



Semanttisen webin hyödyntäminen terveystiedon löydettävyydessä

eHealth 2005 -seminaari 8.-9.11.2005

Petri Lindgren (petri.lindgren@helsinki.fi)

TKK Viestintätekniikka

Semantic Computing Research Group (SeCo)

<http://www.seco.tkk.fi>



Semanttisen laskennan tutkimusryhmä (Semantic Computing Research Group) (SeCo)

- tutkii menetelmiä, joilla tietoa voidaan kuvailla niin, että sen **merkitys** on algoritmisesti prosessoitavissa.
- kehittää sovelluksia, jotka edellä mainittuja kuvailuja hyödyntäen pyrkivät automatisoimaan entistä vaativampia prosesseja.
- SeCo-tutkimusryhmän tutkimusjohtajana toimii professori Eero Hyvönen.
- www.seco.tkk.fi

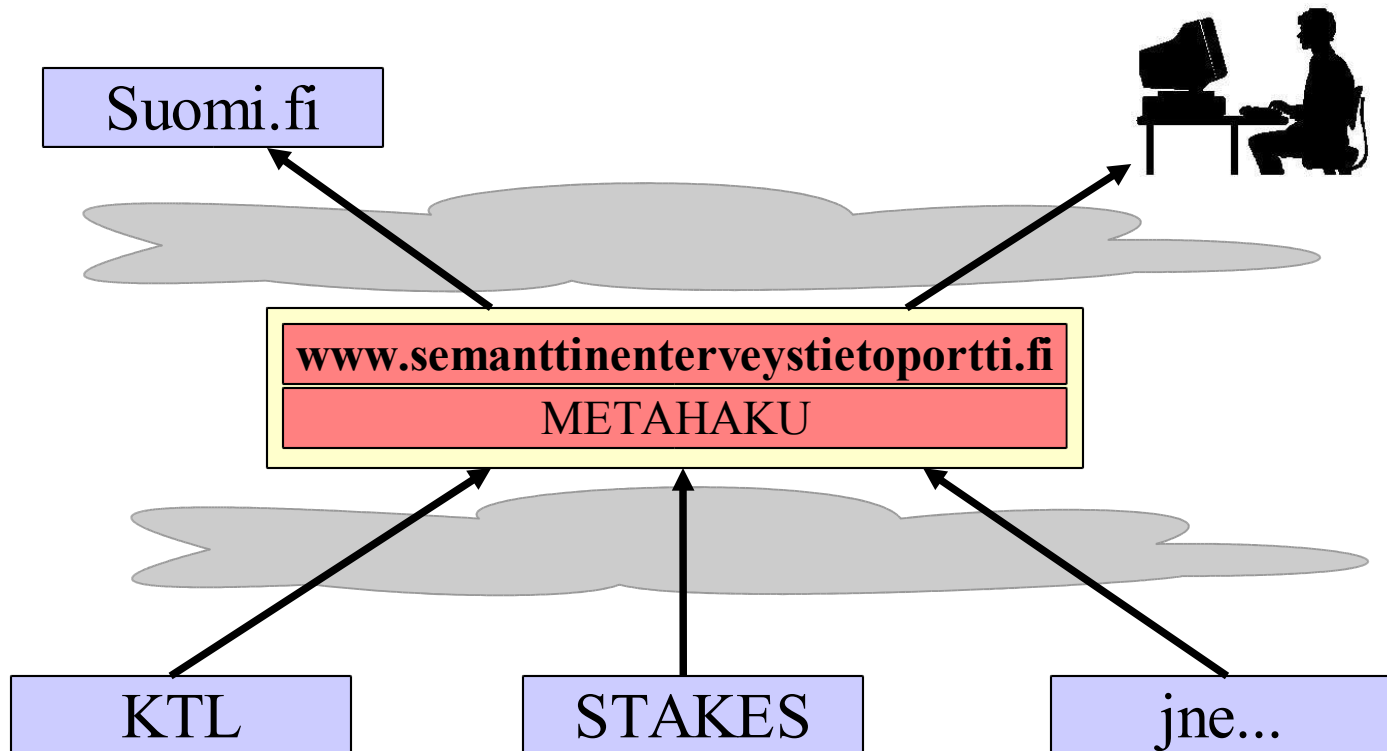
Suomalaiset semanttisen webin ontologiat (FinnONTO) –projekti (2003-2007)

- laaja kansallinen hanke, jossa tutkitaan ja kehitetään suomalaisen semanttisen webin infrastruktuuria ja
- ontologiaperustaista tiedon esittämistä ja hakua.
- **Projektin tavoitteina ovat mm. seuraavat tulokset:**
 - 1. Keskeisten kansallisten ontologioiden kehittäminen
 - **Yleinen suomalainen ontologia YSO:** Yleisestä suomalaisesta asiasanastosta YSA (n. 23.000 termiä) kehitetään Yleinen suomalainen ontologia YSO.
 - **Suomalaiset paikkaontologiat**
 - jne.
 - 2. Ontologioiden kehittämisen ja käytön palvelut
 - Tärkeimpänä ontologiapalveluiden kehityskohteena on **kansallinen ontologiakirjasto ja -palvelin ONKI**
 - Semanttinen terveystietoportti (STP)
 - jne.
- <http://www.seco.tkk.fi/projects/finnonto>

Sosiaali- ja terveysalan osaprojekti (Soster)

■ Tavoitteita:

- Prototyyppi semanttisesta terveystietoportista (STP), jonka avulla kansalaisen on helppo löytää terveyden edistämiseen liittyvää tietoa ja palveluita.
- Tavoitteena tiedon- ja palveluntuottajien tietosisältöjen semanttinen yhdistäminen STP-järjestelmään.

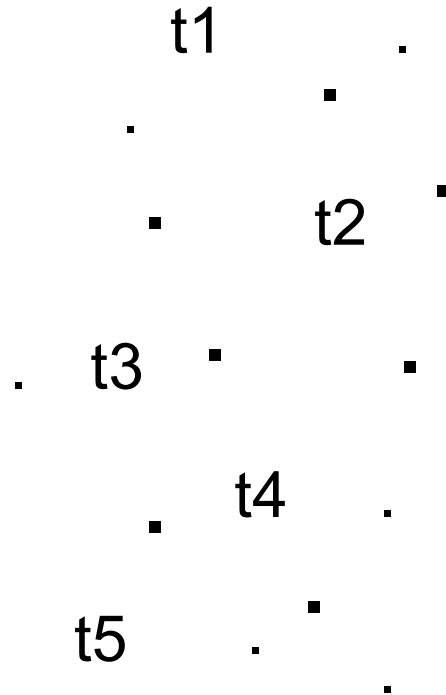


Intelligent Web Services (IWebS)

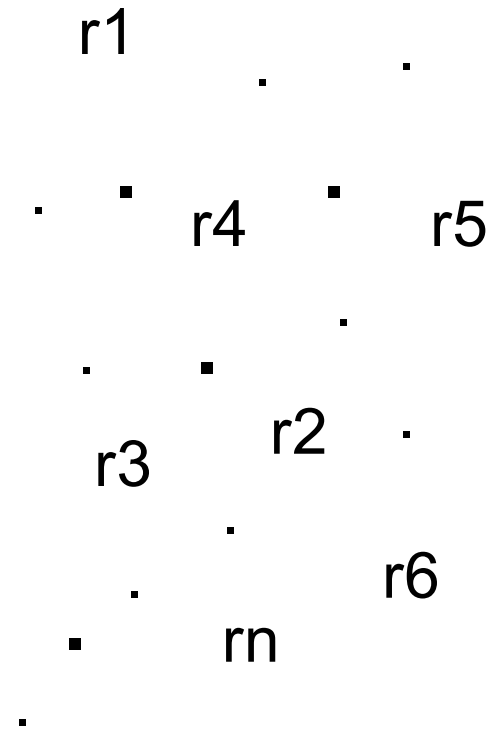
- **Stakes-osaprojekti:** Sosiaali- ja terveystalvet Semantic webissä
- käytännön ongelmana: **vanhusten palveluiden hakeminen.**

Tiedon- ja palvelunhaun ongelma

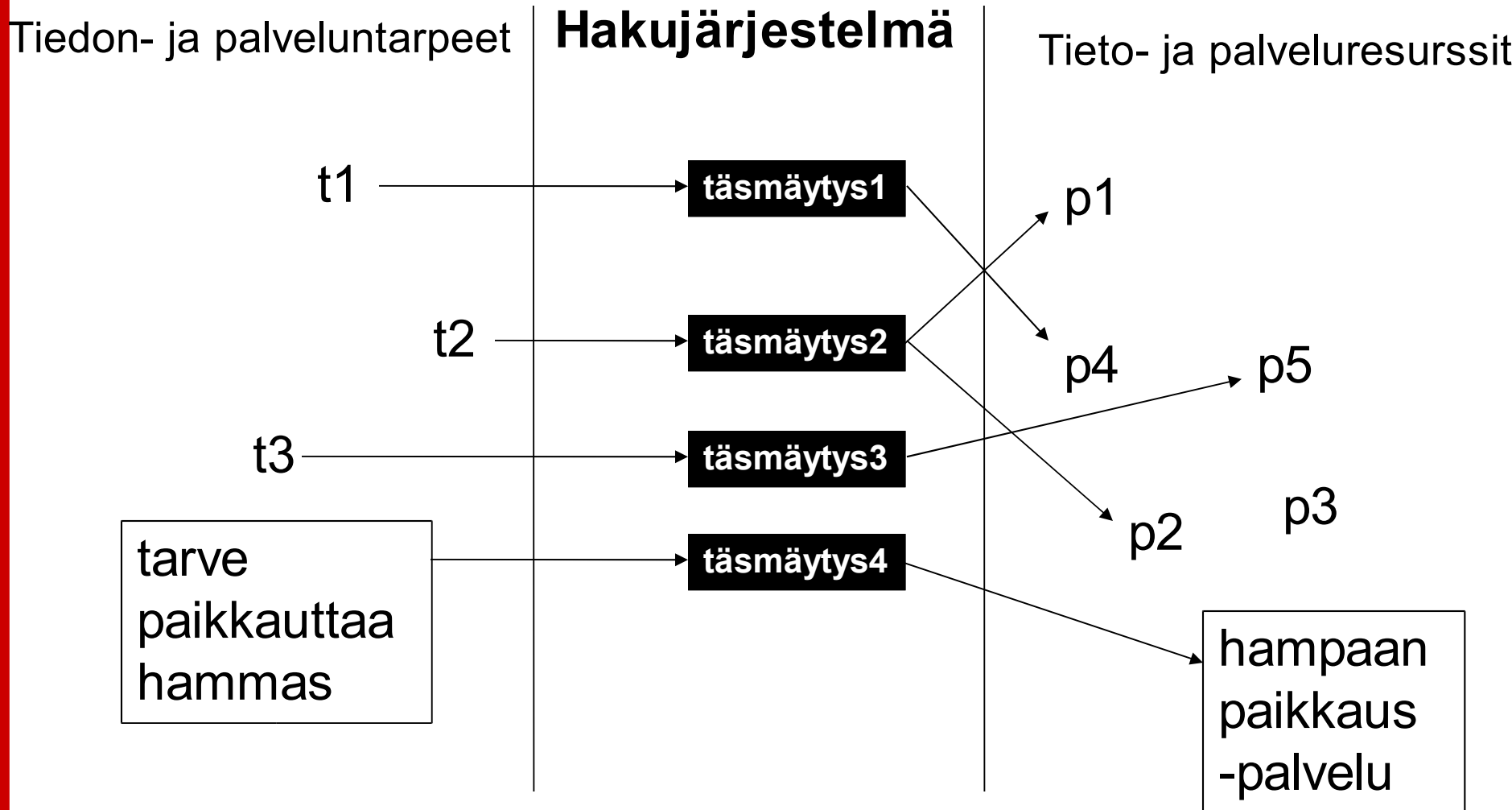
Tiedon- ja palveluntarpeet



Tieto- ja palveluresurssit



Tiedon- ja palvelunhaun ongelma



Tiedon- ja palvelunhaun

osaongelmat

Tarpeet

Resurssit

Kuvailu

Täsmäytys

Täsmäytys

Hakutulos

Järjestys

- Täsmäytyksen mahdollistamiseksi **tarpeet** ja ne mahdollisesti tyydyttävät **resurssit** on kuvailtava.
- Tässä esitelmässä keskitymme resurssien kuvailuun ja erityisesti palveluresurssien kuvailuun.

Nykyisten palvelunhakujärjestelmien ongelmia

- Järjestelmässä olevien palveluiden kuvailutavat
 - **vapaa tekstikuvaus**
 - **kategoriatieto**
- Pääasiallinen hakutapa
 - **vapaa termihaku**
- Täsmäysmenetelmät toimivat vain kuvausten syntaksin tasolla.
- Hakutulosten **saanti** ja erityisesti **tarkkuus** ovat relevanttien hakujen määrään nähden huonoja.
 - Saanti: hakutuloksen relevanttien resurssien suhde kaikkiin järjestelmän sisältämiin relevantteihin resursseihin.
 - Tarkkuus: hakutuloksen relevanttien resurssien suhde kaikkiin löydettyihin resursseihin.

Termihaun ongelmia: synonymymongelma

- Vapaa termihaku joukkojen {"tietokone", "tietokoneen"} ja {"korjaus", "huolto"} tulojoukon järjestetyillä pareilla.
- Fonectan 020202.fi:n hakutulos
 - relevantteja palveluita: **ei**
 - epärelevantteja palveluita: **ei**
- Relevantteja palveluita järjestelmässä: **kyllä**
 - (relevantteja palveluita löytyy esim. hakutermillä "atk huolto")

← synonymymongelma

Termihaun ongelmia: ontologisen tiedon puute

- Eniron yritykset.eniro.fi:n hakutulos
 - relevantteja palveluita: **kyllä**
 - epärelevantteja palveluita: **kyllä**
 - Epärelevantissa palveluilmoituksessa ei kuvailla tietokoneen korjauspalvelua, mutta siinä esiintyvät termit "tietokone" ja "korjaus".
- ← Täsmäytysalgoritmi ei osaa päätellä tai sillä ei ole tietoa termien välisistä suhteista.

Termihaun ongelmia

- Jos palveluun viittaavia termejä on useita ja/tai ne esiintyvät yleisesti muidenkin palveluiden yhteydessä
- Edellä kuvattuja ongelmia

Kategoriahaku

- Kategoriahauksen ideana on tarjota käyttäjälle käsitteitä eli termejä merkityksineen, joilla hän voi suorittaa hakuja.
- Kategoriahauksen toteuttamiseksi palvelut on ryhmiteltävä. Lisäksi on määriteltävä käsitteet, joilla palveluryhmiin viitataan.
- **Kategoriahaku** välttää termihaun ongelmat, mutta synnyttää uusia.

Kategoriahaun ongelmia ja ratkaisuja

- Ongelmana **1** on löytää helposti käsite, joka on tulkittavissa siten, että se kuvailee tarpeen tyydyttävää palvelua.
 - pahimmassa tapauksessa käyttäjä joutuu tulkitsemaan kaikki käsitteet.
- Ongelmana **2** on käsitettä tulkitsemalla saada selkeä ja tarkoitettu kuva niistä palveluista, joihin käsite viittaa.
- Ratkaisu ongelmaan **1**
 - Muokataan kategoriat hierarkiaksi. Helposti tulkittava hierarkia mahdollistaa tehokkaan karsinnan.
- Ratkaisu ongelmaan **2**
 - Kutakin käsitettä vastaavan termin on oltava yksikäsitteinen.
 - Käsitteiden väliset suhteet määrittelevät käsitteen merkitystä
 - esim. käsitteiden hierarkinen ryhmittely

Hierarkisen ryhmittelyn ongelmia

- Nykyisten palveluiden hakujärjestelmien käyttämät ryhmittelyt eivät perustu systemaattisella tavalla ominaisuuksiin ja/tai niiden arvoihin.
 - Ne perustuvat vakiintuneisiin käsitteisiin, joita ihmiset käyttävät eri palveluryhmistä.
- Palveluryhmät eivät varmuudella ole
 - toisiaan poissulkevia,
 - **kaikki sovellusalueen mahdolliset palvelut kattavia.**
- Siirtymiin tasolta toiselle liittyy monesti erityyppisiä merkityksiä ja niitä ei näytetä käyttäjälle.
 - Selkeän logiikan puuttuminen.
- Algoritminen päättely kategorioiden pohjalta erittäin virhealtista tai mahdotonta.

Ratkaisu edellä kuvattuihin ongelmiin

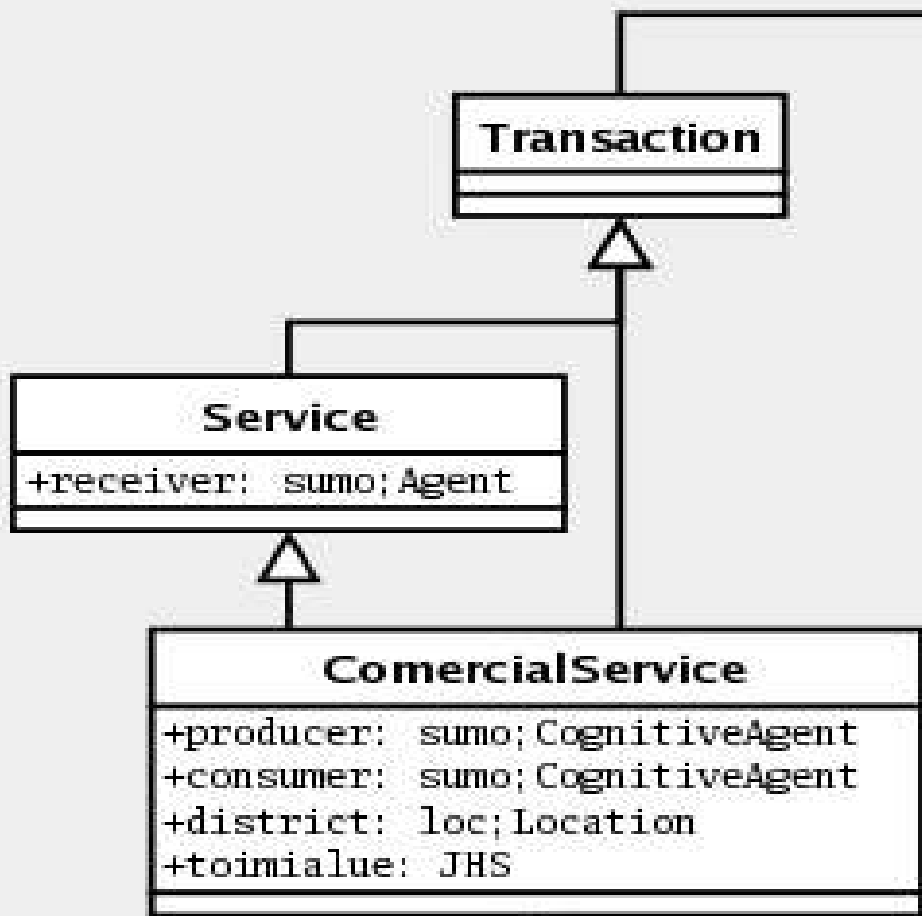
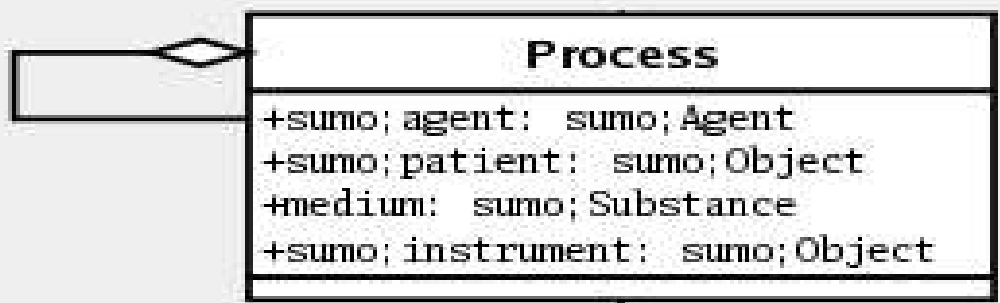
- Ryhmitellään palvelut ja niihin oleellisesti liittyvät entiteetit (niiden ominaisuuksien ja/tai ominaisuuksien arvojen perusteella) systemaattisesti ja täsmällisesti ja kuvaillaan näin muodostetut ryhmät ja niiden ominaisuudet eksplisiittisesti (algoritmisesti prosessoitavalla tavalla).
- Toisin sanoen määritellään formaali **ontologia**, joka kuvailee palvelut ja siihen liittyvät entiteetit.
- **Semanttisen webin** teknologiat tarjoavat valmiita menetelmiä ja välineitä tämän toteuttamiseksi.
- Tärkeimpänä yksittäisenä välineenä se tarjoaa kielen, jolla voi määritellä tietämystä: **ontologioita** ja metatietoa resursseista.

Mitä palvelut ovat?

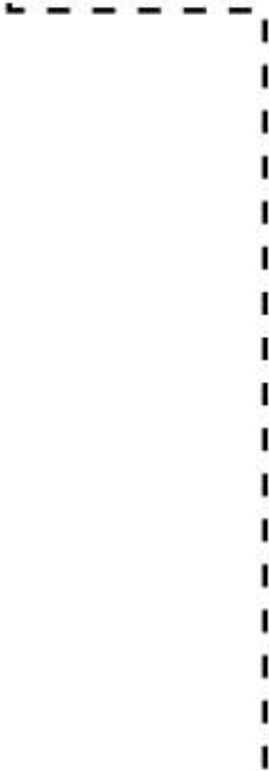
- ”Palvelua ei voi omistaa, ainoastaan palveluoikeuden voi omistaa.”
- ”Palvelu kulutetaan sen hankintahetkellä.”
- ”Palvelun kulutukseen tarvitaan palvelun tarjoajan ja palvelun kuluttajan välistä vuorovaikutusta.”
- ”Organisoidun toiminnan tuloksena syntyvä aineeton hyödyke tarpeiden tyydyttämiseksi.”
- ➔ **Palvelut ovat prosesseja, joissa joku agentti tekee jotakin jonkun agentin tarpeen tyydyttämiseksi.**

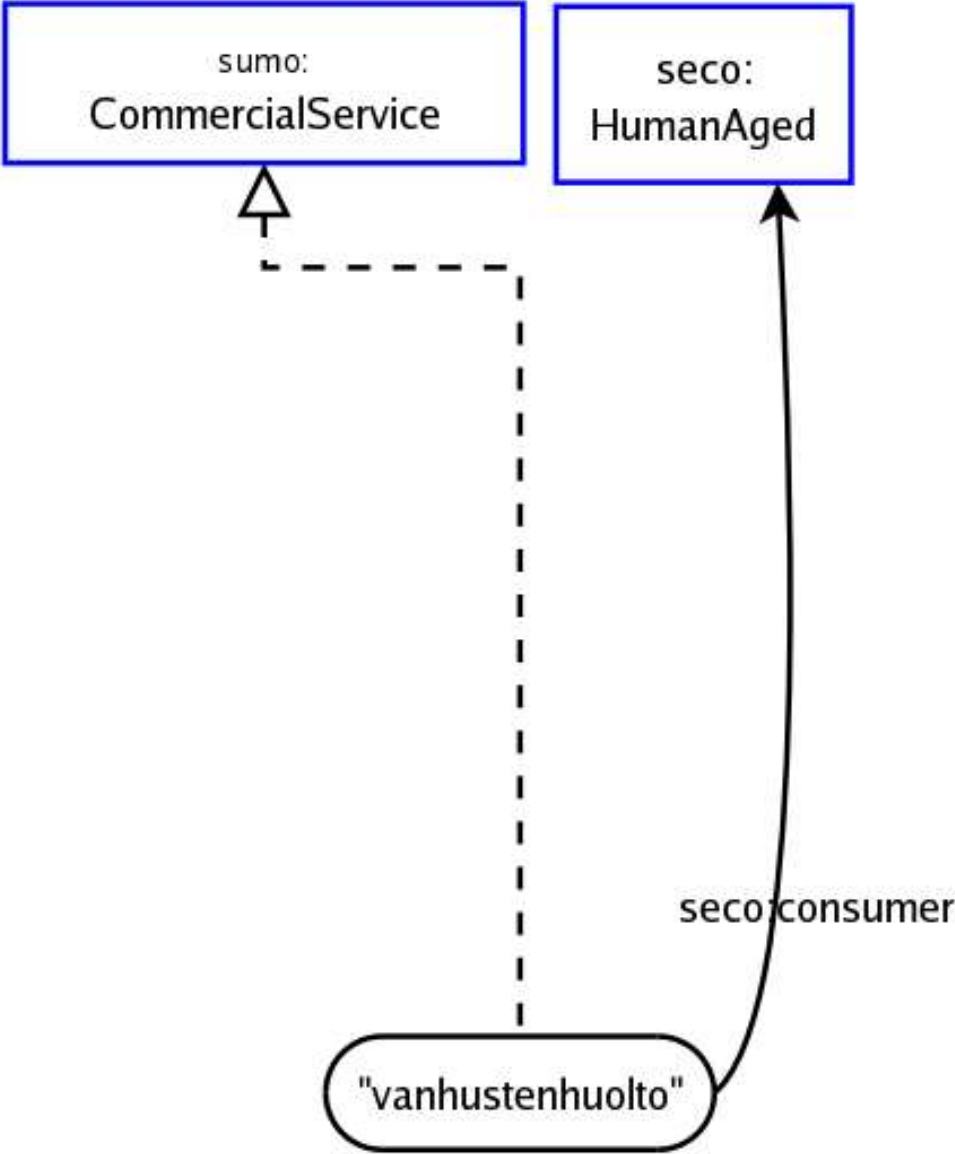
producer-subPropertyOf->agent
consumer-subPropertyOf->receiver
receiver-subPropertyOf->patient

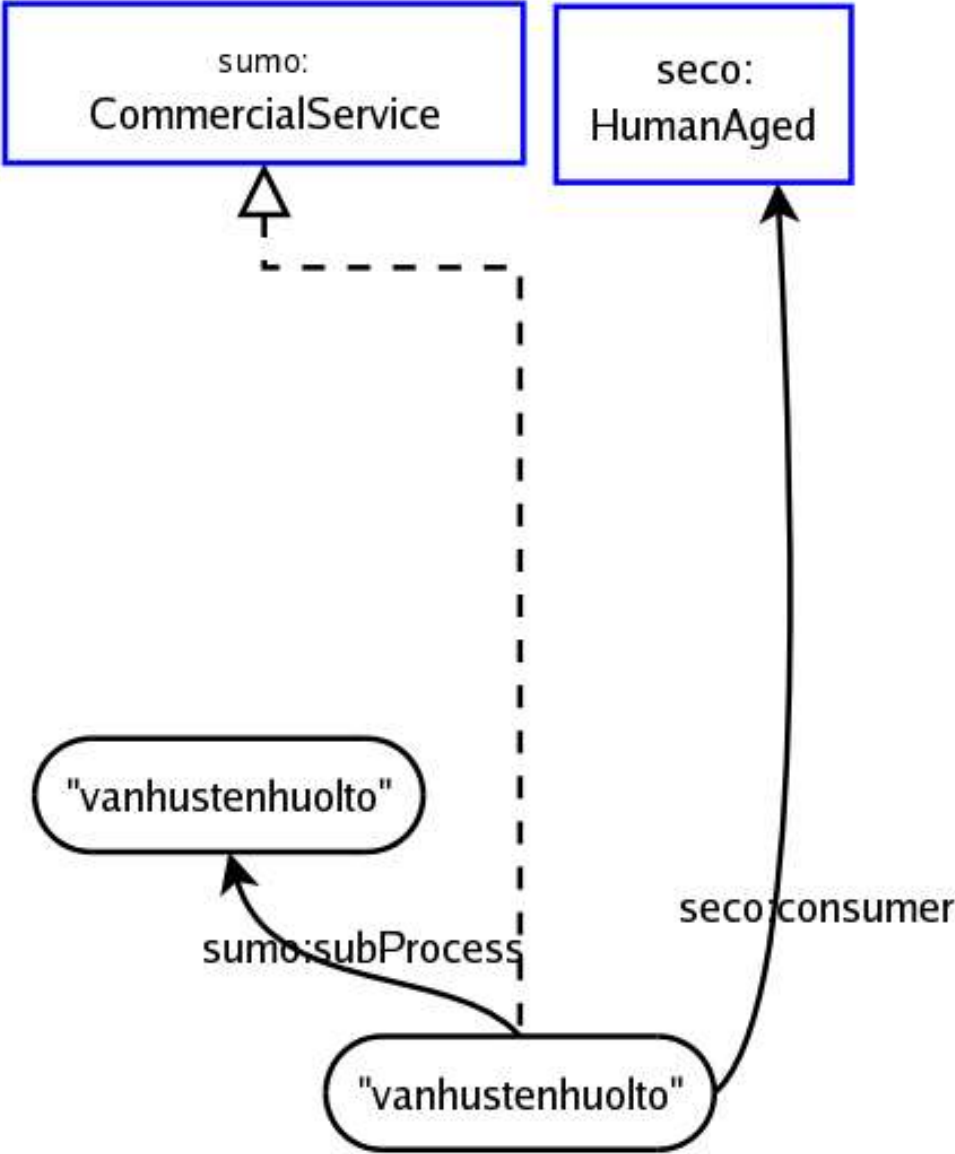
destination-subPropertyOf->goal

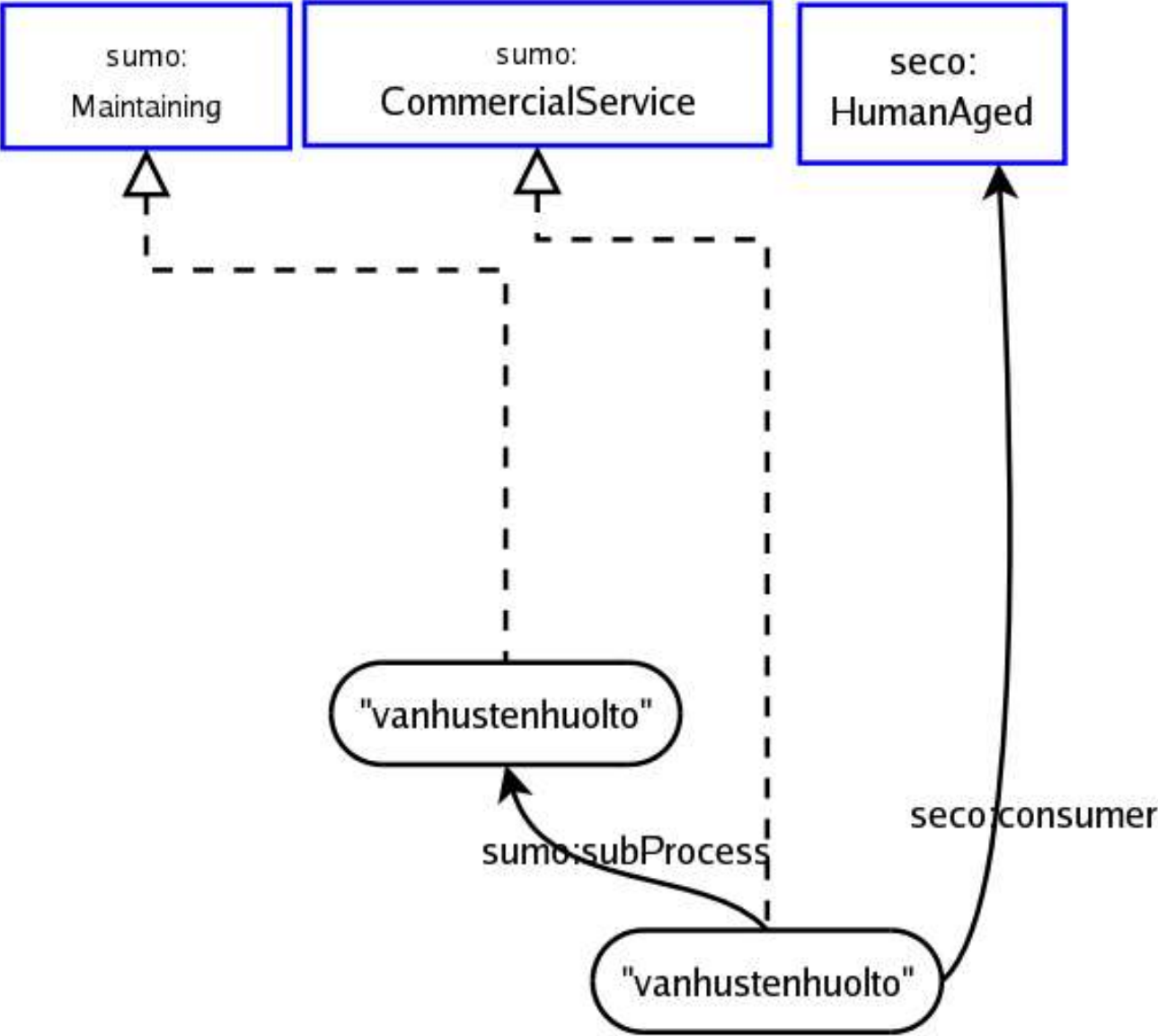


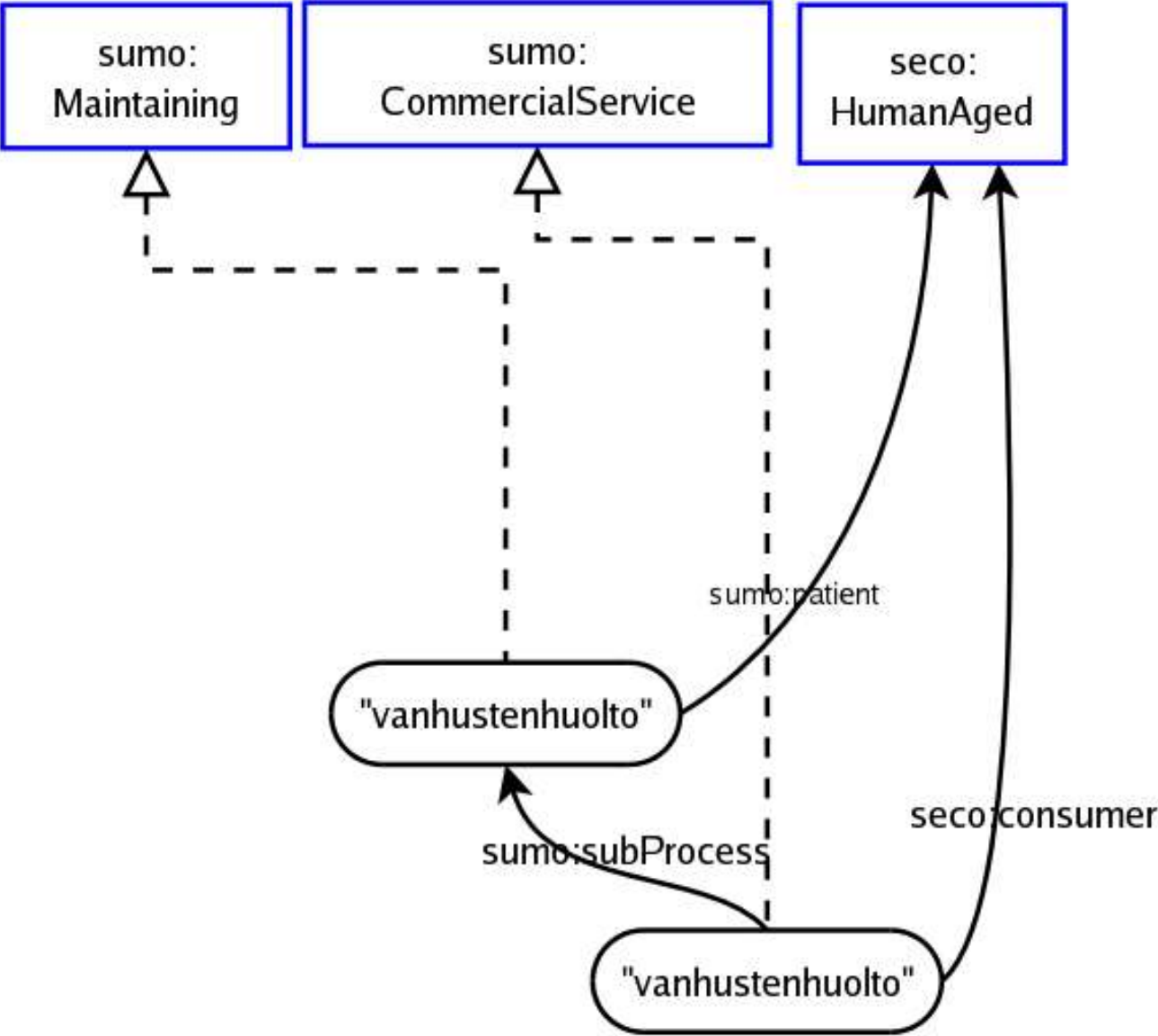
"vanhustenhuolto"

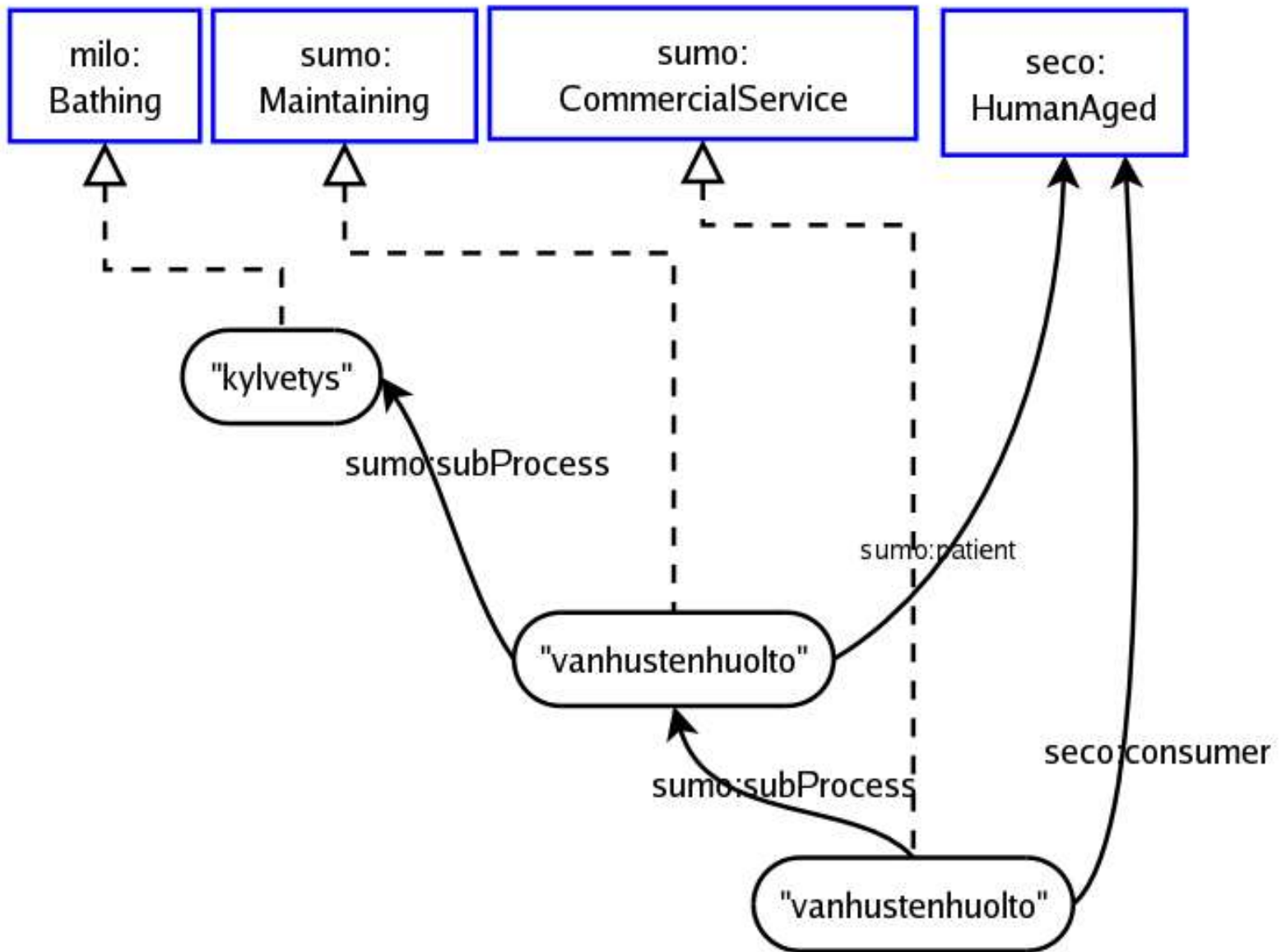


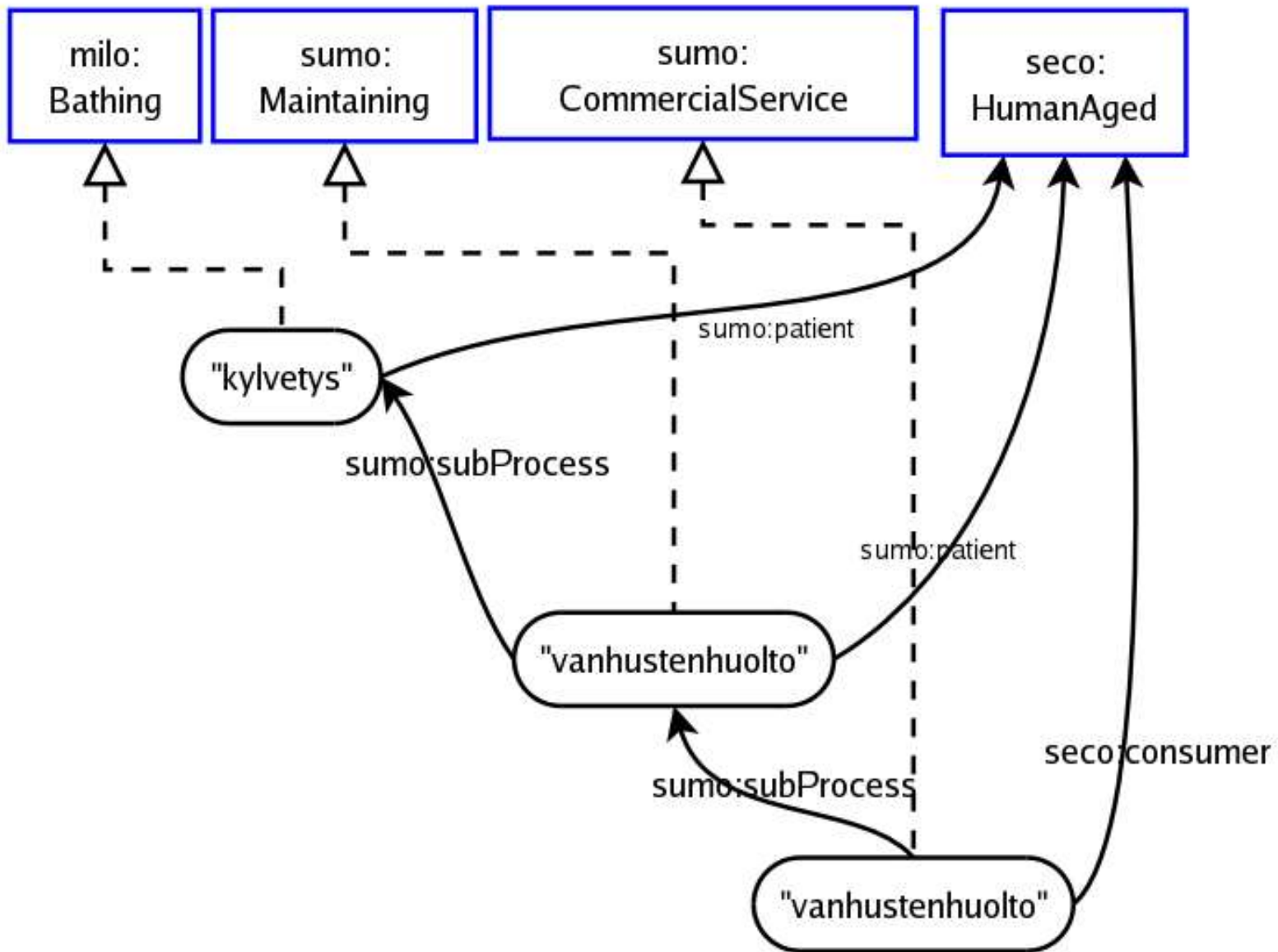


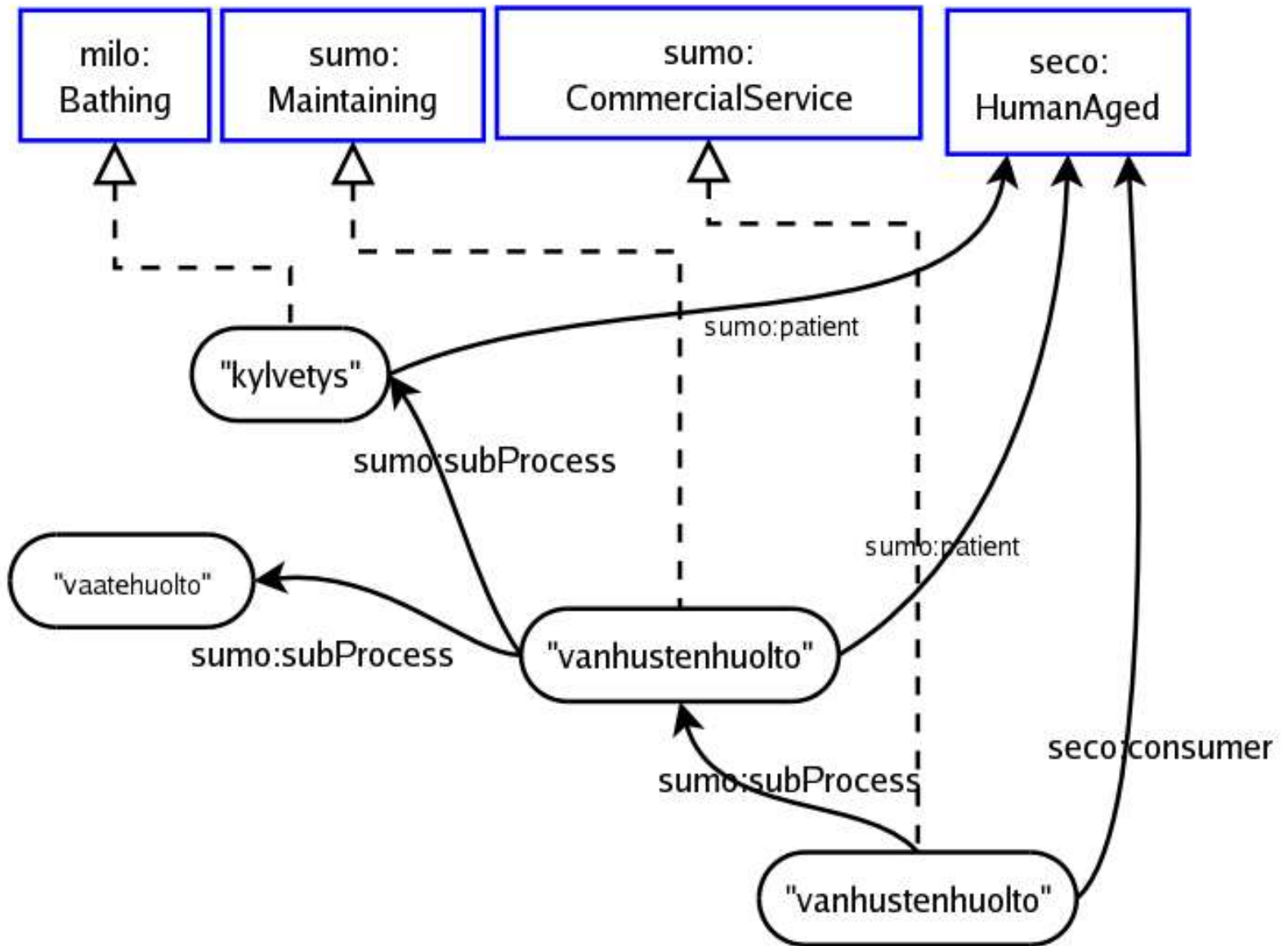


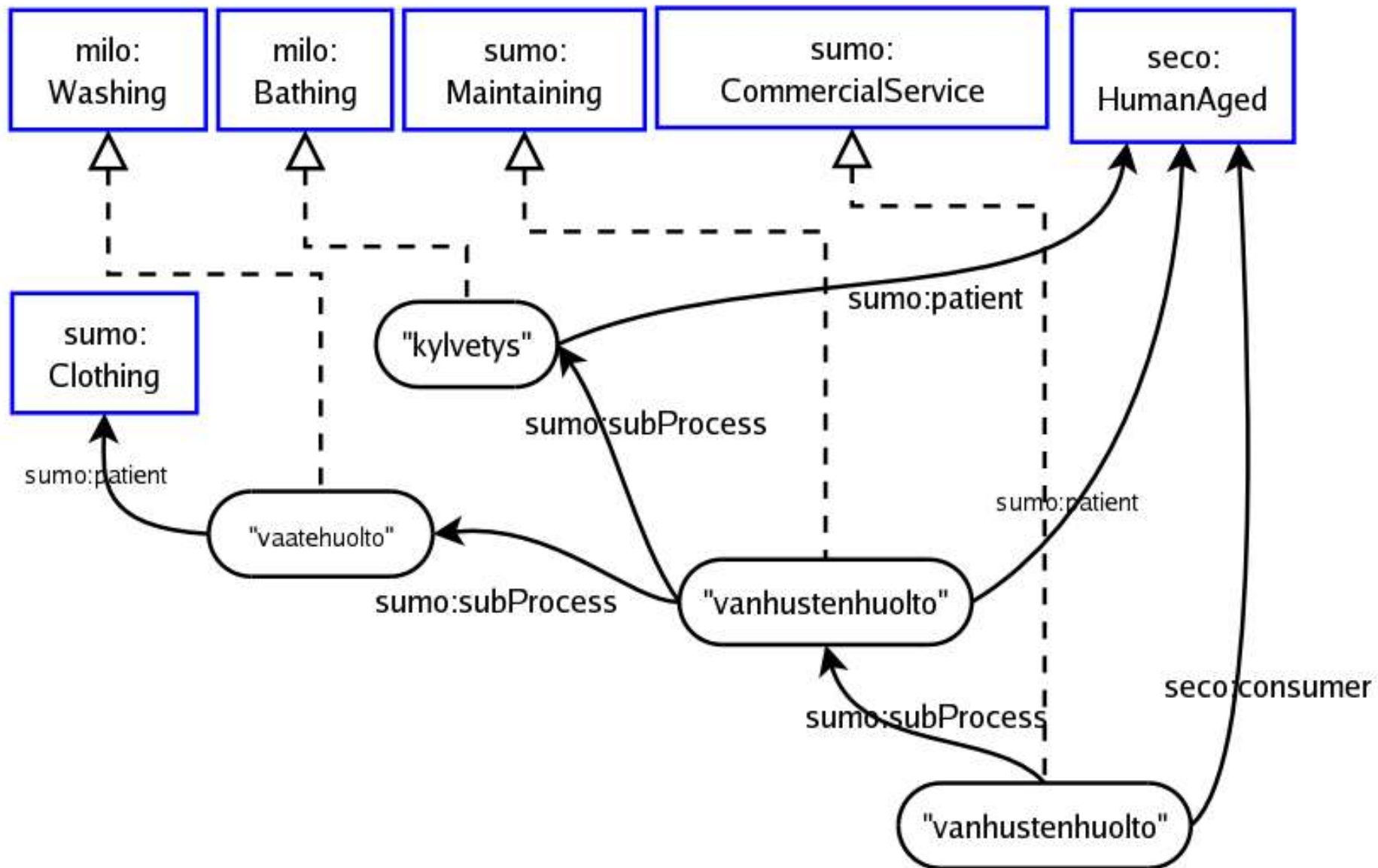


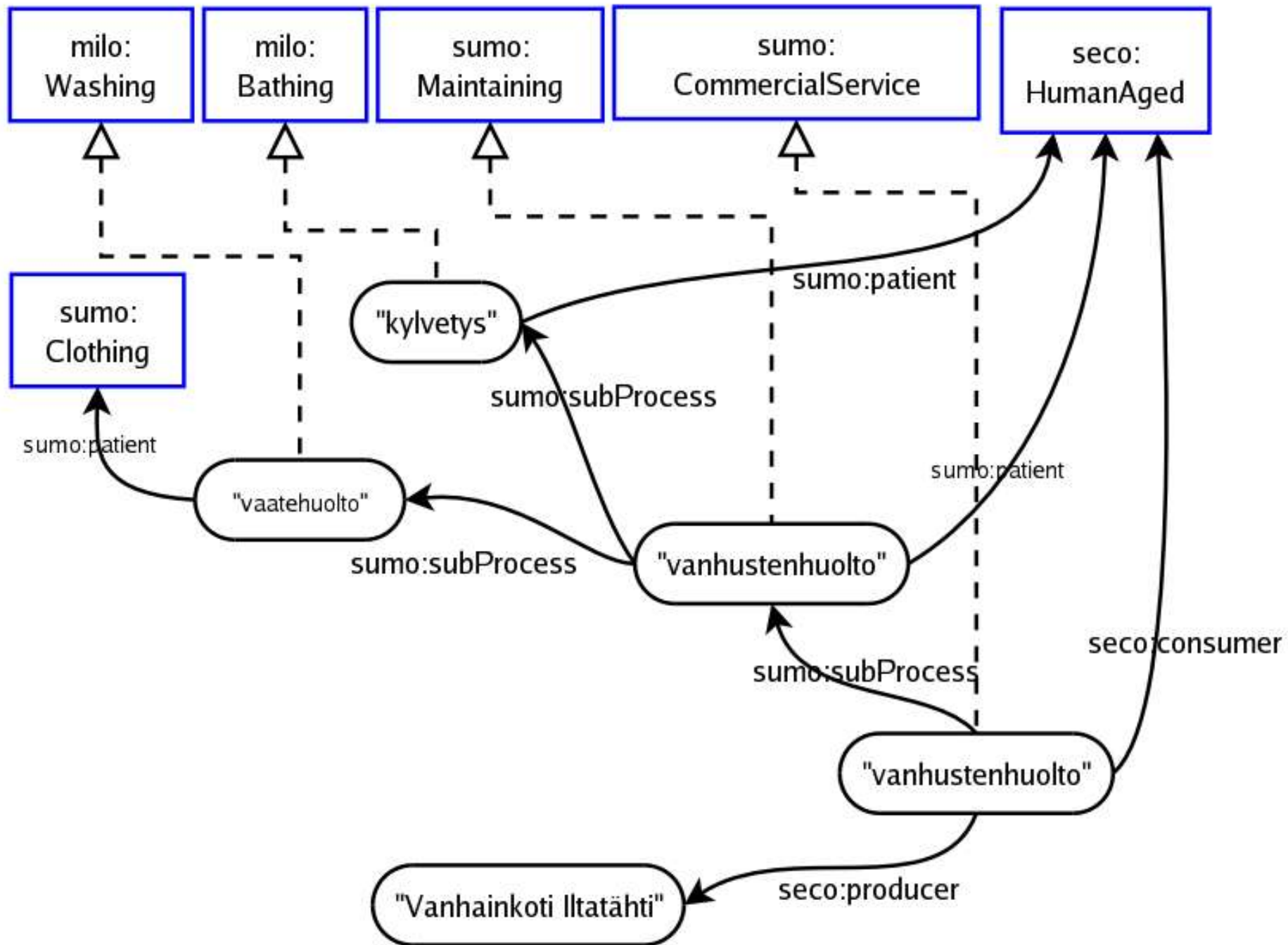


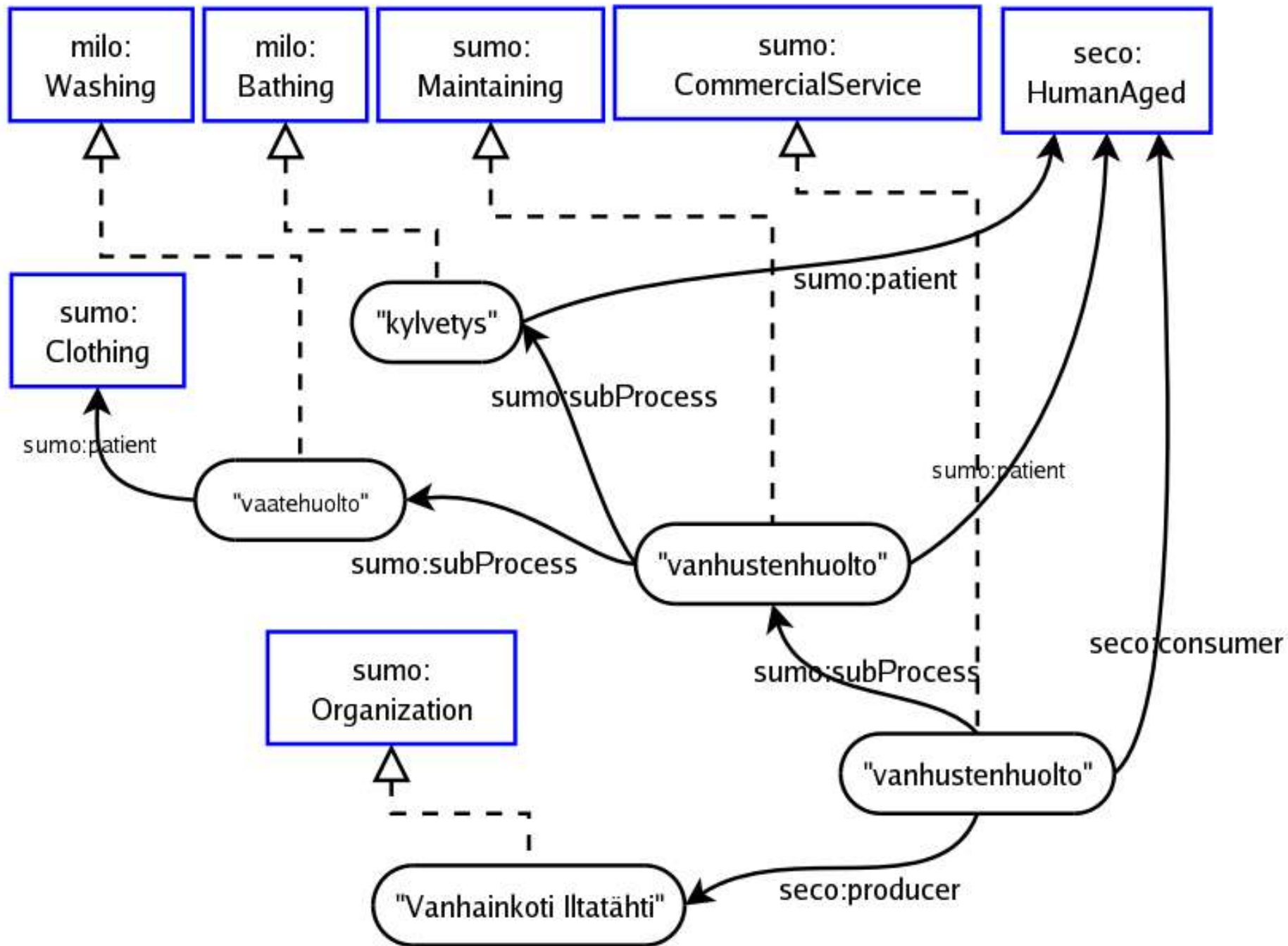


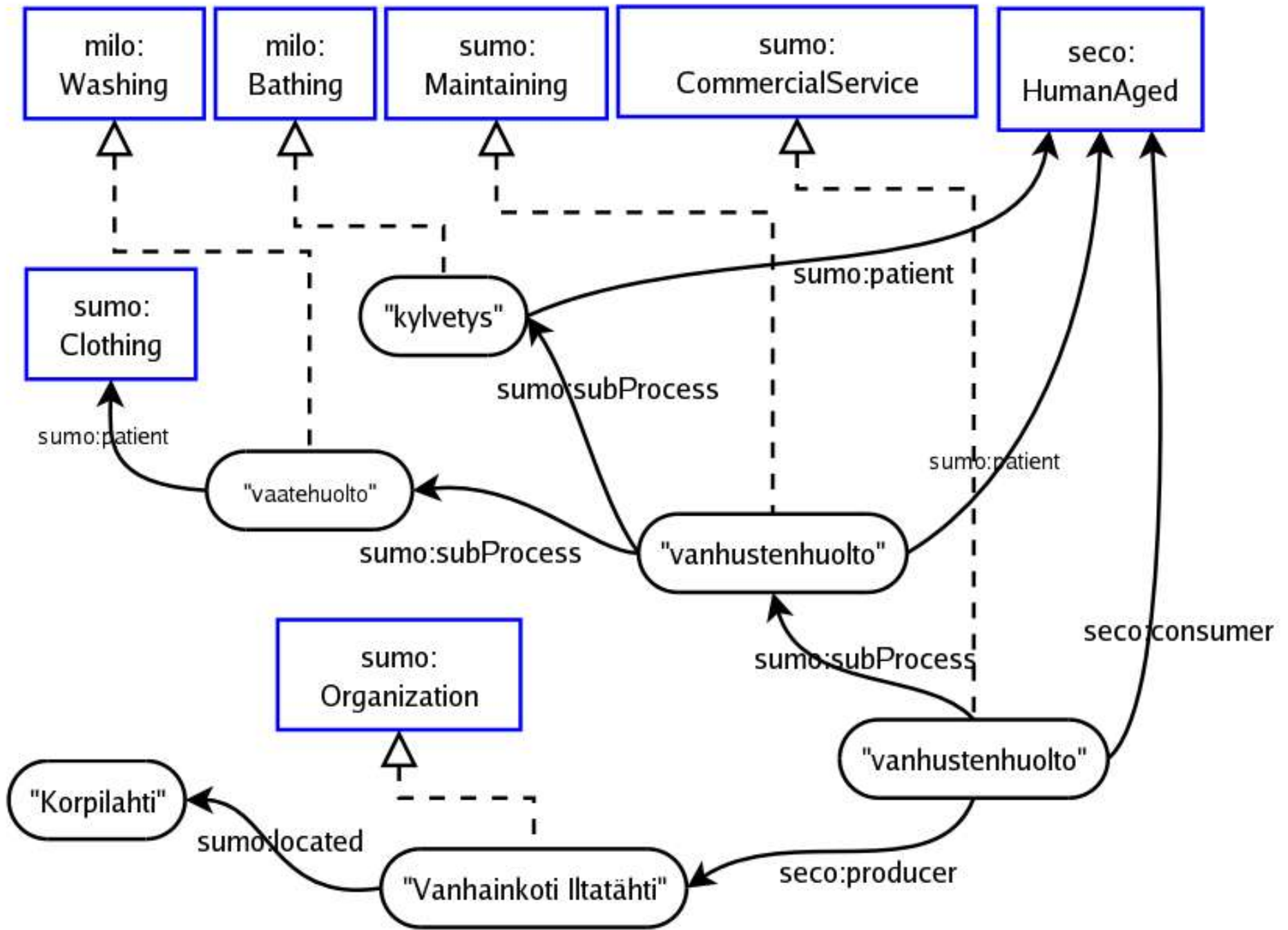


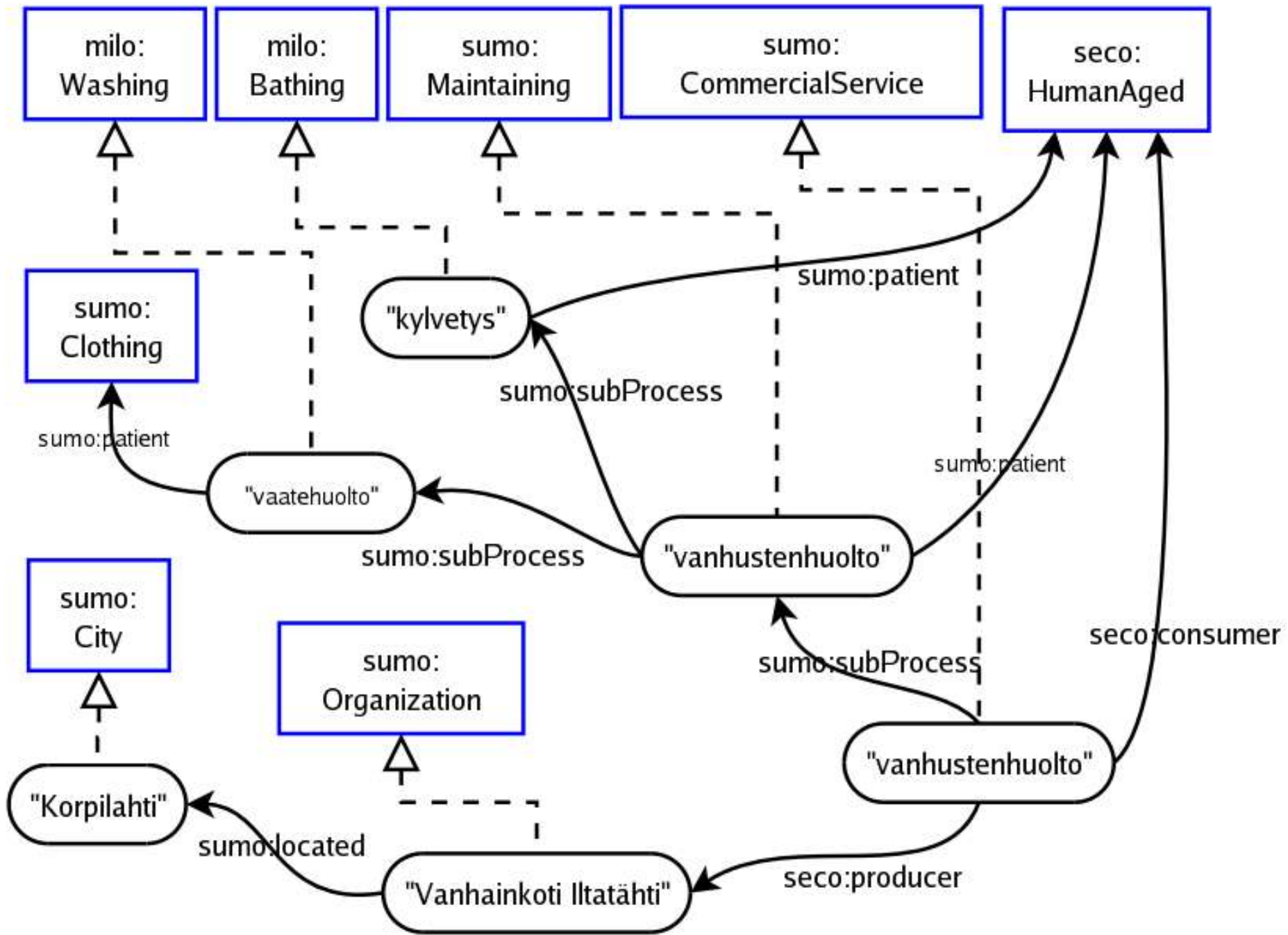


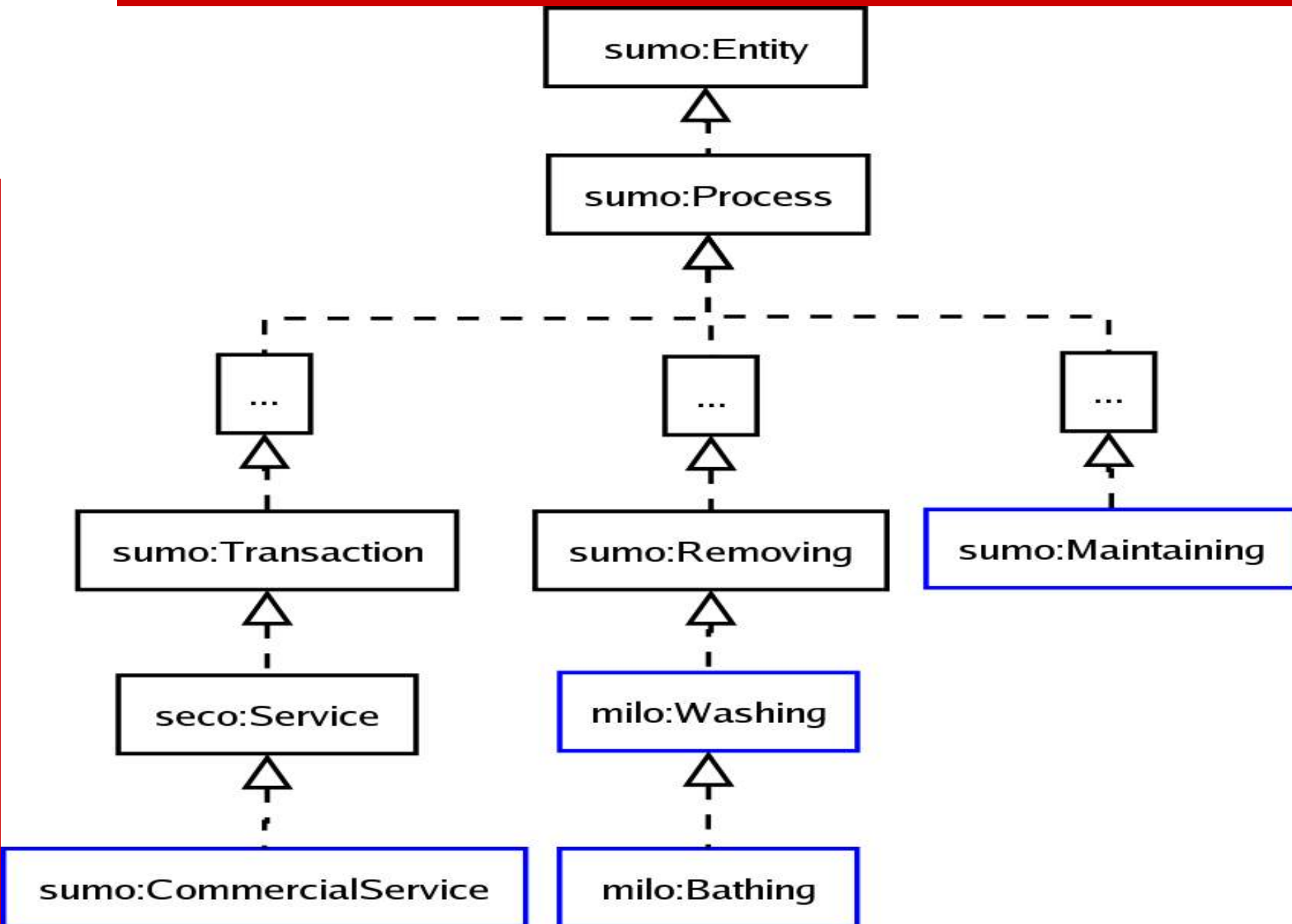


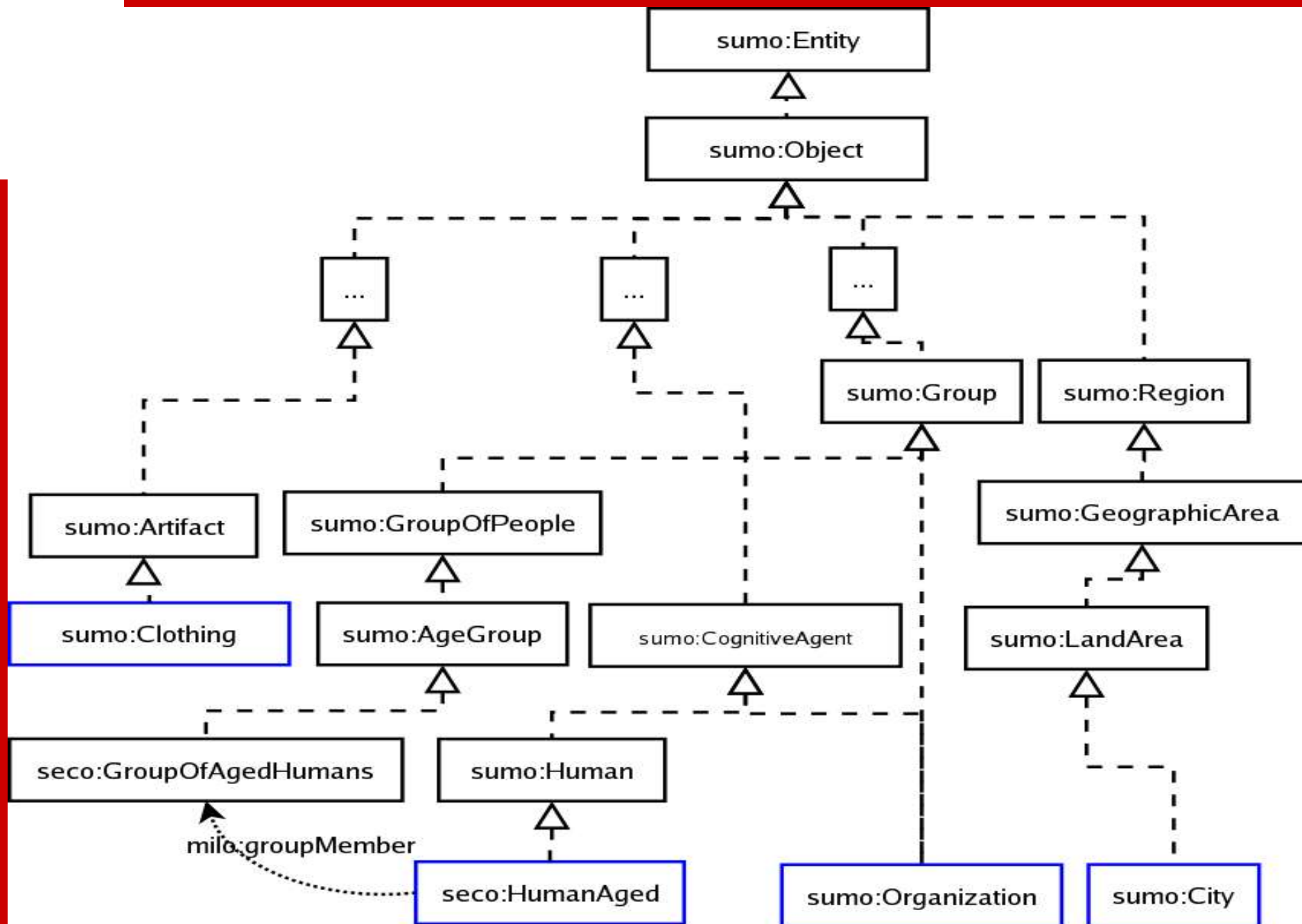




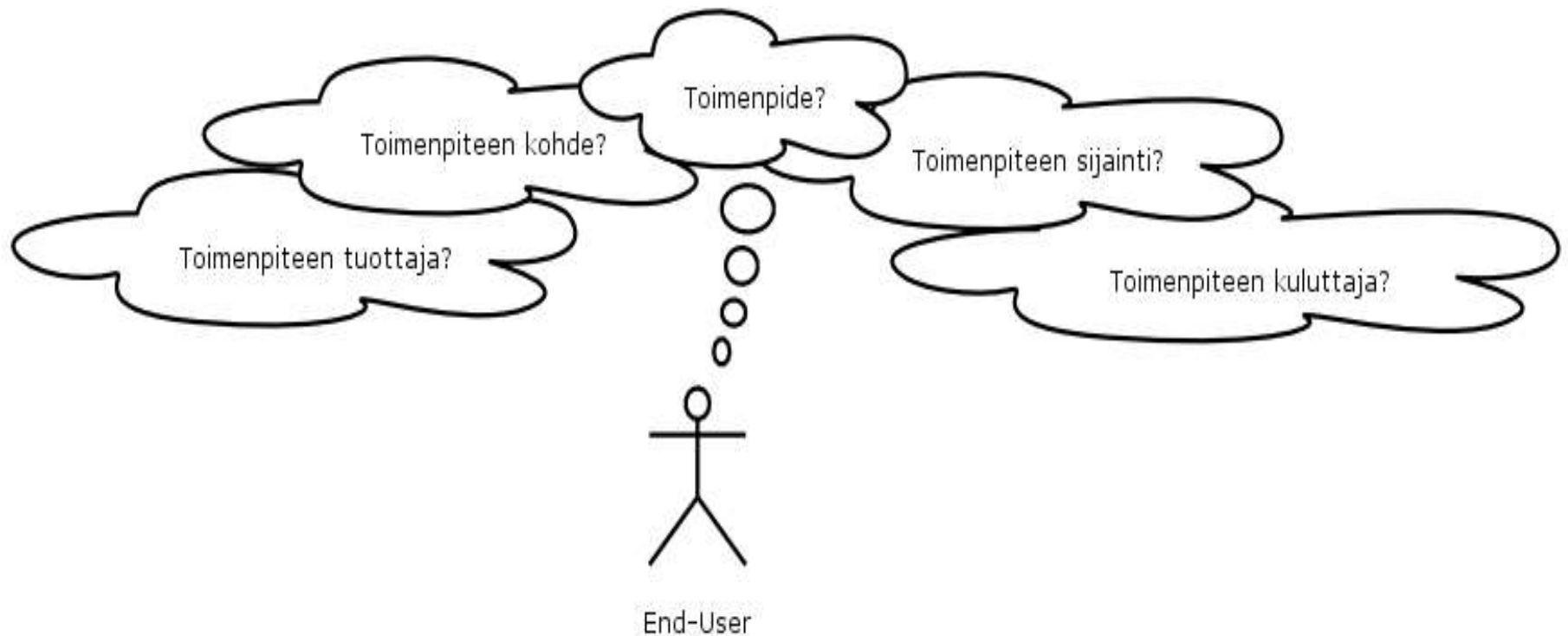








Haku käsitteillä ja niiden välisillä suhteilla



Palveluiden kuvailun osittainen automatisointi

- Palveluntarjoajan näkökulmasta palveluiden kuvailu palvelutietomalliamme käyttäen on varmasti työläämpää, kuin vapaiden kuvailujen laatiminen.
- Kuvailua voi kuitenkin osin automatisoida, jos kuvailtavasta palvelusta on olemassa ryhmittelytietoa.
- Jos näin on, niin voimme käsin luoda jokaista ryhmää koskevat säännöt, joiden perusteella on mahdollista automaattisesti luoda palveluista vähintään alustava palvelutietomallimme mukainen kuvailu.
- Tätä menetelmää käyttäen kuvailimme palvelutietomallillamme Fonectan palveluiden hakujärjestelmän 200 000 ilmoitusta.

[kirjoita hakusana]

Mistä ongelmista haluat tietoa:

- + Emotionaalinen ongelma
- + Käytösongelma
- + Sosiaalinen ongelma
- + Fysiologinen ongelma

Minkälaista tietoa haluat:

- + Testit, itsehoitopopaat
- + Artikkelit
- + Keskustelut ym
- + Video/Ääni
- + Vertaistuki
- + Asiantuntijatieto

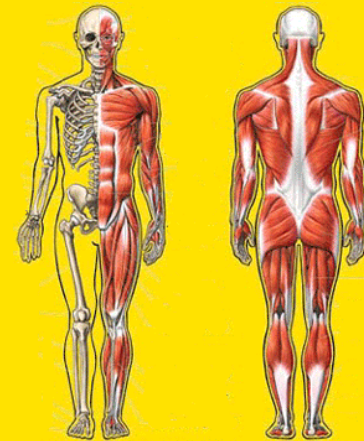
Keihin liittyvää tietoa etsit?

- + Ikä
- + Asuinpaikka
- + Sukupuoli
- + Koulutus
- + Elämäntilanne

Mitä palveluita etsit?

- + Sairaalapalvelut
- + Lääkäripalvelut
- + Sosiaalipalvelut
- + Hammashoito

Valitse ruumiinosa:



Mistä ongelmista haluat tietoa:

- sosiaalinen ahdistus
- koulu ahdistaa
- masentunut mieliala
- viiltely

Minkälaista tietoa haluat:

Keihin liittyvää tietoa etsit?

- 10-15 v
- seutu

PK-

nainen

Mitä palveluita etsit?

Valitse ruumiinosa:

ryhmittelyperuste

sisältötyyppi

järjestelyperuste

relevanssi

Testit	Oppaat	Artikkelit	Keskustelut	Palvelut
1. Testi: Oletko masentunut? (0.86)	1. Voita masennus! (0.97)	1. Raportti: Suomalainen masennus – KTL:n raportti (0.87)	1. Foorumi: Koulussa kärsivien keskusteluryhmä (0.9)	1. Terveyspalvelut: Helsingin terveyspalvelut (0.85)
2. Testi: Stressitesti (0.7)	Itsehoito-opas: Onnelliseksi teiniksi – kolmen askeleen ohjelma (0.94)	2. Tutkimusartikkeli: Ahdistava koulukulttuurimme (0.82)	2. Sähköpostipalsta: Kysy koulukuraattorilta (0.75)	2. e-Palvelut: Löydä kaveri (0.84)
	2. Itsehoito-opas: Kivaa yksinään (0.88)	3. Aikuistuminen (0.80)	3. Chat: Koululaischat (0.73)	2. Yksityiset terapeutit ry. (0.55)
		4. Nuorten itsemurhat (0.78)		

[kirjoita hakusana]

Mistä ongelmista haluat tietoa:

- sosiaalinen ahdistus
- koulu ahdistaa
- masentunut mieliala
- viiltely

Minkälaista tietoa haluat:

Keihin liittyvää tietoa etsit?

- 10-15 v

- seutu

PK-

nainen

Mitä palveluita etsit?

Valitse ruumiinosa:

Unettomuus

Väsymys

Apeus

Itsetuhoisuus

Mielialahäiriöt

Vallilan mielenterveystoimisto

Helsingin nuorisopsykiatrinen poliklinikka

Masennus

Masennus on mielialahäiriö, jota esiintyy kaikissa ikäryhmissä...

Yhteenveto

- Jos sosiaali- ja terveysalan tiedot ja palvelut kuvataan semanttisesti (semanttiseen webiin) niin,
 - hyödyt:
 - tiedot ja palvelut ovat paremmin löydettävissä,
 - tiedot ja palvelut ovat helpommin ja tehokkaammin hallittavissa,
 - järjestelmien yhteentoimivuutta voidaan parantaa merkittävästi,
 - on mahdollista automatisoida entistä vaativampia asioita,
 - terveysala tulee liitetyksi muuhun semanttiseen webiin,
 - uudelleenkäytettävyys,
 - skaalaedut ja
 - kansalaisille on mahdollista toteuttaa parempia palveluita.
- Haasteet:
 - ontologioiden määrittely,
 - semanttisen tiedon tuottaminen ja
 - ontologioiden yhdistäminen
 - kulttuurierot → käsite-erot → ontologiaerot.

Kiitos. Kysymyksiä?

Petri Lindgren (petri.lindgren@helsinki.fi)

TKK Viestintätekniikka

www.seco.tkk.fi