka ovat luovuttaneet tietoja kyseiseen tietueeseen.
Term Date	Kertoo ajankohdan jolloin nimi on ollut käytössä.
Term Sources	Viitteet lähteisiin.
Vernacular	Ilmaisee onko nimi kirjoitettu alkuperäiskielellä.

Taulukko 3: Nimen määrittäviä ominaisuuksia Getty-säätiön ULAN-tietokannassa.

muodostuva rakenne sopii kuvattavaksi esimerkiksi ontologialla.



Kuva 4: Fukazawa ja kumppanit ovat määritelleet rooliontologian suosittelua varten [FNFK06].

Rooleja voidaan hyödyntää useissa eri tarkoituksissa. Niiden avulla voidaan rajata hakuja kansallisuuksien tapaan. Rooliontologioita voidaan hyödyntää esimerkiksi käyttäjän profiloinnissa ja suosittelussa. Fukazawa ja kumppanit [FNFK06] ovat tehneet järjestelmän, joka suosittelee matkapuhelimen käyttäjälle palveluita käyttäjän sen hetkisen roolin mukaan. Fukazawan rooliluokitusta on esitelty kuvassa 4.

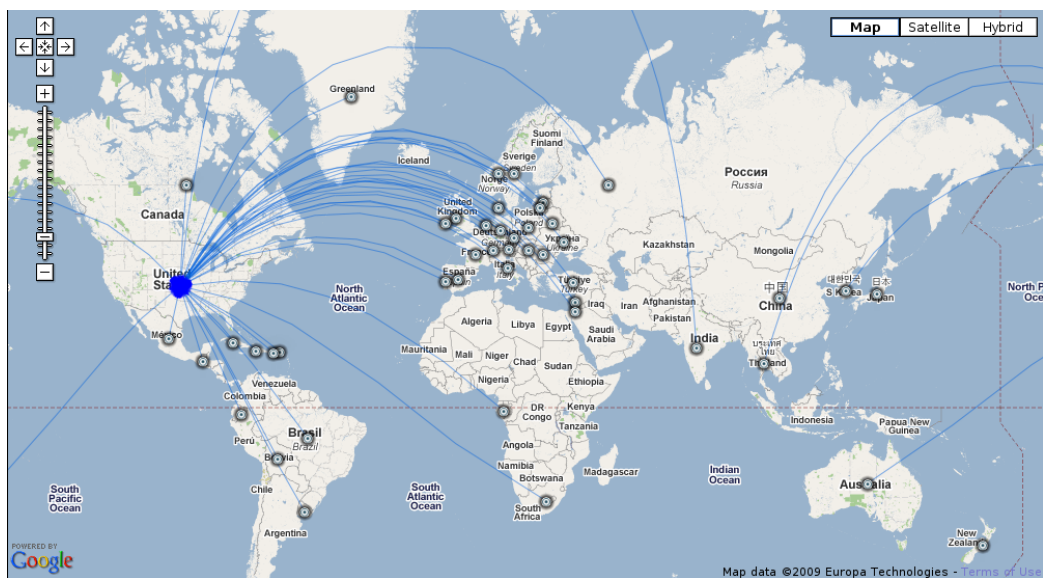
### 3.1.4 Kansallisuus ja paikkatiedot

Toimijalle voidaan määrittellä kansallisuus tai kansallisuuksia. Koska tiedot kytkeytyvät paikkaontologioihin, saadaan paikkojen hallinnointi siirrettyä toimijaa määrittävän ontologian ulkopuolelle. Tämä lisää yhteentoimivuutta muiden järjestelmien kanssa ja tekee toimijaontologioiden ylläpidosta kevyempää. Lisäksi toimijat saadaan liitettyä paikkaan perustuviin, esimerkiksi karttapohjaisiin, haku- ja visualisointityökaluihin.

Kuvissa 5 ja 6 on visualisoitu ULAN-toimijoita Google maps:n avulla. Kuvassa 5 toimijat on aseteltu kartalle syntymäpaikan mukaan. Kuvan 6 avulla voidaan havainnollistaa taitelijoiden muuttoa Yhdysvaltoihin. Kuvassa on esitetty toimijat, jotka ovat kuolleet Yhdysvalloissa, mutta syntyneet jossain muualla. Syntymä- ja kuolinpaikan välille on piirretty kaari visualisoimaan muuttoa.



Kuva 5: ULAN-toimijat sijoitettuna kartalle syntymäpaikan mukaan.



Kuva 6: ULAN-toimijoiden syntymäpaikan suhde kuolinpaikkaan, kun kuolinpaikka on kiinnitetty Yhdysvallat.

### 3.1.5 Aikaan liittyvät tiedot

Toimijatietoihin liitetään usein syntymä- ja kuolinaikatiedot. Myös muita merkittäviä aikaan liittyviä tietoja kuten palkinnon saamisista voidaan kirjata osaksi henkilön historiaa.

Aika esitetään usein literaaliarvona. Samassa järjestelmässä voi nähdä esimerkiksi seuraavanlaisia merkintöjä ”1.2.2008”, ”1.2.-08” tai ”1.2.”. Epämääräisyys aiheuttaa ongelmia aineiston koneellisessa hyödyntämisessä. Parempaan tulokseen päästää, jos käytetään standardimuotoa kuten XML:n DateTime-tietotyyppiä<sup>16</sup>. Tällöinkin ongelmaksi tulee aikavälien tai aikakausien kuvaaminen. Jos aika määritellään ontologisesti välinä, voidaan esittää 1900-luku, 80-luku, talvisota tai kivikausi yhdenmukaisesti. Taulukoissa 4 ja 5 on esimerkit ajan koodaamisesta Kulttuurisampo-portaalissa [HMK<sup>+</sup>09]. Aika on kuvattu nelipisteisenä joukkona, jossa alku- ja loppuajat ovat esitetty välinä. Välien päät on koodattu XML-standardin mukaisessa formaatissa. Lisäksi käyttöliittymässä näytetään ihmisluettava versio ajasta (nimekkä).

aikaisin mahdollinen alkuaika	1949-07-11T00:00:00.0Z
aikaisin mahdollinen loppuaika	1949-07-11T00:00:00.0Z
myöhäisin mahdollinen alkuaika	1949-07-11T23:59:59.999Z
myöhäisin mahdollinen loppuaika	1949-07-11T23:59:59.999Z
nimi	11.07.1949

Taulukko 4: Tarkan ajanhetken esittäminen Kulttuurisampo-portaalissa [HMK<sup>+</sup>09].

aikaisin mahdollinen alkuaika	1800-01-01T00:00:00.0Z
aikaisin mahdollinen loppuaika	1899-12-31T23:59:59.999Z
myöhäisin mahdollinen alkuaika	1800-01-01T00:00:00.0Z
myöhäisin mahdollinen loppuaika	1899-12-31T23:59:59.999Z
nimi	1800-luku

Taulukko 5: Aikakauden esittäminen Kulttuurisampo-portaalissa [HMK<sup>+</sup>09].

<sup>16</sup><http://www.w3.org/TR/xmlschema-2/>

## 3.2 Olemassa olevia malleja

### 3.2.1 MARC

Yhdysvaltain kongressin laajasti käyttämä MARC-tietokantaformaatti<sup>17</sup>. on suunniteltu erityisesti auktoriteettitietojen tallentamiseen. MARC pohjaa 70-luvulla suunniteltuun versioon ja ajan painolasti näkyy melko matalana abstraktiotasona. Tietue koostuu useista riveistä. Rivillä on ominaisuuden numero (esim. 100) sekä arvo (esim. 1#\$aCameron, Simon,\$d1799-1889). Arvokenttään on varsinaisen arvon lisäksi koodattu MARC-syntaksin mukaista metatietoa.

Taulukon 6 esimerkissä tietueen tunnus on talletettu kenttään 010. Risuaitamerkit tarkoittavat tyhjiä arvoja, alussa oleva \$a-merkkiä käytetään Yhdysvaltain kongressin kirjaston tunnuksissa. Varsinainen tunnus on 86114834. Kentässä 005 on viimeisin muokkaus aika 19870121083133.6, kentässä 670 lähdeviitteet ja kentässä 040 tietueen tekijän tunnus. Kentässä 100 on varsinainen tietueen sisältö: henkilön nimi.

MARC:iin liittyvistä ongelmista ja vaihtoehdoista on kirjoitettu useissa yhteyksissä<sup>18</sup>. Eräs ongelma on, että eri maissa on käytössä hieman erilaisia versioita MARC-formaatista. Esimerkiksi kansallisuustietoja koodataan vaihtelevasti kansallisuus- ja kommenttikenttiin. Myös translitterointiskeemat vaihtelevat maakohtaisesti.

### 3.2.2 ULAN

ULAN<sup>19</sup> (Union List of Artist Names) on Getty-säätiön ylläpitämä rekisteri taidehistorian henkilöistä. Taustalla oleva XML-pohjainen malli kuvaa yksityiskohtaisesti toimijan roolit, kansallisuudet, nimien eri kirjoitusasut sekä keskeiset sosiaaliset kontaktit.

Taulukossa 7 on esitetty tietue, joka kuvaa ranskalaista taidemaalaria Jacques Réattua. MARC-tietueeseen verrattuna XML-formaatti on itsensä selittävänä helpompi ymmärtää. Aineistoa voi tutkia myös Getty-säätiön tarjoaman web-käyttöliittymän kautta (kuva 7). Järjestelmä ei kuitenkaan tarjoa avoimia sovellusrajapintoja, eikä toteutus pohjaa avoimeen lähdekoodiin. Lisäksi järjestelmän sisältämä data on tutkimuskäyttöä lukuunottamatta tiukasti lisensoitua.

<sup>17</sup><http://www.loc.gov/marc/>

<sup>18</sup>[http://www.loc.gov/catdir/bibcontrol/tillett\\_paper.html](http://www.loc.gov/catdir/bibcontrol/tillett_paper.html)

<sup>19</sup>[http://www.getty.edu/research/conducting\\_research/vocabularies/ulan/](http://www.getty.edu/research/conducting_research/vocabularies/ulan/)

LDR \* \* \* \* \*nz###22\* \* \* \* \*n##4500

001 <control number>

003 <control number identifier>

005 19870121083133.6

010 ##\$aex#86114834#

040 ##\$a<organization code>\$c<organization code>

100 1#\$aCameron, Simon,\$d1799-1889

670 ##\$aNUCMC data from NJ Hist. Soc. for Bradley, J.P. Papers,  
1836-1937\$b(Simon Cameron)

670 ##\$aLC data base, 1-21-87\$b(hdg.: Cameron,  
Simon, 1799-1889)

670 ##\$aDAB\$b(Cameron, Simon, 1799-1889; Sen. from Pa.  
(Republican boss); financier; Sec. War under  
Lincoln; Min. to Russia; s. Charles & Martha  
(Pfoutz) C.; newspaper editor; owner Harrisburg  
Republican; commis. to settle claims of Winnebago  
Indians; m. Margaret Brua; father of:  
J.D. Cameron (1833-1918))

Taulukko 6: Henkilöä kuvaava MARC-tietue.

```

<Subject Subject_ID="500000010">
  <Record_Type>Person</Record_Type>
  <Associative_Relationships>
    <Associative_Relationship>
      <Relationship_Type>1102/student of</Relationship_Type>
      <Related_Subject_ID>
        <VP_Subject_ID>500019743</VP_Subject_ID>
      </Related_Subject_ID>
    </Associative_Relationship>
  </Associative_Relationships>
  <Biographies>
    <Preferred_Biography>
      <Biography_ID>400000018</Biography_ID>
      <Biography_Text>French painter, 1760-1833</Biography_Text>
      <Birth_Place>4151000241/Arles (Bouches-du-Rhône,
        Provence-Alpes-Côte d'Azur, France)</Birth_Place>
      <Birth_Date>1760</Birth_Date>
      <Death_Place>4151000241/Arles (Bouches-du-Rhône,
        Provence-Alpes-Côte d'Azur, France)</Death_Place>
      <Death_Date>1833</Death_Date>
      <Sex>Male</Sex>
    </Preferred_Biography>
  </Biographies>
  <Nationalities>
    <Preferred_Nationality>
      <Nationality_Code>901600/French</Nationality_Code>
      <Display_Order>1</Display_Order>
    </Preferred_Nationality>
  </Nationalities>
  <Terms>
    <Preferred_Term>
      <Term_Text>Réattu, Jacques</Term_Text>
      <Term_ID>1500000013</Term_ID>
      <Display_Name>Index</Display_Name>
      <Display_Order>1</Display_Order>
    </Preferred_Term>
  </Terms>
</Subject>

```

Taulukko 7: Henkilöä kuvaava ULAN-tietue.







## Union List of Artist Names® Online Full Record Display

[New Search](#)

[Previous Page](#)

Click the  icon to view the hierarchy.

**ID: 500015305**

 **Gallen-Kallela, Akseli** (Finnish painter and graphic artist, 1865-1931)

### Names:

**Gallen-Kallela, Akseli** ([preferred](#), [index](#), [official](#), [V](#))

**Akseli Gallen-Kallela** ([display](#), [V](#))

**Gallén, Axel** ([V](#))

.... until 1904

**Gallén-Kallela, Akseli Valdemar** ([V](#))

**Kallela, Akseli Gallén-** ([V](#))

**Gallén-Kallela, Axel** ([V](#))

**Gallen-Kallela, Akseli Valdemar** ([V](#))

**Gallén, Aksel'** ([V](#))

**Gallèn-Kallela, Akseli** ([V](#))

### Nationalities:

Finnish ([preferred](#))

Swedish

Scandinavian

### Roles:

artist ([preferred](#))

painter

printmaker

architect

graphic artist(s)

designer

**Gender:** male

### Birth and Death Places:

Born: [Pori \(Länsi-Suomen Lääni, Finland\)](#) (inhabited place)

Died: [Stockholm \(Stockholm, Sweden\)](#) (inhabited place)

Kuva 7: Akseli Gallen-Kallelaa kuvaava ULAN-tietue Getty-säätiön web-palvelussa.

### 3.2.3 FOAF

Friend of a Friend<sup>20</sup> (FOAF) on projekti, joka pyrkii määrittelemään yhteisen kehysten henkilötietojen ja erityisesti sosiaalisten kontaktien esittämiseen verkossa.

Kantavana ajatuksena on siirtää henkilötietojen hallinta käyttäjille itselleen. Nykyisin monet verkkopalvelut keräävät tietoja käyttäjistään, mutta käyttäjällä itsellään ei ole mahdollisuutta nähdä tai tietää mitä tietoja hänestä kerätään. Tietoja ei voi myöskään siirtää järjestelmästä toiseen. FOAF-profiilit ovat käyttäjän hallinnoimia ja profiilin voi laittaa esimerkiksi kotisivulleen muiden nähtäväksi. Projektin nykyisenä heikkoutena voi nähdä käyttäjältä vaadittavan teknisen osaamisen. Laajalle levinneitä ja helppokäyttöisiä työkaluja profiilien muodostamiseen ja hyödyntämiseen on niukasti saatavilla.

FOAF-määrittelystä on tehty RDF-perustainen metatietomalli<sup>21</sup>, josta on kehittynyt teollisuusstandardi henkilötietojen esittämiseen semanttisessa webissä. Esimerkkitaulukossa 8 on esitetty FOAF-henkilö RDF/XML-syntaksilla.

```
<foaf:Person>
  <foaf:name>Jussi Kurki</foaf:name>
  <foaf:mbox>jussi.kurki@helsinki.fi</foaf:mbox>
</foaf:Person>
```

Taulukko 8: Esimerkki FOAF-henkilöstä.

### 3.2.4 REL

Relationship-sanasto<sup>22</sup> REL määrittelee yleiselle `foaf:knows`-suhteelle aliominaisuuksia. Sanastoon kuuluvat esimerkiksi ominaisuudet `childOf`, `colleagueOf`, `hasMet`, `friendOf` ja `neighborOf`. Taulukossa 9 on esitetty FOAF-tietue, johon on koodattu sosiaalisia kontakteja REL-sanaston avulla.

---

<sup>20</sup><http://www.foaf-project.org/>

<sup>21</sup><http://xmlns.com/foaf/spec/>

<sup>22</sup><http://vocab.org/relationship/>

```
<foaf:Person>
  <foaf:name>Jussi Kurki</foaf:name>
  <foaf:mbox>jussi.kurki@helsinki.fi</foaf:mbox>

  <rel:friendOf>
    <Person foaf:name="Timo" rdf:about="http://helsinki.fi/timo"/>
  </rel:friendOf>
  <rel:friendOf>
    <Person foaf:name="Sari" rdf:about="http://helsinki.fi/sari"/>
  </rel:friendOf>

  <rel:worksWith>
    <Person foaf:name="John Smith"
      rdf:about="http://seco.hut.fi/u/smith"/>
  </rel:worksWith>

</foaf:Person>
```

Taulukko 9: Esimerkki FOAF-tietueesta, jota on laajennettu REL-sanastolla.

### 3.2.5 BIO

BIO-sanasto<sup>23</sup> on yksinkertainen ontologia biografisten tietojen esittämiseen. Sanasto on laajennos FOAF-määrittelyyn. Määrittely sisältää tekstikentät vapaalle kuvailulle ja asiasanoitukselle.

Syntymä ja kuolema on kuvattu tapahtumina. Tapahtumien käyttö mahdollistaa sanaston suoraviivaisen laajennettavuuden. Yleiselle `bio:Event`-luokalle voidaan määritellä uusia aliluokkia, kuten ”koulusta valmistuminen” tai ”palkinnon voittaminen”. Taulukossa 10 on annettu esimerkki FOAF-tietueesta, johon on liitetty BIO-sanaston avulla yhden rivin biografia, avainsanoja ja syntymäaika.

```
<foaf:Person>
  <foaf:name>Jussi Kurki</foaf:name>
  <foaf:mbox>jussi.kurki@helsinki.fi</foaf:mbox>
  <bio:olb>
    suomalainen tietojenkäsittelytieteen opiskelija, 1982-
  </bio:olb>
  <bio:keywords>
    tietojenkäsittelytiede, semanttinen web
  </bio:keywords>
  <bio:event>
    <bio:Birth>
      <bio:date>1982</bio:date>
      <bio:place>Helsinki</bio:place>
    </bio:Birth>
  </bio:event>
</foaf:Person>
```

Taulukko 10: Esimerkki FOAF-tietueesta, jota on laajennettu BIO-sanastolla.

### 3.2.6 Muita sanastoja

Semantically Interlinked Online Communities<sup>24</sup> (SIOC) määrittelee edellämainittujen kaltaisen sanaston, jolla voi kuvata keskustelufoorumien, blogien ja wikien ra-

<sup>23</sup><http://vocab.org/bio/0.1/>

<sup>24</sup><http://rdfs.org/sioc/spec/>

kennetta ottaen huomioon käyttäjät, ryhmät ja roolit. Taulukossa 11 on annettu esimerkki SIOC-sanaston hyödyntämisestä.

```
<sioc:Post>
  <dcterms:title>Toimijat semanttisessa webissä</dcterms:title>
  <sioc:has_creator>
    <sioc:User rdfs:label="Jussi Kurki"
      rdf:about="http://helsinki.fi/u/p1235123890"/>
  </sioc:has_creator>
  <sioc:content>
    Toimijaontologioilla tarkoitetaan...
  </sioc:content>
  <sioc:topic rdfs:label="Semanttinen web"
    rdf:resource="http://yso.fi/onto/koko/p12345512"/>
</sioc:Post>
```

Taulukko 11: Esimerkki SIOC-sanaston käytöstä semanttista webiä käsittelevän blogi-kirjoituksen esittämisessä.

VCard<sup>25</sup> on puolestaan määrittely virtuaalisille käyntikortteille. Määrittelystä on saatavilla myös W3C:n tekemä RDF-versio<sup>26</sup>.

XHTML-datalle suunnitellussa XHTML Friends Network<sup>27</sup> (XFN) -määrittelyssä on FOAF:n ja BIO:n kaltainen idea. Erilaisilla lisäattribuuteilla voi määrittellä suhteen tyyppin HTML:n linkki-tagissa. Blogien välisissä linkityksissä voidaan tarkentaa onko kyseessä esimerkiksi kaverin tai vanhemman sivusto. Tarkoituksena on rakentaa koneluettavaa sosiaalista verkostoa.

### 3.3 Toimijaontologia

Toimijaontologia TOIMO pohjautuu semanttisen webin olemassa oleviin sanastoihin. Sanastoja kierrättämällä voidaan hyödyntää olemassa olevia työkaluja ja tietovarastoja sekä madaltaa käyttöönottokynnystä. TOIMO pyrkii laajentamaan ja kasaamaan FOAF-pohjaisia sanastoja yhdeksi kokonaisuudeksi, jota voidaan hyödyntää erilaisissa hakujärjestelmissä (luvut 4 ja 5).

<sup>25</sup><http://www.imc.org/pdi/vcardoverview.html>

<sup>26</sup><http://www.w3.org/TR/vcard-rdf>

<sup>27</sup><http://gmpg.org/xfn/>

Aleman-Meza ja kumppanit [AMBB<sup>+</sup>07] ovat käyttäneet samaan tapaan useita eri sanastoja toimijoiden määrittelyyn. He ovat määritelleet asiantuntijahakujärjestelmän muun muassa FOAF-, SIOC- ja SKOS-sanastojen avulla.

Toimijaontologia pohjautuu aiemmin kuvattuihin FOAF, BIO ja REL sanastoihin. Lisäksi on määritelty joukko rooleja ja kansallisuuksia ULAN-skeeman pohjalta. Roolit on linkitetty Yleiseen suomalaiseen ontologiaan [HSVF07]. Toimijaontologia on sanastomainen, eikä se sisällä juurikaan monimutkaisia ontologisia rakenteita kuten syviä luokkahierarkioita tai moniperintää. Toimijaontologian skeema löytyy liitteestä 1.

### 3.3.1 Toimijaontologian URI-tunnisteet

Toimijaontologian URI-tunnisteet on johdettu universaalisti ainutkertaisesta tunnisteesta (Universally Unique IDentifier - UUID<sup>28</sup>), jonka perusteella:

1. URI on muuttumaton  
(ei riipu esim. tietueen datasta)
2. URI on riippumaton kielestä  
(tunniste koostuu satunnaisista numeroista ja kirjaimista)
3. URI-tunnisteesta ei voi päätellä dataa  
(tunnisteita ei jaeta esim. numerojärjestyksessä)
4. URI on universaalisti ainutkertainen  
(ei ole kahta toimijaa, joilla olisi samat tunnisteet)
5. URI-tunnisteita voi generoida hajautetusti  
(kukaan ei hallitse tunnisteavaruutta)

Toimijaontologiassa henkilön URI muodostuu nimiavaruudesta ja yksilöivästä UUID-tunnisteesta.

Toimo-URI on seuraavaa muotoa:

`[toimo-nimiavaruus]/u[UUID-tunniste]`

Tarkistussumman edellä olevalla kirjaimella varmistetaan, ettei paikallinen nimi ala numerolla. Seuraavassa joitakin esimerkkejä toimijaontologian URI-tunnisteista:

<sup>28</sup><http://www.ietf.org/rfc/rfc4122.txt>

<http://seco.tkk.fi/onto/toimo/u9e7ea034-7758-4d50-a171-8d39540b007b>

<http://seco.tkk.fi/onto/toimo/u25baaf70-370a-4797-94d2-2531ac028d9f>

### 3.3.2 Nimikentät

Toimijaontologiassa nimen voi koodata joko `foaf:name`-kenttään tai `foaf:firstName`- ja `foaf:surname`-kenttiin.

Toimijaontologiassa `toimo:term`-kenttä listaa kaikki mahdolliset nimen kirjoitusasut. Esimerkkejä erilaisista muodoista on translitteroinnin, taiteilijanimet, toiset nimet jne. Kun haku kohdistetaan `toimo:term`-kenttään ja saadaan helposti katettua kaikki henkilön nimet. Toinen vaihtoehto olisi käyttää esim. SKOS-sanaston<sup>29</sup> `skos:altLabel`-ominaisuutta.

Toimijalla on tyypillisesti `rdfs:label`- tai `skos:prefLabel`-ominaisuudet, jotka voivat sisältää nimen lisäksi yhteenvedon toimijatiedoista. Ajatus on lainattu ULAN:in nimike-kentästä, jossa nimike on muodostettu kansallisuus- ja roolitiedotiedoista sekä syntymä- ja kuolinajoista. Nimike mukailee seuraavaa muotoa:

Sukunimi, Etunimi ([kansallisuus] [rooli], [syntymävuosi]-[kuolinvuosi])

Seuraavassa joitakin esimerkkejä nimikkeistä:

Gallen-Kallela, Akseli (Finnish painter and graphic artist, 1865-1931)

Gallen-Kallela, Akseli (suomalainen taiteilija, 1865-1931)

### 3.3.3 Rooli ja kansallisuus

Toimijaontologian rooli- ja kansallisuusontologiat on populoitu ULAN-aineistosta. Esimerkiksi taidemaalari on `roles:Painter`. Rooli- ja kansallisuusominaisuudet on nimetty vastaavasti `toimo:role` ja `toimo:nationality`. Esim.

```
<foaf:Person>
  <toimo:role>
    <roles:Painter />
  </toimo:role>
  <toimo:nationality>
    <nationalities:Dutch />
```

<sup>29</sup><http://www.w3.org/2004/02/skos/>

```
<toimo:nationality>
</foaf:Person>
```

Rooli- ja kansallisuusontologiat löytyvät liitteestä 2.

## 4 Toimijoiden hakeminen - ONKI People

ONKI People on keskitetty palvelu henkilö- ja organisaatiotiedon hakemiseen ja hyödyntämiseen [Kur08]. Järjestelmä käyttää tietomallinaan toimijaontologiaa. ONKI People on osa ONKI-palveluperhettä [KVH05, HVTS08], johon kuuluu mm. ONKI SKOS [TFVH09] W3C:n SKOS-muotoisen<sup>30</sup> sanastodatan esittämiseen ja ONKI GEO [Lin08] paikkatiedon esittämiseen.

ONKI People koostuu sekä graafisesta käyttöliittymästä että sovellusrajapinnosta. Graafinen käyttöliittymä on suunniteltu toimijoiden etsimiseen. Käyttöliittymä pohjautuu sekä vapaaseen tekstihakuun että moninäkömähakuun. Toimijoiden verkoston tutkimista helpotetaan graafimaisella sosiaalisen piirin visualisoinnilla.

Tässä luvussa käydään läpi moninäkömähakuparadigmaa sekä sosiaalisten verkkojen visualisointiin liittyviä kysymyksiä. Lisäksi luvussa esitellään ONKI People ja sen taustalla olevat tekniset ratkaisut. Lopuksi pohditaan käyttöliittymäratkaisujen soveltuvuutta hakuongelman ratkaisuun.

### 4.1 Moninäkömähaku

Moninäkömähäussa [HEE<sup>+</sup>02] ideana on jäsentää datajoukko toisistaan riippumattomiksi näkymiksi. Näkymät ovat usein tekstimuotoisia hierarkkisia kategoriapuita, joista valitsemalla käyttäjä voi rajata haun vain tietyn tyyppisiin tai tietyt ominaisuudet täyttäviin kohteisiin. Esimerkiksi Amazon.com tarjoaa näkymät ”osasto”, ”merkki” ja ”megapikselimäärä” helpottamaan ”digitaalinen kamera” -hakusanalla aloitettua hakua (kuva 8).

MultimediaN<sup>31</sup> toteuttaa moninäkömähäun kuka-, mitä-, missä-, milloin-kysymysten kautta (kuva 9). Käyttäjä voi valita kuka on tehnyt teoksen (esim. Rembrandt), millä se on tehty (esim. öljyväreillä), missä teos nykyisin sijaitsee (esim. Rijksmuseum) ja koska teos on tehty (esim. 1630). Haku käynnistyy heti kun jonkin ehdoista on

<sup>30</sup><http://www.w3.org/TR/2009/REC-skos-reference-20090818/>

<sup>31</sup><http://e-culture.multimedian.nl/demo/>



Amazon.com: digital camera

http://www.amazon.com/s/ref=nb\_ss\_0\_8?url=search-alias%3Daps&field-keywords=digital+camera&x=0&y=0&spref=...

Shop All Departments Search All Departments digital camera GO Cart Wish List

**Department**  
**Any Department**  
 Electronics (147,594)  
 Books (41,453)  
 Office Products & Supplies (16,940)  
 MP3 Downloads (5,464)  
 Sports & Outdoors (1,193)  
 Everything Else (899)  
 Home & Garden (881)  
 + See All 30 Departments

**Shipping Option** (What's this?)  
**Any Shipping Option**  
 Prime Eligible  
 Free Super Saver Shipping

**Digital Cameras**  
**Brand**  
 Canon (1,029)  
 Fuji (202)  
 Kodak (378)  
 Nikon (882)  
 Panasonic (326)  
 Polaroid (115)  
 Samsung (248)  
 Sony (555)

**Megapixels**  
**Any Number of Megapixels**  
 5.9 MP & Under (1,097)  
 6 to 7.9 MP (563)  
 8 to 9.9 MP (663)  
 10 to 11.9 MP (1,310)  
 12 MP & Up (1,725)

**"digital camera"**  
**Related Searches:** canon digital camera, digital camera under 100, digital cameras under 100.  
 Select Results from All Departments Choose a Department to enable sorting

**Wearable Camcorders**  
 These specialty shoot-and-shares are designed for capturing point-of-view footage for all to see. Shoot-and-share wearable camcorders are durable, water-resistant, and small enough to mount almost anywhere. [See them all.](#)

- Nikon Coolpix L20 10MP Digital Camera with 3.6 Optical Zoom and 3 inch LCD (Deep Red)** by Nikon  
 Buy new: **\$95.89**  
 53 new 1 used from \$129.99  
 Get it by **Monday, Nov 16** if you order in the next **12 hours** and choose one-day shipping.  
 Eligible for **FREE** Super Saver Shipping.  
 ★★☆☆☆ (73)  
 Electronics: See all 147,594 items
- Amazon Camera & Photo Store**  
 Digital Point & Shoot Cameras Digital SLR Cameras  
 Computer Memory Cards Digital Camera Lenses  
 Camera Cases & Bags Digital Photography Books  
 Camera Batteries Digital Picture Frames
- Nikon Coolpix L20 10MP Digital Camera with 3.6 Optical Zoom and 3 inch LCD (Navy Blue)** by Nikon  
 Buy new: **\$101.98**  
 8 new from \$101.98  
 Get it by **Monday, Nov 16** if you order in the next **10 hours** and choose one-day shipping.  
 Eligible for **FREE** Super Saver Shipping.  
 ★★☆☆☆ (20)  
 Electronics: See all 147,594 items

Kuva 8: Moninäkömähaku Amazon.com-sivustolla. Sivun vasemmasta reunasta käyttäjä voi rajata hakua tietyn osaston, kameravalmistajan tai megapikselinmäärän mukaan.

valinnut. Käyttäjä voi lisätä rajauksia sitä mukaa kun tulosjoukko hahmottuu. Eri-  
laisten tulosjoukkojen näyttäminen auttaa haun lisäksi hahmottamaan saatavilla  
olevaa kokoelmaa.



Kuva 9: Moninäkömähaku MultimediaN-portaalissa on toteutettu kuka-, mitä-,  
missä-, milloin-kysymysten kautta.

MIT Exhibit<sup>32</sup> on Javascript-pohjainen hakukäyttöliittymien generointityökalu. Pal-  
velu mahdollistaa tiedon visualisoinnin hyödyntäen runsaasti esimerkiksi karttoja ja  
aikajanoja. Myös moninäkömähaku on toteutettu (kuva 10). Kuvassa esitetyssä Ex-  
hibitin avulla tehdyssä henkilöiden hakujärjestelmässä näkymiksi on valittu mm.  
henkilön asema, toimiston kerros ja ryhmä, johon henkilö kuuluu.

ONKI People -järjestelmässä moninäkömähaku on yhdistetty vapaaseen sanahakuun  
(kuva 11). Haun voi tehdä valitsemalla yhden tai useamman roolin (esim. arkkitehti)  
ja kansallisuuden (esim. suomalainen). Lisäksi käyttäjä voi syöttää tekstikenttään  
vapaan hakusanan (esim. "napoleon"). Tekstikenttään voi syöttää myös samat tiedot  
mitä näkymistä voi valita, jolloin käyttäjän ei tarvitse selailla näkymiä auki, jos hän  
tietää mitä on hakemassa (esim. "suomalainen arkkitehti").

<sup>32</sup><http://www.simile-widgets.org/exhibit/>

The screenshot shows a web browser window displaying the SIMILE Widgets interface for MIT CSAIL Principal Investigators. The browser address bar shows the URL: <http://www.simile-widgets.org/exhibit/examples/CSAIL-PIs/CSAIL-PIs.html>. The page title is "CSAIL Principal Investigators". Below the title, there is a link to the original data and a link to the exhibit JSON data file.

The interface features a central grid of person cards. The first two cards are for Alan Edelman and Alan Willsky. The third card is for Albert R. Meyer, and the fourth is for Anant Agarwal. Each card displays a photo, name, office, phone number, and email address. To the left of the grid is a "Group" sidebar with a scrollable list of research groups. To the right is a "Positions" sidebar with a scrollable list of job titles. Below the "Positions" sidebar is a "Stata Towers" sidebar with a scrollable list of building names, and below that is an "Office Floors" sidebar with a scrollable list of floor numbers.

Kuva 10: Henkilötietojen hakujärjestelmä MIT Exhibit -alustalla. Näkymiksi on valittu mm. henkilön asema, toimiston kerros ja ryhmä, johon henkilö kuuluu.

The screenshot shows the ONKI People search interface. The search bar contains the text "napoleon". The search results are displayed in a table format. The table has two columns: "nationality" and "role". The "nationality" column lists various nationalities with their respective counts. The "role" column lists various roles with their respective counts. The search results are filtered by "nationality" and "role".

nationality	role
American 2	Architect 2
Belgian 1	Artist 21
Brazilian 1	Cardinal 1
Canadian 2	
Corsican 1	
French 7	
German 2	
Italian 11	
Romanian 1	
Scandinavian 1	
SouthAmerican 1	
Swedish 1	
more...	

The search results are displayed in a table format. The table has two columns: "nationality" and "role". The "nationality" column lists various nationalities with their respective counts. The "role" column lists various roles with their respective counts. The search results are filtered by "nationality" and "role".

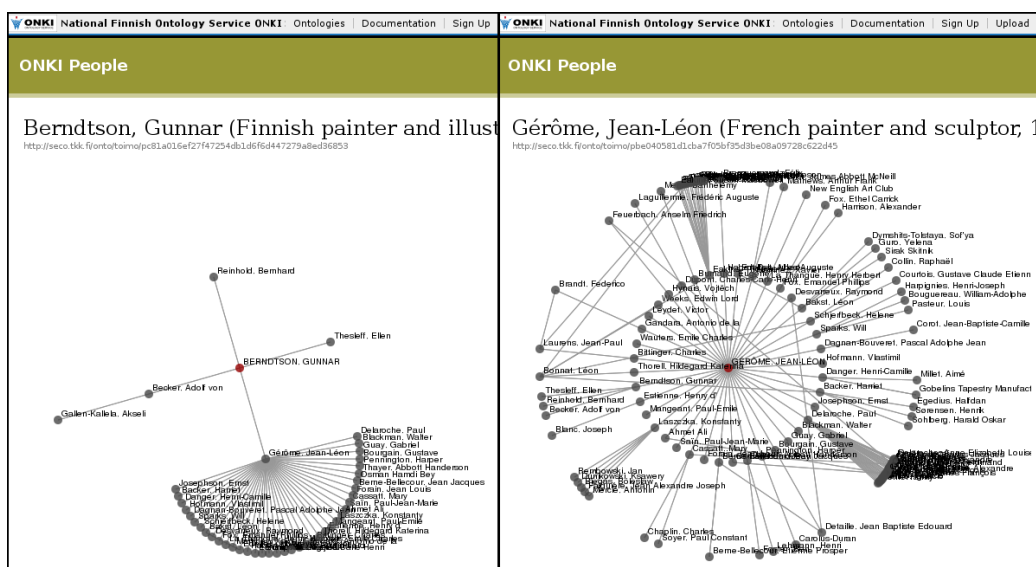
Kuva 11: Moninäkömahaku ja tekstihaku ONKI People -järjestelmässä. Näkymiksi on valittu rooli ja kansallisuus.

## 4.2 Sosiaalisen verkoston visualisointi

Henkilön sosiaalinen piiri muodostuu ystäväistä, työtovereista, sukulaisista jne. Sosiaalista piiriä visualisoimalla saadaan henkilölle hakua helpottavaa kontekstia. Lisäksi henkilön kontaktien tutkailu saattaa olla itsessään kiinnostavaa. Esimerkiksi taiteilijoiden ystäpiiristä löytyy usein toisia kiinnostavia taiteilijoita.

ONKI People -järjestelmässä sosiaalinen piiri on esitetty kahden askeleen päähän, eli myös tuttujen tutut on visualisoitu (kuva 12). Graafissa solmu kuvaa henkilöä tai ryhmää, kaari kontaktia. Graafissa kaikki suhteet on yleistetty ”tuntee”-tyyppisiksi väljiksi yhteyksiksi. Solmut ovat linkkejä, jotka johtavat kyseisten toimijoiden sivuille .

Graafin keskellä on tarkastelussa oleva henkilö. Sisimmällä kehällä on henkilön suorat kontaktit. Uloimmalta kehältä löytyy tuttujen tutut. Kuvan 12 vasemmassa palsassa käyttäjä on tarkastellut Gunnar Berndtsonin sosiaalista piiriä. Klikkaamalla henkilön Jean-Leon Gerome solmua käyttäjä on siirtynyt tarkastelemaan hänen ”lähipiiriään”.



Kuva 12: Sosiaalisen verkoston visualisointi ONKI Peoplessa.

## 4.3 Sovellusrajapinnat

ONKI People toteuttaa yleiset ONKI-rajapinnat [TFVH09, TKVH09] ja sitä voi käyttää sekä Web Service- että AJAX-rajapintojen kautta.

ONKI Selector -rajapinnan<sup>33</sup> avulla ONKI-palvelut voidaan yhdistää olemassa ole-  
viin järjestelmiin Javascript- ja AJAX-tekniikoiden avulla. Kuvassa 13 toimijoita  
on haettu automaattisesti täydentyvän ONKI Selector -hakukentän kautta. Kent-  
tä muodostetaan asiakaspäähän kahdella rivillä Javascript-koodia. Haut kohdistuvat  
ONKI People -palvelimelle.



Kuva 13: ONKI People -järjestelmää voi käyttää ONKI Selector -komponentin kaut-  
ta. Automaattisesti täydentyvään hakukenttään haetaan tiedot AJAX-rajapinnan  
kautta.

#### 4.4 Tekninen toteutus

ONKI People on toteutettu Java-kielellä. Hakumoottorina toimii Apache Lucene<sup>34</sup>.  
Kategoriat muodostetaan käymällä tietyn ominaisuudet tulosjoukko läpi ja indek-  
soimalla henkilölle tietty ominaisuus ja sen arvo (esim. rooli/arkkitehti).

Graafin solmut sommitellaan yksinkertaisella algoritmilla tietyn henkilön näkökul-  
masta. Algoritmi sijoittaa valitun henkilön keskelle, suorat kontaktit ympyrän muo-  
toiselle kehälle tasavälein. Kontaktien kontaktit sijoitetaan toiselle kehälle tasaisen  
kokoisille sektoreille jne. Kyseessä on siis eräänlainen puumainen näkymä. Mukaan  
otetaan puukaarten lisäksi myös muut toimijoiden väliset kaaret, mutta ne eivät  
vaikuta solmujen sijoitteluun.

Sosiaalinen verkko piirretään skaalautuvana vektorigrafiikkana SVG-kuvana<sup>35</sup>. Graa-  
fi piirretään pyynnöstä suoraan HTTP-vastauspakettiin, jolloin säästetään levyha-  
kuun kuluva aika.

<sup>33</sup><http://www.yso.fi/onkiselector/>

<sup>34</sup><http://lucene.apache.org/>

<sup>35</sup><http://www.w3.org/Graphics/SVG/>

## 4.5 Arviointi

Moninäkömäähaussa keskeinen ongelma on näkymien valitseminen ja niiden sisällön esittäminen [HEE<sup>+</sup>02]. Ontologioiden ja sanastojen hierarkkiat eivät välttämättä sovellu suoraan hakuhierarkkioiksi. Esimerkiksi Getty-säätiön taidehistorian sanasto Art and Architecture Thesaurus<sup>36</sup>, asettaa kuninkaan seuraavaan hierarkkiaan:

```

people (agents) (G)
  <people by occupation> (G)
    <people in government and administration> (G)
      rulers (people) (G)
        monarchs (G)
          kings (people)

```

Kuningas on siis kuninkaallinen, hallitsija, henkilö hallitsevassa asemassa jne. Kaikkien tasojen näyttäminen hakukäyttöliittymässä olisi käyttäjälle todennäköisesti vain hämmentävää.

ONKI People -järjestelmässä kategorioista näytetään vain lehtitaso. Näkymien koko tulee kuitenkin ongelmaksi, sillä kumpikaan kansallisuus- tai roolilistoista ei mahdu kerralla ruudulle (kuva 14). Toisaalta pelkkä lista tekee käyttöliittymästä yksinkertaisemmän ja suoraviivaisemmän tilanteessa, jossa vaihtoehtoja on vain kohtalaisen vähän.

Kuitenkin verrattuna Getty-säätiön omaan palveluun ONKI Peoplen käyttäminen on suoraviivaisempaa. Getty:n hakupalvelussa on kolme kenttää, joihin käyttäjä voi syöttää tietoa (kuva 15). Ensimmäiseen syötetään nimi tai sen osa, toiseen rooli ja kolmanteen kansallisuus. Jos syötetty rooli tai kansallisuus ei vastaa kirjoitusasultaan tarkalleen aineistossa olevia arvoja, järjestelmä antaa varoituksen pop-up-ikkunassa ja termi pitää korjata. Käyttäjä voi myös valita roolin tai kansallisuuden listasta, johon pääsee kentän vieressä olevasta linkistä. Kun hakuehdot on syötetty, painetaan ”Search”-nappia. Jos rajauksia haluaa muuttaa pitää, palata takaisin edelliselle sivulle.

ONKI Peoplessa näkymäehtojen (kansallisuus, rooli) muuttaminen heijastuu välittömästi hakutuloksiin. Käyttäjä voi myös arvoida valintojen vaikutuksia rajauslinkkien viereen esilasketun tulosjoukon koon avulla. Lisäksi osittaisten rooli- tai kansallisuusnimien syöttäminen on mahdollista.

<sup>36</sup><http://www.getty.edu/research/tools/vocabularies/aat/>

ONKI Ontologies | Sign Up | Feedback National Finnish Ontology Service ONKI

ONKI People

Search

Try e.g. 'napoleon', 'painter' or 'finnish' and hit Enter.

filters [klik to remove]:  
**napoleon**

**Results (27)**

**nationality**

- American 2
- Belgian 1
- Brazilian 1
- Canadian 2
- Corsican 1
- French 7
- German 2
- Italian 11
- Romanian 1
- Scandinavian 1
- South American 1
- Swedish 1
- more...

**role**

- Architect 2
- Artist 21
- Cardinal 1

Angiolini, Napoleone (Italian painter, 1797-1864)

Bellardel, Napoleon Joseph (French artist, active 19th century)

Bonaparte (French artist, 1811-1832)

Bonaparte, Roland-Napoleon (French photographer, 1858-1924)

Bourassa, Napoléon (Canadian painter, architect, and sculptor, 1827-1916)

Cocchetti, Napoleone (Italian painter, born ca. 1850)

Cocchetti, Napoléon (Italian painter, born ca. 1880)

Delaunois, Alfred (Belgian painter 1876-1941)

Eugen (Swedish painter and printmaker, 1865-1947)

Fiumi, Napoleone G. (Italian artist, born 1898)

Gimbrede, Joseph Napoleon (American engraver, born 1820, active 1841-1860)

Heigel, Franz Napoleon (German portraitist, 1813-1888)

Kuva 14: Suuret kansallisuus- ja roolilistat eivät mahdu kerralla ruudulle vaan käyttäjä joutuu klikkaamaan ”more...” -linkkiä nähdäkseen kaikki täsmäivät kansallisuudet.

The Getty

Home Visit Museum Research Institute Conservation Institute Four

Exhibitions | Conducting Research | Scholarly Activities | About the Research Institute

Research Home | Conducting Research | Learn about the Getty Vocabularies | Union List of Artist Names Online

Union List of Artist Names® Online

GRI CatalogPLUS

Advanced Search

Conducting Research  
Scholarly Activities  
About the Research Institute

Search the ULAN

Find Name or ID:

Role:

Nationality:

Search Clear

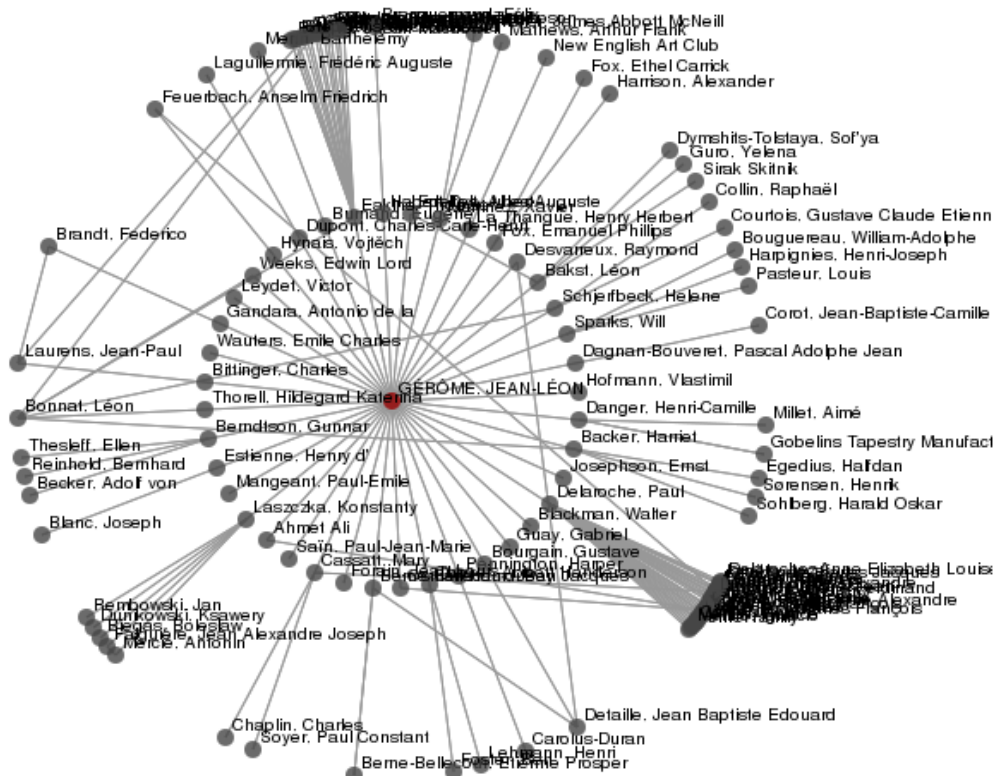
Pop-up Search Browse the ULAN hierarchies

Copyright information

Kuva 15: Getty-säätiön hakujärjestelmä.

Vaikka moninäkökymähaun kategoriat olisi huolellisesti toteutettu, käyttäjälle voi olla turhauttavaa yrittää löytää hyödyllisiä rajausehtoja kategorioista sen sijaan että Google-haun tapaan kirjoittaisi muutamia hakusanoja. Tämän vuoksi kaikki ontologiset käsitteet ovat myös sanahaun piirissä. Tällöin kuitenkin termien monimerkityksellisyys saattaa aiheuttaa ongelmia. Jos käyttäjä hakee esimerkiksi sanalla ”Paris”, tulee tuloksia sekä paikan että henkilön nimen perusteella. Tässä tilanteessa käyttäjä voi kuitenkin rajata haun tarkasti koskemaan vain paikkoja valitsemalla Pariisin paikka-näkymästä.

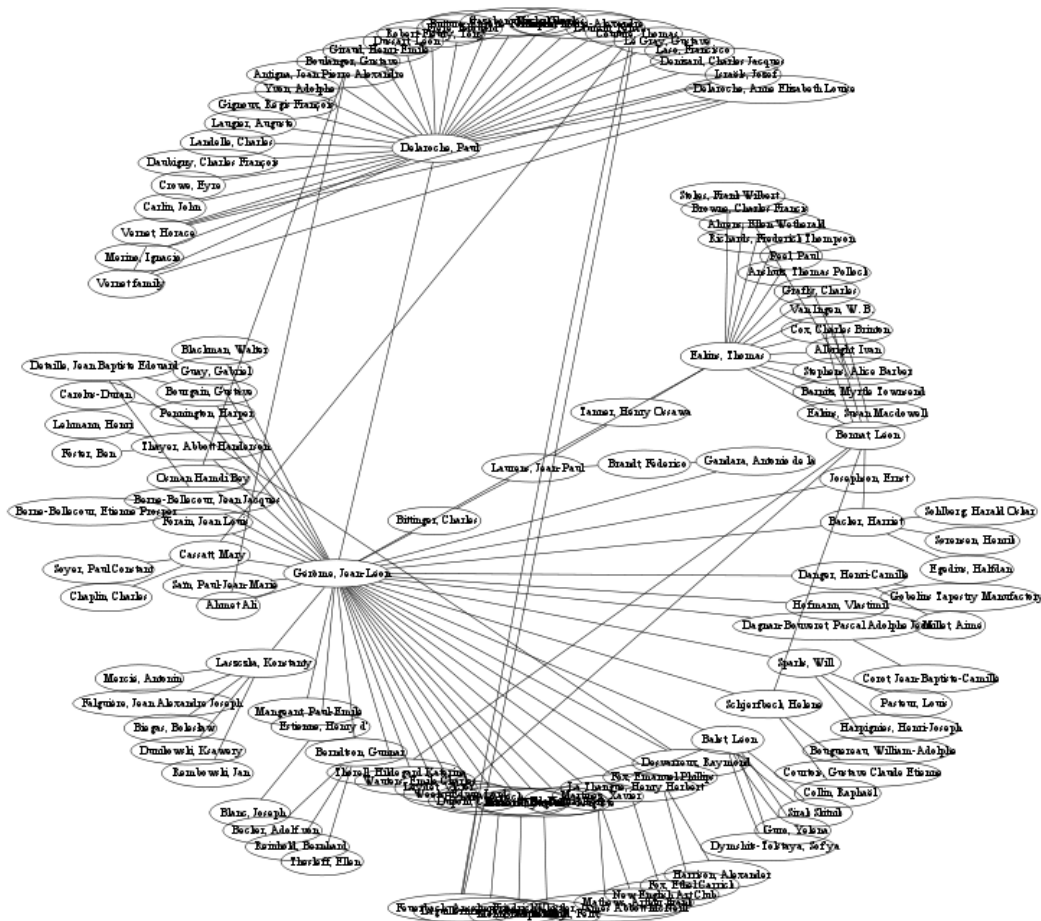
Tulosten visualisointipuolella sosiaalista verkostoa kuvaavan graafin ongelmana on suuren datamäärän näyttäminen. Esimerkiksi Jean-Leon Geromella on kymmeniä kontakteja, joilla puolestaan on kymmeniä lisää, jolloi uloimman kehän henkilöiden nimiä ei enää näe lukea (kuva 16). Tähän saattaisi auttaa solmujen joustavampi sijoittelu, mutta toisaalta silloin voisi olla vaikeaa hahmottaa kokonaisuutta. Vaikka graafin kehä on täynnä, näkee siitä vähintään, että henkilöllä on runsaasti kontakteja ja lisäksi keillä kontakteissa on eniten omia tuttuja.



Kuva 16: Jean-Leon Geromen kontaktit ONKI Peplen graafivisualisoinnissa.



Vertailun vuoksi samasta datasta muodostettiin graafi Graphviz-työkalun<sup>37</sup> twopi-ohjelmalla. Ohjelma sijoittaa solmut niin ikään säteittäin. Lopputulos on kuitenkin varsin vaikealukuinen ja graafi vie paljon tilaa. Kuvassa 17 on esitetty kuvan 16 mukainen Jean-Leon Geromen lähipiiri. Yritimme muodostaa graafia myös Graphvizin muilla algoritmeilla (dot, circo, jne.), mutta kyseisellä datalla graafeista tuli lukukelvottomia.



Kuva 17: Jean-Leon Geromen kontaktit Graphvizin twopi:lla visualisoituna.

ONKI People -hakupalvelun käyttöönoton suurin este on ollut palvelussa käytetyn ULAN-aineiston tiukat lisenssiehdot. Suomessa myös tiukka henkilötietolaki rajoittaa palvelun käyttöönottoa. Palvelun voi pystyttää organisaation sisäverkkoon, mutta silloin tietojen uudelleen käytettävyys on hankalaa ja päädytään todennäköisesti samaan tilanteeseen kuin missä nyt ollaan: eri organisaatiot keräävät ja ylläpitävät päällekkäisiä tietoja samoista henkilöistä. Semanttisen webin ajatus jää toteutumatta ja samankaltaisen järjestelmän voisi toteuttaa millä tahansa tekniikalla.

<sup>37</sup><http://www.graphviz.org/>

Keskitetty reksteri toimijoista, paikoista, aikakausista, teoksista jne. olisi tärkeä komponentti semanttiseen webiin. Nyt tässä roolissa käytetään esimerkiksi Wikipedia-sivuja<sup>38</sup>.

## 5 Tapaustutkimus: Kulttuurisampo

Kulttuurisampo [HMK<sup>+</sup>09] on kansallinen kulttuuriportaali, joka sisältää esimerkiksi taideteoksia, museoesineitä, runoja ja biografioita. Teoksia on koottu useista eri lähteistä, joihin lukeutuu useita suomalaisista museoita ja organisaatioita kuten Kansallismuseo, Ateneum ja Kansalliskirjasto. ULAN-aineistoon pohjautuvat toimijaontologian henkilöt ja ryhmät tuovat oman osuutensa portaalin runsaaseen sisältöön.

Portaaliin on toteutettu useita erilaisia älykkäitä hakukäyttöliittymiä ja suosittelumenetelmiä. Näyttelyn generointi antaa mahdollisuudet tarkastella aineistoa käyttäjän haluamasta kulmasta [MSH07]. Rajausehtoja voi antaa moninäkömahakutyypisesti ja tulokset ryhmitellään virtuaalisiin näyttelyhuoneisiin. Muita hakuominaisuuksia on esim. automaattisesti täydentyvä tekstihaku, joka näyttää tulokset tosiaikaisesti. Lisäksi Kulttuurisamossa on sekä tavallinen karttahaku että ajallisen ulottuvuuden huomioon ottava historiallisia karttoja hyödyntävä haku [KVH08]. KulttuuriSamossa on myös useita teemasivuja, jotka kokoavat aineistoa Kalevalasta saven valamiseen.

Kulttuurisamossa toimijatietoja visualisoidaan ONKI Peoplen graafikomponentilla (luku 4.2). Lisäksi toimijadataa hyödynnetään relaatiohaussa, jossa ideana on valita kaksi henkilöä ja etsiä heitä yhdistävä sosiaalinen polku [KH07].

### 5.1 Toimijatietojen näyttäminen

Toimijan tiedot on koostettu yhdestä tai useammasta lähteestä. Perustietojen lisäksi voidaan näyttää toimijan sosiaalinen verkosto sekä sääntöpohjaisia suosituksia. Kuvassa 18 on Kulttuurisammon muodostama sivu henkilölle Akseli Gallen-Kallela.

Sosiaalinen verkko piirretään ONKI-toimijan visualisointikomponentilla. Kuvassa 19 on esitetty Napoleon ensimmäisen sosiaalinen piiri.

---

<sup>38</sup><http://www.wikipedia.org/>

**Kulttuurisampo**  
suomalainen kulttuuri semanttisessa web 2.0:ssa

Ohje Paasisivu Tietoa portaalista Läheta palautetta

Päähaku:  jäsennelynä Karttahuu ja selauk Yhteyshaku Hae ja jäsennä Kokoelmat Suomen historia Taidot Elämäkerrat Kalevala Karjala

Kieli: suomi svenska English

---

**HENKILÖ: GALLEN-KALLELA, AKSELI (SUOMALAINEN TAITEILIJ, 1865-1931)**

<b>tyyppi</b>	Henkilö
<b>kansallisuus</b>	ruotsalainen, suomalainen, pohjoismaalainen
<b>rooli</b>	graafikot, taiteilijat, taidemaalant, muotoilijat, arkkitehdit
<b>asiasanat</b>	taidemaalari
<b>oppilas on</b>	<a href="#">Simberg, Hugo (suomalainen taiteilija, 1873-1917)</a>
<b>opettaja on</b>	<a href="#">Becker, Adolf von (suomalainen taiteilija, 1831-1909)</a>
<b>tapahtuma</b>	
<b>tietolähde</b>	<a href="#">Kansallisbiografia, SKS, Getty Union List of Artist Names (ULAN)</a>

Kaikki ominaisuudet  
Liitty kohteisiin

**Kulttuurisampo suosittelee henkilön tekemiä teoksia.**

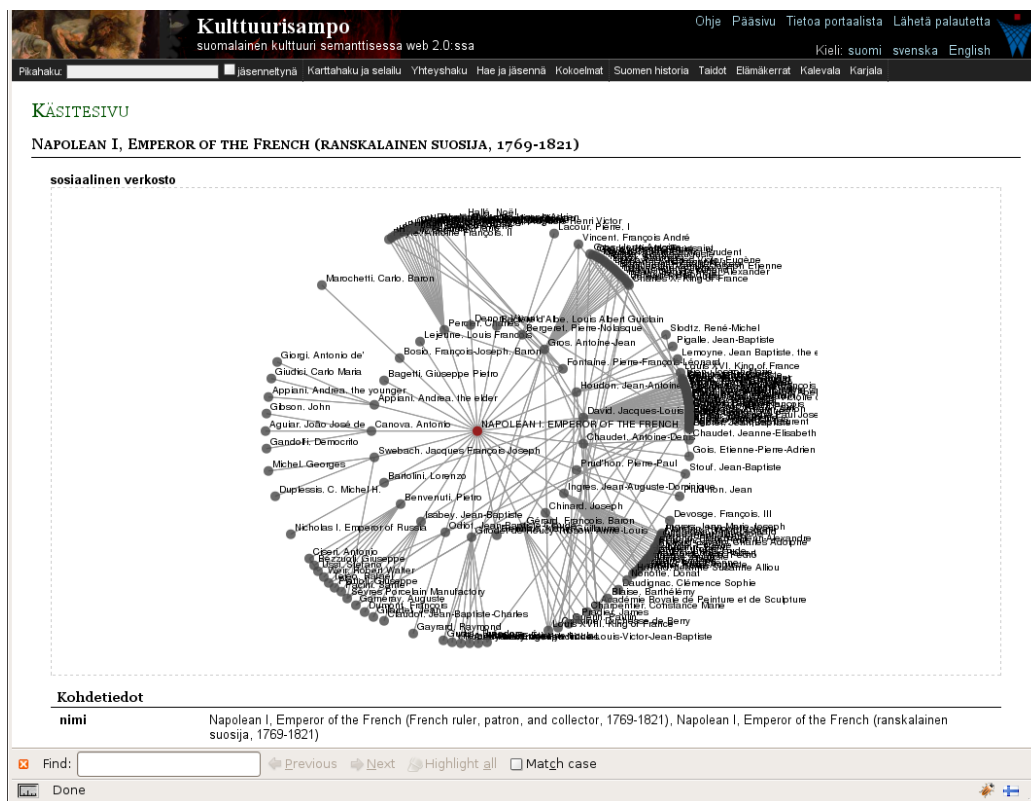

Kuva 18: Kulttuurisampo suosittelee Akseli Gallen-Kallelan sivulla hänen tekemiä teoksia.

## 5.2 Sosiaalisten polkujen hakeminen

Relaatiohaku on eräs Kulttuurisammon uudenlaisista ominaisuuksista. Ajatuksena on rakentaa palvelu, joka pystyy vastaamaan kysymykseen: kuinka asia X liittyy asiaan Y. Vastaukseksi annetaan asioita yhdistävä ketju. Semanttisia yhteyksiä on tutkittu esim. terroristien tunnistamisessa [SAMA<sup>+</sup>05].

Sosiaaliset polut muodostavat erikoistapauksen, jossa yhdistävä polku on muotoa henkilö X tuntee henkilön Y, joka tuntee henkilön Z jne. [KH07]. Sosiaaliset verkostot ovat olleet jo pitkään tutkimuksen alla [WDN02]. Pieni-maailma-ilmio ja väite kuuden askeleen yhteydestä jokaisen ihmisen välillä on useimmille tuttu. ULAN-aineisto tarjoaa käytännön mahdollisuudet kokeilla ilmiötä kuuluisien taidehenkilöiden tapauksessa.

Etsimällä henkilöiden välisiä yhteyksiä voidaan vastata esimerkiksi kysymykseen kuinka Gallen-Kallela liittyy Napoleoniin. Kuvassa 20 käyttäjä on syöttänyt Gallen-Kallelan toiseen kenttään ja Napoleonin toiseen. Relaatiohaku etsii henkilöitä yhdistävän lyhimmän polun. Kuvassa järjestelmä on löytänyt Gallen-Kallelan ja Napoleonin väliltä kuuden henkilön kautta kulkevan yhdistävän polun.



Kuva 19: Napoleon ensimmäisen sosiaalinen piiri Kulttuurisammon visualisoimana.

**KulttuuriSampo**

Teemat Relatiiohaku Tee oma näyttelysi

Name	Name
Gallen-Kallela, Akseli (Finnish painter and graphic artist, 1865-1931)	Knows? Napoleon I, Emperor of the French (French ruler, patron, and collector, 1769-1821)

Gallen-Kallela, Akseli (Finnish painter and graphic artist, 1865-1931)  
 studentOf  
 Becker, Adolf von (Finnish painter, 1831-1909)  
 teacherOf  
 Berndtson, Gunnar (Finnish painter and illustrator, 1854-1895)  
 studentOf  
 Gérôme, Jean-Léon (French painter and sculptor, 1824-1904)  
 teacherOf  
 Burnand, Eugène (Swiss painter and illustrator, 1850-1921)  
 studentOf  
 Menn, Barthélemy (Swiss painter and teacher, 1815-1893)  
 studentOf  
 Ingres, Jean-Auguste-Dominique (French painter and draftsman, 1780-1867)  
 patronWas  
 Napoleon I, Emperor of the French (French ruler, patron, and collector, 1769-1821)  
 Search took: 3ms

Kuva 20: Sosiaalisten polkujen etsiminen.

### 5.2.1 Idea

Relaatiohakuun on tehty selainpohjainen käyttöliittymä. Prosessi datasta käyttöliittymään etenee seuraavasti.

1. Muokataan alkuperäinen data Toimijaontologian muotoon.
2. Rakennetaan datasta muistinvarainen esitys.
3. Vastanotetaan käyttöliittymästä kysely kahden henkilön välisestä yhteydestä.
4. Haetaan muistinvaraisesta tietorakenteesta yhteyksiä perinteisillä hakualgoritmeilla.
5. Analysoidaan löydettyä yhdistävää polkua: tietokannasta haetaan esimerkiksi henkilöiden nimet ja kontaktien tarkemmat määrittelyt.
6. Lähetetään polku näytettäväksi käyttöliittymään.

Kohdat 1 ja 2 tehdään esiprosessointina. Kohdat 3-6 ovat interaktiivisia ja perustuvat käyttöliittymästä saapuviin kyselyihin. Seuraavassa eritellään toteutuksen yksityiskohtia.

### 5.2.2 Tekninen toteutus

ULAN-aineisto on muokattu XSLT-muunnoksen<sup>39</sup> avulla Toimijaontologian muotoon. RDF-verkosta on seuraavaksi listattu kaikki toimijat ja toimijoiden väliset kontaktit. Toimijat on lisätty muistinvaraiseen verkkotietorakenteeseen solmuiksi. Kontaktit on asetettu toimijoiden väliseksi kaariksi. Kaikki kontaktit on yleistetty ”tuntee”-tyyppisiksi.

Muistinvarainen esitys on toteutettu yksinkertaisena Java-tietorakenteena. Hakualgoritmina on käytetty leveysuuntaista hakua (breadth-first search). Aika- ja tilavaativuus ovat vaativuusluokassa  $O(n)$ , missä  $n$  on verkon solmujen lukumäärä. Haku laajenee eksponentiaalisesti, mutta pahimmillaankin joudutaan käymään kaikki solmut vain kerran läpi, sillä jo tutkittuja solmuja ei tarvitse tutkia missään tilanteessa uudestaan. Iteratiivisella syvyysuuntaisella haulla saataisiin muistivaatimukset pienemmiksi, mutta aikavaativuus olisi sama ja toteutus hieman monimutkaisempi.

---

<sup>39</sup><http://www.w3.org/TR/xslt>

Selainpohjainen käyttöliittymä on toteutettu AJAX-komponenteilla ja se on sovel-lusohjelman kaltainen siinä, että sivua ei tarvitse ladata uudestaan kyselyjen välillä. Käyttöliittymässä on kaksi tekstikenttää, joihin syötetään henkilöiden nimet. Ken-tissä on automaattinen täydennys, joka helpottaa toimijan valitsemista. Kun molem-mat nimet on asetettu, haku käynnistyy automaattisesti. Haku kestää käytännössä alle puoli sekuntia ja tulokset näytetään välittömästi käyttöliittymässä.



Kuva 21: Automaattisesti täydentyvä tekstikenttä.

Automaattisesti täydentyvä tekstikenttä (kuva 21) perustuu DOJO-komponenttiin<sup>40</sup>. Komponentti kommunikoi palvelimen kanssa AJAX-kyselyjen kautta. Palvelinpääs-sä kyselyihin vastataan Lucene-indeksin<sup>41</sup> avulla. Näin saadaan tehtyä reaaliaikaises-ti toimiva automaattisesti täydentyvä hakukenttä, joka tukee yli 200000:aa nimeä.

Yhteyksien haussa kommunikointi käyttöliittymän ja palvelimen välillä on toteutet-tu DWR-rajapinnan<sup>42</sup> kautta. DWR julkaisee Java-servletit Javascript-rajapintoina. Servletti hoitaa varsinaisen haun palvelinpäässä ja Javascript-rajapinta mahdollis-taa interaktiivisen kommunikoinnin web-sivun kautta. Dojokomponentilta saadaan siis henkilöiden nimet ja DWR-komponentin kautta tehdään lopullinen yhteyden haku palvelimelle.

### 5.2.3 Tulokset

Haku on nopea. Pisimmätkin yhteydet löytyvät alle puolessa sekuntissa. Käytännös-sä reitinhakukyselyyn liittyvä oheisprosessointi vie enemmän aikaa kuin itse haku. Haun nopeus selittyy osittain datalla. Vaikka ULAN:ssa on noin 120 000 toimijaa,

<sup>40</sup><http://dojotoolkit.org/>

<sup>41</sup><http://lucene.apache.org/>

<sup>42</sup><http://getahead.org/dwr>

yhteyksiä on suhteessa vähän. Suurimmalle osalle henkilöistä ei ole listattu lainkaan kontakteja. Verkko sisältää noin 12000 henkilön ja ryhmän vahvasti kytkeytyneen komponentin, johon haku usein kohdistuu. Tästä komponentista löytyvät niin Gallen-Kallela kuin Napoleonkin.

#### 5.2.4 Arviointi

Yhteyshaussa löydettyjen polkujen pituus herättää kysymyksen yhdistävän tiedon hyödyllisyydestä. Pitkät polut ovatkin lähinnä vain hauskaa tietoa vailla todellista merkitystä. Jos Akseli Gallen-Kallela kytkeytyy Napoleoniin kuuden kontaktin kautta, voi tuskin sanoa, että Gallen-Kallelalla ja Napoleonilla on mitään yhteyttä keskenään. Lyhyemmistä poluista voi sen sijaan löytää myös merkityksellisiä yhteyksiä, esimerkiksi taiteilijoiden yhteisiä suosijoita. Myös sukututkijat saattaisivat olla kiinnostuneita järjestelmän mahdollisuuksista, jos aineisto olisi kattavampi. Aineisto onkin keskeinen mahdollistava tekiä yhteyshaussa. Jatkotutkimuksen kannalta olisi kiinnostavaa laajentaa järjestelmää myös toimijoiden ulkopuolelle, esimerkiksi tapahtumiin tai teoksiin.

## 6 Jatkotutkimus

Tutkielman kaksi keskeistä tulosta ovat toimijaontologia toimijadatan esittämiseen sekä ONKI People tietojen julkaisemiseen ja hyödyntämiseen. Tässä luvussa pohditaan toimijadatan automaattista keräämistä sekä vaihtoehtoa keskitetylle toimijarekisterille.

Tutkielmassa esitellyt toimijatiedot perustuvat asiantuntijoiden muodostamiin rekistereihin. Käyttäjän mallinnus (user modeling) on toinen lähestymistapa profiilitietojen keräämiseen. Siinä pyritään muodostamaan käyttäjää kuvaava profiili käyttäjää tarkkailemalla. Profiilia voidaan hyödyntää esimerkiksi suosituksissa tai haun personoinnissa.

ONKI People on keskitetty palvelu toimijatietojen julkaisuun ja hakuun. Vaihtoehtoinen lähestymistapa keskitetylle rekisterille olisi esimerkiksi vertaisverkkopohjainen hajautettu palvelu. Hajautetussa ratkaisussa vastuu tietojen ylläpidosta ja julkaisusta voitaisiin antaa käyttäjille.

## 6.1 Käyttäjän mallinnus

Skeema määrää muotin, johon toimijatiedot tallennetaan. Se ei ota kantaa tietojen hyödyntämiseen. Käyttäjän mallinnus on uudempi tutkimuskohde [BKW07, SMB07]. Siinä kone yritetään saada ymmärtämään ja ennakoimaan käyttäjän tarpeita. Käyttäjämalli voidaan rakentaa profilin perusteella ja se määrittelee uuden korkeamman abstraktiotason toimijan kuvaamiseen.

Semanttisen webin tekniikoita on hyödynnetty myös käyttäjän mallinnuksessa. Käyttäjämallin esitysmuotona on tutkittu ontologioita [RAM03] ja suosittelussa voidaan hyödyntää ontologisia toimijarekistereitä [MNDR04] vähentämään ns. ”kylmäkäynnistyksen” ongelmaa.

Käyttäjämalleja voi hyödyntää niin suosittelussa kuin käyttöliittymien suunnittelussa ja käsitteet lähestyvät toisiaan adaptiivisen käyttöliittymän muodossa. Käyttöliittymässä suositaan asioita, joita käyttäjä todennäköisimmin tarvitsee, käyttäjälle lähetetään tarjouksia tuotteista, jotka täsmäävät käyttäjän profiliin jne.

Käyttäjän mallinnus muodostuu karkeasti seuraavista vaiheista:

1. Datan kerääminen (kysyminen käyttäjältä, HTTP-kutsujen lokit)
2. Datan tallennus malliin (ontologiat, sanasäkit, henkilö-tuote -matriisit)
3. Päättely (esim. koneoppisen keinoin: päätöspuu, naiivi bayes,...)
4. Käyttöliittymän muodostaminen (suositusten esittäminen, adaptiiviset käyttöliittymät)

Dataa voidaan kerätä joko kyselemällä käyttäjältä tai taustalla käyttäjää tarkkaillen. Kysyminen voi olla esimerkiksi kohteiden pisteyttämistä tai yleisten kiinnostuksen kohteiden listaamista rekisteröintivaiheessa. Käyttäjän keskeyttämistä pidetään yleensä huonona vaihtoehtona sen aiheuttaman lisäkuormituksen takia. Lisäksi kiinnostuksen kohteet vaihtuvat ja voi olla vaikeaa itsekään tietää mikä kiinnostaa.

Käyttäjän näkymättömässä tarkkailussa keskeisin keino on palvelimen lokitiedot. Tiedoista voidaan tutkia esim. kuinka usein käyttäjä on katsonut jotain tiettyä kohdetta, kuinka pitkään kohdetta on katsottu ja koska viimeksi. Tietojen tulkinta on kuitenkin hankalaa. Jos käyttäjä katsoo sivua 20 minuuttia, lukeeko hän pitkää artikkelia vai onko hän kahvitauolla?

Käyttäjämalli voi olla esimerkiksi sanalista, käyttäjä-kohde matriisi tai käyttäjäontologia [GCP03, MNDR04, HHTS05, NSP03]. Ontologioita voidaan hyödyntää esim.



painottamalla käyttäjäontologian tiettyjä käsitteitä katsottujen kohteiden asiansanoihin täsmäämällä. Datasta voidaan tehdä päätelmiä myös perinteisin koneoppimisen keinoin.

Eräs kiinnostava malli on yhteisöllinen suodatus, jossa päätelmiä laajennetaan etsimällä samankaltaisia henkilöitä. ”Naapurille” voidaan suositella omia samankaltaisia, mutta ennestään tuntemattomia kohteita. Samankaltaisuus määrittyy subjektiivisen ihmisen tekemän arvion kautta, eikä esim. kohteen ominaisuuksien perusteella.

Menestyneimpiä esimerkkejä käyttäjämalleja hyödyntävistä järjestelmistä ovat verkko-kauppa Amazon<sup>43</sup> ja sosiaalinen musiikkiportaali Last.fm<sup>44</sup>.

## 6.2 Hajautettu palvelu

Tässä tutkielmassa esitetty ONKI People on keskitetty rekisteri toimijatietojen hallintaan. Keskitetty rekisteri on ongelmallinen esimerkiksi tietosuoja ja lisensointi kysymysten takia. Vaihtoehtoinen lähestymistapa keskitetylle palvelulle olisi järjestelmä, jossa käyttäjät itse julkaisisivat omat tietonsa. Hajautetun palvelun voisi toteuttaa esimerkiksi vertaisverkkojen (Peer to Peer networks) avulla, jolloin käyttäjät eivät olisi riippuvaisia tietystä palveluntarjoajasta kuten Facebookista<sup>45</sup> tai MSN Messengeristä<sup>46</sup>.

Vertaisverkoissa tehokas tiedon reititys ilman keskittettyä palvelintä voidaan toteuttaa esimerkiksi hajautettujen hajautustaulujen avulla [PRR99, ZHS<sup>+</sup>04, RD01]. Tällä hetkellä kenties tunnetuin vertaisverkkoihin ja hajautettuun hajautustauluun perustuva toteutus on BitTorrent-tiedostonjakojärjestelmä ja sen hyödyntämä Kademlia-algoritmi [MM02]. Vaikka reititys on tehokasta, tiedonhaku ilman keskitettyjä indeksejä on hankalaa. Tähän on esitetty ratkaisuksi muun muassa sanojen alkuosiin perustuvaa hakua ja vertaisverkon tilastollista analysointia [AS03, BMPC07].

Henkilötietojen hakupalvelun voisi rakentaa myös sosiaalisten verkostojen päälle. Ystävien ja tuttujen kautta muodostuisi sosiaalinen verkko, jossa viestien välitys tapahtuisi vain luotettavien kontaktien kautta [LD06].

---

<sup>43</sup><http://www.amazon.com/>

<sup>44</sup><http://last.fm/>

<sup>45</sup><http://facebook.com>

<sup>46</sup><http://messenger.msn.com/>

## 7 Yhteenveto

Tässä tutkielmassa käsiteltiin henkilö- ja organisaatitietojen esittämistä semanttisen webin keinoin. Pääpaino oli datan mallintamisessa, haussa sekä visualisoinnissa. Mallia ja työkaluja tarkasteltiin kirjastoissa käytettyjen auktoriteettijärjestelmien näkökulmasta tärkeimpinä päämäärinä toimijoiden yksilöiminen ja hakeminen. Jatkotutkimusosiossa pohdittiin käyttäjämalleja ja hajautettua henkilörekisteriä.

Luvussa 2 määriteltiin toimijan käsite ja sitä määrittävät keskeiset ominaisuudet. Kuten koko tutkielmassa, näkökulmana oli henkilötietojen hyödyntäminen webmaailmassa. Keskeisin ongelma on toimijoiden identiteettien hallinta. Tärkeiksi toimijaa kuvaaviksi ominaisuuksiksi nostettiin mm. hakua helpottavat sijainti ja rooli.

Luvussa 3 käsiteltiin toimijatietojen esittämistä käytännössä. Keskeisinä kysymyksenä oli tunnisteen ja nimien esittäminen ja hallitseminen. Näiden tietojen lisäksi pohdittiin sekä roolien että kansallisuuksien esittämistä. Luvussa käsiteltiin myös ajallisen tiedon kuten syntymä- ja kuolinaikojen tai taiteilijan aikakausten esittämistä. Lisäksi esiteltiin runsaasti olemassa olevia malleja niin kirjastomaailmasta kuin semanttisen webin kontekstista. Lopuksi määriteltiin edellä esiteltyihin sanastoihin perustuva toimijaontologia toimijatietojen esittämiseen.

Luvussa 4 esiteltiin ONKI People -palvelua toimijatietojen hyödyntämiseen. Palvelu koostuu hakukäyttöliittymästä (moninäkömähaku) ja visualisointikomponentista (sosiaalisen verkon visualisointi) sekä konerajapinnoista (web-palvelut ja AJAX-rajapinnat). Hakuliittymiä havainnollistettiin ja verrattiin olemassa oleviin palveluihin (Amazon.com ja MIT Exhibit). Lopuksi kuvattiin lyhyesti ONKI Peoplen keskeisiä ominaisuuksia toteutusnäkökulmasta.

Lopuksi luvussa 5 kuvattiin toimijatietojen hyödyntämistä Kulttuurisampo-portaalissa. Yhteyshaku ja graafivisualisaatio ovat portaalin tärkeimmät toimijatietoa hyödyntävät komponentit. Yhteyshaussa ideana on etsiä kahta henkilöä yhdistävä sosiaalinen polku yhteisten tuttavien kautta. Sosiaalisen piirin visualisointikomponenttia on suoraan sovellettu ONKI People -järjestelmästä.

Toimijatieto keskeisessä roolissa niin perinteisessä kirjastotiedonhaussa kuin sosiaalisissa web-palveluissa. Kirjastomaailmassa henkilöiden kautta haetaan teoksia, sosiaalisissa palveluissa trendejä ja ideoita. Kummassakin tapauksessa tärkeää on identiteettien hallinta, toimijoiden löytäminen ja toimijoiden hyvä linkittyvyys muihin entiteetteihin kuten toimijoihin, teoksiin, paikkoihin ja aikoihin.

Kirjastomaailmassa perinteinen lähtökohta on ollut rakentaa ns. auktoritettijärjes-

telmä, jossa jokainen entiteetti tunnustetaan kontrolloidulla nimenmuodolla. Järjestelmä on luotettava ja hyväksi havaittu, mutta vanhanaikainen. Ylläpito on työlästä ja vaatii suurta asiantuntemusta. Lisäksi tiedot linkittyvät huonosti järjestelmän ulkopuolelle. Myös monikielisyys hallinta on ongelmallista käyttäessä nimeen perustuvia tunnisteita ja kansallisia variaatioita esimerkiksi translitteroinneissa.

Web-maailmassa lähtökohtana on ollut joustavuus. Joissain palveluissa nimi on suoraan tunniste (esim. wikipedia ja last.fm). Tällöin samalle entiteetille saattaa syntyä useita sivuja, josta seuraa, että kaikkea entiteettiin liittyvää tietoa on vaikea löytää. Ongelmia on koetettu paikata esimerkiksi linkittämällä resursseja ristiin ja lisäämällä käsitteitä erottelevia sivuja (esim. wikipedia disambiguation-sivut).

Semanttinen web yrittää yhdistää kirjastomaailman jämäkkyyttä webin joustavuuteen. Entiteetit tunnustetaan yleismaailmallisella muuttumattomalla URI-tunnisteella, joka tyypillisesti viittaa johonkin web-resurssiin. Tällöin URI toimii myös URL:na. Tieto kuvataan W3C:n suosittelemalla RDF-formaatilla, joka on yksinkertainen mutta ilmaisuvoimainen. Tavoitteena on muodostaa yhteisiä sanastoja ja ontologioita parantamaan koneluettavuutta ja uudelleenkäytettävyyttä. Monikielisten sanastojen avulla voidaan rakentaa myös kielirajat ylittäviä hakujärjestelmiä. Lisäksi data pyritään tarjoamaan erilaisten sovellusrajapintojen kautta (esim. web-palvelut, REST-rajapinnat jne.) sen sijaan, että tieto olisi saatavilla vain ihmisille tarkoitettuina web-sivuina.

Tässä tutkielmassa määriteltiin malli henkilö- ja organisaatietietojen esittämiseen. Lisäksi esiteltiin prototyyppipalvelu tietojen hakemiseen ja hyödyntämiseen niin ihmis- kuin sovellusrajapintojen kautta. Semanttisen webin perustana on pidetty jaettuja sanastoja ja ontologioita. Nimetyt entiteetit ovat saaneet vähemmän huomiota. Jaetut ja avoimet rekisterit toimijoista, paikoista ja teoksista tuovat tärkeän lisän semanttisen webin työkalupakkiin.

## Kiitokset

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

Haluan kiittää professori Eero Hyvöstä työn ohjaamisesta ja hyvistä kommentteista. Erityiset kiitokset myös huonetovereille Eetu Mäkelälle, Reetta Sinkkilälle ja Panu Paakkariselle sekä Robin Lindroosille hyvästä seurasta.

## Lähteet

- AMBB<sup>+</sup>07 Aleman-Meza, B., Bojars, U., Boley, H., Breslin, J., Mochol, M., Nixon, L., Polleres, A. ja Zhdanova, A., Combining RDF vocabularies for expert finding. *Proceedings of the 4th European conference on The Semantic Web (ESWC 2007)*, Innsbruck, Austria, kesäkuu 2007, Springer-Verlag.
- AS99 Amato, G. ja Straccia, U., User profile modeling and applications to digital libraries. *Proceedings of the 3rd European Conference on Research and Advanced Technology for Digital Libraries (ECDL 1999)*, Paris, France, syyskuu 1999, Springer-Verlag.
- AS03 Awerbuch, B. ja Scheideler, C., Peer-to-peer systems for prefix search. *Proceedings of the Symposium on Principles of Distributed Computing (PODC 2003)*, Boston, MA, USA, heinäkuu 2003, ACM Press.
- BKW07 Brusilovsky, P., Kobsa, A. ja Wolfgang, N., toimittajat, *The Adaptive Web: Methods and Strategies of Web Personalization*. Lecture Notes in Computer Science. Springer-Verlag, 2007.
- BMPC07 Bender, M., Michel, S., Parreira, J. X. ja Crecelius, T., P2P web search: Make it light, make it fly. *Proceedings of 3rd Biennial Conference on Innovative Data Systems Research (CIDR 2007)*, Asilomar, CA, USA, tammikuu 2007.
- BPSV09 Bouquet, P., Palpanas, T., Stoermer, H. ja Vignolo, M., A conceptual model for a web-scale entity name system. *Proceedings of the 4th Asian Semantic Web Conference (ASWC 2009)*, Shanghai, China, 2009, Springer-Verlag.
- Bro00 Broder, A. J., Data mining, the internet, and privacy. *Revised Papers from the International Workshop on Web Usage Analysis and User Profiling, WEBKDD '99*, London, UK, 2000, Springer-Verlag, sivut 56–73.
- BS92 Borgman, C. ja Siegfried, S., Getty's synonyme and its cousins: A survey of applications of personal name-matching algorithms. *Journal of the American Society for Information Science and Technology*, 43,7(1992), sivut 459–476.

- Che76 Chen, P. P.-S., The entity-relationship model: Toward a unified view of data. *ACM Transactions on Database Systems*, 1,1(1976), sivut 9–36.
- CRA99 Cranor, L. F., Reagle, J. ja Ackerman, M. S., Beyond concern: Understanding net users' attitudes about online privacy. Tekninen raportti TR 99.4.3, AT&T Labs-Research, 1999.
- FNFK06 Fukazawa, Y., Naganuma, T., Fujii, K. ja Kurakake, S., Construction and use of role-ontology for task-based service navigation system. *Proceedings of the 5th International Semantic Web Conference (ISWC 2006)*, Athens, GA, USA, marraskuu 2006, Springer-Verlag.
- FPS97 French, J. C., Powell, A. L. ja Schulman, E., Automating the construction of authority files in digital libraries: A case study. *Proceedings of the 1st European Conference on Research and Advanced Technology for Digital Libraries (EDLC 1997)*, Pisa, Italy, syyskuu 1997, Springer-Verlag.
- FPS00 French, J., Powell, A. ja Schulman, E., Using clustering strategies for creating authority files. *Journal of the American Society for Information Science*, 51,8(2000), sivut 774–786.
- Fra03 Franklin, R., Re-inventing subject access for the semantic web. *Online Information Review*, 2,27(2003), sivut 94–101.
- GCP03 Gauch, S., Chaffee, J. ja Pretschner, A., Ontology-based personalized search and browsing. *Web Intelligence and Agent System*, 1,3(2003), sivut 219–234.
- GMA07 Galvez, C. ja Moya-Anegon, F., Approximate personal name-matching through finite-state graphs. *Journal of the American Society for Information Science and Technology*, 58,13(2007), sivut 1960–1976.
- Gra73 Granovetter, M., The strength of weak ties. *American Journal of Sociology*, 78,6(1973), sivut 1360–1380.
- Gua92 Guarino, N., Concepts, attributes and arbitrary relations. *Data Knowledge Engineering*, 8,3(1992), sivut 249–261.
- HEE<sup>+</sup>02 Hearst, M., Elliott, A., English, J., Sinha, R., Swearingen, K. ja Yee, K., Finding the flow in web site search. *Communications of the ACM*, 45,9(2002), sivut 42–49.

- HHTS05 Haase, P., Hotho, A., Thieme, L. ja Sure, Y., Collaborative and usage-driven evolution of personal ontologies. *Proceedings of the 2nd European Semantic Web Conference (ESWC 2005)*, Heraklion, Greece, toukokuu 2005, Springer-Verlag.
- HMK<sup>+</sup>09 Hyvönen, E., Mäkelä, E., Kauppinen, T., Alm, O., Kurki, J., Ruotsalo, T., Seppälä, K., Takala, J., Puputti, K., Kuittinen, H., Viljanen, K., Tuominen, J., Palonen, T., Frosterus, M., Sinkkilä, R., Paakkari, P., Laitio, J. ja Nyberg, K., Culturesampo – a national publication system of cultural heritage on the semantic web 2.0. *Proceedings of the 6th European Semantic Web Conference (ESWC 2009)*, Heraklion, Greece, toukokuu 2009, Springer-Verlag.
- HSVF07 Hyvönen, E., Seppälä, K., Viljanen, K. ja Frosterus, M., Yleinen suomalainen ontologia YSO – kohti suomalaista semanttista webiä. *Tietolinja*, toukokuu 2007.
- HVTS08 Hyvönen, E., Viljanen, K., Tuominen, J. ja Seppälä, K., Building a national semantic web ontology and ontology service infrastructure – the FinnONTO approach. *Proceedings of the 5th European Semantic Web Conference (ESWC 2008)*, Tenerife, Spain, kesäkuu 2008, Springer-Verlag.
- KH07 Kurki, J. ja Hyvönen, E., Relational semantic search: Searching social paths on the semantic web. *Poster Proceedings of the 6th International Semantic Web Conference (ISWC 2007)*, Busan, Korea, marraskuu 2007, Springer-Verlag.
- KSS97 Kautz, H., Selman, B. ja Shah, M., Combining social networks and collaborative filtering. *Communications of the ACM*, 40,3(1997), sivut 63–65.
- Kur08 Kurki, J., Finding people and organizations on the semantic web. *Proceedings of the 13th Finnish Artificial Intelligence Conference (STeP 2008)*, Espoo, Finland, elokuu 2008, Finnish Artificial Intelligence Society.
- KVH05 Komulainen, V., Valo, A. ja Hyvönen, E., A collaborative ontology development and service framework ONKI. *Poster Proceedings of the 2nd*

*European Semantic Web Conference (ESWC 2005)*, Heraklion, Greece, toukokuu 2005, Springer-Verlag.

- KVH08 Kauppinen, T., Väättäinen, J. ja Hyvönen, E., Creating and using geospatial ontology time series in a semantic cultural heritage portal. *Proceedings of the 5th European Semantic Web Conference (ESWC 2008)*, Tenerife, Spain, kesäkuu 2008, Springer-Verlag.
- LD06 Li, J. ja Dabek, F., F2F: Reliable storage in open networks. *Proceedings of the 5th International Workshop on Peer-to-Peer Systems (IPTPS 2006)*, Santa Barbara, CA, USA, helmikuu 2006.
- Lin08 Lindroos, R., Paikkatiedon ontologiapalvelu. Pro gradu, Helsinki University of Technology (TKK), toukokuu 2008.
- Loe03 Loebe, F., An analysis of roles: Towards ontology-based modelling. Tekninen raportti 6, Onto-Med Research Group, University of Leipzig, 2003.
- MM02 Maymounkov, P. ja Mazières, D., Kademia: A peer-to-peer information system based on the XOR metric. *Revised Papers from the 1st International Workshop on Peer-to-Peer Systems, IPTPS 2002*, London, UK, 2002, Springer-Verlag, sivut 53–65.
- MNDR04 Middleton, S., Nigel, R. ja De Roure, D., Ontological user profiling in recommender systems. *ACM Transactions on Information Systems*, 22,1(2004), sivut 54–88.
- MSH07 Mäkelä, E., Suominen, O. ja Hyvönen, E., Automatic exhibition generation based on semantic cultural content. *Proceedings of the Cultural Heritage on the Semantic Web Workshop at the 6th International Semantic Web Conference (ISWC 2007)*, Busan, Korea, marraskuu 2007, Springer-Verlag.
- MVB<sup>+</sup>04 Masolo, C., Vieu, L., Bottazzi, E., Catenacci, C., Ferrario, R., Gange-mi, A. ja Guarino, N., Social roles and their descriptions. *Proceedings of the 9th International Conference on Principles of Knowledge Representation and Reasoning (KR 2004)*, Whistler, Canada, kesäkuu 2004, AIII Press.



- NSP03 Nebel, I., Smith, B. ja Paschke, R., A user profiling component with the aid of user ontologies. *Proceedings of the Learning-Teaching-Knowledge-Adaptivity (LLWA 03)*, Karlsruhe, Germany, lokakuu 2003.
- PMB<sup>+</sup>07 Parreira, J. X., Michel, S., Bender, M., Crecelius, T. ja Weikum, G., P2P authority analysis for social communities. *Proceedings of the 33rd International Conference on Very Large Data Bases (VLDB 2007)*, Vienna, Austria, syyskuu 2007, VLDB Endowment.
- PRR99 Plaxton, G., Rajaraman, R. ja Richa, A., Accessing nearby copies of replicated objects in a distributed environment. *Theory of Computing Systems*, 32,3(1999), sivut 241–280.
- RAM03 Razmerita, L., Angehrn, A. A. ja Maedche, A., Ontology-based user modeling for knowledge management systems. *Proceedings of the 9th International Conference on User Modeling (UM 2003)*, Pittsburgh, Johnstown, USA, kesäkuu 2003, Springer-Verlag, sivut 213–217.
- RD01 Rowstron, A. ja Druschel, P., Pastry: Scalable, decentralized object location and routing for large-scale peer-to-peer systems. *Proceedings of the IFIP/ACM 18th International Conference on Distributed Systems Platforms (Middleware 2001)*, Heidelberg, Germany, marraskuu 2001, Springer-Verlag.
- SAMA<sup>+</sup>05 Sheth, A., Aleman-Meza, B., Arpinar, I. B., Bertram, C., Warke, Y., Ramakrishnan, C., Halaschek, C., Anyanwu, K., Avant, D., Arpinar, F. S. ja Kochut, K., Semantic association identification and knowledge discovery for national security applications. *Journal of Database Management on Database Technology*, 16,1(2005), sivut 33–53.
- SCDT00 Srivastava, J., Cooley, R., Deshpande, M. ja Tan, P.-N., Web usage mining: discovery and applications of usage patterns from web data. *ACM SIGKDD Explorations Newsletter*, 1,2(2000), sivut 12–23.
- SMB07 Sieg, A., Mobasher, B. ja Burke, R., Web search personalization with ontological user profiles. *Proceedings of the 16th ACM Conference on Information and Knowledge Management (CIKM 2007)*, Lisboa, Portugal, marraskuu 2007, ACM Press.
- Tay06 Taylor, A., *Introduction to Cataloging and Classification*. Library and Information Science Text Series. Libraries Unlimited, 2006.

- TFVH09 Tuominen, J., Frosterus, M., Viljanen, K. ja Hyvönen, E., Onki skos server for publishing and utilizing skos vocabularies and ontologies as services. *Proceedings of the 6th European Semantic Web Conference (ESWC 2009)*, Heraklion, Greece, toukokuu 2009, Springer-Verlag, sivut 768–780.
- Til04 Tillett, B., Authority control: State of the art and new perspectives. *Cataloging & classification quarterly*, 38,3(2004), sivut 23–41.
- TKVH09 Tuominen, J., Kauppinen, T., Viljanen, K. ja Hyvönen, E., Ontology-based query expansion widget for information retrieval. *Proceedings of the 5th Workshop on Scripting and Development for the Semantic Web (SFSW 2009), 6th European Semantic Web Conference (ESWC 2009)*, Heraklion, Greece, toukokuu 2009, Springer-Verlag.
- VA08 V., P. ja A., G., Identity of resources and entities on the web. *International Journal on Semantic Web and Information Systems*, 4,2(2008), sivut 49–72.
- WC05 Williams, J. P. ja Copes, H., ”How Edge Are You?” Constructing authentic identities and subcultural boundaries in a straightedge internet forum. *Symbolic Interaction*, 28,1(2005), sivut 67–89.
- WDN02 Watts, D., Dodds, P. ja Newman, M., Identity and search in social networks. *Science*, 296,5571(2002), sivut 1302–1305.
- WJ95 Wooldridge, M. ja Jennings, N. R., Intelligent agents: Theory and practice. *The Knowledge Engineering Review*, 10,2(1995), sivut 115–152.
- ZDLW10 Zhang, J., Duan, H., Liu, W. ja Wu, J., A light-weighted extension of anonymous communications in ipv6 network. *Proceedings of the 1st International Conference on Green Circuits and Systems (ICGCS 2010)*, Shanghai, China, kesäkuu 2010.
- Zha01 Zhang, S. L., Planning an authority control project at a medium-sized university library. *College and Research Libraries*, 62,5(2001), sivut 395–405.
- ZHS+04 Zhao, B., Huang, L., Stribling, J., Rhea, S., Joseph, A. ja Kubiawicz, J., Tapestry: A resilient global-scale overlay for service deployment.

*IEEE Journal on Selected Areas in Communications*, 22,1(2004), sivut  
41–53.

## Liite 1. Toimijaontologian skeema

Toimijaontologiaan sisältyy seuraavat ominaisuudet ja luokat. Sanasto on esitetty Turtle-muodossa<sup>47</sup>.

```
@prefix rdf:      <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs:     <http://www.w3.org/2000/01/rdf-schema#> .
@prefix owl:   <http://www.w3.org/2002/07/owl#> .
@prefix foaf:    <http://xmlns.com/foaf/0.1/> .
@prefix bio:     <http://purl.org/vocab/bio/0.1/> .
@prefix rel:     <http://purl.org/vocab/relationship/> .
@prefix toimo:   <http://seco.tkk.fi/onto/toimo/> .
@prefix roles:   <http://seco.tkk.fi/onto/toimo/roles/> .
@prefix nationalities: <http://seco.tkk.fi/onto/toimo/nationalities/> .
```

```
toimo:
  rdf:type owl:Ontology ;
  rdfs:comment "Toimijaontologia" .
```

```
foaf:Agent
  rdf:type owl:Class .
```

```
foaf:Person
  rdf:type owl:Class ;
  rdfs:label "Person"@en , "Henkilö"@fi ;
  rdfs:subClassOf foaf:Agent .
```

```
foaf:Group
  rdf:type owl:Class ;
  rdfs:label "Group"@en , "Ryhmä"@fi ;
  rdfs:subClassOf foaf:Agent .
```

```
foaf:Organization
  rdf:type owl:Class ;
  rdfs:label "Organization"@en , "Organisaatio"@fi ;
  rdfs:subClassOf foaf:Agent .
```

```
foaf:name
  rdf:type owl:DatatypeProperty ;
  rdfs:domain foaf:Agent ;
  rdfs:label "Nimi"@fi , "Name"@en .
```

---

<sup>47</sup><http://www.w3.org/TeamSubmission/turtle/>

```
foaf:firstName
    rdf:type owl:DatatypeProperty ;
    rdfs:domain foaf:Person ;
    rdfs:label "Etunimi"@fi , "First name"@en .

foaf:member
    rdf:type owl:ObjectProperty ;
    rdfs:domain foaf:Group ;
    rdfs:range foaf:Agent .

foaf:gender
    rdf:type owl:DatatypeProperty ;
    rdfs:domain foaf:Person ;
    rdfs:label "Sukupuoli"@fi , "Gender"@en ;
    rdfs:range rdfs:Literal .

foaf:surname
    rdf:type owl:DatatypeProperty ;
    rdfs:domain foaf:Person ;
    rdfs:label "Surname"@en , "Sukunimi"@fi .

foaf:knows
    rdf:type owl:ObjectProperty ;
    rdfs:domain foaf:Person ;
    rdfs:range foaf:Person .

rel:parentOf
    rdf:type owl:ObjectProperty ;
    rdfs:domain foaf:Person ;
    rdfs:range foaf:Person ;
    rdfs:subPropertyOf foaf:knows ;
    owl:inverseOf rel:childOf .

rel:colleagueOf
    rdf:type owl:ObjectProperty ;
    rdfs:domain foaf:Person ;
    rdfs:range foaf:Person ;
    rdfs:subPropertyOf foaf:knows .

rel:spouseOf
    rdf:type owl:ObjectProperty ;
    rdfs:domain foaf:Person ;
    rdfs:range foaf:Person ;
    rdfs:subPropertyOf foaf:knows .
```

```
rel:grandparentOf
    rdf:type owl:ObjectProperty ;
    rdfs:domain foaf:Person ;
    rdfs:range foaf:Person ;
    rdfs:subPropertyOf foaf:knows ;
    owl:inverseOf rel:grandchildOf .

rel:grandchildOf
    rdf:type owl:ObjectProperty ;
    rdfs:domain foaf:Person ;
    rdfs:range foaf:Person ;
    rdfs:subPropertyOf foaf:knows ;
    owl:inverseOf rel:grandparentOf .

rel:friendOf
    rdf:type owl:ObjectProperty ;
    rdfs:domain foaf:Person ;
    rdfs:range foaf:Person ;
    rdfs:subPropertyOf foaf:knows .

rel:siblingOf
    rdf:type owl:ObjectProperty ;
    rdfs:domain foaf:Person ;
    rdfs:range foaf:Person ;
    rdfs:subPropertyOf foaf:knows .

rel:childOf
    rdf:type owl:ObjectProperty ;
    rdfs:domain foaf:Person ;
    rdfs:range foaf:Person ;
    rdfs:subPropertyOf foaf:knows ;
    owl:inverseOf rel:parentOf .

toimo:studentOf
    rdf:type owl:ObjectProperty ;
    rdfs:domain foaf:Person ;
    rdfs:label "Opettaja"@fi , "Student of"@en ;
    rdfs:range foaf:Person ;
    rdfs:subPropertyOf foaf:knows ;
    owl:inverseOf toimo:teacherOf .

toimo:teacherOf
    rdf:type owl:ObjectProperty ;
```

```
rdfs:domain foaf:Person ;
rdfs:label "Oppilas"@fi , "Teacher of"@en ;
rdfs:range foaf:Person ;
rdfs:subPropertyOf foaf:knows ;
owl:inverseOf toimo:studentOf .
```

toimo:masterOf

```
rdf:type owl:ObjectProperty ;
rdfs:domain foaf:Person ;
rdfs:range foaf:Person ;
rdfs:subPropertyOf foaf:knows ;
owl:inverseOf toimo:apprenticeOf .
```

toimo:apprenticeOf

```
rdf:type owl:ObjectProperty ;
rdfs:domain foaf:Person ;
rdfs:range foaf:Person ;
rdfs:subPropertyOf foaf:knows ;
owl:inverseOf toimo:masterOf .
```

toimo:assistedBy

```
rdf:type owl:ObjectProperty ;
rdfs:domain foaf:Person ;
rdfs:range foaf:Person ;
rdfs:subPropertyOf foaf:knows ;
owl:inverseOf toimo:assistantOf .
```

toimo:assistantOf

```
rdf:type owl:ObjectProperty ;
rdfs:domain foaf:Person ;
rdfs:range foaf:Person ;
rdfs:subPropertyOf foaf:knows ;
owl:inverseOf toimo:assistedBy .
```

toimo:workedWith

```
rdf:type owl:ObjectProperty ;
rdfs:domain foaf:Person ;
rdfs:range foaf:Person ;
rdfs:subPropertyOf foaf:knows .
```

toimo:collaboratedWith

```
rdf:type owl:ObjectProperty ;
rdfs:domain foaf:Person ;
rdfs:range foaf:Person ;
```

```
rdfs:subPropertyOf foaf:knows .
```

```
toimo:employeeWas
```

```
  rdf:type owl:ObjectProperty ;  
  rdfs:domain foaf:Person ;  
  rdfs:range foaf:Person ;  
  rdfs:subPropertyOf foaf:knows ;  
  owl:inverseOf toimo:employeeOf .
```

```
toimo:employeeOf
```

```
  rdf:type owl:ObjectProperty ;  
  rdfs:domain foaf:Person ;  
  rdfs:range foaf:Person ;  
  rdfs:subPropertyOf foaf:knows ;  
  owl:inverseOf toimo:employeeWas .
```

```
toimo:uncleOf
```

```
  rdf:type owl:ObjectProperty ;  
  rdfs:domain foaf:Person ;  
  rdfs:range foaf:Person ;  
  rdfs:subPropertyOf foaf:knows .
```

```
toimo:nephewOf
```

```
  rdf:type owl:ObjectProperty ;  
  rdfs:domain foaf:Person ;  
  rdfs:range foaf:Person ;  
  rdfs:subPropertyOf foaf:knows .
```

```
toimo:nieceOf
```

```
  rdf:type owl:ObjectProperty ;  
  rdfs:domain foaf:Person ;  
  rdfs:range foaf:Person ;  
  rdfs:subPropertyOf foaf:knows .
```

```
toimo:halfsiblingOf
```

```
  rdf:type owl:ObjectProperty ;  
  rdfs:domain foaf:Person ;  
  rdfs:range foaf:Person ;  
  rdfs:subPropertyOf foaf:knows .
```

```
toimo:cousinOf
```

```
  rdf:type owl:ObjectProperty ;  
  rdfs:domain foaf:Person ;  
  rdfs:range foaf:Person ;
```



```
rdfs:subPropertyOf foaf:knows .
```

```
toimo:siblingInLaw
```

```
  rdf:type owl:ObjectProperty ;
  rdfs:domain foaf:Person ;
  rdfs:range foaf:Person ;
  rdfs:subPropertyOf foaf:knows .
```

```
toimo:role
```

```
  rdf:type owl:ObjectProperty ;
  rdfs:domain foaf:Agent ;
  rdfs:range roles:Role ;
  rdfs:subPropertyOf rdf:type .
```

```
toimo:nationality
```

```
  rdf:type owl:ObjectProperty ;
  rdfs:domain foaf:Agent ;
  rdfs:range nationalities:Nationality ;
  rdfs:subPropertyOf rdf:type .
```

```
bio:date
```

```
  rdf:type owl:DatatypeProperty ;
  rdfs:comment "The date of an event."@en ;
  rdfs:domain bio:Event ;
  rdfs:label "Date"@en ;
  rdfs:subPropertyOf <http://purl.org/dc/elements/1.1/date> .
```

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

```
  rdf:type owl:DatatypeProperty .
```

```
bio:olb
```

```
  rdf:type owl:DatatypeProperty ;
  rdfs:comment "A one-line biography of the person."@en ;
  rdfs:domain foaf:Person ;
  rdfs:label "One-line bio"@en , "Lyhyt biografia"@en ;
  rdfs:range rdfs:Literal .
```

```
bio:place
```

```
  rdf:type owl:DatatypeProperty ;
  rdfs:comment "The place of an event."@en ;
  rdfs:domain bio:Event ;
  rdfs:label "Place"@en .
```

```
bio:event
```

```
    rdf:type owl:ObjectProperty ;  
    rdfs:domain foaf:Person ;  
    rdfs:range bio:Event .
```

```
bio:Event  
    rdf:type owl:Class .
```

```
bio:Birth  
    rdf:type owl:Class ;  
    rdfs:subClassOf bio:Event .
```

```
bio:Death  
    rdf:type owl:Class ;  
    rdfs:subClassOf bio:Event .
```

```
toimo:WinnerOfContest  
    rdf:type owl:Class ;  
    rdfs:subClassOf bio:Event .
```

```
toimo:Babtism  
    rdf:type owl:Class ;  
    rdfs:subClassOf bio:Event .
```

```
toimo:Exhibited  
    rdf:type owl:Class ;  
    rdfs:subClassOf bio:Event .
```

```
toimo:Active  
    rdf:type owl:Class ;  
    rdfs:subClassOf bio:Event .
```

## Liite 2. Rooli- ja kansallisuusontologiat

Ontologiat on esitetty Turtle-muodossa<sup>48</sup>.

```
@prefix rdf:      <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs:    <http://www.w3.org/2000/01/rdf-schema#> .
@prefix owl:   <http://www.w3.org/2002/07/owl#> .
@prefix foaf:    <http://xmlns.com/foaf/0.1/> .
@prefix toimo:   <http://seco.tkk.fi/onto/toimo/> .
@prefix roles:   <http://seco.tkk.fi/onto/toimo/roles/> .
@prefix nationalities: <http://seco.tkk.fi/onto/toimo/nationalities/> .
```

```
roles:Role
  rdf:type owl:Class ;
  rdfs:label "Role"@en , "Rooli"@fi .
```

```
roles:Architect
  rdf:type owl:Class ;
  rdfs:label "arkkitehti"@fi , "architect"@en ;
  rdfs:subClassOf roles:Role ;
  owl:equivalentClass <http://www.yso.fi/onto/yso/p4299> .
```

```
roles:Designer
  rdf:type owl:Class ;
  rdfs:label "designer"@en , "suunnittelija"@fi ;
  rdfs:subClassOf roles:Role ;
  owl:equivalentClass <http://www.yso.fi/onto/yso/p12245> .
```

```
roles:Chemist
  rdf:type owl:Class ;
  rdfs:label "kemisti"@fi , "chemist"@en ;
  rdfs:subClassOf roles:Role .
```

```
roles:Modeler
  rdf:type owl:Class ;
  rdfs:label "modeler"@en , "muotoilija"@fi ;
  rdfs:subClassOf roles:Role ;
  owl:equivalentClass <http://www.yso.fi/onto/yso/p12245> .
```

```
roles:Professor
  rdf:type owl:Class ;
  rdfs:label "professori"@fi , "professor"@en ;
```

---

<sup>48</sup><http://www.w3.org/TeamSubmission/turtle/>

```
rdfs:subClassOf roles:Role ;
owl:equivalentClass <http://www.yso.fi/onto/yso/p17044> .

roles:Woodcutter_s_
  rdf:type owl:Class ;
  rdfs:label "woodcutter(s)"@en , "puunleikkaaja"@fi ;
  rdfs:subClassOf roles:Role .

roles:Caricaturist
  rdf:type owl:Class ;
  rdfs:label "caricaturist"@en , "pilapiirtäjä"@fi ;
  rdfs:subClassOf roles:Role ;
  owl:equivalentClass <http://www.yso.fi/onto/yso/p11154> .

roles:GraphicDesigner
  rdf:type owl:Class ;
  rdfs:label "graphic designer"@en , "graafinen suunnittelija"@fi ;
  rdfs:subClassOf roles:Role ;
  owl:equivalentClass <http://www.yso.fi/onto/yso/p7529> .

roles:Collector
  rdf:type owl:Class ;
  rdfs:label "collector"@en , "ker?ilij?"@fi ;
  rdfs:subClassOf roles:Role ;
  owl:equivalentClass <http://www.yso.fi/onto/yso/p2286> .

roles:Poet
  rdf:type owl:Class ;
  rdfs:label "poet"@en , "runoilija"@fi ;
  rdfs:subClassOf roles:Role ;
  owl:equivalentClass <http://www.yso.fi/onto/yso/p3490> .

roles:Sitter
  rdf:type owl:Class ;
  rdfs:label "sitter"@en , "taiteilijan malli"@fi ;
  rdfs:subClassOf roles:Role ;
  owl:equivalentClass <http://www.yso.fi/onto/yso/p9286> .

roles:Writer
  rdf:type owl:Class ;
  rdfs:label "kirjailija"@fi , "writer"@en ;
  rdfs:subClassOf roles:Role ;
  owl:equivalentClass <http://www.yso.fi/onto/yso/p8970> .
```

```
roles:Sculptor
  rdf:type owl:Class ;
  rdfs:label "kuvanveistäjä"@fi , "sculptor"@en ;
  rdfs:subClassOf roles:Role ;
  owl:equivalentClass <http://www.yso.fi/onto/yso/p13252> .

roles:Dealer
  rdf:type owl:Class ;
  rdfs:label "j?lleenmyyj?"@fi , "dealer"@en ;
  rdfs:subClassOf roles:Role ;
  owl:equivalentClass <http://www.yso.fi/onto/yso/p996> .

roles:Printmaker
  rdf:type owl:Class ;
  rdfs:label "printmaker"@en , "taidegraafikko"@fi ;
  rdfs:subClassOf roles:Role ;
  owl:equivalentClass <http://www.yso.fi/onto/yso/p7529> .

roles:Stonemason
  rdf:type owl:Class ;
  rdfs:label "kivenhakkaaja"@fi , "stonemason"@en ;
  rdfs:subClassOf roles:Role ;
  owl:equivalentClass <http://www.yso.fi/onto/yso/p10419> .

roles:Artist
  rdf:type owl:Class ;
  rdfs:label "taiteilija"@fi , "artist"@en ;
  rdfs:subClassOf roles:Role ;
  owl:equivalentClass <http://www.yso.fi/onto/yso/p7017> .

roles:Illustrator
  rdf:type owl:Class ;
  rdfs:label "illustrator"@en , "kuvittaja"@fi ;
  rdfs:subClassOf roles:Role ;
  owl:equivalentClass <http://www.yso.fi/onto/yso/p20004> .

roles:Draftsman
  rdf:type owl:Class ;
  rdfs:label "draftsman"@en , "piirt?j?"@fi ;
  rdfs:subClassOf roles:Role ;
  owl:equivalentClass <http://www.yso.fi/onto/yso/p11154> .

roles:FurnitureDesigner
  rdf:type owl:Class ;
```

```
rdfs:label "huonekalusuunnittelija"@fi , "furniture designer"@en ;  
rdfs:subClassOf roles:Role .
```

```
roles:InteriorDecorator  
  rdf:type owl:Class ;  
  rdfs:label "interior decorator"@en , "sisustusarkkitehti"@fi ;  
  rdfs:subClassOf roles:Role ;  
  owl:equivalentClass <http://www.yso.fi/onto/yso/p7636> .
```

```
roles:Painter  
  rdf:type owl:Class ;  
  rdfs:label "taidemaalari"@fi , "painter"@en ;  
  rdfs:subClassOf roles:Role ;  
  owl:equivalentClass <http://www.yso.fi/onto/yso/p16021> .
```

```
roles:Photographer  
  rdf:type owl:Class ;  
  rdfs:label "photographer"@en , "valokuvaaja"@fi ;  
  rdfs:subClassOf roles:Role ;  
  owl:equivalentClass <http://www.yso.fi/onto/yso/p12307> .
```

```
nationalities:Scottish_Scots_  
  rdf:type owl:Class ;  
  rdfs:label "skotlantilainen"@fi , "Scottish (Scots)"@en ;  
  rdfs:subClassOf nationalities:Nationality .
```

```
nationalities:Nationality  
  rdf:type owl:Class ;  
  rdfs:label "Nationality"@en , "Kansallisuus"@fi .
```

```
nationalities:Swedish  
  rdf:type owl:Class ;  
  rdfs:label "Swedish"@en , "ruotsalainen"@fi ;  
  rdfs:subClassOf nationalities:Nationality .
```

```
nationalities:British  
  rdf:type owl:Class ;  
  rdfs:label "brittiläinen"@fi , "British"@en ;  
  rdfs:subClassOf nationalities:Nationality .
```

```
nationalities:Danish  
  rdf:type owl:Class ;  
  rdfs:label "Danish"@en , "tanskalainen"@fi ;  
  rdfs:subClassOf nationalities:Nationality .
```

```
nationalities:Austrian
    rdf:type owl:Class ;
    rdfs:label "itävaltalainen"@fi , "Austrian"@en ;
    rdfs:subClassOf nationalities:Nationality .

nationalities:Finnish
    rdf:type owl:Class ;
    rdfs:label "Finnish"@en , "suomalainen"@fi ;
    rdfs:subClassOf nationalities:Nationality .

nationalities:Canadian
    rdf:type owl:Class ;
    rdfs:label "Canadian"@en , "kanadalainen"@fi ;
    rdfs:subClassOf nationalities:Nationality .

nationalities:Japanese
    rdf:type owl:Class ;
    rdfs:label "japanilainen"@fi , "Japanese"@en ;
    rdfs:subClassOf nationalities:Nationality .

nationalities:German
    rdf:type owl:Class ;
    rdfs:label "saksalainen"@fi , "German"@en ;
    rdfs:subClassOf nationalities:Nationality .

nationalities:Ukrainian
    rdf:type owl:Class ;
    rdfs:label "Ukrainian"@en , "ukrainalainen"@fi ;
    rdfs:subClassOf nationalities:Nationality .

nationalities:Norwegian
    rdf:type owl:Class ;
    rdfs:label "Norwegian"@en , "norjalainen"@fi ;
    rdfs:subClassOf nationalities:Nationality .

nationalities:American
    rdf:type owl:Class ;
    rdfs:label "American"@en , "amerikkalainen"@fi ;
    rdfs:subClassOf nationalities:Nationality .

nationalities:Swiss
    rdf:type owl:Class ;
    rdfs:label "sveitsiläinen"@fi , "Swiss"@en ;
```

```
    rdfs:subClassOf nationalities:Nationality .

nationalities:French
    rdf:type owl:Class ;
    rdfs:label "ranskalainen"@fi , "French"@en ;
    rdfs:subClassOf nationalities:Nationality .

nationalities:Polish
    rdf:type owl:Class ;
    rdfs:label "Polish"@en , "puolalainen"@fi ;
    rdfs:subClassOf nationalities:Nationality .

nationalities:Belgian
    rdf:type owl:Class ;
    rdfs:label "Belgian"@en , "belgialainen"@fi ;
    rdfs:subClassOf nationalities:Nationality .

nationalities:English
    rdf:type owl:Class ;
    rdfs:label "English"@en , "englantilainen"@fi ;
    rdfs:subClassOf nationalities:Nationality .

nationalities:Estonian
    rdf:type owl:Class ;
    rdfs:label "Estonian"@en , "virolainen"@fi ;
    rdfs:subClassOf nationalities:Nationality .

nationalities:Russian
    rdf:type owl:Class ;
    rdfs:label "Russian"@en , "ven?l?inen"@fi ;
    rdfs:subClassOf nationalities:Nationality .

nationalities:Italian
    rdf:type owl:Class ;
    rdfs:label "Italian"@en , "italialainen"@fi ;
    rdfs:subClassOf nationalities:Nationality .

nationalities:Spanish
    rdf:type owl:Class ;
    rdfs:label "Spanish"@en , "espanjalainen"@fi ;
    rdfs:subClassOf nationalities:Nationality .

nationalities:Dutch
    rdf:type owl:Class ;
```



```
rdfs:label "hollantilainen"@fi , "Dutch"@en ;  
rdfs:subClassOf nationalities:Nationality .
```

```
nationalities:Hungarian  
  rdf:type owl:Class ;  
  rdfs:label "unkarilainen"@fi , "Hungarian"@en ;  
  rdfs:subClassOf nationalities:Nationality .
```

```
nationalities:Netherlandish  
  rdf:type owl:Class ;  
  rdfs:label "alankomaalainen"@fi , "Netherlandish"@en ;  
  rdfs:subClassOf nationalities:Nationality .
```

```
nationalities:Lithuanian  
  rdf:type owl:Class ;  
  rdfs:label "Lithuanian"@en , "liettualainen"@fi ;  
  rdfs:subClassOf nationalities:Nationality .
```

```
nationalities:Irish  
  rdf:type owl:Class ;  
  rdfs:label "Irish"@en , "irlantilainen"@fi ;  
  rdfs:subClassOf nationalities:Nationality .
```