

# **Ontologia: katteeton lupaus vai todellinen työkalu?**

Kimmo Kajas

Tampereen yliopisto  
Tietojenkäsittelytieteiden laitos  
Tietojenkäsittelyoppi  
Pro gradu -tutkielma  
Ohjaaja: Hannu Kangassalo  
Kesäkuu 2008

Tampereen yliopisto  
Tietojenkäsittelytieteiden laitos

Tietojenkäsittelyoppi  
Kimmo Kajas: Ontologia: katteeton lupaus vai todellinen työkalu?  
Pro gradu -tutkielma, 68 sivua

Kesäkuu 2008

---

Ontologia on teoria olemassa olevasta, yhteisesti sovitusta muodollisesta tietokone-lukuisesta käsitteistöstä. Monet pitävät ontologiaa ratkaisuna tietomassojen hallintaan, kun taas toiset kokevat sen olevan liian monimutkainen tapa järjestää tietoa. Tässä tutkimuksessa eritellään ontologian puolestapuhujien ja vastustajien näkemyksiä, jonka jälkeen niiden pohjalta yritetään luoda käsitys siitä, onko ontologia hyödyllinen työkalu tietojärjestelmien suunnitteluun.

Esittelen tässä tutkimuksessa ensin ontologioiden rakennetta, luokittelua ja teorioita. Tämän jälkeen paneudun ontologioiden vastustajien esille tuomaan kritiikkiin sekä kritiikin perusteluihin. Lopuksi havainnollistan ontologioiden käyttöä esittelemällä tunnetuimpia ontologioita ja niiden käyttökohteita. Tutkimus pohjautuu aikaisempiin tutkimuksiin ja kirjallisuuteen ontologiasta. Teorioita ja luokittelua selvennän lisäksi esimerkeillä, jotka liittyvät suurimmaksi osaksi kiinteistönvälitysalaan.

Kiinteistönvälitysala on käsitteistöltään sekä käsitteiden suhteiden osalta monipuolinen ja monimutkainen ja tästä syystä hedelmällinen aihepiiri löytää esimerkkejä. Lisäksi uskon, että kiinteistönvälitysalan termistö on useimmille tuttua, mikä osaltaan lisää teorioiden ymmärrettävyyttä.

Avainsanat ja -sanonnat: ontologia, ydinontologia, teoria, käytäntö, aika, kritiikki ontologioita kohtaan, ontologioiden koostaminen, ontologioiden sovellusalueet.

## Sisällys

|        |   |    |
|--------|---|----|
| 1.     | Johdanto .....                                      | 1  |
| 2.     | Ontologioiden perusta .....                         | 3  |
| 2.1.   | Määritelmä .....                                    | 3  |
| 2.2.   | Luokitukset .....                                   | 6  |
| 3.     | Ontologioiden rakenneosat .....                     | 7  |
| 3.1.   | Käsite .....  | 7  |
| 3.2.   | Ilmentymä .....                                     | 8  |
| 3.3.   | Suhde .....   | 9  |
| 3.4.   | Ominaisuus .....                                    | 10 |
| 3.5.   | Rajoite .....                                       | 11 |
| 3.6.   | Aksiooma .....                                      | 11 |
| 3.7.   | Muita rakenneosia .....                             | 12 |
| 4.     | Ontologinen luokittelu ja ontologioiden tasot ..... | 14 |
| 4.1.   | Uscholdin ja Gruningerin luokittelu .....           | 14 |
| 4.2.   | Guarinon luokittelu .....                           | 15 |
| 4.3.   | Gómez Pérezin ja Benjaminsin luokittelu .....       | 16 |
| 4.4.   | Polin luokittelu .....                              | 18 |
| 4.5.   | Ontologian tasot .....                              | 19 |
| 5.     | Ydinontologiat ontologioiden taustalla .....        | 21 |
| 6.     | Aika ontologiassa .....                             | 24 |
| 6.1.   | McTaggartin aikateoriat .....                       | 25 |
| 6.2.   | Ajan käsite .....                                   | 26 |
| 7.     | Ontologiat käytännössä .....                        | 28 |
| 7.1.   | Ontologioiden koostaminen .....                     | 28 |
| 7.2.   | Ontologioiden sovellusalueet .....                  | 30 |
| 8.     | Ontologiat epäilyksen alaisena .....                | 31 |
| 8.1.   | Clay Shirkyn ontologiavastainen näkemys .....       | 31 |
| 8.1.1. | Argumentit .....                                    | 32 |
| 8.1.2. | Shirkyn argumenttien kommentointia .....            | 35 |
| 8.2.   | Boris Wyssusekin kritiikki .....                    | 38 |
| 8.2.1. | Näkemykset .....                                    | 38 |

|   |    |
|---|----|
| 8.2.2. Ontologiakeskustelua .....                     | 41 |
| 8.2.3. Bunge-Wand-Weber -ontologian rakenteesta ..... | 43 |
| 9. Ontologioiden käyttökohteita.....                  | 48 |
| 9.1. WordNet.....                                     | 49 |
| 9.2. CYC .....  | 50 |
| 9.3. Bunge-Wand-Weber (BWW) -ontologia .....          | 52 |
| 9.4. Suggested Upper Merged Ontology (SUMO).....      | 54 |
| 9.5. OWL .....  | 55 |
| 10. Yhteenveto ontologioista.....                     | 58 |
| 11. Lopuksi.....                                      | 60 |
| Viiteluettelo .....                                   | 63 |

## 1. Johdanto

Lisääntyvät tietomäärät ovat johtaneet tilanteeseen, jossa tietojärjestelmien suunnittelijat ovat joutuneet etsimään uusia mahdollisuuksia tietojen järjestämiseen. Tietomassoja on kyettävä tallentamaan sellaisissa muodoissa, että niitä voidaan käyttää uudelleen ja että tieto olisi löydettävissä mahdollisimman helposti ja varmasti. Ongelmaan on tarjottu aikaisemmin ratkaisuksi tesauksia ja asiasanastoja, jotka ovat sittemmin osoittautuneet riittämättömiksi yksinkertaisuutensa vuoksi. Tesauksilla ja asiasanastoilla voidaan ainoastaan helpottaa tunnetun asian löytämistä, mutta vain sellaisissa tilanteissa, jossa tiedonhakija tuntee hyvin aihepiirin. Jos tiedonhakija ei tunne haettavaa asiaa hyvin, niin tarvitaan huomattavasti kehittyneempi järjestelmä. Tiedon järjestämiseen liittyvien indeksointi- ja hakuongelmien ratkaisuna on yleisesti alettu käyttää ontologiaan pohjautuvia sovelluksia, jotka ovat ominaisuuksiltaan perinteisiä asiasanastoja monipuolisempia.

Ontologiat ovat saaneet osakseen paljon huomiota, ja monet eri tahot ovat julkaisseet tutkimuksia ontologioiden teorioista ja hyödyntämisestä tiedonhallinnan piirissä [ks. Wyssusek, 2004]. Tästä huolimatta yksimielisyyteen teorioista ei ole aina päästy. Myös ontologioiden käyttämisen hyödyistä on kiistelty. [ks. esim. Guarino, 1998.] Ontologioiden käyttämisen hyödyllisyydestä tietojärjestelmissä on kiistelty; puolestapuhujat ovat luotelleet hyötyjä ja vastustajat ovat yrittäneet kumota niitä.

Tutkimuksessani tarkastelen ontologioiden ja niiden käytön saamaa kritiikkiä sekä sitä, millä tasolla ontologioita pystytään todellisuudessa hyödyntämään. Esittelen eri ontologiateorioita ja pohdin niiden käytettävyyttä ja soveltamiskelpoisuutta. Tuon esiin tutkimuksia ja artikkeleita sekä ontologioiden puolesta että niitä vastaan. Tavoitteenani on myös esitellä muutamia käytännön toteutuksia, joissa ontologioiden hyödyntäminen on joko onnistunut tai epäonnistunut.

Ovatko julkisuudessa esiintyneet ontologioiden käytön vastustajat perusteettomasti väittäneet ontologioita hyödyttömiksi? Ovatko ontologiat vain yksi uusi luokittelu, joka on ainoastaan edeltäjiään monimutkaisempi vai onko tietämyksen rakenne saatu

sellaiseen muotoon, että tietokoneohjelmat voivat hakea suurista tietomassoista kysytyä tietoa arvioiden tietosisällön relevanssia älykkäästi? Näihin kysymyksiin pyrin löytämään vastauksen tutkimuksessani.

Tässä tutkimuksessa tullaan havainnollistamaan ontologioiden ontologioihin liittyviä teoreettisia tapauksia suurimmaksi osaksi kiinteistönvälitysalaan liittyvin esimerkein. Kiinteistönvälitysala sopii esimerkkeihin hyvin sen monipuolisuuden ja monimutkaisuuden puolesta sekä siksi, että uskon alan termistön olevan tuttua useille lukijoille.

## 2. Ontologioiden perusta

### 2.1. Määritelmä

Ontologia on filosofian tutkimushaara, joka tutkii kaiken olevaisen rakenteita ja olemusta kaikilla todellisuuden osa-alueilla. Barry Smith [2003] on määritellyt ontologian tavoitteeksi tarjota vastaus kysymykseen, mitä luokkia tarvitaan kaikkien olioiden, ominaisuuksien, tapahtumien, prosessien ja suhteiden täydelliseen kuvaukseen. Tarkemmin sanottuna ontologia on kohteen yksityiskohtainen kuvaus tai määrittely eikä ainoastaan selitys. Lisäksi ontologia on kaiken aikaa muuttuva eikä pysyvä, stabiili tila. Useita aiemmin totena pidettyjä argumentteja on osoitettu vääriksi, ja näin ollen vanhoja ontologioita on jouduttu muuttamaan. On hyvin todennäköistä, että myös nykyään totena pidettäviä argumentteja todistetaan vääriksi ja kumotaan, kun uutta tutkimusta tehdään, joten argumenttien muuttumisen lisäksi myös ontologiat tulevat mahdollisesti muuttumaan. [Gruber, 1993a]

Ontologia on Thomas Gruberin [1993a] mukaan käsitteellistyksen yksiselitteinen määritelmä (engl. "An ontology is an explicit specification of a conceptualization."). Gruberin määritelmässä lähtökohtana on se, että ontologia pohjautuu filosofiaan ja tarkoittaa alkuperäisessä kontekstissään systemaattista olemassaolon selitystä. Kun filosofiassa ontologia tarkoittaa olemassaoloa, tekoälyn piirissä ontologia on keino, jolla olemassaoloa voidaan kuvailla tietokoneen ymmärtämässä muodossa. Ontologia siis mahdollistaa sanaston, jolla voidaan esittää reaalimaailman tapahtumat, objektit ja niiden väliset suhteet tietämyksen hallintaohjelmalle. Ontologiaa voidaankin pitää loogisen teorian ilmiönä.

Nicola Guarino on kehitellyt Gruberin määritelmän pidemmälle ja luonnehtii ontologiaa joukoksi loogisia aksioomia, jotka on suunniteltu selventämään sanaston aiottua tarkoitusta. Guarino tarkentaa Gruberin määritelmää toteamalla myös, että ontologia ei ole pelkästään käsitteellistyksen määritelmä, vaan se on teoria selventämään formaalin sanaston tarkoitettua merkitystä eli ontologista sitoutumista tiettyyn käsitteellistykseen. Tällöin loogisen kielen tarkoitettut mallit rajoittuvat ontologiseen sitoutumiseen. [Guarino, 1998]

Guarino [1998] lisää mielestään tärkeän kohdan Gruberin määritelmään: Ontologia on kielestä riippuvainen, kun taas käsitteellistäminen on kielestä riippumaton. Guarinon lisäys määritelmään selventää mallia huomattavasti, jolloin käsitteellistäminen voidaan mieltää kuvana tai ajatuksena kohdealueesta, ja ontologia onkin jonkin loogisen kielen kuvaus ajatuksesta. Looginen kieli rajoittaa ilmaisua oman sanastonsa ja rakenteensa puolesta, jolloin ontologian pitää sitoutua kielen sanastoon ja rakenteeseen. Kielen rajoitteiden vuoksi sillä ei kyetä muodostamaan täydellisesti ajatusta kohdealueesta, vaan tarkoitettu malli jää jonkin verran suppeammaksi kuin alkuperäinen ajatus tai kuva on ollut.

Toisin kuin Gruber ja Guarino, Jeff Heflin [2004] on esitellyt ontologian [www.w3.org](http://www.w3.org) -sivuilla hieman kansantajuisemmin: Ontologia määrittelee tunnetun toimialueen kuvailuun ja esittelyyn käytetyt *ehdot* (engl. terms). Ontologiat sisältävät tietokonelukuiset määritelmät toimialueen peruskäsitteistä ja niiden välisistä suhteista [Heflin, 2004]. Määritelmä sopii hyvin ontologiakielen esittelyyn, mutta se ei ole tarkoitettu teoreettisessa mielessä ontologian määritelmäksi, koska se esittelee vain pintapuolisesti ontologian ideaa.

Kaikki edellä mainitut tutkijat ovat kuitenkin yksimielisiä siitä, että ontologiat ilmaistaan logiikkaan pohjautuvalla kielellä. Tämä mahdollistaa sen, että luokkien, ominaisuuksien ja suhteiden väliset erot pysyvät johdonmukaisina, ja lisäksi logiikkaan pohjautuvien kielten muuntaminen tietokoneohjelmien logiikaksi on huomattavasti helpompaa kuin luonnolliseen kieleen pohjautuvien sääntöjen. Gruberin ja Guarinon määritelmien pohjimmainen tarkoitus eroaa huomattavasti Heflinin määritelmästä, sillä Gruber ja Guarino pyrkivät määrittelemään yleistä ontologiaa, kun taas Heflin käsittelee kohdealueeseen sidottua ontologiaa.

Kuten jo ontologioiden määritelmistä voidaan havaita, eri tutkijat ovat määritelleet ontologiaa erilaisista lähtökohdista. Niinpä ontologioita käsiteltäessä on oltava erityisen tarkka siitä, mitä ontologialla kulloinkin tarkoitetaan.

Ensinnäkin ontologiaa voidaan käsitellä teoriana ja tarkastella sitä yhtenäisenä filosofisena tutkimusalueena, jolloin sitä käsitellään yleisenä *toimialueesta*



*riippumattomana* (engl. domain independent) ja jolloin se käsittelee yleisiä sääntöjä, jotka ovat yhteisiä kaikille mahdollisille ontologioille. Tutkimuskohteena on myös se, millainen ontologian rakenteen tulisi olla ja miten sitä pitäisi käyttää.

Vaihtoehtoisesti ontologiaa voidaan pitää *toimialueesta riippuvana* (engl. domain dependent), jolloin luokitteluun otetaan mukaan termejä ja ominaisuuksia, jotka esiintyvät vain tutkittavassa kohdealueessa. Tämä jaottelu voidaan kuvata myös *ydin-/ylempänä* (engl. top-level/upper ontology) sekä *toimialueen ontologiatasona* (engl. domain ontology), jolloin ensimmäinen on kohdealueesta riippumaton ja jälkimmäinen kohdealueesta riippuvainen. [Smith, 2003; Poli, 2002] Lisäksi ontologialla voidaan tarkoittaa jotakin valmista mallia tai kieltä, jonka mukaan termejä ja niiden suhteita rakennetaan. Tällaisia malleja voidaan käyttää valmiissa tietokoneohjelmissa, jolloin käyttäjällä on käyttöliittymä, jolla hän voi tehdä ontologian muodostustyön. Hierarkia, joka on jotain mallia käyttäen muodostettu kohdealueen olemuksesta, on myös ontologia.

Roberto Poli [2002] on puolestaan rajannut ontologiaa määrittelemällä eron ontologisen, semioottisen ja semanttisen tiedon välille. Semiotiikka käsittelee merkkien ja sanojen merkitystä sekä syntaksia eli yhdistelysääntöjä. Semantiikka analysoi sanojen ja niiden yhdistelmien merkitystä ja luokittelua. Ontologinen tieto on jotain olevaa, kuten tapahtuma tai kohde, jolla on sijainti. Sijainti pitää kyetä määrittelemään jollain järjestelmällä, mutta järjestelmää ei tarvitse määritellä spesifisti. Epistemologinen tieto määrittelee sen, millaista tieto on. Tieto voi olla epistemologiselta kannalta uskomuksia, epävarmaa tieto tai väärää tietoa.

Ontologiat eivät ole vain luokitteluita. Ontologian rakenteista puhuttaessa voidaan helposti sekoittaa keskenään *luokitukset* (engl. taxonomy) ja *partonomiat* (engl. partonomy), jotka käsittelevät asioiden koostumista osasista. Ontologioissa käsitellään joissain tapauksissa myös partonomian tapaisia rakenteita.

Luokitukset eivät kuitenkaan ole välttämättä täysimääräisiä ontologioita, vaikka niissä onkin ontologisia piirteitä, koska ontologia on semanttisesti selvästi monipuolisempi. Luokituksissa lajittelu ennalta määrättyihin luokkiin ja mahdollisten luokkien välillä

tapahtuva valinta on keskeinen tavoite, kun taas ontologiassa pääpaino sijoittuu suuren asiayhteyksien ja asiakokonaisuuden rakenteen hallintaan.

## 2.2. Luokitukset

Käsitteet ja sisältymissuhteet muodostavat hierarkian, jota kutsutaan luokituksiksi. Tarkemmat ja yksityiskohtaisemmat käsitteet sisältyvät abstraktimpiin ja yleisempiin käsitteisiin, jolloin muodostuu matemaattisessa mielessä puumallinen hierarkia. [Gómez-Pérez ja Corcho, 2002]

Smith [2003] on luetellut kolme periaatetta, joiden pitäisi toteutua ontologioiden luokitusten kohdalla. Nämä periaatteet itsessään selventävät hyvin luokitusten tarkoitusta. Ensimmäisessä periaatteessa luokituksen tulisi muodostaa matemaattinen puumalli. Toisessa periaatteessa luokituksen tulisi muodostaa mahdollisimman matala puurakenne. Lopuksi kolmas periaate vaatii, että luokituksen tulisi olla yhtenäinen siten, että luokituksella olisi vain yksi juuri, jolloin kaikki kategoriat sijoittuisivat tämän juuren alle.

Käsitteet luokitus ja partonomia sekoitetaan helposti keskenään. Toisin kuin luokittelu, partonomia käsittelee kokonaisuutta, joka voidaan jakaa pienempiin kokonaisuuksiin. Partonomia ei kuitenkaan ole luokittelun vastakohta vaan se käsittelee kokonaisuutta eri näkökulmista. Partonomia koostuu kolmesta käsitteestä: kokonaisuus, osat ja osien osat. Partonomia ei kuitenkaan ole synonyymi mereologialle, joka on teoria osakokonaisuussuhteesta. Mereologiaa pidetään usein pohjana ontologialle tai ainakin pääasiallisena ontologian teoriana [ks. esim. Winston *et al.*, 1987].

### 3. Ontologioiden rakenneosat

Ontologiakieliä on kehitetty jo joitakin vuosia ja lähtökohtia on ollut useita. Kielet pohjautuvat erilaisiin jo olemassa oleviin tekniikoihin ja kieliä hyödynnetään ohjelmissa, joilla puolestaan muodostetaan ja prosessoidaan ontologioita. Näin ollen myös ontologian rakennetta kuvaavat termit ontologiakielissä ovat usein muodostuneet omanlaisikseen osittain tekniikoiden ja perusperiaatteiden pohjalta.

Yleisesti ontologioiden yhteydessä käytetään melko epäyhtenäisiä termejä komponenteista, jotka esittelen seuraavissa alaluvuissa, ja pyrin antamaan niistä asiaa selventäviä esimerkit.

#### 3.1. Käsite

*Käsite* (engl. concept) tarkoittaa sellaista termin merkitystä, jonka ryhmä vastuullisia henkilöitä on hyväksynyt. Tosin tämä ei ole ainoa määritelmä, jolla käsitettä on kuvailtu. Realistit, konseptualistit ja nominalistit filosofian koulukuntina ovat käyneet keskustelua käsitteen määritelmästä jo keskiajalta lähtien ja jättäneet jälkensä käsitteen määritelmään. Lisäksi psykologia, kielitiede ja tietotekniikka ovat termin runsaalla käytöllä muokanneet määritelmää. [Klein ja Smith, 2005] Nykyisellään käsitettä luonnehditaan abstraktiksi tai konkreettiseksi, yksittäiseksi tai yhdistetyksi, todelliseksi tai fiktiiviseksi. Se voi olla myös tehtävän, funktion, toiminnan, strategian tai päättelyketjun kuvaus. [Gómez-Pérez ja Corcho, 2002]

Tietämyksen esityksen kirjallisuudessa on kutsuttu lajittelevaa universaalia käsitteeksi [Guarino, 1994]. Lajitteleva universaali on jokseenkin samaa tarkoittava kuin olio, josta voidaan luoda todellisuuden ilmentymä.

Klein ja Smith [2005] painottavat englannin kielen termin 'concept' käytön vaikeutta ja tästä syystä he esittävät vaihtoehtoisia termejä käytettäväksi termin 'concept' sijasta. He

esittelevät filosofiasta tunnetun ongelman universaaleista ja niiden olemassaolosta. Realismi myöntää universaalien eli yleiskäsitteiden olemassaolon mielessä tai jopa todellisuudessa. Konseptualismin mukaan yleiskäsitteitä on olemassa, mutta niitä ei ole ilman niitä vastaavia ilmentymiä. Toisin sanoen universaali toteutuu ilmentymiensä yhtäläisyyksissä. Nominalismin perusteella universaaleja ei ole lainkaan olemassa, vaan universaalit ovat vain mielen muodostama yhteinen nimittäjä joukolle olioita. Klein ja Smith [2005] ehdottavatkin, että käsite jätettäisiin lähinnä tarkoittamaan ideaa, mielen käsitettä tai termin selitystä, ja korvattaisiin ontologiassa käytetty olevan yksikkö termillä *olio* (engl. entity), joka vastaisi molempia fyysistä oliota ja abstraktia yleistystä. Kleinin ja Smithin ehdotus fyysisen olion termiksi on *ilmentymä* (engl. instance), jota kutsutaan myös termeillä 'individual' tai 'particular'. Abstraktin yleistyksen he nimeävät *tyypiksi* (engl. type). Tyyppi vastaa filosofian termiä universaali ja lisäksi ontologioita käsittelevässä kirjallisuudessa termejä 'class', 'kind', 'category' ja 'taxon'. Klein ja Smith kuitenkin tekevät eron käsitteen ja tyyppin välille niin, että käsitettä käytetään jokapäiväisessä elämässä käytettävistä asioista, kuten haamu, vesivahinko tai ongelmaton rakentaminen. Käsitteet ovat sellaisia olioita, joilla ei ole ilmentymiä todellisessa elämässä tai niitä ei voida osoittaa mihinkään tiettyyn tyyppi-luokkaan.

Useat tutkijat käyttävät käsitteestä termiä 'object', jota käytetään kohteesta, joka vastaa Kleinin ja Smithin tyyppiä. Kuten Gruber [1993a] asian ilmaisee: Objektien joukko ja niiden väliset esitettävät suhteet kuvaavat esityskielen avulla tietämyksen. Lisäksi termiä käytetään käsitteen ominaisuudessa, jolloin puhutaan yleisesti olioista tai ontologian rakenneosista.

### 3.2. Ilmentymä

Barry Smith [2003] mukaan *ilmentymä* (engl. instance) on jokin, joka luo universaalin. Kuten jo aiemmin on esitelty, universaalit ovat asioiden tyyppejä tai lajeja, joita voidaan luokitella omiksi ryhmikseen. Universaalin voi luoda joko *pysyviksi objekteiksi* (engl. continuant) tai *ajallisiksi tapahtumiksi* (engl. occurent). Universaali voidaan siis määritellä olevan se, minkä voi luoda ja ilmentymä on kaikki se, mikä luo universaalin.

Smith kuitenkin huomauttaa, etteivät kaikki partikulaarit ole ilmentymiä. Jokin partikulaari voi olla niin ainutlaatuinen, että siitä ei voida muodostaa järkevää universaalia, josta voisi muodostaa uuden partikulaarin, jolla taas olisi samoja ominaisuuksia kuin ensimmäisellä partikulaarilla. [Smith, 2003]

Universaali on juuri se yksikkö, joita ontologiassa yritetään esittää mahdollisimman yksiselitteisesti. Ilmentymät ovat niitä todellisen maailman kohteita, joita ontologian rakentajat ympärillään näkevät ja yrittävät muodostaa sääntöjä, joiden mukaan voitaisiin aukottomasti luoda universaaleja ja niiden välisiä suhteita. Tallennetun informaation lukemisessa on tärkeää, että kuka tahansa, joka osaa lukea ontologian sisältöä, ymmärtäisi samalla lailla universaalien ja partikulaarien välisen ilmentymäsuhteen.

### 3.3. Suhde

*Suhde* (engl. relation) voi olla käsitteiden ja toisen käsitteen, ominaisuuden tai ilmentymän välillä, joista tärkein on kahden käsitteen välinen suhde. Ydinontologiassa pyritään määrittelemään eksplisiittisesti suhteet, joita ontologioissa saatetaan käyttää. Jos ontologiat eivät noudata mitään ydinontologiaa, myös suhteet voivat olla sovellusalueeseen liittyviä. Ontologiassa suhde kokoaa asiakokonaisuuksia yhteen ja selittää asioiden rakennetta. Kuten Smith [2004] artikkelissaan esittelee, suhde voi olla sisältämissuhde, jolloin yksityiskohtainen käsite sisältää yleisemmän käsitteen. Tällaisesta suhteesta esimerkkinä toimii ”tupakeittiö on keittiö” - tai ”tupakeittiö kuuluu keittiöihin” – suhde, jossa tupakeittiöön sisältyy keittiön ominaisuudet. Käsite ’keittiö’ ei siis pidä sisällään kaikkien keittiöiden ominaisuuksia, vaan vain ne, jotka tekevät keittiöstä juuri keittiön. Sisältymissuhteen lisäksi Smith [2004] esittelee osasuhteen. Tällöin toinen käsite kuuluu tärkeänä osana toiseen käsitteeseen tai toisin sanoen toinen käsite on toisen osakokonaisuus. Esimerkiksi talon ja oven suhteen voi ilmaista ”ovi on osa taloa” tai ”talossa on ovi”. Nämä kaksi suhdetta ovat yleisimpiä, mutta ontologioissa käytetään usein myös muita suhteita. Smith [2004] esittelee suhteet ”aiheuttaa”, ”mielikuva jostakin”, ”osallistuu”, ”kehittää” ja ”johtua jostakin”.

*Funktion* (engl. function) sanotaan olevan erityinen suhde yhden tai useamman olion välillä. Funktio saa arvoikseen  $n$ -kappaletta arvoja, joista  $n+1$  arvo on funktion muodostama tulos. Tapauksessa, jossa funktio saa vain yhden arvon, muodostaa funktio tuloksen saamastaan arvosta. [Gómez-Pérez ja Corcho, 2002]. Esimerkiksi kiinteistön omistajan ilmaiseva funktio saa arvokseen vain kiinteistön ja funktio palauttaa omistajan. Asunnon tai tontin hinnan laskeva funktio määrittää hinnan esimerkiksi sijainnin, kysynnän ja muiden hintaan vaikuttavien arvojen perusteella.

### 3.4. Ominaisuus

*Ominaisuus* (engl. property) tarkoittaa käsitettä. Käsitteen kuuluessa johonkin luokkaan sillä on luokalle ominaisia ominaisuuksia. Kuitenkin samaan luokkaan kuuluvilla ilmentymillä on samat luokan periyttämät ominaisuudet, vaikka todellisuudessa ilmentymät eroavat jollain tavoin toisistaan. [Laurence ja Margolis, 1999] Ominaisuudella lisätään ilmentymälle erityispiirteitä, joita se ei saa miltään ylemmältä luokalta. Synonyymejä ominaisuudelle on *lokeri* (engl. slot), *funktio* (engl. function) ja *attribuutti* (engl. attribute).

*Tietämyksen esittämisen* (engl. Knowledge representation) aihealueen kirjallisuudessa ominaisuutta käytetään tarkoittamaan *luonnehtivaa universaalia* (engl. characterising universal) [Guarino, 1994]. Ominaisuus siis tarkoittaa käsitettä samaan tapaan kuin Kleinin ja Smithin 'tyyppi'. Lisäksi Mario Bunge on käsitellyt ominaisuutta [Bunge, 1977], ja lisäksi hän esittelee ontologiaa termeillä *asia* (engl. thing) ja *asian ominaisuus* (engl. property of thing). Bunge tarkoittaa asian ja asian ominaisuuden merkitystä seuraavasti: asioiden ja asioiden ominaisuuksien ero on siinä, että asia on konkreettinen ja asian ominaisuus on abstraktinen, eikä kumpikaan voi esiintyä ilman toista. Jokainen ominaisuus lisää käsitteen tarkkuutta ja lokeroi käsitettä tarkempaan luokkaan tai antaa sille tarkemman tyypin.

*Luomisella* (engl. instantiation) tarkoitetaan yleisestä mallista yksityiskohtaisen esimerkin tuottamista. Esimerkki *luomisesta*: universaalista 'talo' on luotu ilmentymä "naapurin punainen omakotitalo, jossa katto vuotaa". Jos kyseessä on partikulaari, siitä

ei voi luoda ilmentymää, koska esimerkiksi kaikkien ihmisten joukko on lueteltavissa oleva lista ihmisiä eikä ole mahdollista muodostaa ihmistä, joka täyttäisi kaikkien ihmisten ominaisuudet.

Gómez-Pérez ja Corcho [2002] ovat esitelleet erilaisia ominaisuuksia, jotka ovat ontologian käsitteille yhteisiä omilla tavoillaan. *Ilmentymäominaisuuden* (engl. instance attribute) arvot voivat olla eriävät käsitteiden ilmentymien kesken. *Luokkaominaisuus* (engl. class attribute) taas on yhteinen kaikille käsitteen ilmentymille ja ominaisuuden arvo on myös yhteinen. Useammalla käsitteellä voi olla samannimiset ominaisuudet jolloin niitä kutsutaan *paikallisiksi ominaisuuksiksi* (engl. local attribute). Lopuksi Gómez-Pérez ja Corcho esittelevät *globaalin ominaisuuden* (engl. global attribute). Globaali ominaisuus voidaan liittää kaikkiin ontologian käsitteisiin, jolloin kaikilla ontologian käsitteillä on samanniminen ominaisuus.

### 3.5. Rajoite

*Rajoite* (engl. constrain) mahdollistaa ontologiaan tallennettavan tiedon rajoittamisen siten, että arvot ovat kontekstiin sopivia. Tällainen rajoite on *tyyppirajoite* (engl. type constraint), jolla määritetään arvon tyyppi. Tyyppirajoite varmistaa, että arvoa syötettäessä päivämääräkenttään ei syötetä esimerkiksi rahallista arvoa. *Kardinaalirajoitteella* (engl. cardinality constraint) määritellään arvojen oikea lukumäärä. Lisäksi ontologioihin voidaan lisätä *eheysrajoitteita* (engl. integrity constraint), joilla voidaan varmistaa, ettei loogisesti selvästi virheellisiä tietoja tai arvoja tallenneta. [Gómez-Pérez ja Corcho, 2002]

### 3.6. Aksioma

*Aksiomia* (engl. axiom) lisätään ontologioihin kuvaamaan käsitteiden välisiä suhteita, joissa ei ole kysymys tavanomaisista sisältymissuhteista, vaan pikemminkin rajoittavista suhteista [Gómez-Pérez ja Corcho, 2002]. Esimerkiksi kiinteistönvälittäjä

on henkilö, joka on suorittanut hyväksytysti kiinteistövälityskokeen. Esimerkin tapauksessa ei riitä, että kiinteistönvälittäjä määrittellään henkilönä, joka välittää asuntoja, koska asuntoa myyvä henkilö ei saa esiintyä kiinteistönvälittäjänä, jos hän ei ole saanut tutkintotodistusta.

### 3.7. Muita rakenneosia

Robert Poli [2002] tarkastelee ontologian rakenneosia aiemmin esiteltyyn verrattuna hieman erilaisesta näkökulmasta. Hänen lähtökohtanaan ovat ontologiaa käsittelevät erilaiset teoriat. Polin ontologian tärkein rakenneosa on teoria *substanssista* (engl. substance), johon kuuluvat osateoriat: teoria partikulaareista, todellisuuden taso ja teoria osa-kokonaisuus suhteesta. Teoria partikulaareista käsittelee kohteita, jotka voidaan jakaa objekteihin, prosesseihin, ryhmiin ja muihin vastaaviin asioihin, joita voidaan käsitellä yksittäisinä kohteina tai kokonaisuuksina. Polin käyttämät käsitteet ja termit eroavat jokseenkin esimerkiksi Guarinon käyttämistä termeistä. Poli käsittelee yksityiskohtaisemmin olemassaolon eri osa-alueita, kun taas esimerkiksi Guarino pyrkii käsittelemään ontologiaa mahdollisimman yleisesti.

*Todellisuuden taso* (engl. level of reality) ja *ontologinen kerros* (engl. ontological stratum) ovat harvinaisempia käsitteitä ontologioissa. Ontologinen kerros määrittää kohteen abstraktiuden tasoa. Jokaisella kerroksella on omat periaatteensa, lakinsa ja kategoriansa, ja kerroksen luonteen voi ymmärtää vain tarkastelemalla kyseisen kerroksen omia käsitteitä. Poli on luetellut kolme ontologian kerrosta: materiaallinen, psykologinen ja sosiaalinen kerros. Esimerkiksi tunne 'epätoivo' on psykologisella kerroksella eikä tunnetta voi olla ilman, että sitä kantaa jokin olento. Olento taas on materiaalisella kerroksella. Todellisuuden tasolla puolestaan määrittellään tietyn kerroksen sisäistä rakennetta. Esimerkiksi materiaalisella tasolla kudos koostuu soluista, solut koostuvat molekyyleistä ja molekyylit koostuvat atomeista. [Poli, 2002]

Polin käsittelemä aihe ei ole mitenkään poikkeava muusta aihepiirissä tehdystä tutkimuksesta, mutta hänen käyttämänsä asiakokonaisuudet ovat hankalia yhdistää muihin ylemmän tason ontologioista tehtyihin tutkimuksiin, koska hän käyttää termejä



hieman eri tavoin. Poli käsittelee enemmänkin periaatteellisella tasolla ontologioita kuin Smith tai Guarino, joiden tutkimus tukeutuu matemaattiseen ja perinteiseen filosofiaan.

## 4. Ontologinen luokittelu ja ontologioiden tasot

Ontologioita on luokiteltu monin eri perustein. Ontologian määritelmä on jo itsessään niin väljä, että yksiselitteistä luokittelua ei ole pystytty tekemään. Seuraavissa alaluvuissa esitellään tunnetuimpia ontologioiden luokitteluja. Luokitteluja havainnollistetaan esimerkein käyttäen kunkin luokitteluperusteen yhteydessä samoja kiinteistökauppaan liittyviä käsitteitä.

### 4.1. Uscholdin ja Gruningerin luokittelu

Mike Uschold ja Michael Gruninger jakavat ontologiatyypit formaaliuden mukaan neljään ryhmään: *erittäin vapaamuotoinen* (engl. highly informal), *puoliksi vapaamuotoinen* (engl. semi-informal), *puoliksi formaali* (engl. semi-formal) ja *tiukan formaali* (engl. rigorously formal). *Erittäin vapaamuotoinen* kuvaillaan väljästi luonnollisen kielen keinoin. *Puoliksi vapaamuotoinen* kuvaillaan luonnollista kieltä muistuttavalla kielellä, joka on rajoitettua ja rakenteistettua, jolloin moniselitteisyys vähenee ja selvyys lisääntyy. *Puoliksi formaali* esitetään keinotekoisella kielellä, joka on formaalisti määritelty. *Tiukan formaalissa* ontologiassa termit on määritelty huolellisesti käyttäen formaalia semantiikkaa ja teoreemia. [Uschold ja Gruninger, 1996]

Esimerkkinä edellä mainituista ontologioista voidaan pitää kiinteistökaupan alalla 'asunnon myynnin' luokittelua. Erittäin vapaamuotoisen ontologian mukaan esimerkki esitettäisiin 'asunnon myynnissä asunnon omistaja vaihtuu'. Puoliksi vapaamuotoisen vastaava esimerkki voi olla 'asunnon myynti -> omistajan vaihtuminen'. Puoliksi formaalissa esimerkki näyttäisi puolestaan funktiolta 'asunnon\_myynti(vanha\_omistaja, uusi\_omistaja) = omistajan\_vaihdos'. Tiukan formaalissa ontologiassa esimerkissä funktio ja sen määreet määritettäisiin huomattavasti tarkemmin ja lisäksi omistajan vaihdoksessa tapahtuvat henkilörekisteri ja varallisuuteen liittyvät muutokset ilmaistaisiin tarkemmin.

#### 4.2. Guarinin luokittelu

Nicola Guarinin neljä ontologiaa ovat *ydinontologiat* (engl. top-level ontologies), *toimialueontologiat* (engl. domain ontologies), *tehtäväontologiat* (engl. task ontologies) ja *sovellusontologiat* (engl. application ontologies). Ydinontologiat ovat toimialueesta riippumaton menetelmä kuvaa käsitteitä, kuten tilaa, aikaa, ainetta, objekteja, tapahtumia ja tekoja. Ydinontologiat kuvaavat yleisiä käsitteitä, jotka ovat yhteisiä suurelle määrälle käyttäjiä. Toimialue- ja tehtäväontologiat käsittelevät tietyn sovellusalueen tai yleisten tehtävien ja aktiviteettien ominaista käsitteistöä laajentaen ja tarkentaen ydinontologiassa määriteltyjä käsitteitä. Sovellusontologia puolestaan kuvailee käsitteitä, jotka ovat yhteisiä sekä toimialue- että tehtäväontologioista. [Guarino, 1998]

Esimerkkinä edellä mainituista ontologioista voidaan pitää kiinteistökaupan alalla 'asunnon myynnin' luokittelua. Ydinontologiatasolla määritellään objekti, tapahtuma ja tapahtuman sisältämä aika. Luokittelua tarkennetaan aihealuetoimialue- ja tehtäväontologioiksi, jolloin luokiteltavia käsitteitä olisivat ”asunto” (aihealuetoimialueontologia) ja ”myynti” (tehtäväontologia). Sovellusontologiassa käsitteitä tarkennetaan yhdistelemällä jo luokiteltuja käsitteitä uusiksi käsitteiksi, esimerkiksi 'asunnon myynti'.

Guarinin [1998] mielestä on tehtävä ero sovellusontologioiden ja tietämuskantojen välille. Ontologiaan rinnastetaan tiukasti termi *tietämuskanta* (engl. knowledge base), jolla tarkoitetaan todellisen elämän tietoja kohdealueesta. Guarino erittelee ontologian informaatioksi, joka ei ole tilanteeseen tai tilaan sidonnaista, toisin kuin tietämuskannassa oleva informaatio. Tietämuskannassa voi olla tallennettuna esimerkiksi yrityksen henkilöstön osaaminen, kun taas ontologia määrittelee, kuinka henkilöstön osaaminen tallennetaan niin, että sitä voidaan hyödyntää useissa järjestelmissä.

### 4.3. Gómez Pérezin ja Benjaminsin luokittelu

Asunción Gómez Pérez ja V. Richard Benjamins [1999] ovat jakaneet ontologiat useampaan alaluokkaan kuin aiemmin mainitut Uschold, Gruninger ja Guarino. Gómez Pérezin ja Benjaminsin luokittelu on yksityiskohtaisempi kuin edellä esitetyt luokittelut. Mahdollisesti juuri tämän seikan vuoksi Uscholdin ja Gruningerin sekä Guarinon luokittelut ovat nykyisin käytettävämpiä: Gómez Pérezin ja Benjaminsin luokittelut ovat mahdollisesti päällekkäisiä, jolloin yksi ontologia saattaa kuulua useampaan kuin yhteen luokkaan. Liian yksityiskohtainen luokittelu voi näin ollen vaikeuttaa ontologian luokituksen määrittämistä.

*Tiedon esitysentologia* (engl. knowledge representation ontologies) on tarkoitettu vangitsemaan tiedon esitysprimitiivit, joita on käytetty formalisoimaan tieto esitysparadigmoissa. *Yleiset ontologiat* (engl. general/common ontologies) sisältävät esimerkiksi asioihin, tapahtumiin, aikaan, avaruuteen, kausaalisuuteen, käyttäytymiseen ja funktioihin liittyvää sanastoa. *Ydinontologioihin* (engl. top-level ontologies) puolestaan liittyvät ja sisältyvät kaikki muut yleiset ontologiat, ja *metaontologiat* (engl. meta-ontologies/generic ontologies/core ontologies) ovat toimialueesta riippumatta yleiskäyttöisiä ontologioita, jotka sopivat yhteen muiden ontologioiden kanssa. *Toimialueontologiat* (engl. domain ontologies) ovat toimialuekohtaisesti uudelleen käytettäviä ja niiden sanasto koostuu toimialueen käsitteistä ja niiden suhteista sekä toiminnoista. *Tehtäväontologiat* (engl. task ontologies) ovat tehtäväkeskeisiä mutta eivät välttämättä yhteensopivia toimialueen sisällä. Sanastollisesti ne sisältävät muiden muassa yleisiä nimiä, verbejä ja adjektiiveja. *Toimialue–tehtäväontologiat* (engl. domain–task ontologies) ovat yhteensopivia tietyn toimialueen tehtävissä mutta eivät eri toimialueen kesken, ja *metodiontologiat* (engl. method ontologies) sisältävät määritelmät esittää päättelyprosessin olennaiset käsitteet ja suhteet. *Sovellusontologiat* (engl. application ontologies) puolestaan sisältävät tarvittavan tiedon tietyn sovelluksen mallintamiseen. [Gómez Pérez ja Benjamins, 1999]

Tiedon esitysentologia määrittää esimerkiksi kiinteistökauppaontologian ja sen yhteydessä käytettävien ontologioiden yhteiset merkintätavat (ontologiakielestä riippuen yksilöt, ominaisuudet, luokat, ilmentymät ja lokerot), jotta kyseisiä ontologioita voitaisiin käyttää rinnakkain. Esitysentologia voidaan mieltää

tallennuskielen määritelmänä, kuten esimerkiksi RDF-kieli [ks. <http://www.w3.org/RDF/>]

Kaikkein yleisimmät käsitteet kuvataan yleisillä ontologioilla, joiden käsitteinä voi olla esimerkiksi aika, omistaminen, sääntö, laki, muutos ja tarve. Käsitteitä voidaan käyttää uudelleen hyvin laaja-alaisesti muissa ontologioissa ja malleissa, jolloin käsitteiden määritelmässä ei käytetä tarkkoja sovellutusalueita, vaan pyritään määrittelemään käsitteet yleiskäyttöisiksi.

Ydinontologioissa määritellään, millaisista osista yleisontologiat koostuvat. Tällaisia rakenneosia voivat olla esimerkiksi suhteet, luokat ja ominaisuudet, joita voidaan käyttää jokaisessa yleisessä ontologiassa. Määritelmässä käytetään predikaattilogiikan ja joukko-opin ilmaisuja.

Metaontologioiden tarkoituksena on tarkastella muiden ontologioiden tai ontologian rakenteiden semiotiikkaa. Toisin sanoen ontologia määrittelee, millaisilla merkinnöillä toiset ontologiat esittävät sisältönsä. Esimerkkinä metaontologiasta voidaan pitää graafista kuvausta spesifimmän ontologian rakenteesta.

Toimialueontologioiden esimerkkikäsitteet kiinteistökaupan alalla voivat olla kiinteistö, myynti, osto, välitys ja asunto. Toimialueontologia on kuitenkin uudelleen käytettävissä tietyn toimialueen sisällä, mikä tarkoittaa, että kiinteistökauppaan, asumiseen ja ihmisten elämiseen liittyvien toimialueiden kesken voidaan käyttää samaa toimialueontologiaa.

Tehtäväontologia tarkentuu jo pelkästään kiinteistökauppaan ja esimerkiksi myynti-, osto- ja välitystilanteisiin. Ontologia keskittyy ainoastaan tapahtumien ja tehtävien nimeämiseen ja niiden semanttisten yhteyksien esittämiseen.

Toimialue-tehtäväontologia on yhteensopiva samanlaisten tehtävien kesken, mutta se ei ole yhteensopiva samankaltaisten toimialueiden kesken, jolloin tässä esimerkissä kiinteistökauppa toimii tehtävänä, mutta omakotikiinteistö ja kerrostalokiinteistö esittävät erilaisia toimialueita. Tapauksessa, jossa kiinteistönvälitystilanteessa

esimerkiksi kerrostalo vaihtaa omistajaa, käytetään tiettyä ontologiaa, jolloin tämä ontologia ei ole enää yhteensopiva omakotitalon kauppaa tehtäessä.

Metodiontologiat täsmentävät päättelyä erityistapauksissa. Yhdistelmä 'ostaja, kiinteistö ja myyjä' määrittävät ostajan ja myyjän roolit. Jos ostajaa ja myyjää yhdistävä kiinteistö vaihtuu, voi myös vaihtua henkilöiden roolit. Omakotitaloa myynyt voi esiintyä kesämökin ostajana.

Sovellusontologia määrittelee tietyille sovellutusalueelle ominaisia käsitteitä. Esimerkiksi ontologia KIVI-kiinteistönvälitysovelluksesta sisältää käsitteet 'välittäjä', 'kauppari' ja 'toimeksianto'. Ontologia käsittelee lähinnä sovelluksessa käytettyjä käsitteitä ja näiden käsitteiden määrittelemiseen vaadittavia käsitteitä. Tämä ontologia ei kuitenkaan sisällä koko kiinteistönvälitykseen liittyvää käsitteistöä.

#### 4.4. Polin luokittelu

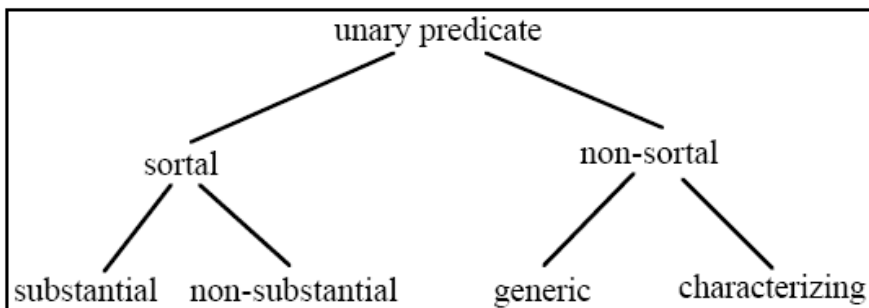
Roberto Poli on puolestaan jakanut ontologiat deskriptiivisiin, formaaleihin ja formalisoiuihin. Lisäksi kukin näistä ontologioista voi esiintyä kohdealueesta riippuvana tai riippumattomana [2002]. Polin jaottelu ei kuitenkaan ole saanut kannatusta, kun taas Guarinon neljään ryhmään jako on selvästi tunnetumpi.

Poli näyttäisi tutkivan ontologiaa enemmänkin käytännönläheisestä näkökulmasta, sillä hän muodostaa ontologioiden luokittelun käytännön esimerkkien pohjalta. Hän esittelee käsitteen *osittainen ontologia* (engl. quasi-ontological category), joka sisältää esimerkiksi ontologisen luokituksen tekijän eli luokittelijan tiedot kuten esimerkiksi allekirjoituksen. Näin voidaan arvioida ontologian osien yhteensopivuutta. [2002]

#### 4.5. Ontologian tasot

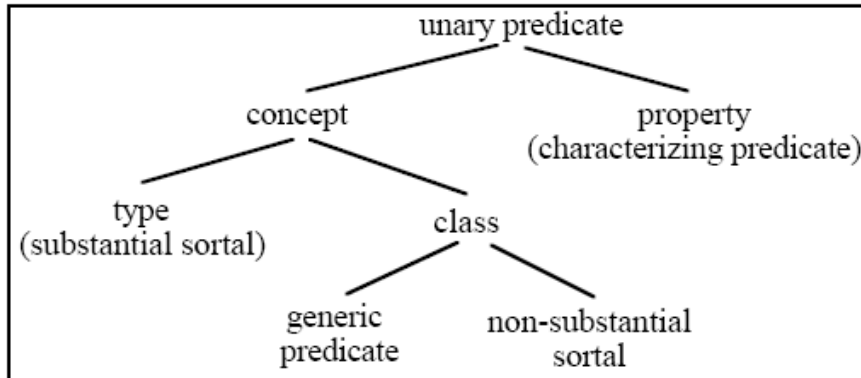
1970-luvun lopulla Ron Brachman esitti, että tietämyksen esityksessä käytetyt primitiivit voidaan luokitella neljään eri tasoon: toteutuksen tasoon, loogiseen tasoon, käsitteelliseen tasoon sekä lingvistiseen tasoon. Myöhemmin havaittiin, että luokittelu on vajaa, ja sen täydentämiseksi esitettiin epistemologista tasoa määrittelemään käsitteellisten yksilöiden ja niiden suhteiden formaalia rakennetta. [Guarino, 1994] Nicola Guarino kuitenkin pitää tätäkin viisitasoista luokittelua riittämättömänä kuvaamaan tietämyksen esityksen tarkkaa ontologista perustaa ja esittää, että epistemologisen ja käsitteellisen tason väliin tulee lisätä vielä yksi taso, ontologinen taso. Tämä taso olisi merkityksen taso, kun aiemmin lisätty epistemologinen taso olisi nimenomaan rakenteen taso. [Guarino, 1994]

Yksittäispredikaattityyppien perusontologian luokittelu koostuu Guarinon [1994] mukaan seuraavista osista: *Yksittäispredikaatit* (engl. unary predicate) vastaavat käsitteitä. Ne voidaan jakaa *laskettaviin yleisnimiin* (engl. sortal) ja *ei-laskettaviin*, *luonnehtiviin adjektiiveihin* (engl. non-sortal). Laskettavat yleisnimet voidaan puolestaan jakaa edelleen *konkreettisiin* (engl. substantial) ja *abstrakteihin* (engl. non-substantial) ominaisuuksiin. Ei-laskettavat, luonnehtivat adjektiivit voidaan myös luokitella tarkemmin *yleisiin asioihin* (engl. generic) ja *luonnehtiviin ominaisuuksiin* (engl. characterizing).



Kuva 1: Yksittäispredikaattityyppien perusontologia [Guarino, 1994]

Tämän pohjalta Guarino [1994] ehdottaa tietämyksen esitykseen formaalin esityksen ontologisen tason terminologiaksi seuraavaa:



Kuva 2: Ehdotus tietämyksen esityksen terminologiaksi [Guarino, 1994]

Tässä mallissa yksittäispredikaatit voidaan jakaa käsitteiksi (engl. concept) ja ominaisuuksiksi (engl. property). Käsitteet voidaan määritellä tarkemmin tyypeiksi (engl. type) ja luokiksi (engl. class), ja luokat voidaan jakaa edelleen yleisiin predikaatteihin (engl. generic predicate) ja abstrakteihin yleisnimiin (engl. non-substantial sortal). Ontologian tulisi olla logiikkaan pohjautuva, jota perusontologia ei ole. Guarinon [1994] ehdotus on formaalimpi ja ontologiset viittaukset ovat helpommin ymmärrettävissä, sekä ontologian uudelleen käyttö parantuu. Jälkimmäinen käsitteistön jaottelu on huomattavasti intuitiivisempi ja vastaa tietorakenteita, joita on totuttu tietojärjestelmissä käyttämään.



## 5. Ydinontologiat ontologioiden taustalla

Ydinontologioilla voidaan yhdistää kaikki olemassa olevat ontologiat ja niiden piirteet [Gómez Pérez ja Benjamins, 1999]. Ydinontologian tarkoitus on olla mahdollisimman yleinen ja täysin irrallinen yksittäisistä kohdealueista, jolloin sen antama rakenne mahdollistaa useiden saman aihepiirin ontologioiden käyttämisen samassa sovelluksessa. Jotta voidaan ymmärtää ontologioiden erot ja muodostaa formaali runko ydinontologialle, on tunnettava hyvin ydinontologioille yhteiset termit ja metodologia. Formaalin rungon muodostamiseen tarvitaan *toimialueneutraali* (engl. domain-neutral), formaali ontologia [Gangemi *et al.* 2001].

Formaali ontologia käsittelee Barry Smithin [1998] mukaan sellaisten formaalien objektien ominaisuuksia, jotka ovat formaaleja siksi, että niitä voidaan periaatteessa havainnollistaa todellisen maailman objektien avulla. Tämän määritelmän pohjalta Gangemi *et al.* [2001] toteavat, että formaaleja suhteita voidaan pitää kaikkia toimialueiden olioita koskevinä, jolloin formaalit suhteet voidaan ymmärtää yleisinä. Formaalien suhteiden avulla voidaan puolestaan muotoilla yleiset rajoitteet kaikkien toimialueiden olioille sekä määrätä toimialueen yleinen rakenne ja muodostaa erot olioiden välille, eli aikaansaada ydinontologia. Yleiset formaalit suhteet ja ominaisuudet esitellään seuraavaksi:

*Partikulaarit* ovat olioita, joista ei voida muodostaa ilmentymiä, koska ne ovat itse ilmentymiä, kun taas universaalit ovat olioita, joiden pohjalta voidaan muodostaa partikulaareja tai toisia universaaleja. *Ilmentymien muodostaminen* (engl. instantiation) on binäärisuhde partikulaarien ja universaalien välillä [Smith *et al.*, 2006]. Gangemi *et al.* [2001] mukaan ilmentymien muodostamissuhde yhdistää ne oliot, jotka kuuluvat eri loogisille tasoille. Toimialueen oliot ovat siis ominaisuuksien ilmentymiä, mistä johtuen ilmentymä onkin ontologisesti moniselitteinen, jos sen loogista tasoa ei ole määritelty: teorian A perusteella voidaan muodostaa metatason teoria B, jonka vakiosymbolit vastaavat A teorian predikaatteja ja jonka tarkoitettu toimialue on sama kuin A teorian ominaisuuksien toimialue.

Ilmentymien muodostamista ei tulisi sekoittaa *jäsenyyteen* (engl. membership). Nämä kaksi suhdetta eroavat toisistaan siten, että ilmentymä muodostetaan ominaisuudesta eli universaalista kohteesta, kun taas jäsenyys määritellään ilmentymien joukkona ja itsessään tämä joukko on partikulaari. [Gangemi *et al.* 2001]. Esimerkki jäsenyydestä voidaan esittää kiinteistökaupan alan termein seuraavasti: asunto kuuluu kaikkien kiinteistöjen joukkoon. Kuitenkin voidaan sanoa, että ”asunnolla on ominaisuus että se on kiinteistö”. Jäsenyydestä ei voida luoda uutta ilmentymää, kun taas ominaisuuden ’kiinteistö’ pohjalta voidaan luoda tietty ’asunto’.

*Osallisuussuhde* (engl. parthood) on intuitiivisesti melko yksinkertainen ja se pohjautuu teoriaan mereologiasta [ks. Casati and Varzi, 1999]. Tällä tieteenalalla tutkitaan koosteita ja niiden osia. Osallisuussuhteella voidaan analysoida kohteen sisäisestä rakenteesta se, onko kohde jakamaton tai jaettavissa oleva. Osallisuussuhteeseen liitetään usein piirteitä. Kun halutaan tarkastella koosteeseen liittyviä mielikuvia, otetaan käyttöön *yhteyden* (engl. connection) suhde. Sen avulla on tarkoitus saada selville joitakin perusmielikuvia, jotka liittyvät käsitykseen kokonaisuudesta. Yhteyden suhdetta pidetään yksinkertaisempänä kuin osallisuussuhdetta [Gangemi *et al.* 2001].

*Sijainti* (engl. location) kuvaa kohteen olemassaoloa yleensä avaruudessa, mutta Gangemi *et al.* [2001] mielestä suhde voidaan yleistää koskemaan myös yhteyttä sattumanvaraisten kohteiden ja neliulotteisten tilojen kesken.

*Laajennos* (engl. extension) on joukko kohteita, joilla on kaikilla yhteinen ominaisuus, jonka laajennos joukko on. Laajennosta pidetään suhteena, jolla yleistetyksi voidaan tarkastella suhdetta sattumanvaraisesti poimittujen olioiden ja neliulotteisten alueiden välillä. Laajennoksesta käytetään suomen kielessä myös nimitystä ekstensio. Ekstensiolla voidaan tosin tarkoittaa myös olioiden luokkaa, joihin jokin käsite viittaa.

*Riippuvuus* (engl. dependence) on suhde, joka on formaalin logiikan ja teorian formalisoinnin kannalta monitahoinen ja vaikeasti määriteltävä, mutta intuitiivisesti selvä. Riippuvuus on mikä tahansa suhde, jossa toinen tai molemmat osapuolet vaikuttavat toisen osapuolen olemassa olemiseen. Gangemi *et al.* [2001] toteavatkin, että riippuvuussuhde vaikuttaa mahdollisesti kaikkiin ontologian olioihin.

*Ominaisuudet* ovat ontologioiden toinen pääasiallinen rakenneosia ja sen vuoksi ne liittyvät läheisesti suhteisiin. Suhteita voidaankin käyttää ominaisuuksien määrittelyssä, kuten myös Gangemi *et al.* [2001] tekevät. Esittelen ominaisuudet seuraavaksi lyhyesti. *Konkreettisuus* (engl. concreteness) ja *abstraktisuus* (engl. abstractness) voidaan esitellä laajennoksen avulla niin, että ominaisuudesta laajennettu kohde on konkreettinen ja laajentamaton kohde on abstraktinen. Kuitenkaan ominaisuuden abstrahointia kohteiden joukosta ei pidä sekoittaa abstraktisuusominaisuuteen. Gangemi *et al.* [2001] esittelevät formaalien suhteiden lisäksi formaalit ominaisuudet.

*Yksiköllisyys* (engl. unity) ja *monikollisuus* (engl. plurality), sekä *riippuvuus* (engl. dependence) ja *riippumattomuus* (engl. independence) ovat formaaleja ominaisuuksia, joilla kohteille voidaan liittää yleisiä ominaisuuksia. Edellä esitetyillä suhteilla ja ominaisuuksilla voidaan kuvailla kohteita yleisesti, eivätkä kohdealueen erityispiirteet tuota ongelmia ontologialle.

*Ekstensionaalisuuden* (engl. extensionality) Gangemi *et al.* [2001] määrittelevät seuraavasti: Olio on ekstensionaalinen, jos ja vain jos minkä tahansa muun olion osat ollessa täsmälleen samat ovat oliot identtisiä toisilleen. Ekstensionaalisia olioita ovat esimerkiksi alueet ja aineiden määrät. Ekstensionaalisuudella tarkoitetaan, että ominaisuudet ovat ekstensionaalisia, jos ja vain jos kaikki sen ilmentymät ovat ekstensionaalisia. Näin ollen ominaisuudella on identiteetin ekstensionaalinen tunnusmerkki. Ekstensionaalisuudella tarkoitetaan alan kirjallisuudessa tosin myös sitä, että joukot ovat ekstensionaalisia kun ne ovat identtisiä ja kun niillä on samat jäsenet. Ominaisuudet ovat ekstensionaalisia kun samojen ilmentymien ominaisuudet ovat identtisiä.

## 6. Aika ontologiassa

Tavallisesti aika on sitä, mitä rannekellosta näkyy tällä hetkellä. Aika merkitsee ihmisille tätä hetkeä tai etäisyyttä tästä hetkestä eteenpäin tai taaksepäin. Aika on toisaalta jotain, joka kuluu tai kulkee tulevaisuudesta nykyisyyteen ja nykyisyydestä menneisyyteen [Oaklander, 2004] Ajoittain käsittelemme tapahtumia verraten niitä toisiinsa ja samalla vertailemme tapahtumahetkiä toisiinsa. Harvemmin kuitenkin vertailemme suurta joukkoa tapahtumia ja järjestämme niitä kronologiseen järjestykseen.

Aikajana jaetaan tavallisesti ennalta sovittuihin aikayksiköihin. Jotta voidaan ilmaista päivämääriä, tarvitaan kalenteri. *Encyclopaedia Britannica* määrittelee kalenterin ”laitteeksi, jolla lasketaan aikaa käyttämällä säännöllisiä rajoja tai jaksoja ja jota käytetään ajoittamaan tapahtumia.” Goralwalla *et al.* [2001] ovat tutkineet muun muassa kalentereiden esittämistä. Kalenterit ovat ihmisen tapa esittää fyysinen aika helpommin ymmärrettävässä muodossa. Edelleenkin maailmassa ei käytetä ainoastaan yhtä kalenterijärjestelmää, vaan erilaisia kalentereita on useita. Useinkaan ei tulla ajatelleeksi, että Suomessakin käytetään useampaa kalenterijärjestelmää. Gregoriaanisen kalenterin lisäksi meillä on käytössä organisaatioiden asettamia ajanlaskumalleja. Yrityksissä ja koulutuslaitoksissa ovat omat kalenterinsa, joilla aikaa jaksotetaan. Ajallisten ilmiöiden tallentaminen järjestelmään tuottaakin ongelmia, koska kalentereiden seuranta pakottaa toteuttamaan järjestelmään useita päällekkäisiä kalentereita. Käytössä olevien kalentereiden ajan rakeisuutta onkin kyettävä muuntamaan keskenään, niin että kalentereita voidaan käyttää rinnakkain. Vertaaminen onkin melko helppoa yhteiskunnan käyttämien kalentereiden kesken, koska pienimmät yksiköt ovat samanmittaisia. [Goralwalla *et al.*, 2001.]

Ajan käsite tuo kuitenkin mukanaan kaksi olennaista ongelmaa: epistemologisen ja ontologisen ongelman. Epistemologinen kysymys pyrkii löytämään vastauksen siihen, miten voidaan ymmärtää ajan luonnetta ja kuinka voidaan erottaa yksi aika toisesta ajasta. Ontologinen kysymys puolestaan vaatii selityksen siihen, ovatko ajat olevia, tulevia vai menneitä sekä milloin aika lakkaa olemasta. [Chisholm, 1996]

### 6.1. McTaggartin aikateoriat

Chisholm [1996] mukaan filosofit ovat yleisesti sitä mieltä, että on olemassa aikao-  
lioita, joiden avulla voimme identifioida ja yksilöidä tapahtumia; ajan käsite liittyy siis  
olennaisesti tapahtuman käsitteeseen. [Chisholm, 1996] Kun tapahtumat tietyssä ajassa  
ymmärretään *menneeksi* (engl. past), *nykyiseksi* (engl. present) ja *tulevaksi* (engl.  
future), puhutaan McTaggartin A-sarjasta tai ajan A-teoriasta. Tämän teorian  
kannattajien mukaan ajallisten suhteiden kuvaaminen tapahtumien välillä voidaan  
muuttaa kielestä toiseen ilman, että merkitys kärsii, koska *ajallinen muutos* [engl.  
temporal becoming] on metafyyysiseltä luonteeltaan perustavamman laatuinen kuin  
*ajalliset suhteet* (engl. temporal relations). Ajallisia suhteita ovat puolestaan *ennen*  
(engl. before), *jälkeen* (engl. after) ja *samanaikaisesti kuin* (engl. simultaneous with), ja  
ne muodostavat McTaggartin B-sarjan tai ajan B-teorian. Ajallisten suhteiden teorian  
kannattajien mukaan tapahtumien paikat A-sarjassa voidaan kuitenkin muuttaa tai  
määritellä käyttämällä ajallisia suhteita, koska ne ovat heidän mukaansa  
perustavamman laatuista kuin ajallinen muutos. Kiistelystä huolimatta kaikki yritykset  
muuttaa suhteet *ennen* ja *jälkeen* käsitteiksi *mennyt*, *nykyinen* ja *tuleva* ovat  
epäonnistuneet, koska jos yritetään muuttaa ajallinen suhde

*a* on aikaisempi kuin *b*,

ja kun *a* ja *b* voidaan määritellä käsittein seuraavasti

*a* on mennyt ja *b* on nykyinen TAI

*a* on nykyinen ja *b* on tuleva TAI

*a* on mennyt ja *b* on tuleva,

huomataan, että muutos on mahdoton suorittaa, koska sekä *a* että *b* voidaan kumpikin määrittellä kahdella eri käsitteellä. [Oaklander, 2004]

McTaggartin mukaan ontologia, joka perustuu vain ajan B-teorian mukaisiin suhteisiin, ei voi pitää paikkaansa. Väite perustuu siihen, että ajallisilla suhteilla ei voida kuvata ajan suuntaa tai muutosta eikä tilan ja ajan välistä eroa. Toisaalta McTaggart toteaa, että mikäli hyväksytään ainoastaan ajan A-teoria ja väitetään ajallisten suhteiden olevan määrittely- ja analysointikelpoisia, kohdataan ristiriita, jossa tapahtumalla on yhteensopimattomia ominaisuuksia ja jossa niiden poistaminen johtaisi loputtomaan ja ratkaisemattomaan jatkumoon. [Oaklander, 2004]

McTaggartin aikateoriat ja niihin liittyvät argumentit ovat saaneet aikaiseksi runsaasti keskustelua. Sekä A- että B-teorialla on kannattajansa ja vastustajansa, joista kaikki yrittävät jollain tapaa ratkaista ajallisen muutoksen ongelmia: Miten yhdellä ja samalla tapahtumalla voi olla keskenään yhteensopimattomia sisäisiä ajallisia ominaisuuksia? Miten yhdellä ja samalla oliolla voi olla yhteensopimattomia ominaisuuksia, jotka kuitenkin ovat selvästi yhteydessä keskenään? [Oaklander, 2004]

## 6.2. Ajan käsite

Aikaa käsitteenä voidaan tutkia monesta eri näkökulmasta. Patrick J. Hayes [1995] esittelee englannin kielen sanan *time* merkityksiä: *ajan fyysikaalinen ulottuvuus* (time-dimension), *aikajana* l. *aika-avaruus* (time-line), *ajanjakso* (time-interval), *aikapiste* (timepoint), *kesto* (duration) ja *ajankohta* (timeposition). Näistä termeistä ajanjaksoa, aikapistettä, kestoa ja ajankohtaa käytetään yleensä temporaalisen tiedon tallentamisen yhteydessä.

Aikapisteet muodostavat ajanjakson osoittamalla alku- ja päätepisteet jaksolle. Aikapiste ja ajanjakso ovatkin käsitteinä yhteensopivia keskenään. Sen sijaan tapahtuman kesto ei ole täysin yhteensopiva käsite aikapisteen ja ajanjakson kanssa. Tästä huolimatta keston määrittelyyn tarvitaan kuitenkin jonkinlaista ajanjaksoa, kuten Hayesin [1995] mainitsemaa standardiajanjaksoa, esimerkiksi silmänräpäys tai päivä.

Yhteensopivuus syntyykin käsitteiden abstraktiotasoeroista. Kesto on omalla abstraktiotasollaan ja ajankohta on kestoä konkreettisemmalla tasolla. Ajankohdalla tarkoitetaan tiettyä nimettyä hetkeä aikajanalla, kuten esimerkiksi 12.00 tai 24.12.2003. Ajankohdat voivat määrätä keston samoin kuin aikapisteetkin voivat. Kestoa voidaan käyttää abstraktina käsitteenä puhuttaessa esimerkiksi ajanjakson kestosta. Kahden ajankohdan välistä eroa voidaan myös kutsua kestoksi.

Tietojärjestelmien suunnittelussa on jouduttu huomaamaan, että ajalla on erittäin ratkaiseva merkitys, kun tallennetaan suurta määrää esimerkiksi mittaustietoa. Enää ei riitä pelkästään tapahtuman tapahtumahetki, vaan tarvitaan myös tallennushetki, voimaantuloaika ja voimassaolon päättymishetki.

## 7. Ontologiat käytännössä

Käytännössä ontologia on kuitenkin matemaattinen puurakenne. Hierarkkinen puurakenne on saanut alkunsa Aristoteleen kategoria-ajatuksista. Sittemmin Roderick M. Chisholm esittelee ontologian yksinkertaisena ja yksiselitteisenä puurakenteena, joka alkaa juuresta, joita on yksi ja josta lähtee yksi tai useampia oksia, jotka taas jakautuvat yhdeksi tai useammaksi oksaksi, niin kauan että oksa päättyy lehteen. Juuresta lehtiä kohden mentäessä kaksi oksaa ei kuitenkaan voi yhdistyä yhdeksi oksaksi. Ontologia ei kuitenkaan rakennu ainoastaan yhdestä yksinkertaisesta puurakenteesta, vaan useista puista tai *puuperheestä* (engl. family of trees). Puuperhe voidaan käsittää niin, että jokainen puu on saman asian oma luokittelu, mutta luokittelu on tehty eri näkökulmasta. Useat tutkijat ovat laiminlyöneet sen tosiasian, että kaikkia olioita ei saada mallinnettua samaan puurakenteeseen, vaan mallinnuksessa pitää päättää näkökulma. Mallinnus tehdään valitusta näkökulmasta ja jos yhdestä näkökulmasta tehty malli ei ole riittävä, niin mallinetaan kohdealue uudelleen toisesta näkökulmasta ja käytetään näitä malleja niille ominaisissa tilanteissa. [Smith, 2003]

### 7.1. Ontologioiden koostaminen

Ontologiaan sitoutuminen terminä kuulostaa triviaalilta, vaikka se todellisuudessa onkin tärkeä huomioitava asia. Tietyn sanaston käyttö ei pelkästään tarkoita ontologiaan sitoutumista, vaan on tehtävä ero lingvististen termien eli sanojen ja epistemologisen käsitteen välille. Tällä tarkoitetaan sitä, että kaksi eri kieltä ei voi jakaa samaa ontologiaa [Guarino, 1998]. Esimerkiksi suomen kieltä puhuvien ihmisten ajatusrakenne on hieman erilainen kuin esimerkiksi englannin kieltä äidinkielenään puhuvien. Ontologian ja epistemologian erottelu on tärkeää, koska ensimmäisen käsitteitä ovat objektit, prosessit, partikulaarit, individuaalit, kokonaisuudet, osat, tapahtumat, ominaisuudet, laadut ja tilat, kun taas jälkimmäisen ovat uskomus, oikea tieto, väärä tieto, epävarma tieto ja tarkistettu tieto.



Gruber [1993a] on todennut, että ontologia on sopimus käyttää tiettyä sanastoa yhdenmukaisesti ja sovitusti. Gruber painottaa ontologian suunnittelussa tiettyjä kriteerejä, joita pitäisi käyttää ontologian arvioimisessa. Ontologian laatua ei pitäisi arvioida omien tuntemusten pohjalta, vaan ennalta määriteltyjen kriteerien perusteella. Gruber [1993a] on esitellyt seuraavat kriteerit ontologian laadulliseen arvioimiseen:

1. Ontologian tulee olla *selkeä* (engl. clarity). Ontologian pitäisi kyetä selittämään sisältämillään termeillä se kokonaisuus, jota varten ontologia on luotu. Määritelmien pitäisi olla objektiivisia eli niiden ei pidä ottaa kantaa vallitseviin elämäntilanteisiin tai tietotekniikan tuomiin rajoitteisiin. Tämä tavoite voidaan saavuttaa formalismin avulla. Näin ollen kaikissa mahdollisissa tilanteissa määritelmät pitäisi esittää loogisilla aksioomilla. Kaikki määritelmät pitäisi esittää kokonaisina, niin että niissä on esitelty molemmat, vaadittavat ja riittävät tilanteet. Kaikki määritelmät pitäisi esittää myös luonnollisella kielellä.
2. Ontologian tulee olla *yhtenäinen* (engl. coherent). Ontologian pitäisi hyväksyä ainoastaan sellaiset päätelmät, jotka ovat määritelmien mukaisia. Sen lisäksi, että määritelmien aksioomat ovat loogisesti yhtenäisiä, pitäisi luonnollisella kielellä kirjoitetut määritelmät ja esimerkit olla yhtenäisiä loogisten aksioomien kanssa. Luonnollisen kielen esitys ja aksiomaattinen määrittely eivät saa olla ristiriidassa keskenään, ettei ontologiasta tule epäjohdonmukainen.
3. Ontologian tulee olla *laajennettava* (engl. extendibility). Ontologian tulee ennakoida yhteiskäyttö. Sen pitäisi mahdollistaa käsitteellinen perusta oletettavien toimenpiteiden varalle, jolloin ontologiaa voidaan kehittää ja tarkentaa ilman, että koko ontologiaa tarvitsisi korjata. Toisin sanoen olemassa olevan sanaston pohjalta pitäisi voida luoda uusia termejä ja niiden määritelmiä.
4. Ontologiassa *merkinnälliset ennakko-odotukset tulee minimoida* (engl. minimal encoding bias). Käsitteellistäminen pitäisi tapahtua *tietämyskerroksessa* (engl. knowledge-level) ilman, että *symboliikan kerros* (engl. symbol-level) vaikuttaa käsitteellistämiseen. Merkinnöistä johtuva väärinkäsitys johtuu yleensä siitä, että käsite määritellään pelkästään toteutuksen ja määrittelykielen

näkökulmasta. Tiedon jakamiseen toteutettu agentti on mahdollisesti tarkoitettu toisenlaiseen tiedon esitykseen. Tietämys pitää siis esittää mahdollisimman yleiskäyttöisellä esitysmuodolla.

5. Ontologinen *sitoutuminen tulee pitää mahdollisimman vähäisenä* (engl. minimal ontological commitment). Ontologian tulisi vaatia mahdollisimman vähän mallinnettavalta maailmalta. Vähäinen ontologinen sitoutuminen mahdollistaa ontologian laajemman uudelleen käytön useissa erilaisissa erikoistuneimmissa tilanteissa. Sitoutumista voidaan vähentää niin, että käytetään mahdollisimman yksinkertaista teoriaa ontologian muodostamiseen, jolloin tärkeimmäksi sitouttajaksi muodostuu yhtenäinen sanasto.

## 7.2. Ontologioiden sovellusalueet

Ontologiat on otettu käyttöön tekoälyn, tietokoneavusteisen lingvistiikan sekä tietokantatutkimuksen alueilla [Guarino, 1998]. Tekoälyn piirissä ontologioita tarvitaan päättelyketjujen muodostamiseen, tarkasteluun ja arviointiin. Tekoälyn päättelyketjut koostuvat tietämystä ja järkeä vastaavista väittämistä, jotka on kuvattu tietokonelukuisessa muodossa. [Smith, 2003] Lingvistiikassa voidaan luokitella sanoja sanaluokkien ja ryhmien mukaan tai sanojen muodostumisen historiaan liittyvien sukujen mukaan. Kun tietokoneavusteisessa lingvistiikassa hyödynnetään ontologioita, voi ontologian sääntöjen mukainen kielioppiin liittyvä sanojen luokittelu mahdollistaa täydellisen oikolukuohjelman. Toisaalta ontologiaa voidaan käyttää sanojen merkityssuhteiden kuvaamiseen, kuten anaforisen suhteen esimerkissä ”Auto on kulkuneuvo” [Cimiano ja Handschuh, 2003] on myös *intensionaalinen sisältämissuhde*. Intensionaalilla suhteella tarkoitetaan yleisesti suhteen ja olioiden välistä rakennetta käsitteellisessä tasolla [ks. Guarino, 1998]. Tietokantojen tutkimuksessa puolestaan ontologiaa voidaan käyttää uuden järjestelmän vaatimusmäärittelyä tehtäessä ja käsitteellisen mallinnuksen aikana. Lisäksi ontologiaa voidaan käyttää järjestelmien integroimisessa. Ontologia mahdollistaa eri järjestelmissä käytettyjen termien ja käsitteiden vertailemisen ja auttaa löytämään tilanteet, joissa toisistaan eroavia käsitteitä nimitetään samalla termillä ja tästä johtuvat ongelmat.

## 8. Ontologiat epäilyksen alaisena

Ontologia-sanaa käytetään nykyään yhä enemmän, mutta itse ontologioita ei niin usein. Ontologia sekoitetaan hyvin usein käsitteellistämiseen, jolloin toimialueen käsitteiden ja termien muodostamaa hierarkiaa kutsutaan ontologiaksi. Kuitenkin ontologia on huomattavasti yleisempi ja abstraktimpi. Ontologioiden rakentaminen on hankalaa. Ontologioiden rakentamisesta tekee vaikeaa erityisesti sen ajanvievyys ja siten kalleus varsinkin silloin, kun tavoitteena on saada aikaiseksi riittävän formaali ontologia. Riittävä formaalius mahdollistaa ontologian käsittelyn automaattisesti agenttien ja tiedonkäsittelyohjelmien avulla. Formaalin ja yhtenäisen rakenteen ylläpitoon tarvitaan yhteisö, joka käyttää ontologiaa ja sopii keskenään ontologian rakenteesta. Kaikki eivät jaa samaa näkemystä yhteisön kanssa ja he muodostavat omat ontologiansa tai muuttavat jo olemassa olevia ontologioita. [Grüninger ja Lee, 2002] Ontologiat ovat teorioita olemassa olevasta, ja niiden taustalla ovat filosofiset koulukunnat, jotka tuovat ontologiaan oman näkökulmansa. Tietämysjärjestelmien suunnittelussa ei oteta huomioon, että kaikkien toteutukseen osallistuvien pitäisi sitoutua samaan näkökulmaan toteutettavasta toimialueesta. Lisäksi ontologiaa pidetään usein liian abstraktina, jotta siitä voitaisiin saada hyötyä pienemmissä tietojärjestelmissä. Suuremmissa tietojärjestelmissä sen käyttöön uskotaan, koska se tuntuu ainoalta mahdolliselta keinolta hallita valtavia tietomääriä. Tavoitteenani on käydä lyhyesti läpi kriittisiä puheenvuoroja ontologioita vastaan ja muodostaa kuva siitä, miksi ontologioiden käyttöä yritetään vältellä tai niitä jopa halveksitaan.

### 8.1. Clay Shirky'n ontologiavastainen näkemys

Clay Shirky on amerikkalainen kirjailija ja konsultti. Hän toimii apulaisprofessorina New Yorkin Yliopistossa (NYU). Tavallisesti Shirky kirjoittaa kolumneja Internetin käyttäjien sosiaalisesta käyttäytymisestä. Ontologioihin Shirky on tarttunut keväällä 2005 pitämässään puheissa, joissa hän arvostelee kategorisoinnin ja ontologian käytettävyyttä.

Vaikka Clay Shirky onkin populistinen ”tutkija”, hänen aikaansaamansa kohu ja julkisuus ovat saattaneet vaikuttaa alkavien tietojärjestelmähankkeiden suunnitteluun, niin että ontologiset ratkaisut on haudattu Shirkyin kommenttien johdosta tarkastelematta niitä tapauskohtaisesti. Vaikka Clay Shirkyin kommentteissa on melko kevyt tiedetausta, tuodaan hänen vaikutuksensa esiin tässä tutkimuksessa.

### 8.1.1. Argumentit

Shirkyin [2005] mukaan digitaalisessa maailmassa laajalti käytetty luokittelu on yliarvostettua. Luokittelun ongelmakohta on hänen mukaansa juuri luokkien muodostamisessa etukäteen, kun luokiteltava aineisto ei sovikaan enää myöhemmin luokkiin, jotka on määrätty aikaisemmin. Luokat muodostetaan nykyisen tiedon pohjalta, eikä tulevaisuuden tarpeita voida ennustaa niin, että ne palvelisivat tulevia luokittelutarpeita.

Alkuaineiden jaksollinen järjestelmä on ehkä tunnetuin käytössä oleva luokittelu. Siinä törmätään Shirkyin mielestä juuri luokkien luomisen ongelmaan. Hän tuo ilmi seikan, kuinka jalokaasut olivat taulukkoa luotaessa tunnetusti kaasuja, mutta nykyään tiedetään, etteivät ne ole kaikissa lämpötiloissa kaasuja, vaan niillä kaikilla on kaikki kolme tavallista olomuotoa. Shirky toteaa myös kirjastoissa käytetyssä luokittelujärjestelmässä, Deweyn luokittelussa (engl. Dewey Decimal Classification), olevan ongelmia, koska se on vuodelta 1876. Shirky osoittaa, että tuonaikainen luokittelu epäonnistui esimerkiksi uskontojen ja maantiedon luokittelussa. Luokittelu on selvästikin tehty sen perusteella, miten paljon kirjallisuutta on ollut saatavilla kyseisistä aihepiireistä. Uskonnoista kristinusko on saanut useita luokkia muiden uskontojen kustannuksella, ja vastaavasti maantiedossa Aasia, Afrikka ja Balkanin niemimaa ovat saaneet kukin oman luokkansa. Kuitenkin Balkanin niemimaa verrattuna Afrikkaan ja Aasiaan on saanut samantasoisien luokan, vaikka Afrikka ja Aasia ovat asukasluvuiltaan ja pinta-aloiltaan huomattavasti suurempia. Samantasoisien luokan saaminen johtui siitä, että 1800-luvun lopulla Balkanin niemimaasta oli kirjallisuutta suurin piirtein saman verran kuin Afrikasta ja Aasiasta. [Shirky, 2005]

Kirjastojen luokitusjärjestelmien kritisointi on ehkä hieman kaukaa haettu esimerkki käytettäväksi silloin, kun kritisoidaan ontologioita. Tietämystä muodostuu lisää kaiken aikaa, sen rakenne muuttuu ja itse tietämys vaihtuu silloin, kun uusi tietämys korvaa vanhentuneen. Kaikenlaisissa tietojärjestelmissä on otettava huomioon tietämyksen kehittyminen. Shirky [2005] sanoo, että ontologia on teoreettinen luokittelu. Tästä näkemyksestä voidaan huomata, että Shirky ei ole sisäistänyt ontologian olemusta, sillä ontologia ei nimenomaan ole luokittelu vaan teoria. Kirjastojen luokitusjärjestelmä tähtää kirjojen fyysiseen sijoitteluun niin, että niitä voidaan tutkia kävelemällä hyllyjen välissä. Ontologiaa voidaan käyttää kirjojen luokittelemisessa, mutta ontologian kaltainen luokittelu on niin monitahoinen, että se ei sovi kirjojen järjestämiseen fyysisesti. Kirja voi olla vain yhdessä paikassa, jolloin on valittava mihin aiheeseen se suurimmaksi osaksi kuuluu. Ontologisessa luokittelussa kirja kuuluu erilaisiin ryhmiin näkökulmasta tai ajasta riippuen.

Kirjoja on luokiteltu sen mukaan, miten paljon tietystä aihealueesta on saatavilla kirjoja, kun taas Internetissä luokittelua eivät rajoita fyysiset esteet kuten kirjahyllyt. Tietokokonaisuudet voidaan luokitella intuitiivisesti sellaisiksi kokonaisuuksiksi kuin nähdään tarpeellisena. Shirky ei tosin kerro, mitä tehdään, kun tietämys toimialueesta muuttuu. Shirky ottaa esimerkikseen Yahoo!-sivuston, joka julkaisi ensimmäisen tunnetun sivustohakemiston. [Shirky, 2005]

Yahoo!-esimerkissään hän esittelee Yahoo!-sivuston käyttämää hakemistoa ja sen rakennetta. Hakemiston rakenne on hierarkkinen puurakenne, jossa on lisäksi linkkejä puun haaroista toisiin. Hakemisto on tarkoitettu aihepiirien selailemiseen siten, että kategoriasta päästään navigoimaan aina kerros tai oksa kerrallaan yhä tarkempaan kategoriaan. Shirky arvostelee juuri kategorian valinnan vaikeutta jollekin uudelle asialle. Hänen mielestään Yahoo!n ei ole oikeutettua päättää, mihin haaraan kunkin asian tulee sijoittua. Hakemiston käyttäjä ei voi navigoida omien näkemystensä pohjalta, vaan hänen on seurattava Yahoo!n näkemystä asioiden sijoittumisesta hakemistoon. Lisäksi hän vastustaa hakemiston haarasta toiseen tehtyjä linkkejä, koska ne ohjaavat ja muokkaavat selailijan käsitystä asioiden rakenteista. Linkki ei ole samanarvoinen kuin hierarkian solmusta oksaa pitkin toiseen solmuun siirryttäessä, vaan linkkiä käytetään lähinnä oikeaan paikkaan ohjaamisessa. Tällaista linkittämistä Shirky pitää arveluttavana. Shirky paneutuu käytännön esimerkkinä kirjojen ja

kirjallisuuden sijoittumiseen Yahoo!n luokittelussa. Kirjat ja kirjallisuus eivät sijaitse viihteessä, vaan viihteestä löytyy linkki humanisiin tieteisiin, jossa kirjat ja kirjallisuus –luokka oikeasti sijaitsee. [Shirky, 2005]

Shirkyn mukaan Yagoon hakemistoa ei pitäisi kutsua ontologiaksi enää siinäkään vaiheessa, kun siihen lisätään viitteitä hierarkian haarasta toiseen. Viittaukset voidaan ymmärtää osoitukseksi epäonnistuneesta luokittelusta. Viittaukset ovat osoitus juuri siitä, mitä seuraa luokittelusta, jossa yritetään mallintaa kohdealuetta useista eri näkökulmista. Kirjat ja kirjallisuus –esimerkki osoittaa, että romaaneja etsivälle kirjat ovat viihdettä, kun taas kirjallisuutta tutkivalle opiskelijalle kirjat ja kirjallisuus liittyvät melkeinpä itsestään selvästi humanisiin tieteisiin. [Shirky, 2005]

Tiedostojärjestelmät ovat hierarkkisia siten, että on olemassa ylätaso ja sillä alihakemistoja, joilla on lisäksi alihakemistoja rajattomasti. Shirky havainnollistaa, kuinka esimerkiksi Yahoo! käyttää rajoitetusti linkkejä hierarkiarakenteen joukossa. Tim Berners-Lee, joka loi World Wide Webin vuonna 1989, osoitti, kuinka linkkejä voi olla paljonkin. Näin päästiinkin tilanteeseen, jossa voitiin luopua kokonaan hierarkiasta, koska linkit riittivät kattamaan sivujen tai tietojen yhdistelyn tarpeen. Shirky vertaa Yahoo!ta Googleen osoittaakseen, että varsinkin Internetissä hierarkkinen rakenne on turhaa, koska linkit riittävät ja ovat helpommin hallittavissa. Hän väittää satunnaisten linkkien riittävän, koska käyttäjät eivät voi tietää, kuinka joku toinen ihminen on järjestänyt asiat, vaan käyttäjän on helpompi hakea sivustoja tai tietoja pelkillä hakusanoilla. [Shirky, 2003; Shirky, 2005]

Clay Shirky kuitenkin myöntää, että ontologioita voi käyttää joissain tapauksissa. Ontologia on Shirkyn mielestä käyttökelpoinen pienten, formaalien ja täsmällisten aineistojen kanssa, joiden oliot ovat pysyviä. Tämän lisäksi ontologian käyttäjienkin pitäisi olla hänen mielestään sellaisia, jotka kykenevät käyttämään ontologiaa. Sopivia henkilöitä Shirkyn mukaan voisivat olla asiantuntijat omalla alallaan tai luokittelussa, sekä käyttäjät, jotka käyttävät ontologiaa keskenään samassa käyttötarkoituksessa. Aineiston muuttuessa suuremmaksi ja epämääräisemmäksi tai käyttäjien muuttuessa naiivimmiksi ontologian käyttö muodostuu mahdottomaksi. Shirky mainitsee tästä esimerkkinä suuret katumaastoautot, jotka eivät ole Yhdysvaltain hallituksen määritelmän mukaan autoja vaan kuorma-autoja. Esimerkissä korostuu

nimeämiskäytännöissä olevan ongelman ydin: joitain asioita nimetään ristiriitaisesti ja siksi asioiden luokittelu on hankalaa tai luokitteluperusteita on vaikea ymmärtää, jos ei ole perehtynyt aihealueeseen. [Shirky, 2005]

Clay Shirky [2005] käyttämässä esimerkissä katumaastoautojen luokittelusta ei mielestäni ole kuitenkaan ongelmaa. Tavalliset ihmiset, jotka tutkivat asiaa ulkonäön ja vaadittavan kuljettajan tutkinnon näkökulmasta, näkevät asian eri lailla, kun verrataan hallituksen näkökulmaan. Hallituksen näkökulmaan vaikuttavat katumaastoauton polttoaineen kulutus, saasteen tuotto ja turvallisuusseikat. Suuri auto on täysin erilainen kuljettaja kuin pieni henkilöauto. Asiaan perehtymättömät maallikot eivät useinkaan tule ajatelleeksi suuren massan vaikutusta hidastuvuuteen jarrutus- tai törmäystilanteissa. Hallituksella on siis mitä luultavimmin hyvin aksiomaattinen tapa luokitella autot henkilöautoihin ja kuorma-autoihin.

Shirky'n kanta on pohjimmiltaan se, että kategorisointi on liian kankeaa, jos kohteena on laaja ja hajanainen kohde, kuten esimerkiksi WWW-sivustot. Hän esittääkin, että URL-linkkejä käytettäisiin yksilöivinä avaimina ja kuka tahansa käyttäjä voisi antaa WWW-sivulle *nimen* (engl. tag), jolloin suuri joukko ihmisiä nimeää sivustoja omien mieltymystensä pohjalta. Shirky osoittaa del.icio.us-sivuston tutkimuksen mukaan, että nimeäminen tapahtuu jollain asteella yhtenäisesti käyttäjämassan kesken. Käyttäjät nimeävät toisistaan tietämättä sivustoja osittain samalla lailla, jolloin voidaan muodostaa kategorioita useimmin käytettyjen nimien mukaan. Toimintamalli on päinvastainen kuin kategorisoinnilla perinteisesti ymmärretään esimerkiksi kirjastoissa. Kirjastot käyttävät yleisiä kategorioita, joista uudelle kirjalle valitaan luokka. Ongelmaksi Shirky'n mukaan valmiissa kategorioissa muodostuu se, että luokiteltava kohde saattaa sopia useampaan kategoriaan tai toisaalta kohde ei sovi oikeastaan mihinkään kategoriaan. [Shirky, 2003; Shirky, 2005]

### **8.1.2. Shirky'n argumenttien kommentointia**

Ontologiat ovat lupaavia tarjotessaan mahdollisuuden linkittää keskenään yhteiset käsitteet toisiinsa, sillä ne eivät ole ainoastaan hierarkkinen puumalli. Useinkaan

tutkimuksissa tai artikkeleissa ei keskitytä ontologian rakenteeseen, vaan lähinnä sen ominaispiirteisiin. Ontologiassa täsmennetään käsitteiden suhteita aksiomilla ja suhteiden ominaisuuksilla, jolloin saadaan ymmärrys käsitteiden yhteydestä toisiinsa sekä käsitys siitä, miksi suhde on olemassa. Ontologiaa ei näin ollen saisi rinnastaa kirjastoissa käytettyihin luokkiin, koska kirjat luokitellaan ammattitaitoon pohjautuvan tulkinnan mukaan. Ontologia on mielestäni enemmän tieteellisesti perusteltu rakenne, jolloin jokainen käsitteen liitos toiseen käsitteeseen sisältää loogisen perustelun sille, miksi käsite sijaitsee juuri tietyssä paikassa toisiin käsitteisiin nähden. Kun kirjastoissa tai WWW-sivustoilla käsitellään suuria kokonaisuuksia, ontologiassa jokainen kirja tai WWW-sivu on suuri joukko käsitteitä. Ontologian lävitse voidaan etsiä esimerkiksi kirjoja, jolloin haku voidaan suorittaa Shirkin [2005] esittämän nimeämismallin mukaisesti käsitteiden nimiä vertailemalla. Kirjassa käytetyt käsitteet tai kirjaan liittyvät käsitteet muodostavat joukon, jota voidaan pitää kirjalle annettuina niminä. Tällöin huomataan, ettei olla kaukana Shirkin nimeämismallista.

Ontologian yhteydessä voidaan hyötyä lisäksi käsitteiden välisten suhteiden tuottamasta merkitysverkosta. Verkkoa voidaan käyttää uusien käsitteiden oppimisessa, jolloin hakua voidaan täsmentää. Ontologian hyöty on tällaisessa tapauksessa mahdollisuus etsiä etsittyjen käsitteiden yhdistelmällä jotain aihepiiriä, joka on hakua tehneelle henkilölle ehkä vielä tuntematon.

Tärkeässä tavoitehakuaisessa työssä ontologia antaa lisäarvoa, mutta kuten Internetissä, jossa kaikki voivat kirjoittaa mitä tahansa, järjestelmällisyydellä ei ole väliä. Tilanteessa, jossa etsittäviä asioita on vain vähän, ontologia tuo hyötynsä esiin, koska ontologian rakenteessa voidaan helpommin selailta yhteyksiä. Ontologia ei kuitenkaan ole vain erittäin hyvin määritelty kortisto, josta etsitään asioita, vaan ontologia auttaa sellaisissakin hakutilanteissa, joissa käyttäjä ei varsinaisesti tiedä, millä hakusanalla asiaa pitäisi etsiä, mutta tietää mihin asiakokonaisuuksiin asia liittyy [ks. Vallet *et al.*, 2005]. Ontologia tuo myös lisää tietämystä aiheesta, josta käyttäjällä ei ole vielä tietoa riittävästi. Kokonaan uuden asian oppimisessa itse vapaasti linkittäminen tai asioiden nimilaputtaminen ei tuo asiaan perehtymättömälle uutta tietoa. Perusteellinen ryhmittely ja asioiden välisten suhteiden ominaisuudet sen sijaan tuottavat lisää ymmärrystä uudesta aihepiiristä. Ontologiaa voidaan pitää eräänlaisena



käyttöliittymänä, joka selittää toimialueen rakennetta ja jonka avulla voidaan suurta tietomassa tarkastella samaan tapaan kuin kirjan hakemistorakennetta.

Omassa työssäni järjestelmien integroinnissa olen huomannut, kuinka ontologian käyttäminen erilaisten järjestelmien integroinnissa voisi olla hyödyksi. Työssäni Etuovi.comissa ylläpidän järjestelmiä, joiden avulla kiinteistönvälittäjät voivat esitellä myynnissä tai vuokrattavissa olevia kiinteistöjä tai huoneistoja. Työhöni kuuluu muun muassa järjestelmien välisten aineistosiirtojen luominen ja ylläpito. Kiinteistöjen tiedot lähetetään palvelimellemme erityyppisissä tiedostomuodoissa, jonka jälkeen jokaista toisistaan eroavaa tiedostomuotoa varten luotua ohjelmaa käytetään siirtotiedostojen muuntamiseen oman järjestelmämme ymmärtämään muotoon. Haasteelliseksi työstä tekee se, että saamme joiltakin välittäjiltä kohdetietoja, joiden tiedot ovat puutteellisia. Tällaisten kohteiden kohdalla ohjelma tekee oletuksia muiden tietojen pohjalta ja joskus puuttuva tieto korvataan oletustiedolla, jonka toivotaan olevan oikein suurimmassa osassa tapauksia. Kansainvälistyminen tuo lisäksi kansallisten toimintatapojen välisten erojen mukanaan tuomia ongelmia:

Asiakkaanamme on Etelä-Euroopassa lomakiinteistöjä välittävä yritys. Tehtyämme XSL-muunnoksen [ks. <http://www.w3.org/TR/xslt20/>] heiltä saamamme XML-tiedoston [ks. <http://www.w3.org/XML/>] ja oman aineistosiirto-XML-tiedoston välille jouduimme muunnoksessa muuntamaan myös eteläeurooppalaisessa kiinteistönvälityksessä käytettyjen käsitteiden määritelmät suomalaiseen kiinteistönvälitysjärjestelmään sopiviksi. Suomessa loma-asunnot ovat huoneistoja silloin, kun kyseessä on yhtenäinen rakennus, jossa on useita huoneistoja, joihin johtaa oma ulko-ovi. Kiinteistö on sen sijaan sellainen asumus, jolla on oma tontti ja jota ei voida jakaa useamman toisistaan eriävän ryhmän samanaikaiseen käyttöön. Espanjassa vastaava jaottelu on täysin erilainen. Siellä lähes kaikki myytävänä olevat loma-asunnot ovat kiinteistöjä, vaikka ne sopisivatkin edellä esiteltyyn huoneiston määritelmään. Etuovi.com-palvelussa oletetaan, että käyttäjät ovat suomalaisia ja ajattelevat suomalaisten käytäntöjen mukaisesti, jolloin kun he tekevät hakuja ulkomaalaisiin kohteisiin, he osaavat tehdä haut oletetun mallin mukaisesti. Esimerkiksi omaa rauhaa arvostavat hakevat kiinteistöjä, kun taas halvempaa loma-asuntoa etsiville riittää huoneiston tyylinen asunto.

Tällaiset muunnokset olisivat automaattisia, jos apuna käytettäisiin ontologiaa. Kuten useissa ontologian määritelmässä todetaan, ontologia sitoutuu kieleen. Monissa teksteissä todetaan myös, että on lähes mahdotonta tehdä ontologiaa, joka on täysin kielestä riippumaton. [ks. esim. Guarino, 1998] Kansallisuuksien välillä olevat kielelliset erot pitäisi mallintaa ontologiaan siten, että ontologioilla olisi jokaista kieltä vastaava kieliversio, jotka taas liitettäisiin toisiinsa suhteilla, jolloin käsitteet kuvautuisivat kielistä toisiin erilaisiksi rakenteiksi. Saman kielen sisällä esiintyy lisäksi kohdealueista riippuvaisia termistöjä, joten kieli yksistään ei riitä erottamaan ontologioita toisistaan: Kiinteistökaupassa ja autokaupassa käytetään eri termistöjä, jotka molemmat ovat suomenkielisiä ja liittyvät kaupankäyntiin. Näin ollen siis ontologiaa voidaan käyttää vain hyvin tarkasti määritellyssä kohdealueessa, jonka myös Shirky myöntää.

## 8.2. Boris Wyssusekin kritiikki

### 8.2.1. Näkemykset

Kuten Shirky, on myös tiedeyhteisö kritisoinut ontologioiden käyttökelpoisuutta. Ontologiaa on tutkittu pitkään filosofiassa, ja filosofian ontologiasta on muodostettu jokseenkin yhtenäinen teoriapohja, jonka ansiot on huomattu myös tietojärjestelmätieteiden piirissä. Ontologia onkin otettu käyttöön tietojärjestelmätieteissä juuri sen hyvin määritellyn tiedepohjansa vuoksi. Kuitenkin filosofian ja tietojärjestelmätieteiden soveltamisalueiden huomattavat erot ovat muodostaneet tietojärjestelmätieteissä rajoitteita ja aiheuttaneet kiistoja ontologian alkuperäisen teorian tulkitsemisesta. Seuraavaksi tutkimuksessa esittelen Boris Wyssusekin julkaisussaan *Ontology and Ontologies in Information Systems Analysis and Design: A Critique* [2004] esiin tuomaa kritiikkiä ontologioiden käyttökelpoisuudesta.

a. *“Information systems researchers are seldom philosophers.”* (Tietojärjestelmien tutkijat ovat harvoin filosofeja.) Kiinnostus ontologiaan johtuu pitkälti ontologian vankasta ja kurinalaisesta filosofisesta pohjasta sekä sen hyvästä maineesta

tiedeyhteisöissä. Kuitenkin kaikkien tietojärjestelmätieteiden asiantuntijoiden pitäisi olla riittävästi perehtyneitä filosofisiin teorioihin ja filosofian historiaan pystyäkseen täysin hyödyntämään alkuperältään filosofispohjaista ontologiaa, jossa tarkat määritelmät ovat harvinaisia ja termien ymmärtäminen vaatii käyttäjältään ymmärryksen myös termien taustateorioista. Näiden filosofian teorioiden käyttö johtaa väistämättä myös kauaskantoisiin seuraamuksiin kuten Wyssusek havainnollistaa Mario Bungen ontologiateorian avulla. Kuten aiemmin on todettu, Bungen teoria pohjautuu hyvin konkreettisiin käsitteisiin, joiden oletetaan olevan olemassa eli toisin sanoen realistiseen näkemykseen maailmasta. Tällainen maailmankuva rajoittaaakin teorian hyödyntämistä tietojärjestelmissä, joissa tarkastellaan esimerkiksi sosiologiaan liittyviä tai muuten abstrakteja ongelmia. Kuitenkaan näiden ongelmien valossa tietojärjestelmiä ei pitäisi suunnitella ainoastaan teknisestä tai käytännönläheisestä näkökulmasta, vaan huomioiden yksilölliset kognitiiviset rajoitteet. [Wyssusek, 2004] Koska filosofiset teoriat ovat vanhakantaisia, niitä ei pitäisi käyttää nykyaikaisissa tietojärjestelmäteorioissa suoraan, vaan ne pitäisi kyetä uudistamaan huomioiden nykyaikaiset vaatimukset ja muuttuneet lähtökohdat. Näin voitaisiin mahdollisesti välttää vanhojen teorioiden tuomat rajoitteet tietojärjestelmissä.

b. *“If ontology is everything, maybe it is nothing?”* (Jos ontologia on kaikkea, ehkä se ei ole mitään?) Viimeisen kahdenkymmenen vuoden aikana termi 'ontologia' on saanut osakseen suuren määrän huomiota, joka kuitenkin ei ole kohdistunut täsmälleen samaan asiakokonaisuuteen. Termiä 'ontologia' on käytetty kirjallisuudessa huomattavasti toisistaan eriävissä tarkoituksissa. Ontologian sanakirjamääritelmä viittaa filosofiseen tutkimukseen olevasta tai olemassa olemisesta. Perinteinen ontologian filosofinen tutkimus kohdistuu olemassa olemisen selvittämiseen ja siihen, mitä vaaditaan jonkin olemassa olemiseen. Tähän tutkimukseen joudutaan ottamaan mukaan kiista siitä, miten kukin filosofian koulukunnista tulkitsee olemassa olemisen. Realistit väittävät kohteen olevan olemassa riippumattomana ihmisen havaintokyvystä, kun taas idealistit väittävät täysin eriävästi kohteen olevan olemassa vain ihmisen havainnoista riippuvana. [Wyssusek, 2004] Kiista ei ole pelkästään tiedeyhteisössä käytävä pohdinta, vaan perimmäinen ongelma tulee esiin tietojärjestelmien suunnittelussa silloin, kun mallinnetaan uskomuksia eikä ainoastaan fyysisiä todellisen maailman kappaleita. Realismin ja idealismin välinen kamppailu tapahtuu jopa tietojärjestelmän suunnittelijan päässä, kun tämä yrittää päättää, mitä käsitteitä tietojärjestelmään

mallinnetaan ja mitkä käsitteet jätetään järjestelmästä pois, eli toisin sanoen mitkä käsitteet eivät ole olemassa järjestelmässä.

Yleisesti siis oletetaan, että ontologia pystyy kuvaamaan formaalilla tavalla käsitteet ja suhteet kaikesta mahdollisesta olevasta. Tietojärjestelmästä pyritään suunnitteluvaiheessa muodostamaan mahdollisimman mukautuva, jolloin järjestelmä ei itsessään aseta rajoitteita. Lisäksi siitä yritetään tehdä kaiken kattava kokonaisuus, jolloin kärjistetyksi voidaan jopa sanoa ontologian olevan kaikkea.

Ontologioiden yleisesti tunnustetusta kattavuudesta huolimatta Wysseck [2004] kyseenalaistaa ontologian T. Gruberin [1993b] ja W. V. O. Quinen [1961] ontologiapohdintojen perusteella tekemänsä päättelyketjun avulla. Wysseckin mukaan ontologian ollessa kuvaus jonkin toimijan tai toimijaryhmän käsitteistä ja suhteista se on myös kuvaus, jota käytetään tekemään ontologisia sitoumuksia. Ontologisella sitoumuksella tarkoitetaan sitoutumista käyttämään jotain ennalta määriteltyä mallia eli ontologiaa. Käytännössä ontologiset sitoumukset ovat sopimuksia käyttää jotain sanastoa, joka muodostuu jonkin puhutun kielen käyttämistä sanoista. Tältä pohjalta Wysseck vetää johtopäätöksen, että ontologia on lingvistinen konventio puhutun kielen sanaston käyttämisestä, eikä se itse asiassa kerro mitään ympäröivästä maailmasta. [Wysseck, 2004.] Wysseckin päättelyketjun mukaan ontologia kertookin lähinnä kielen rakenteesta.

c. *“We do not speak about ‘biologies’ just because different scientists have a different understanding of biology.”* (Emme puhu ’biologioista’ vain siitä syystä, että eri tutkijat ymmärtävät biologian eri tavalla.) Klassisessa filosofiassa ontologiasta ei esiinny monikollista muotoa ja Wysseck [2004] havainnollistaa sitä biologialla, joita ei ole olemassa monia vain siksi, että eri tutkijat ymmärtävät sanan ’biologia’ eri tavalla. Ontologia onkin tiukasti sitoutunut kieleen. Useiden tutkijoiden mielestä eri kieltä käyttävät ihmiset muodostavat maailmankuvansa eri lailla. Lisäksi Wysseck [2004] tuo ilmi, että jotkut tutkijat kuitenkin ottavat huomioon, että heidän muodostamansa ontologiat eroavat filosofian määrittelemästä ontologiasta, jolloin uusi määritelmä on ’käsitteellistämisen määrittely’. Käsitteellistämisen määrittely tarkoittaa lopulta sitä, että ontologia on kieleen pohjautuva sopimus. Kielessä on kuitenkin olemassa synonyymejä, jotka tarkoittavat samaa, vaikka kirjoitusasu olisikin eriävä. Toisaalta

sama kirjoitusasu voi myös tarkoittaa kahta eri käsitettä riippuen asiayhteydestä, jolloin puhutaan homonyymeistä. Näin ollen ontologian keinotekoinen toimija ei voi tulkita asiasisältöä vaan toimijan on tukeuduttava kielelliseen semantiikkaan. Käsitteet pysyvät siis täydellisinä ainoastaan ihmisen mielessä, eikä niitä ole mahdollista täydellisesti mallintaa. [Wyssusek, 2004]

Vaikka ontologioita on useita ja niissä käytetty kieli muodostaa rajoitteita niiden yhteiselle käytölle, ei se mielestäni estä ontologioiden käyttöä. Voidaan todeta, että ongelma on myös ihmisten välillä. Ihmisten on välillä vaikea ymmärtää toisiaan, koska he käyttävät samaa kieltä hieman eri lailla ja siksi ymmärtämisessä voi esiintyä ongelmia. Ontologia voisi olla tilaisuus käyttää kieltä ennalta määritellyllä tavalla, jolloin tulkinnalle jäisi mahdollisimman vähän tilaa.

### **8.2.2. Ontologiakeskustelua**

Edellä esitellyn kritiikin lisäksi Wyssusek on myös muulloin kyseenalaistanut ontologioita. Toisessa artikkelissaan Boris Wyssusek [2006] kritisoi yleensä ontologioita ja erityisesti Bunge-Wand-Weber (BWW) -ontologiaa siitä, että ontologioita ei ole tarkasteltu riittävän huolellisesti eikä niitä ole aiemmin kritisoitu.

Tietokantasuunnittelijat osaavat muodostaa tietokantamalleja ja siirtää kahden erilaisen tietokantamallin välillä tietoja, mutta tällaisiin tehtäviin tarvitaan aina tietokantoihin perehtynyt henkilö. Tarvittaisiin siis mallinnustapa, jolla voitaisiin kuvata tietokantamallit keskenään yhtenäisellä tavalla. Tietokantoja suunnittelevilla tutkijoilla on tietokantarakenteesta määräytyvä malli, jonka pohjalta tutkijat haluaisivat muodostaa käsitteellisen mallin, mutta teorioita käsitteellisistä malleista ei ole riittävästi. Wyssusekin [2006] mielestä tällaisia käsitteellisiä malleja on muodostettu ilman riittävästi kritiikkiä ja tieteellistä pohjaa. Yksi tällainen pyrkimys luoda malli käsitteellistämiseen on paljon keskustelua herättänyt BWW-ontologia.

BWW-ontologia pohjautuu Mario Bungen ontologiaan. Bunge julkaisi filosofiasta kahdeksan teoksen sarjan, joista ontologiaa käsittelee kaksi 1970-luvun loppupuolella

kirjoitettua osaa. Wyssusek tuo esiin, että Bungen ajatukset noudattavat dialektista materialismia. Materialisti selittää maailmansa materiaalilla ja konkreettisilla tapahtumilla, jotka on kyettävä todistamaan esimerkiksi matemaattisilla yhtälöillä. ”Objektiivinen tieto on mahdollista, jos se perustuu tieteellisiin metodeihin, koska tällöin se edustaa todellista maailmaa. Tämän vuoksi totuus on mahdollista, mutta ainoastaan tieteen nimissä, sillä tieteellinen metodi on ainoa tapa saavuttaa totuus.” [Wyssusek, 2006] Wyssusek korostaakin Bungen olevan äärimmäisen materialistinen ajatuksissaan eikä siksi anna tilaa esimerkiksi subjektivismille, jonka mukaan todellisuuden olemassaolo riippuu tajunnasta.

Wyssusek [2006] toteaa siteeraamalla Bungea, että ontologia käsittelee ainoastaan konkreettisia, olemassa olevia asioita. Kuitenkin ihmisen kognitioiden tutkiminen esimerkiksi käytettävyyden tai tekoälyn kannalta voidaan mieltää vähemmän materialistisena tieteenä ja näin ollen abstrakteja asioita tutkivana. Kognitiotieteissä ontologiaa voidaankin käyttää kognitioiden rakenteiden mallintamiseen, toisin kuin Bunge kuvaili ontologian käyttökohteita. Lisäksi Guarino ja Guizzardi [2006] ovat huomanneet, että Bungen tarkoittamat *konkreettiset asiat* (engl. things) ovat joukko asioita, jotka ovat osa maailmaa, mutta kuitenkin Bungen käyttämät universaalit taas eivät ole osa maailmaa. Myös Bunge käyttää erityisesti esimerkeissään ja teorioissaan käsitettä ’laki’, joka ei ole konkreettinen asia, mikä puoltaa myös ontologioiden käyttämistä abstraktien asioiden kuvaamiseen. Guarinon ja Guizzardin päätelmistä voidaan todeta, että ontologia ei voi koostua vain konkreettisista asioista, ja vastaavanlaiseen päätelmään päätyivät myös Wand ja Weber [2006] todetessaan, että Bungen ontologialla voidaan mallintaa vaikutelmia. Jokainen ihminen käsittää todellisen maailman omien ”tosiuskomustensa” pohjalta. Näiden havaintojen pohjalta vaikuttaakin siltä, että Wyssusek on tulkinnut Bungea hieman liian sanatarkasti.

Wyssusekin [2006] tulkintaa puoltaa toisaalta se, että konkreettisia kappaleita ja abstraktisia asioita on vaikea erottaa toisistaan. Esimerkkinä tästä ovat kappaleiden pienimmät osat: pilkottaessa kappaletta yhä pienemmäksi päädytään osiin, jotka eivät ole muuta kuin teoreettisia malleja, joiden uskotaan olevan totta. Toisena esimerkkinä voidaan kuvitella kiinteistönvälityksestä tunnettu käsite ’raja’. Rajapyykki on konkreettinen joko kivinen tai metallinen kappale maastossa, jonka sijainti on tunnettu.

Rajapyykkien välissä oleva rajaviiva taas on maastossa täysin abstrakti, mutta toisaalta kartalla konkreettinen kahden pisteen välille piirretty viiva.

Guarino ja Guizzardi [2006] osoittavat, kuinka heidän mielestään ontologialla voidaan mallintaa myös käsitteitä eikä vain konkreettisia asioita. Wussysek [2006] väittää, että Bungen ontologia noudattaa pelkästään materialistisia oppeja. Kuitenkin ontologia voi noudattaa deskriptiivistä tai revisionääristä metafysiikkaa. Deskriptiivinen metafysiikka pyrkii käsittelemään ihmisten luonnollista kieltä ja luonnollisia käsityksiä todellisuudesta sekä muodostamaan niistä täsmällisiä. Inhimillisestä tiedosta johtuen kategoriat saattavat olla jonkin verran riippuvaisia ihmisen havaintokyvystä ja kulttuurillisista eroista. Kuitenkin näiden kategorioiden olemassaolon voi todistaa kognitiivisiin tieteisiin pohjaavalla tutkimuksella. Revisionäärinen metafysiikka päinvastoin tavoittelee havaintojen tekijöistä itsenäistä totuutta maailman luonteesta. Guarinon ja Guizzardin mukaan näiden kahden voidaan olettaa olevan yhtenäiset, jos rakeisuuden erot otetaan huomioon.

Epäkohta BWW-ontologiassa, jonka Wyssusek [2006] tuo esiin, on Wandin ja Weberin tapa valita Bungen ontologia. Wand ja Weber valitsivat Wyssusekin mielestä Bungen ontologian perusteettomasti vertailematta muita tarjolla olevia ontologioita. Wandin ja Weberin [2006] mielestä ontologian rakenteet ja termit muistuttivat heidän mielestään tietojärjestelmissä käytettyjä rakenteita ja termejä. Lisäksi Wand ja Weber mainitsevat valintansa perusteeksi sen, että Bungen ontologia on paras mahdollinen ontologia, jonka he löysivät. Tällainen toteamus ei Wyssusekin mielestä ole millään lailla tieteellinen peruste valita filosofista teoriaa ontologian pohjaksi. Kuitenkaan Wyssusek ei artikkelissaan tuo yhtään uutta filosofista teoriaa tilalle, kuten Guarino ja Guizzardi [2006] toteavatkin.

### **8.2.3. Bunge-Wand-Weber -ontologian rakenteesta**

Bunge-Wand-Weber -ontologian nimi muodostuu sen kehittämien keskeisimpien tutkijoiden nimistä. Teoreettisena pohjana ontologialle on Bungen 1970-luvulla kirjoittamat kirjat, joita Wand on tutkinut jo vuodesta 1997 lähtien [ks. Parsons and

Wand, 1997]. Bungen teorian rakenteiden osien termejä ovat *asia* (engl. thing), *asian ominaisuus* (engl. property of a thing), *laki* (engl. law), *tila* (engl. state), *tapahtuma* (engl. event), *kytkentä* (engl. coupling), *ysteemi* (engl. system) ja *taso* (engl. level). Rakenteet ovat aiheuttaneet paljon keskustelua Wandin ja Weberin otettua ne käyttöönsä omassa ontologiassaan. Wyssusek [2006] väittää, että Wand ja Weber olisivat ottaneet rakenteet käyttöönsä, mutta jättäneet sitoutumatta ontologisesti Bungen teorioihin. Wyssusekin mielestä rakenteet eivät ole tästä syystä enää ontologisia, vaan rakenteiden ja termien käytöstä ei ole enää muuta hyötyä kuin kielellinen kokonaisuus ilman tarkoitettua sisältöä. Wand ja Weber [2006] puolustautuvat Wyssusekin hyökkäystä vastaan myöntämällä ottaneensa käyttöön termit ja rakenteet, mutta väittävät myös käyttävänsä niitä niiden oikeissa merkityksissä, jotka Bunge on niille tarkoittanut. Wand ja Weber ottivat käyttöön Bungen ontologiasta kuitenkin omiin tarkoituksiinsa sopivia rakenteita. Näiden rakenteiden filosofinen kanta ei ole tärkein, vaan tärkeintä on ontologinen kanta eli se, millaista ontologiaa Wand ja Weber ovat muodostamassa. Wand ja Weber myöntävät, että Bungekin on sanonut rakenteiden muuttavan merkitystään silloin, kun ne siirretään toiseen teoriaan, mutta samalla puolustautuvat sanoen, että tieteen tekeminen on pääasiassa vanhojen teorioiden parantelua ja toisten teorioiden osien liittämistä omaan uuteen teoriaan. Lisäksi Wand ja Weber ovat huomanneet, että myös Bunge myöntää, että rakenteet eivät muutu merkityksettömiksi, jos rakenteita käytetään kontekstin ulkopuolella.

Wyssusek vastustaa myös Wandin ja Weberin tapaa valita kategoriat ontologiaansa perusteena vain omakohtaiset kokemukset siitä, mitkä kategorioista sopisivat tietojärjestelmätieteisiin. Wyssusek [2006] täsmentää vielä, että Wand ja Weber sovelsivat valitsemiensa kategorioiden semantiikan omiin tarpeisiinsa sopivalla tavalla, ilman että heidän tarvitsi sitoutua Bungen ontologiaan kokonaisuudessaan. Tällaiset valintakriteerit vievät selvästi tieteellisen perustan Wandin ja Weberin ontologialta, koska Wand ja Weber eivät osoita tutkimustuloksia valintojensa tueksi vaan myöntävät, että valinnat on tehty vain ammatilliseen kokemukseen pohjautuvalla tuntemuksella.

Wyssusek [2006] arvelee, että Wand ja Weber ovat sekoittaneet ontologian termit käsitteellisen mallintamisen käsitteisiin. Wyssusek lainaa Bungea todistaakseen, että Wand ja Weber ovat ymmärtäneet väärin ontologioiden perimmäisen tarkoituksen käyttäessään käsitteitä ontologioiden tarkasteluun. Käsitteellisessä mallintamisessa on



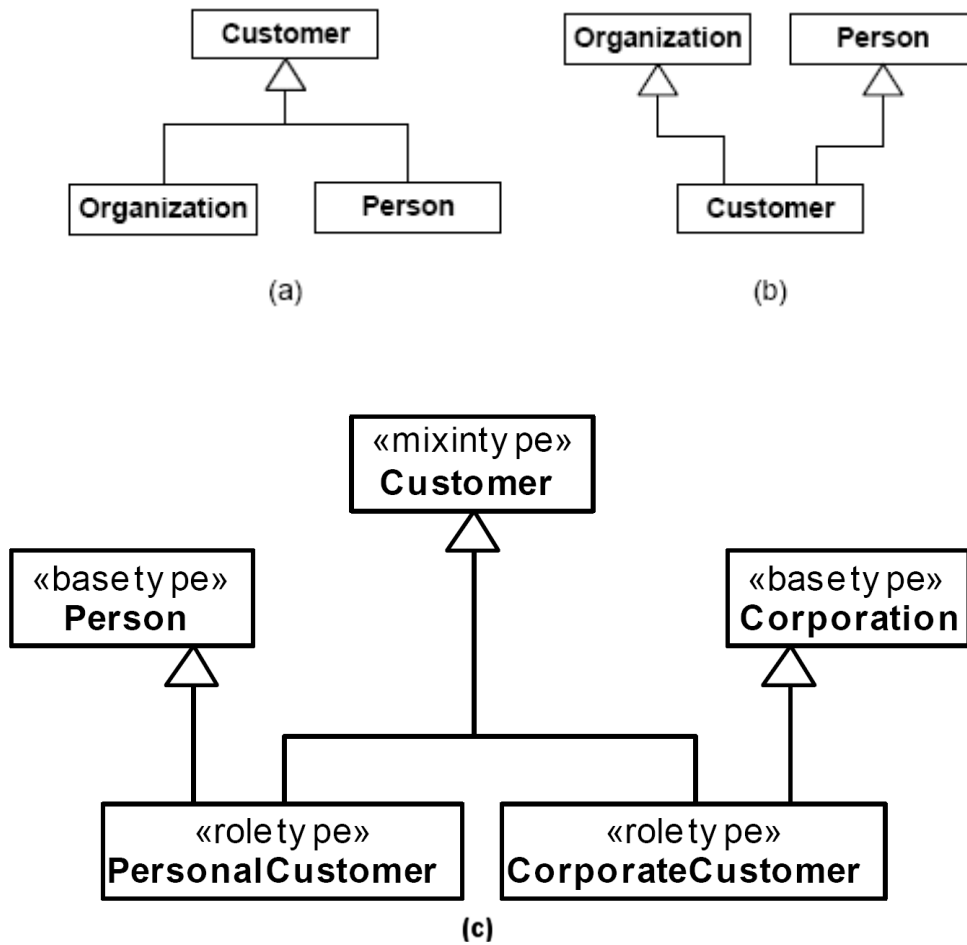
kyse ihmisen tietämyksen rakenteesta ja ontologioissa taas on kyse konkreettisista objekteista, eivätkä käsitteet ole konkreettisia objekteja, tulkitsee Wyssusek Bungea. Guarino ja Guizzardi [2006] väittävät tätä vastaan sanoen, että ”ontologisia kysymyksiä ei voi välttää ihmisten tavassa rakentaa tietämystään”. Guarino ja Guizzardi antavatkin useita esimerkkejä käsitteellisen mallintajan mahdollisista ontologisista kysymyksistä:

- Ovatko kaikki objektit yksilöitävissä?
- Sisältyykö jokin tyyppi useampaan ylätyyppiin?
- Onko olemassa ominaisuuden ominaisuutta?
- Onko jokin sisältymissuhde transitiiivinen?
- Onko yhteisö muuta kuin jäsentensä joukko?

Yksilöitävyyden ja lukumäärien ongelmaa ovat havainnollistaneet Wieringa ja de Jonge [1993] esimerkissään bussimatkustajien lukumäärän laskemisesta. Esimerkissä tutkitaan viikon aikana matkustaneiden matkustajien laskemista ja sitä, kuinka lukumäärä muuttuu näkökulmasta riippuen. Kun lasketaan matkustajia, lukumääräksi saadaan 4000 kappaletta, mutta kun lasketaan henkilöitä, niin saadaankin vain 1000 kappaletta. Tämä tietenkin johtuu siitä, että useat henkilöt matkustavat esimerkiksi työmatkansa linja-autolla, jolloin heidät lasketaan useina matkustajina viikossa. Wieringa ja de Jonge esittelevätkin objektiluokan rinnalle rooliluokan, jolla voidaan ratkaista bussimatkustajien ongelma. Henkilö on objektiluokka, jolloin kukin todellinen henkilö on henkilöluokan esiintymä ja matkustajista tulee rooliluokka, jolloin esimerkiksi opiskelija ja matkustaja ovat henkilöluokan esiintymän tiloja. Näin ollen Kimmoniminen henkilö voi saada tiloihinsa opiskelija- ja bussimatkustajaroolit tiettyinä ajanjaksoina.

Esimerkkinä käsitteellisen mallintamisen ongelmaan useammasta ylätyypistä käytetään usein yrityksen kahdenlaisia asiakkaita. Oletetaan, että yrityksellä on yksityisasiakkaita eli henkilöitä ja yritysasiakkaita eli yrityksiä. Tällaiset rakenteet saatettaisiin mallintaa useimmiten kuten UML/MOF [ks. <http://www.omg.org/mof/>] -muotoisen kuvan 3 kaavioissa (a) ja (b). Kuitenkaan henkilöt tai yritykset eivät voi olla asiakkaan alityyppejä, joten ’henkilöstä’ muodostetaan alityyppi ’yksityisasiakas’ ja yrityksestä alityyppi ’yritysasiakas’, kuten kaaviossa (c) on esitetty. Henkilö ei voi olla asiakas, koska on olemassa muitakin henkilöitä, jotka eivät ole asiakkaita. Yrityksien kohdalla

pätee myös sama mallinnussääntö. [Guizzardi ja Wagner, 2004] Asiakkaan olemuksen ongelma on selkeästi ontologinen, sillä samojen käsitteiden mallinnusmuodosta riippuen voidaan todellisuus ymmärtää täysin eri tavoilla.



Kuva 3: Käsitteellisen mallintamisen ongelma useammasta ylätyypistä [Guizzardi ja Wagner, 2004]

Käsitteellisessä mallintamisessa kuvataan usein käsitteiden ominaisuuksia ja sisältymissuhteita samaan tapaan kuin ontologiassa. Ihmiset muodostavat luontaisesti suuren osan tietämyksestään kokonaisuuksista, jotka sisältävät pienempiä kokonaisuuksia.

## 9. Ontologioiden käyttökohteita

Huolimatta ontologioiden saamasta kritiikistä ja kysymyksestä, onko niiden käyttö todella perusteltua vai ei, on ontologioita nykyisin käytössä jo paljon ja niitä hyödynnetään laaja-alaisesti. Ontologioiden käyttötapoja on useita, joista tärkein on tiedon jäsentäminen konelukaiseksi ja hyvin järjestetyksi. Ontologiaa voidaan myös käyttää kommunikointiin, jolloin ontologia määrittelee käytettävät käsitteet ja niiden asiayhteyden. Kommunikointi voi tapahtua tietojärjestelmien välillä, ihmisten välillä tai tietojärjestelmien ja ihmisen välillä. Tietojärjestelmien välillä ontologiaa voidaan käyttää erilaisten tietorakenteiden välillä muuntamaan tieto rakenteesta toiseen automaattisesti niin, että semanttinen merkitys säilyy muunnoksessa.

Hakukoneiden tehostaminen on eräs mielenkiintoinen osa-alue, jossa ontologiaa käytetään. Ontologiaa hyväksikäyttäen voidaan löytää puutteellisilla hakumäärittelyillä oikeita dokumentteja. Hakuehdot voidaan hakukoneessa täydentää ontologian semantiikan mukaisiksi asiayhteyden arvauksien pohjalta vastaamaan hakuehtoja joilla etsitty dokumentti on mahdollista löytää.

Elektronisessa kaupankäynnissä puolestaan ontologiaa voidaan käyttää myytävien tuotteiden aineistoliikenteen selventämiseen. Tuotteet voidaan kuvata sellaisiksi, että niitä voidaan siirtää usean järjestelmän kesken automaattisesti. Tässäkin tapauksessa aineiston läpikäyminen onnistuu agenttiohjelmilla automaattisesti.

Asunto- ja kiinteistökaupan nettipalvelussa Etuovi.com:issa on esillä noin 50 000 kohdetta, joista suurin osa lähetetään erilaisista välittäjien välitysjärjestelmistä Etuovi.com -palvelun tietojärjestelmään. Aineistosiirtojärjestelmä vastaanottaa ja muuntaa aineistot palvelussa käytettyyn aineistomuotoon ja ajaa aineiston palveluun, niin että kuluttajat voivat tarkastella kohteita web-sivuilta. Järjestelmässä on ollut pitkään ongelmana aineistojen semanttisten erojen hallinnan vaikeus. Tiettyjen käsitteiden kohdalla kuluttaja joutuu arvaamaan merkintöjen tarkoituksen ja kysymään kohdetta välittäväältä yritykseltä täsmentäviä tietoja kohteesta. Sekaannuksien lisäksi ongelmia tuottaa myös täysin uudenlaisen aineiston vastaanottaminen, koska jokainen

uusi aineistomuunnos tehdään alusta alkaen manuaalisesti. Muunnoksen tekijä tuntee hyvin palvelun aineistomuodon ja sen semanttisen sisällön, mutta uuden aineiston kohdalla lähes kaikki perustuu arvaukseen. Varsinkin ulkomaisten kohteiden kohdalla kulttuuri- ja kielierot aiheuttavat ylitsepääsemättömiä ongelmia käsitteiden määrittelyiden puuttuessa. Toisaalta suomalaisessa kiinteistönvälityksessä ei käytetä yhtä kirjavaa asuntojen luokittelua kuin ulkomailla. Luokittelun laajentaminen palvelussa ei ole suuresta työmäärästä tai käytettävyyden rappeutumisesta johtuen mahdollista. Näin ollen palvelun luokittelusta eroavat luokittelut joudutaan sulauttamaan palveluun niin, että kohteita voidaan löytää palvelusta olemassa olevan luokittelun pohjalta. Tässäkin tapauksessa aineistomuunnoksen tekijä päättää ammattitaitoonsa tukeutuen, miten vieraat luokitukset muuntuvat palveluun. Kohteiden välittäjät ja kuluttajat osaltaan pitävät palvelun laatua yllä ilmoittaessaan palvelussa oletetun kohteen puuttumisesta tai siitä, ettei kohde löydy riittävän helposti. Tällaisessa tapauksessa kohteiden luokittelumuunnosta voidaan muuttaa. Lähimpänä ontologiaa järjestelmästä löytyy muistilistan tyyppisesti tehty aineistodokumentaatio, jossa kerrotaan hyvin summittaisesti kohdetietojen semantiikka. Koko järjestelmän laajuinen ontologia voisi auttaa myös edellä kuvailtua kohdeaineistoliikennettä huomattavasti automatisoidumpaan suuntaan.

Asunto- ja kiinteistökaupan laajuinen ontologia voisi auttaa ihmisiä asuntokaupan eri vaiheissa ymmärtämään täsmällisesti asuntokauppaan liittyvää termistöä ja lainalaisuuksia. Tietenkin välittäjän tulisi toimia edelleen kiinteistökaupassa henkilönä, joka selittää sopimusteknisiä ja lainsäädännöllisiä termejä, jos kaupan osapuolten tietotaso sitä vaatii. Tavallisesti kuluttajilla ei ole aikaa perehtyä syvällisesti aihealueeseen, joten termien selitykset lyhyillä esimerkkilauseilla voisi olla halutumpi tapa laajentaa ymmärrystä aihepiiristä. WordNet on hyvä esimerkki tällaisesta apuvälineestä.

## 9.1. WordNet

WordNet [ks. <http://wordnet.princeton.edu/>] on englanninkielinen synonyymisanakirja, jonka tietokannassa on 150 000 yksilöllistä sanaa, joille löytyy selkokieliset selitykset.

Sanakirjalle otettiin mallia ihmisen leksikaalisen muistin psykolingvistisistä teorioista. Sanasto on ryhmitelty substantiiveihin, verbeihin, adjektiiveihin ja adverbeihin, jotka ovat järjestetty käsitteittäin sekä niiden semanttisten suhteiden perusteella. Seuraavaksi esittelen sanaston järjestämisessä käytettyjä semanttisia suhteita:

Antonymia: Kielellisen merkityksen vertailua sisältävä vastakkaisuus eli vastakohtasana

Hyponymia: Merkityssuhde, jossa jonkin sanan merkitys sisältyy kokonaan toisen sanan merkitykseen

Meronymia: Merkityssuhde osien ja kokonaisuuden välillä

Synonymia: Samamerkityksisyys

Morfologinen suhde: Suhde, jolla vähennetään sanojen kirjoitusasuja eli löydetään sanan vartalo

WordNetin suosio johtuu osaksi siitä, että sitä voi käyttää ilmaiseksi Internetissä. Lisäksi se erottuu pelkästä sanakirjasta semanttisten suhteittensa vuoksi sekä käsitteidensä suuren määrän vuoksi. Kuitenkaan WordNetissä käsitteiden semanttiseen määrittelyyn ei ole käytetty formaaleja suhteita, vaan käsitteiden semantiikka määräytyy ainoastaan luonnollisen kielen termein. Formaalien määritysten puuttuessa WordNet ei ole konelukuinen, vaan se onkin pääosin lingvistiikkaan keskittynyt. Tämän vapaan rakenteensa vuoksi voidaan kiistellä siitä, onko WordNet niinkään ontologia.

## 9.2. CYC

Toisin kuin WordNet, CYC [ks. <http://www.cyc.com/>] on formaali tietämuskanta, jonka taustalla on ontologia. CYC:n kehittämisen lähtökohtana oli muokata *terveen järjen*

*tieto* (engl. common-sense knowledge) tietokoneella käsiteltävään muotoon. Tarvetta tällaiselle sovellukselle oli huomattu olevan lähes kaikilla tekoälyn sovellusalueilla. Ihminen kykenee terveellä järjellä päättämään, mitä havainnoista pitää oppia. Tätä puutetta aloitettiin korjata formalisoimalla ympäröivää maailmaa ja lisäämällä siihen semanttiset viitteet. Satoja tuhansia käsitteitä on määritelty ja näitä määriä kuvaamaan on tallennettu miljoonia aksioomia, sääntöjä ja väitteitä, jotka rajaavat käsitteitä vastaavia kohteita. Termien kuvailemiseen käytetään CycL-nimistä ensimmäisen kertaluvun predikaattilogiikkaan pohjautuvaa kuvauskieltä. Lisäksi kielessä on useita korkeamman kertaluvun predikaattilogiikan ominaisuuksia ja tämä mahdollistaa formaalien tietämuskantojen kehittämisen sekä sen, että kaikki tieto tietämuskannassa kuvataan samalla kielellä kuin termitkin. Yhdenmuotoinen rajapinta tietoon mahdollistuu, kun kaikki tieto kuvataan predikaattilogiikasta tutulla tavalla ja silti kieli säilyy ilmaisuvoimaisena.

CYC ryhmittelee käsitteet mikroteorioihin, jotka taas muodostavat itse ontologian. Mikroteorioiden tehtävä on kuvata asiayhteyteen liittyvä tieto. Yhteen mikroteoriaan kuuluvat termit ja loogiset aksioomat, jotka kuvailevat termien semantiikkaa ja suhdetta muihin termeihin. Näin ollen yksittäinen mikroteoria on itsenäinen osa koko ontologiaa, mikä helpottaa ontologian ylläpitoa. CYC-ontologia koostuu vakioista, predikaateista, muuttujista, kaavoista, *lukumäärää ilmaisevista määrittäjistä* (engl. quantification), funktiosta ja mikroteorioista. [Fensel, 2001]

*Vakiot* (engl. constants) kuvaavat yksilöitäviä tietokokonaisuuksia, joita käytetään aksioomissa. Vakio voi ilmaista yksilöitä, muiden käsitteiden kokoelmaa tai mielivaltaista predikaattia, joka muodostaa yhteyden ilmaisun vakioiden tai funktioiden kesken. Vakioilla pitää olla aina yksilöivä nimi. *Predikaatit* (engl. predicates) kuvaavat termien välisiä suhteita. Suhteita kuvaavat loogiset lausekkeet, joiden väittämien pitää olla tautologioita. *Muuttujat* (engl. variables) ovat yksilöimättömiä termejä tai kaavoja, jotka voivat esiintyä samoissa paikoissa, joissa termit tai kaavatkin esiintyvät. *Kaavat* (engl. formulas) muodostavat termeistä mielekkäitä kokonaisuuksia. Kaava muodostuu sulkujen sisällä olevista kvanttoreista tai pilkulla erotettavista paikoista, joista ensimmäisessä voi olla predikaatti eli looginen operaatio. Loput paikoista voivat sisältää termejä eli vakioita, muuttujia, numeroita, merkkijonoja tai muita kaavoja. Koska kaavaan saa annettua syötteenä toisen kaavan, on myös rekursio mahdollista.

*Lukumäärän määrittäjät* (engl. quantification) eli *kvanttorit* (engl. quantifier) voivat olla universaalikvanttoreita tai eksistenssikvanttoreita. Hyvin määritellyssä kaavassa kaikkiin muuttujiin on määritelty kvanttori. *Funktiot* (engl. functions) kuten predikaatitkin saavat tietyn määrän syötteitä. Syötteiden tyypit on myös määriteltävä CycL-tietämuskantaan. *Mikroteoriat* (engl. microtheories) ovat joukko tietämuskannan kaavoja, jotka ovat aina ainakin yhden mikroteorian osana. Jokaisella kaavalla on oma totuusarvonsa jokaisessa mikroteoriassaan ja totuusarvoja voi olla yhteensä viisi. Totuusarvot ovat *tautologisesti tosi* (engl. monotonically true), tosi, *tautologisesti epätosi* (engl. monotonically false), epätosi tai tuntematon. Tautologisesti tosi tarkoittaa sitä, että väite on joka tapauksessa tosi, kun taas *oletettu totuus* (engl. default true) on totta vain suuressa osassa tapauksia, mutta se voi muuttua tietyssä asiayhteydessä. [Fensel, 2001]

CycL on hyvä esimerkki formaalista ontologian mallintamiskielestä. Kielen rakenne on hyvin lähellä predikaattilogiikan rakennetta ja saattaa siksi olla syntaksinsa puolesta varsin vaikeasti luettavaa. Kuitenkaan ontologian merkintäkielen ei ole tarkoituksaan olla helppolukuista vaan ennemminkin konelukuista. Kielen käyttökelpoisuus syntyykin sen ilmaisuvoimasta. Kieltä tulkitsevan ohjelmiston ratkaisut tuovat järjestelmälle omaksuttavuuden ja käytettävyyden.

### 9.3. Bunge-Wand-Weber (BWW) -ontologia

Yksi tunnetuimmista ontologioista on Bunge-Wand-Weber (BWW) -ontologia. Sen kehitys alkoi vuonna 1986, jonka jälkeen useat julkaisut nostivat esille sen erityisen lähestymistavan tiedonhallintaa kohtaan. BWW-ontologia tuodaan esille silloin, kun verrataan tietojärjestelmien analyysi- ja suunnittelusanastoa, mallinnussanaston ontologista vertailua ja useita muita tietojärjestelmien suunnitteluun liittyviä vaatimuksia. Vaikka BWW-ontologiaa onkin käytetty hyvin formaaleissa käyttökohteissa, sitä on käytetty myös käytännönläheisissä tapauksissa kuten yritysmaailmassa, jossa sen puutteet on huomattu. Ontologian on huomattu sopivan parhaiten mallintamaan konkreettisia asioita kuten fyysisiä kappaleita. Lähtökohdiltaan BWW-ontologia on Bungen teorioihin pohjautuva ja kuten on aiemmin todettu, hänen



materialistinen näkemyksensä on jättänyt siihen jälkensä, ja siksi esimerkiksi yrityksen organisaation mallintaminen on vaikeaa ilmaista. Yrityksen mallintamisessa suurimpana vaikeutena ovat juuri epäformaalit osa-alueet kuten sosiaaliset rakenteet. [Rosemann ja Wyssusek, 2005]

Rosemann ja Wyssusek [2005] epäilevät puutteiden johtuvan Wandin ja Weberin tavasta lainata Bungen ontologian käsitteitä hyvin reduktionistisesti, jolloin ontologiasta jää pois esimerkiksi systeemin hierarkian käsite. Tästä johtuvat ongelmat tulevat ilmi varsinkin silloin, kun ontologiaa käytetään empiirisissä ympäristöissä. Rosemann ja Wyssusek esittelevät myös BWW-ontologian keskeisimmät käsitteet:

*Asia* (engl. thing): Maailma rakentuu asioista, joilla on ominaisuuksia

*Asiakokonaisuus* (engl. composite thing): Asiakokonaisuus koostuu useista muista asioista

*Mahdollinen tila* (engl. conceivable state): Tilojen joukko, joissa asia voi koskaan olla

*Asioiden muuttuminen* (engl. transformation of a thing): Asioiden liittäminen asiayhteydestä toiseen asiayhteyteen

*Stabiili asian tila* (engl. stable state of a thing): Tila, jossa asia pysyy siihen asti, kunnes asian tai ulkopuolisen tapahtuman toiminta muuttaa asian tilaa

*Ominaisuus* (engl. property): Asioiden ominaisuudet määrittävät, millainen asia todellisuudessa on

*Yhteinen ominaisuus* (engl. mutual property): Yhteinen ominaisuus on olemassa vain silloin, kun se voidaan liittää kahteen tai useampaan asiaan

*Järjestelmä* (engl. system): Järjestelmä koostuu asioiden joukosta, jossa jollakin sen osalla on yhteys toisen osajärjestelmän asian kanssa

*Osajärjestelmä* (engl. subsystem): Järjestelmä, jonka osat ja rakenne ovat jonkin toisen järjestelmän osien tai rakenteen osajärjestelmä

*Luonnonlait* (engl. natural law): Luonnonlait ovat vakiintuneita luonnon asettamia lakeja

Näiden BWW-ontologian abstraktien rakenneosien pohjalta voidaan huomata, että empiiriset rakenteet kuten yritysmaailman organisaatorakenteet voivat olla vaikeita

mallintaa. Hierarkiarakenneosan puuttuminen estää esimerkiksi luomasta mallia, jossa ekosysteemi muodostaa sisäkkäisten hierarkioiden ketjun. Esimerkkinä voidaan tutkia seuraavaa ketjua: ”ekosysteemi koostuu populaatioista, jotka koostuvat organismeista, jotka koostuvat elimistä, jotka koostuvat soluista, jotka koostuvat molekyyleistä, jotka koostuvat atomeista”. [Rosemann ja Wyssusek, 2005]

#### 9.4. Suggested Upper Merged Ontology (SUMO)

SUMO [ks. <http://www.ontologyportal.org/>] on yritys muodostaa *standardi ylätason ontologia* (engl. Standard Upper Ontology), joka lyhennetään SUO. SUO oli alkujaan sähköpostijakelulista, jonne tutkijat lähettivät ehdotuksia standardiksi ylätason ontologiasta. SUMO perustettiin sähköpostilistalle kertyneiden ehdotusten ja muiden julkisesti saatavilla olevien ylätason ontologioiden pohjalta. Näin syntyi esitys sulautetusta ylätason ontologiasta eli SUMOsta, jota nykyisin työstää IEEE-rahoitteinen työryhmä avoimen lähdekoodin standardiksi. Ontologiaan sulautettiin muun muassa kirjastojen ontologioita Ontolingua-palvelimelta, John Sowan ylätason ontologia, Russellin ja Norvigin ylätason ontologia, James Allenin temporaaliaksiomat, Casatin ja Varzin formaali teoria aukoista, Barry Smithin ontologia rajoista sekä Nicola Guarinon formaali mereotologia. [Niles ja Pease, 2001]

Useiden ontologioiden sulauttamisessa ongelmana on Nilesin ja Peasen [2001] mukaan päällekkäisten rakenneosien yhdistäminen, vertaaminen ja poistaminen. Rakenneosien yhdistäminen ja poistaminen väistämättä vähentää jo tehdyn sisällön yhtenäisyyttä ja sisältö muuttaa hieman merkitystään. He toteavat, että joitain päällekkäisiä rakenneosia on jätettävä, sillä olemassa oleva sisältö tukeutuu rakenneosien ominaisuuksiin niin tiukasti, että rakenneosaa käyttävää sisältöä ei voi muuntaa käyttämään toista rakenneosaa. Esimerkiksi luokan ja *joukon* (engl. set) käsitteistä Nilesille ja Peaseille muodostui tällainen ongelma. Nämä käsitteet käyttäytyvät samalla lailla, mutta niiden perustana olevat teoriat eivät ole riittävän samankaltaisia.

SUMOn ylin taso alkaa eli sen juurisolmu on, kuten useimmissa ontologioissa, olio. Olion alapuolella ovat fyysinen ja abstrakti. Fyysisiin olioihin lukeutuvat kaikki, joilla on sijainti aika-avaruudessa ja abstraktiin kategoriaan sisältyy kaikki muu.

SUMO on alkujaan suunniteltu teknisen aihealueen tarpeisiin, eikä teknisen ja teoreettisen sisällön yhdistäminen onnistu ilman ongelmia. Yhtenä suurimmista ongelmista SUMOa työstäessään Niles ja Pease [2001] ovat havainneet haasteeksi päättää, onko SUMO 3D- vai 4D-suuntautunut ontologia. Vaihtoehtoina on jättää temporaaliset ominaisuudet pois ja yhdistää erillisillä moduuleilla puuttuvat ominaisuudet. Tällainen vaihtoehto loisi useita ontologian tasoja, jotka olisivat keskenään epäyhteensopivia, mutta ylätasoon ontologian kautta yhteensopivia. Hyvinä puolina Niles ja Pease esittelevät SUMOa helposti omaksuttavana ja käytännönläheisenä, sekä verrattaessa Cycorpin Cyc -ontologiaan SUMOa voidaan pitää parempana, koska Cyc on tuotteistettu ja suurimmaksi osaksi maksullinen tuote. Cycorp ei ole julkaissut koko ontologiaa julkisesti eikä sitä ole kunnolla *vertaisarvioitu* (engl. peer reviewed). SUMOa saatetaan käyttää tulevaisuudessa yleisesti, koska IEEE:n mukana olo antaa vankan pohjan standardin luomiselle.

## 9.5. OWL

Protégé-OWL [ks. <http://protege.stanford.edu/overview/protege-owl.html>] on toteutus Web Ontology Language -kielestä (OWL), joka puolestaan on mahdollisesti tunnetuin standardoitu ontologiakieli. OWL-kieltä kehittää World Wide Web Consortium (W3C) [ks. <http://www.w3.org/TR/owl-features/>]. Tässä tutkimuksessa Protégé-OWL-sovelluksesta käytetään nimitystä Protégé. Protégéä käytetään OWL-ontologioiden muodostamiseen.

Protégé-OWL-oppaassa [Horridge *et al.*, 2004] OWL-ontologian kerrotaan rakentuvan *yksilöistä* (engl. individuals), *ominaisuuksista* (engl. properties) ja *luokista* (engl. classes). Protégéssa vastaavat termit ovat *ilmentymä* (engl. instance), *lokero* (engl. slot) ja *luokka*.

OWLssa *yksilöt* ja Protégéssa *ilmentymät* tarkoittavat kohdealueen kiinnostuksen kohteena olevia yksilöllisiä objekteja. Nämä objektit on kyettävä nimeämään, ja nimeämiskäytäntö onkin erilainen Protégéssa ja OWLssa. Protégéssa objekteilla voi olla useampia nimiä, jotka ovat keskenään synonyymejä, kun taas OWLssa yksilöllä vain yksi yksilöivä nimi.

*Ominaisuuksilla* on kaksoissuhde yksilöjen välillä eli ominaisuus liittää kaksi yksilöä toisiinsa [Horridge *et al.*, 2004]. Ominaisuuden yksilöiden välinen suhde voi olla kaksisuuntainen, jolla on varsinkin englanninkielisissä ontologioissa merkitystä, esimerkiksi ”An apartment is owned by Matti” ja ”Matti owns an apartment”. Kuten voidaan huomata, ensimmäisessä esimerkissä lähtökohtana on asunnon omistaja, kun taas jälkimmäisessä tarkastellaan sitä, mitä Matti omistaa. Suomenkielisessä esimerkissä eroa tuskin huomaa: ”Asunnon omistaa Matti”, jonka vastakkainen suhde on ”Matti omistaa asunnon”. Horridgen *et al.* [2004] mukaan yksilöiden välinen suhde voi olla funktionaalinen, transitiivinen tai symmetrinen. Funktionaalinen suhde on yksiarvoinen, jolloin on kyse ominaisuudesta, jota yksilöllä voi olla vain yksi (yhdessä myyntitilanteessa voi toteutua vain yksi kauppahinta.). Transitiivinen suhde tarkoittaa välillistä suhdetta  $\forall a, b, c \in A | ((aRb) \wedge (bRc)) \Rightarrow (aRc)$ , jossa yksilöiden a ja c suhde selittyy yksilöiden a ja b sekä b ja c suhteilla. Esimerkkinä transitiivisesta suhteesta voidaan pitää asunnon omistajahistoriaa: Henkilö A omisti asunnon ennen henkilöä B, joka myi asunnon henkilölle C. Henkilö A on asunnon entinen omistaja paitsi henkilölle B myös henkilölle C välillisesti henkilön B kautta. Symmetrinen suhde puolestaan kuvaa suhdetta kahden yksilön välillä, joiden välinen ominaisuus on sama molempiin suuntiin, kuten asunnon myynnin tapauksessa ostaja on myyjän kauppakumppani ja päinvastoin.

*Luokat* ovat kohdealueen yksilöiden muodostamia ryhmiä, joilla on yhteinen nimittäjä jokaisen luokan yksilön kesken [Horridge *et al.*, 2004]. Esimerkiksi kiinteistökaupan kiinteistöihin kuuluvat uudiskohteet, kesämökkit, omakotitalot, tontit sekä kerrostalot. Luokat voivat myös olla sisäkkäisiä eli järvenrantatontti on tontti ja lisäksi kiinteistö. Joskus luokkia saatetaan sanoa käsitteiksi, jolloin luokka on käsitteen konkreettinen esitys [Horridge *et al.*, 2004]. Luokka on joukko todellisen maailman objekteja, joilla on samat ominaisuudet. Käsite taas on luokan objektien yhteinen nimittäjä eli käsitteellä ilmaistaan luokan objektien yhteisen olemuksen idea.

Yleisesti tietojenkäsittelyssä käytetty OWL Web Ontology Language -kieli on käytännön esimerkki ontologiasta sekä ontologian tallennukseen käytettävästä esityskielestä. OWL helpottaa webin konetulkintaa ja sen osakielet tarjoavat lisäsanastoa sekä formaalia semantiikkaa niiden sovellusten käytettäväksi, jotka nimenomaan käsittelevät tiedon sisältöä.

## 10. Yhteenveto ontologioista

Eri ontologioiden vertailu keskenään on hankalaa, koska ontologioiden käsitteistö ei ole vakiintunutta ja koska eri ontologiat määrittelevät asioita eri tavoilla. Olen koonnut seuraavaan taulukkoon yleisimpiä ontologioissa käytettyjä primitiivejä:

| <b>BWW-ontologia</b>  | <b>OWL</b>   | <b>CYC</b>  | <b>SUMO</b>                               | <b>WordNet</b>                    |
|---|--|---|---|-----------------------------------|
| Asia<br>(engl. Thing)   | Yksilö<br>(engl. Individual)                       | Asia<br>(engl. Thing)                             | Käsite<br>(engl. Object)                  | Olio<br>(engl. Entity)            |
| Ominaisuus ja Attribuutti<br>(engl. Property and Attribute)   | Ominaisuus ja Luokka<br>(engl. Property and Class) | Luokka<br>(engl. Class)                           | Attribuutti<br>(engl. Attribute)          | Omistaminen<br>(engl. Possession) |
| Asian tila tietyssä hetkessä, asian attribuuteilla on arvoja<br>(engl. State of a Thing, the attributes of a thing have values) |  | Ajallinen asia<br>(engl. Temporal Thing)          |   | Tila<br>(engl. State)             |
| Tapahtuma: tilan muutos asiassa<br>(engl. Event: change of state in a thing)  |  | Tapahtuma<br>(engl. Event)                        |   | Tapahtuma<br>(engl. Event)        |
| Asian historia: tapahtumajärjestys asiassa<br>(engl. History of a Thing: a sequence of events in a thing)                       |  | Toiminta<br>(engl. Action)                        | Prosessi<br>(engl. Process)               | Ilmiö<br>(engl. Phenomenon)       |
| Tyyppi/Luokka ja Ala-tyyppi/Alaluokka<br>(engl. Type/Class and Subtype/Subclass)  | Luokka<br>(engl. Class)                            | Joukko tai Ko-koelma<br>(engl. Set or Collection) | Joukko tai Luokka<br>(engl. Set or Class) | Ryhmä<br>(engl. Group)            |
| Yhdistetty asia: muodostettu muista asioista kuin itsestään<br>(engl. Composite thing: is composed of things other than itself) |  | Suhde<br>(engl. Relation)                         |   |                                   |

|  |  |                                       |  |  |
|--|--|---------------------------------------|--|--|
|  |  | Mikroteoria<br>(engl.<br>Microtheory) |  |  |
|--|--|---------------------------------------|--|--|

Taulukko 1: Yhteenveto ontologioista

Tämän tutkimuksen puitteissa kaikille käsitteille ei löytynyt vastaavuuksia eri ontologioiden kesken. On kuitenkin täysin mahdollista, että joillakin käsitteillä ei ole lainkaan vastaavaa rakenneosaa toisissa ontologioissa. Yhteisen ydinontologian luominen jälkikäteen on näin ollen vaikeaa. Lähtökohta pitäisikin ennemmin olla, että kaikki ontologiat sitoutuisivat yhteiseen ydinontologiaan, jolloin päästäisiin lähemmäksi kaiken kattavaa ontologiaa.

## 11. Lopuksi

Ontologia on teoria toimialueen olemuksesta. Ontologia on laaja kokonaisuus, jossa toisaalta tutkitaan yläluokan ontologioiden käsitteistöä ja toisaalta huomattavasti erikoistuneempia ontologioita. Näiden lisäksi teoriaa ja itse sanastoa tai tietämuskantaa [ks. Ullman, 1988], joka tuotetaan kohdealueen ilmentymien termeistä, kutsutaan molempia virheellisesti ontologioiksi. Ontologiat mielletään usein monimutkaisuutensa vuoksi liian vaikeiksi ottaa käyttöön, ja näin ollen niiden ei uskota maksavan itseään takaisin. Raja tietämuskannan muodostavan sanaston ja sovellusalueen ontologian välillä on hankala ymmärtää, kun jotkut tutkijat tarkastelevat ontologioita matemaattis-filosofisesta näkökulmasta, toiset taas oman sovellusalueensa näkökulmasta. Ontologian rakentaminen on lisäksi työlästä ja kallista sovelluksissa, joissa tieto pysyy hallinnassa jonkin muun tekniikan avulla. Kuitenkin nykyään tutkitaan ontologioiden rakentamista puoliautomaattisesti tai jopa täysin automaattisesti [Taylor *et al.*, 2007; Wang, 2006]. Automaattinen ontologian rakentaminen ei kuitenkaan onnistu vaivatta, vaan siinä tarvitaan jo olemassa oleva ontologia, johon agenttiohjelma lisää käsitteitä. Agentti vertaa lisättävää käsitettä sekä jo olemassa olevan ontologian rakennetta keskenään päätelläkseen uuden käsitteen tulevan sijainnin. Automaatiolla voitaisiin helpottaa ontologioiden muodostamista ja käyttöönottoa.

Julkisuudessa on usein kritisoitu ontologiaa pyörän uudelleen keksimisestä. Ontologian väitetään olevan uudelleen keksitty luokittelu, joka on jo käytössä kaikkialla. Ontologian todellinen olemus on kuitenkin paljon enemmän: se on lähempänä tietämystä tai ymmärrystä kuin luokiteltua tietoa. Ontologia kykenee parhaimmillaan vastaamaan kysymykseen ”*Miksi?*”. Vastaus kysymykseen on ontologiassa eräänlainen puun oksa, joka koostuu käsitteistä, aksioomista ja suhteista. Kun verrataan ontologiaa ja luokittelua vastattaessa kysymykseen, miksi maapallo on pyöreä, havaitaan olennainen ontologian ja luokittelun ero. Ontologiassa määriteltäisiin muiden muassa gravitaation vaikutus materiaan sekä isostasian vaikutus maan pintakerrokseen, mikä ajan kuluessa tasoittaa planeetan pinnan kauttaaltaan lähes yhtä etäälle keskipisteestä. Gravitaatio ja isostasia määriteltäisiin matemaattisesti ja selkokielellisesti, jolloin ontologian rakennetta läpikäymällä saataisiin selitys kysymykseen. Pelkällä luokittelulla voitaisiin järjestää esimerkin gravitaation ja isostasian selitykset tai



määrietykset helposti löydetävään rakenteeseen. Luokittelussa tieto on kokonaisuutena, ja tämän vuoksi luokitellusta tiedosta löydetään vain eksplisiittinen vastaus.

Clay Shirkyin ontologioihin kohdistama kritiikki saattaa pohjautua juuri luokitteluiden ja ontologioiden sekoittamiseen keskenään. Shirkyin kritiikki kohdistuu lähinnä asioiden luokittelun ja nimeämisen väliseen vertailuun. Boris Wysusek ei myöskään suoranaisesti kiellä, etteivätkö ontologiat olisi hyödyllisiä, vaan hän kritisoi ontologian suunnittelussa ja muodostamisessa käytettyjen teorioiden löyhää käyttämistä. Wysusekin mielestä on ensiarvoisen tärkeää sitoutua niihin filosofisiin teorioihin, joihin suunnittelun aikana on tukeuduttu.

Tässä tutkimuksessa esiteltyjen teorioiden pohjalta teen sen johtopäätöksen, että ontologiat ovat perusteellista paneutumista vaativa tieteenala, mutta silti tietyissä tapauksissa erittäin käyttökelpoinen apuväline tietomassojen hallinnassa. Itseäni kiinnostaa ontologian käyttö aineiston siirrossa järjestelmien välillä sekä ontologiapohjaisten hakujen mahdollisuudet. Juuri tällaisissa tilanteissa sovelluksen suunnittelu helpottuu, jos tietomassa on formaalissa muodossa. Ongelmana ontologian käyttöönotolle saattaa olla riittävän yksityiskohtaisen ontologian löytyminen, koska sellaista harvoin löytyy valmiina ja uuden ontologian muodostaminen on vaikeaa ja kallista. Valmiin ontologian hyödyntäminen tuokin jo esiin ontologian hyödyn, jos sitä käytetään tunnollisesti ja sen käyttöön sitoudutaan aidosti. Sitoutuminen ontologian käyttöön saattaa olla haaste järjestelmissä, joita kehitetään nopeasti tilanteiden muuttuessa, eikä ontologiaan sitoutumista tai sen päivittämistä muisteta tehdä. Näin ollen jatkossa pitäisi tutkia vielä lisää ontologioiden muodostamisen ja käytön aikaisten apuvälineiden jatkokehitystä ja suunnitella käytäntöjä, jotka helpottavat ontologian ja järjestelmän yhtenäisyyden pysymistä yllä.

Tämän tutkimuksen yhteydessä kävi ilmi myös, että ontologioiden laadun ylläpito on tärkeä ja mielenkiintoinen aihealue. Kun sovellusta käytetään dynaamisessa maailmassa ja tietämys kehittyy, on ontologian pysyvä muutos mukana. Ontologian muuttuessa sovelluksen on myös muututtava, jotta se kykenee palvelemaan muuttuneita tiedon tarpeita. Muutoksen yhteydessä ontologian sisällön ja teorian vastaavuuden tulisi säilyä. Ontologian laadun arvioiminen on vielä melko kehittymätöntä eikä käytännön sovelluksissa kiinnitetä juurikaan huomiota laatuun. Laatua tarkastellaan lähestulkoon

ainoastaan ontologian muodostusvaiheessa, joten uskon tämän aiheen tarvitsevan vielä lisätutkimusta.

**Viiteluettelo**

- [Bunge, 1977] Mario Augusto Bunge, *Treatise on Basic Philosophy Vol 3, Ontology I: The Furniture of the World*. Reidel, Dordrecht, Boston, (1977)
- [Casati and Varzi, 1999] Roberto Casati and Achille C. Varzi, *Parts and Places: The Structures of Spatial Representation*. MIT Press, Cambridge MA, 1999.
- [Chisholm, 1996] Roderick M. Chisholm, *A Realistic Theory of Categories*, Cambridge, 1996.
- [Cimiano and Handschuh, 2003] Philipp Cimiano and Siegfried Handschuh, Ontology-based linguistic annotation. In: *Proceedings of the ACL 2003 workshop on Linguistic annotation: getting the model right* (2003), 14-21.
- [Fensel, 2001] Dieter Fensel, *Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce*, Springer-Verlag, Berlin, 2001.
- [Gangemi *et al.*, 2001] Aldo Gangemi, Nicola Guarino, Claudio Masolo and Alessandro Oltramari, Understanding top-level ontological distinctions, In: *Proceedings of IJCAI-01 Workshop on Ontologies and Information*, 2001, 26-33.
- [Gómez Pérez and Benjamins, 1999] Asunción Gómez Pérez and V. Richard Benjamins, Overview of Knowledge Sharing and Reuse Components: Ontologies and Problem-Solving Methods. 1999. Available as <http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-18/1-gomez.pdf>
- [Gómez-Pérez and Corcho, 2002] Asunción Gómez-Pérez and Oscar Corcho, Ontology Languages for the Semantic Web, *IEEE Intelligent Systems* **17**, 1 (2002), 54-60.

- [Goralwalla *et al.*, 2001] Iqbal A. Goralwalla, Yuri Leontiev, M. Tamer Özsu and Duane Szafron, Temporal granularity: Completing the puzzle, *Journal of Intelligent Information Systems* **16**, 1 (2001), 41-63.
- [Gruber, 1993a] Thomas Gruber, Toward principles for the design of ontologies used for knowledge sharing. Originally in N. Guarino and R. Poli (Eds.), *International Workshop on Formal Ontology*, Padova, Italy. Revised August 1993. In: *International Journal of Human-Computer Studies* **43**, 5-6 (Nov./Dec. 1995), 907-928.
- [Gruber, 1993b] Gruber, T. R. What is an Ontology? (1993) Available as <http://www.ksl.stanford.edu/kst/what-is-an-ontology.html>
- [Grüninger and Lee, 2002] Michael Grüninger and Jintae Lee, Ontology Applications and Design. *Communications of the ACM* **45**, 2 (Feb. 2002), 39-41.
- [Guarino, 1994] Nicola Guarino, The Ontological Level. In: R. Casati, B. Smith and G. White (eds.), *Philosophy and the Cognitive Sciences*. Vienna, 1994, 443-456.
- [Guarino, 1998] Nicola Guarino, Formal Ontology and Information Systems. In: *Proc. of Formal Ontologies in Information Systems, 1st International Conference on Formal Ontologies in Information Systems FOIS'98*, (1998), IOS Press, 3-15.
- [Guarino and Guizzardi, 2006] Nicola Guarino and Giancarlo Guizzardi, In the Defense of Ontological Foundations for Conceptual Modeling, *Scandinavian Journal of Information Systems* (2006), 115-126.
- [Guizzardi and Wagner 2004] Giancarlo Guizzardi and Gerd Wagner, A Unified Foundational Ontology and some Applications of it in Business Modeling, *CAiSE Workshops* **3** (2004), 129-143.

- [Hayes, 1995] Patrick Hayes, A Catalog of Temporal Theories, University of Illinois, Tech report **UIUC-BI-AI-96-01**, 1995. Available as <http://www.ihmc.us/users/phayes/Pub/TimeCatalog.pdf>
- [Heflin, 2004] Jeff Heflin, OWL Web Ontology Language Use Cases and Requirements. February 10, 2004. Available as <http://www.w3.org/TR/2004/REC-webont-req-20040210/>
- [Horridge *et al.*, 2004] Matthew Horridge, Holger Knublauch, Alan Rector, Robert Stevens and Chris Wroe, A Practical Guide To Building OWL Ontologies Using The Protégé-OWL Plugin and CO-ODE Tools Edition 1.0. The University Of Manchester, 2004.
- [Klein and Smith, 2005] Gunnar O. Klein and Barry Smith, Concept Systems and Ontologies. [updated 2006 Oct 6; cited 2006 Dec 4]. Available as <http://ontology.buffalo.edu/concepts/ConceptsandOntologies.pdf>
- [Laurence and Margolis, 1999] Stephen Laurence and Eric Margolis, Concepts and Cognitive Science. In E. Margolis, S. Laurence (eds.). *Concepts: Core Readings*. Cambridge, 1999, 3-81.
- [Niles and Pease, 2001] Ian Niles and Adam Pease, Towards a Standard Upper Ontology. In: Chris Welty and Barry Smith (eds.), *Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (2001)*, 2-9.
- [Oaklander, 2004] L. Nathan Oaklander, *The Ontology of Time*. Prometheus Books, New York, 2004.
- [Obrst *et al.*, 2007] Leo Obrst, Werner Ceusters, Inderjeet Mani, Steve Ray and Barry Smith, The Evaluation of Ontologies: Toward Improved Semantic Interoperability, In: Christopher J. O. Baker and Kei-Hoi Cheung (eds.), *Semantic*

*Web: Revolutionizing Knowledge Discovery in the Life Sciences*. Springer, USA, 2007.

[Parsons and Wand, 1997] Jeffrey Parsons and Yair Wand, Using objects for systems analysis. *Communications of the ACM* **40**, 12 (Dec. 1997), 104-110.

[Poli, 2002] Roberto Poli, Ontological methodology. *IJHCS* **56**, 6 (June 2002), 639-664.

[Quine, 1961] W. V. O. Quine, *From a Logical Point of View: 9 Logico-Philosophical Essays (second edition)*, Cambridge, 1961.

[Rosemann and Wyssusek, 2005] Michael Rosemann and Boris Wyssusek, Enhancing the expressiveness of the Bunge–Wand–Weber ontology. In: *Proceedings of the Americas Conference on Information Systems*. Omaha, USA, 2005, 2803–2810.

[Shirky, 2003] The Semantic Web, Syllogism, and Worldview. (2003) Available as [http://www.shirky.com/writings/semantic\\_syllogism.html](http://www.shirky.com/writings/semantic_syllogism.html)

[Shirky, 2005] Ontology is Overrated: Categories, Links, and Tags. (2005) Available as [http://www.shirky.com/writings/ontology\\_ouerrated.html](http://www.shirky.com/writings/ontology_ouerrated.html)

[Smith, 1998] Barry Smith, Basic Concepts of Formal Ontology. In: Nicola Guarino (ed.), *Formal Ontology in Information Systems*, IOS Press, 1998, 19-28.

[Smith, 2003] Barry Smith, Ontology. In: L. Floridi (ed.), *Blackwell Guide to the Philosophy of Computing and Information*, Oxford, Blackwell, 2003, 155–166.

[Smith, 2004] Barry Smith, Beyond Concepts: Ontology as Reality Representation. In *Proc. of FOIS 2004. International Conference on Formal Ontology and Information Systems*, 2004, 73-84.

- [Smith *et al.*, 2006] Barry Smith, Waclaw Kusnierczyk, Daniel Schober and Werner Ceusters, Towards a Reference Terminology for Ontology Research and Development in the Biomedical Domain. In *Proc. of KR-MED 2006*, 57-66.
- [Taylor *et al.*, 2007] Matthew E. Taylor, Cynthia Matuszek, Bryan Klimt and Michael J. Witbrock, Autonomous Classification of Knowledge into an Ontology. In: *Proceedings of the Twentieth International Florida Artificial Intelligence Research Society Conference*, Florida, USA (2007), 140-145.
- [Ullman, 1988] Jeffrey D. Ullman, *Principles of Database and Knowledge-Base Systems, Volume 2*. Computer Science Press, 1988.
- [Uschold and Gruninger, 1996] Mike Uschold and Michael Gruninger, Ontologies: Principles, Methods and Applications. *KER* **11**, 2 (June 1996), 93-155.
- [Vallet *et al.*, 2005] David Vallet, Miriam Fernández, Pablo Castells: An Ontology-Based Information Retrieval Model. *European Semantic Web Conference*, Heraklion, Greece, (2005), 455-470
- [Wand and Weber, 2006] Yair Wand and Ron Weber, On Ontological Foundations of Conceptual Modeling: A Response to Wyssusek, *Scandinavian Journal of Information Systems* (2006), 127-138.
- [Wieringa and de Jonge, 1993] Roel Wieringa and Wiebren de Jonge, The Identification of Objects and Roles. Vrije Universiteit, Faculty of Mathematics and Computer Science, Technical Report **TR-267**, 1993.
- [Winston *et al.*, 1987] Morton Emanuel Winston, Roger Chaffin, Douglas Herrmann, A taxonomy of part-whole relations. In: *Cognitive Science: A Multidisciplinary Journal*, **11**, 4 (Jan. 1987), 417 - 444.

[Wyssusek, 2004] Boris Wyssusek, Ontology and Ontologies in Information Systems Analysis and Design: A Critique. In: *Proc. Of Association for Information Systems, Tenth Americas Conference on Information Systems (2004)*, 4303–4308.

[Wyssusek, 2006] Boris Wyssusek, On Ontological Foundations of Conceptual Modelling, *Scandinavian Journal of Information Systems (2006)*, 63-80.