

Kansalliskirjasto
Raportteja ja selvityksiä 3 / 2014

ASIASANASTON MUUTTAMINEN ONTOLOGIAKSI

Yleinen suomalainen ontologia
esimerkkinä FinnONTO-hankkeen mallista

Katri Seppälä, Sanastokeskus TSK
Eero Hyvönen, Aalto-yliopisto ja Helsingin yliopisto



Kansalliskirjasto, Helsinki 2014

ISBN 978-952-10-9883-3 (pdf)

Julkaisija	Kansalliskirjasto
Julkaisun päivämäärä	1.8.2014
Tekijä	Katri Seppälä, Sanastokeskus TSK Eero Hyvönen, Aalto-yliopisto ja Helsingin yliopisto
Julkaisun nimi	Asiasanaston muuttaminen ontologiaksi Yleinen suomalainen ontologia esimerkkinä FinnONTO-hankkeen mallista
Julkaisun osat	
Sarjan nimi ja numero	Raportteja ja selvityksiä 3/2014
ISSN ISBN	978-952-10-9883-3
URL URN	URN:ISBN:978-952-10-9883-3
Kokonaissivumäärä	68
Kieli	Suomi
Avainsanat – YSA (Suomi)	FinnONTO; ontologiat; yhteensopivuus, asiasanastot; semanttinen web; käsiteanalyysi
Tiivistelmä	Tämän raportin tavoitteena on edistää tätä työtä kiteyttämällä ja dokumentoimalla suomen kielellä niitä periaatteita, joita FinnONTO-hankkeessa kehitettiin suomalaisten asiasanastojen muuttamisesta semanttisen webin ontologioiksi, sekä tästä työstä saatuja kokemuksia. Esimerkkinä periaatteiden soveltamisesta käytäntöön, esitellään tapaus-tutkimuksena Yleisen Suomalaisen Asiasanaston YSA muunnosprosessia Yleiseksi Suomalaiseksi Ontologiaksi YSO, joka muodostaa KOKO-ontologian ylärakenteen ja tärkeimmän komponentin
Lisätietoja	

BESKRIVNING SIDA

Utgivare	Nationalbiblioteket
Utgivningsdatum	1.8.2014
Författare	Katri Seppälä, Sanastokeskus TSK Eero Hyvönen, Aalto-yliopisto och Helsingfors universitet
Publication	Från ämnesordsregister till ontologi Den allmänna finländska ontologin som exempel på FinnONTO-projektets arbetsmodell
Publikationens delar	
Seriens namn och nummer	Rapporter och utredningar 3/2014
ISSN ISBN	978-952-10-9883-3
URL URN	URN:ISBN:978-952-10-9883-3
Sidoantal	68
Språk	Finska
Nyckelord (Allärs)	FinnONTO; ontologier; sammanjämkning; ämnesordsregister; semantisk webb; egreppsanalys
Sammandrag	Målet för rapporten är att främja omvandlingsarbetet genom att på finska sammanfatta och dokumentera principerna för och erfarenheterna av omvandlingen av finländska ämnesordsregister till ontologier för den semantiska webben som utvecklades i projektet FinnONTO. Som exempel på hur principerna tillämpas i praktiken presenteras som fallstudie YSA:s (den allmänna tesaursen på finska) omvandlingsprocess till den allmänna finländska ontologin ALLFO, vilken utgör KOKO-ontologins överbyggnad och viktigaste komponent.
Tilläggsuppgifter	

Sisältö

ESIPUHE.....	9
1 KANSALLINEN ONTOLOGIAINFRASTUKTUURI FINNONTO.....	11
1.1 Ontologiat yhdistävät tietoa ja tarkentavat sisällönkuvailua	11
1.2 Kansallinen haaste: ontologiat	12
1.3 Ontologiainfrastruktuurin osat.....	14
2 ONTOLOGIATYÖN TAVOITTEET, PERIAATTEET JA RAJOITTEET	17
2.1 KOKO - joukko toisiinsa yhdistettyjä eri alojen ontologioita	17
2.2 Yhteentoimivuus olemassa oleviin asiansastoihin	20
2.3 Keskittyminen ensin olennaisimpiin ontologisiin periaatteisiin.....	21
2.4 Luokkahierarkia ja sen täydentäminen.....	23
2.5 Luokkien ja yksilöiden erottaminen.....	25
2.6 Laajempi/suppeampi termi -suhteiden tarkentaminen	27
2.7 Yksilö-luokka-suhteiden transitiivisuuden tarkistaminen.....	28
2.8 Monimerkityksisten sanojen taustalla olevien käsitteiden erottelu.....	28
2.9 Standardiformaatin valinta: RDF(S), OWL ja SKOS	29
2.10 Käsitteiden nimeäminen	29
2.11 Moniarvoisuus	30
2.12 Avoin yhdistetty tieto.....	31
3. ASIANSASTOSTA ONTOLOGIAKSI: CASE YSO.....	32
3.1 Hierarkian pääluokat.....	32
3.1.1 Abstrakti (abstract)	34
3.1.2 Pysyvä (endurant)	34
3.1.3 Muuttuva (perdurant).....	38
3.2 Hierarkian rakentaminen	40
3.2.1 Hierarkiapalojen yhteensovittaminen	40
3.2.2 Koostumussuhteiden käsittely	41
3.2.3 Korvattujen termien käsittely.....	41
3.2.4 Assosiativisten suhteiden käsittely	42
3.2.5 Asiansaketjujen käsittely.....	42

3.2.6	Hierarkkisten alakäsitteiden ryhmittely.....	42
3.3	Assosiatiivisten suhteiden käyttö	45
3.3	Roolien käyttö.....	46
3.5	Monimerkityksisten asiasanojen käsittely	46
3.5.1	Koostekäsitteet.....	49
3.6	Monikielisen ontologian rakentaminen.....	49
3.6.1	Monikielisen ontologian haasteet	51
3.6.2	Ruotsinkieliset asiasanat	52
3.6.3.	Asiasanojen englanninkieliset vastineet	54
3.7	Ontologioiden päivitys	55
3.7.1	Asiasanaa koskevat muutokset	55
3.7.2	Käsitteen hierarkiaa koskevaa muutokset	57
3.7.3	Käsitteen assosiatiivisissa suhteissa tapahtuneet muutokset	57
3.7.4	Ontologian käsittekokonaisuuteen liittyvät muutokset	57
3.8.	Ehdotuksia jatkotoimenpiteiksi	58
4	YHTEENVETO.....	60
5	KIRJALLISUUTTA	62
6	LIITE: KESKEISIÄ KÄSITTEITÄ.....	66

Esipuhe

Aalto-yliopiston ja Helsingin yliopiston Semanttisen laskennan tutkimusryhmän (SeCo) vetämässä FinnONTO-hankkeiden sarjassa luotiin vuosina 2003-2012 prototyyppi kansallisesta semanttisen webin infrastruktuurista, joka sisältää eri alojen ontologioita, ONKI-ontologiapalvelun ja vision prosessista, jolla monialainen ontologiatyö voitaisiin Suomessa jatkossa järjestää ja vakinaistaa. Hankkeen eri vaiheita rahoitti Tekes ja yhteensä n. 50 suomalaista yritystä ja julkisten sektorin organisaatiota.

Ontologioiden kehittämisestä sekä ONKI-ontologiapalvelusta vastasi vuoden 2013 loppuun Aalto-yliopiston mediatekniikan laitoksella ja Helsingin yliopiston tietojenkäsittelytieteen laitoksella Semanttisen laskennan tutkimusryhmä SeCo. Vuoden 2014 alussa keskeinen osa FinnONTO:n ontologiastyöstä työstä ja ontologioista siirtyi Kansalliskirjaston ylläpidettäväksi ja edelleen kehitettäväksi Finto-nimisenä uutena palveluna. Sen teknologiapohjana on FinnONTOssa kehitetty ONKI Light -palvelin ja keskeisimpään tietosisältöön kuuluu joukko FinnONTOssa kehitettyjä ontologioita, erityisesti asiasanastoista kehitetyt yleiskäsitteitä sisältävä Yleinen Suomalainen Ontologia YSO ja siihen liittyvät erikoisalojen ontologiat eli ns. KOKO-ontologiapilvi. FinnONTOssa ehdotetun ONKI-konsortiomallin mukaan ontologioiden kehittämistä voitaisiin jatkossa organisoida yhteisöllisesti siten, että Kansalliskirjasto kehittää järjestelmän ytimessä olevaa YSOa ja siihen liittyviä omia ontologioitaan ja koordinoi yhteistyötä, jossa erikoisontologioita kehitetään alakohtaisten asiantuntijaryhmien toimesta.

Tämän raportin tavoitteena on edistää tätä työtä kiteyttämällä ja dokumentoimalla suomen kielellä niitä periaatteita, joita FinnONTO-hankkeessa kehitettiin suomalaisten asiasanastojen muuttamisesta semanttisen webin ontologioiksi, sekä tästä työstä saatuja kokemuksia. Esimerkkinä periaatteiden soveltamisesta käytäntöön, esitellään tapaustutkimuksena Yleisen Suomalaisen Asiasanaston YSA muunnosprosessia Yleiseksi Suomalaiseksi Ontologiaksi YSO, joka muodostaa KOKO-ontologian ylärakenteen ja tärkeimmän komponentin. Tämä FinnONTO-ontologiatyön kuvaus on laadittu Kansalliskirjaston erillisestä toimeksiannosta, ja Kansalliskirjaston Finto-palvelun kehittäjät ovat osallistuneet kuvauksen laatimiseen antamalla palautetta kirjoitusprosessin aikana. Kiitokset kuuluvat raportin kirjoittamista aktivoineelle Tuomas Paloselle sekä raporttia ystävällisesti kommentoineille henkilöille: Mirja Anttila, Esa-Pekka Keskitalo, Susanna Nykyri, Mikko Lappalainen, Liisa Virtanen ja Henri Ylikotila. Mahdollisista tekstiin jääneistä epätarkkuuksista ja virheistä samoin kuin raportissa esitetyistä ehdotuksista vastaavat kuitenkin kirjoittajat.

Raportin pohja-aineistona on käytetty FinnONTO:n aikana muistiin kirjattuja havaintoja ja ohjeita. Katri Seppälä osallistui keskeisellä tavalla kaikkien FinnONTO-hankkeessa kehitettyjen ontologioiden kehitystyöhön, joiden perustana olivat asiasanastot, ja aivan erityisesti YSA:n muuttamiseen YSOksi. Eri osaontologioiden kehittäminen tapahtui usein yhteistyössä hankkeeseen osallistuneiden organisaatioiden asiantuntijoiden kanssa siten, että FinnONTO:n tutkijat opastivat kehitystyössä ja tarjosivat sille teknisen tuen. Sisältötyötä koskevaa ohjeistusta on monikielisten ontologioiden osalta tuolloin kirjoittanut myös Tuomas Palonen. FinnONTO-hanke syntyi Eero Hyvösen aloitteesta ja hän myös johti hanketta alusta loppuun saakka.

Lukuisat SeCo-tutkimusryhmän tutkijat Aalto-yliopistossa ja Helsingin yliopistossa sekä hankekonsortion eri organisaatioissa ovat olleet mukana tässä raportissa esiteltävässä ontologiatyössä, kuten (aakkosjärjestyksessä) Mirja Anttila (YSO), Johanna Eerola (TERO), Maria Forsén (ALLSO), Eeva-Liisa Haanpää (KITO), Outi Hietanen (MERO), Kaisa Hypén (KAUNO), Esa Hämäläinen (PUHO), Anne Isomursu (VALO), Ellen Karhulampi (TAO), Suvi Kettula (MAO), Anna-Liisa Kristiansson (KTO), Eeva-Liisa Leppänen (YSO), Päivi Lipsanen (AFO), Katariina Nyberg (MUSO), Kristiina Näyhö (KULO), Tuomas Palonen (JUHO), Ringa Sandelin (ALLSO), Satu Savia (VALO), Alex Spåre (KAUNO), Timo Viitanen (ALLSO) ja Anu Ylisalmi (YSO, LIITO). Tekniseen kehittämiseen osallistuivat eri aikoina mm. Matias Frosterus, Tomi Kauppinen, Ville Komulainen, Eetu Mäkelä, Sini Pessala, Tuukka Ruotsalo, Reetta Sinkkilä, Osma Suominen, Jouni Tuominen, Arttu Valo ja Kim Viljanen. Eeva Kärjen apu YSA-sanastoon liittyvissä kysymyksissä oli hankkeen onnistumisen kannalta tärkeää. Kiitos kaikille yhteistyöstä ja mukanaolosta.

Espoossa 13.3.2014

Katri Seppälä ja Eero Hyvönen

1 Kansallinen ontologiainfrastruktuuri FinnONTO

1.1 Ontologiat yhdistävät tietoa ja tarkentavat sisällön- kuvailua

Asiasanoja käytetään yleisesti tiedon kuvailussa kirjastoissa, museoissa, arkistoissa, mediayhtiöissä jne. Kontrolloidun sanaston käytöllä voidaan parantaa ja yhdenmu-kaistaa tiedon indeksointia, jolloin tiedon haussa (Baeza-Yates, Ribeiro-Neto, 2011) päästään myöhemmin parempaa tarkkuuteen (precision) ja saantiin (recall). Asia-sanoja on organisoitu perinteisesti merkitysten perusteella asiasanastoiksi (tesau-ruksiksi) (Aitchison et al., 2000), jolloin termeihin merkityksen perusteella liittyvät toiset termit voidaan löytää helpommin kuin esimerkiksi aakkosellisesta hakemis-tosta.

Asiasanastoja on alettu kehittää ontologioiksi (Staab, Studer, 2009) erityisesti se-manttisen webin (Semantic Web) tutkimuksen piirissä. Semanttinen web (Antoniou, van Harmelen, 2008) on webin kehitystä koordinoivan W3C-järjestön v. 2001 käyn-nistämä kehityssuunta, jonka pyrkimyksenä on täydentää perinteistä, sivujen ver-kosta muodostuvaa webiä siinä esiintyvien käsitteiden ja (meta)tietojen muodosta-malla semanttisella tietoverkolla¹. Kehityssuunnasta käytetään nykyisin myös nimi-tyksiä Web of Data ja Linked Data (Heath, Bizer, 2011). Ideana on, että yhteisten kansainvälisten standardien ja tällaisen tietojen verkon avulla voidaan yhdistää eri alojen tietosisältöjä yhä laajemmiksi ja rikkaammiksi kokonaisuuksiksi. Semanttisen webin käsitteverkkoa ymmärtävät paitsi ihmiset myös verkkosovellukset, ts. tietoko-neet, mikä mahdollistaa aiempaa älykkäämpien ja keskenään yhteentoimivampien (interoperable) verkkosovellusten kehittämisen. "Semanttinen" tarkoittaa tässä yh-teydessä tietoa, joka on esitetty sellaisessa muodossa, että sen tulkinta on mahdolis-ta tietokoneella hyvin määritellyllä tavalla (tässä käytetään hyväksi logiikkaa). Se-manttisen webin keskeisen komponentin ja "selkärangan" muodostavat yhdistetyn tiedon (Linked Data) sisällönkuvailussa käytettävät ontologiat, joiden kautta aineis-tot yhdistyvät toisiinsa.

Oletetaan, että etsit "helsinkiläisiä ravintoloita" tietojärjestelmästä. Perinteinen tie-donhakujärjestelmä toimii siten, että se vertailee hakutermejä "helsinki" ja "ravinto-la" tietokantaan tiedon tallennusvaiheessa indeksoituihin kuvailutermeihin ja pa-lauttaa hakutuloksena vastaavat tietueet. Haku perustuu merkkijonojen vertailuun;

¹ <http://www.w3.org/standards/semanticweb/>

jos hakusanaa vastaavaa merkkijonoa ei löydy tietueesta sopivasta paikasta saadaan tyhjä tulos. Jos esimerkiksi tietokantaan on talletettu indeksointivaiheessa vain "Kalliossa" oleva "pizzeria", ei se löydy hakusanoilla "Helsinki" ja "ravintola", ellei käyttäjä osaa muotoilla kyselyään tarkemmin kaupunginosa- ja ravintolatyypikkoh-taisesti, mikä on usein hankalaa. Tässä ei auta sanojen taivutusmuotojen tai syno-nyymien huomioiminenkaan. Myös kielimuurit ovat perinteisen merkkijonohaun haasteena: "ravintolalla" haettaessa ei "restaurant" löydy.

Ontologiaperustainen hakupalvelu sen sijaan ymmärtää tarjota tuloksena "Kalliossa" sijaitsevaa "pizzeriaa", koska indeksoinnin ja/tai haun apuna käytetään maailman käsitteitä kuvaavia ontologioita. Ne kertovat, että "Kallio" on osa Helsinkiä ja että "Kallio" tässä yhteydessä ei tarkoita luonnonmuodostelmaa tai presidentti Kyösti Kalliota. Samoin "pizzeria" voidaan tunnistaa eräänlaiseksi "ravintolan" alatyypiksi. Myös kielimuurit luhistuvat, sillä ontologinen indeksointi ja haku perustuvat kieliriippumattomiin tunnisteisiin eikä kielellisiin ilmauksiin. Ontologisten viittausten kautta tieto yhdistyy luontevasti toisiin tietoihin: esimerkissämme järjestelmä voi tarjota lisätietoa pizzakulttuurista Wikipedian kautta, mikäli hakupalvelussa käytetään samoja käsitteitä kuin Wikipedian ontologisessa versiossa DBpedia². Voidaan myös luoda linkki lähellä olevan teatterin Italia-aiheista revyytä käsittelevään artikkeliin, koska pizza on italialainen ruokalaji, tai tarjota matkapuhelimeen navigointipalvelu Kallioon.

Tällaisten älykkäiden palveluiden kehittämisen edellytyksenä on, että verkossa toimivien järjestelmien käytettävänä on yhteisesti sovittuja, tietokoneen "ymmärtämässä" muodossa määriteltäviä käsitteitä ja näiden välisiä suhteita, kuten käsiteluo-ka "kaupunginosa" ja siihen kuuluva yksilö "Kallio", joka on osa kaupunki-luokan yksilöä nimeltä "Helsinki". Käsitteitä ja niiden välisten suhteiden määrittelyverkkoja kutsutaan ontologioiksi

1.2 Kansallinen haaste: ontologiat

Suomessa semanttisen webin ontologioita alettiin kehittää laajamittaisemmin MuseoSuomi-järjestelmän³ yhteydessä v. 2002 alussa (Hyvönen et al., 2005). Työssä tarvittiin käsitteistöjä mm. museoesineille, materiaaleille, henkilöille, paikoille ja tapahtumille. Hankkeessa kävi heti ilmi, ettei käytettävissä ollut ontologioita ja ettei kansainvälisiä järjestelmiä voitu sellaisinaan hyödyntää suomen kielestä sekä kansallisista sisällöistä ja käytännöistä johtuen. Tästä havainnosta kypsyi ajatus siitä, että maamme tarvitsee avoimeen dataan perustuvan semanttisen webin ontologisen sisältöinfrastruktuurin, jonka varaan esimerkiksi MuseoSuomen kaltaiset julkiset ja yksityiset palvelut voitaisiin rakentaa kustannustehokkaasti. Työ voitaisiin perustaa eri aloilla sisällönkuvailua varten käytössä oleviin asia- ja muihin sanastoihin, kuten tässä tapauksessa Museoalan asiasanastoon MASA (Leskinen, 1998). Näin ontologi-

² <http://dbpedia.org/>

³ <http://www.museosuomi.fi/>

at olisivat heti yhteentoimivampia aiemmin indeksoitujen tietokantojen sisältöjen kanssa.

Kokemuksesta viisastuneena Suomessa käynnistyi v. 2003 semanttisen webin tietoinfrastruktuurihanke FinnONTO⁴, joka oli tiettävästi ensimmäinen kansallisen tason hanke maailmassa. Sen yhtenä ydintavoitteena oli eri alojen ydinsanastojen ontologisoiminen ja siltaaminen (align, link) eri alat ylittäväksi harmoniseksi kokonaisuudeksi, joka olisi kaikkien käytettävissä avoimena datana erityisten ontologiakirjastopalveluiden kautta. Hanketta toteuttamaan koottiin aluksi 14 yrityksen ja julkisen organisaation rahoittajakonsortio, ja kehitystyö käynnistyi Tekesin tuella. Vuosien kuluessa FinnONTO-hankkeen koko kasvoi suurimmillaan 39 rahoittajaan tehden projektista tiettävästi Tekesin historian laajimman hankkeen rahoittajakonsortion koolla mitattuna. FinnONTO-hanketta veti Aalto yliopiston ja Helsingin yliopiston Semanttisen laskennan tutkimusryhmä SeCo⁵, jonka suojissa myös pääosa tutkimustyöstä on tehty.

Webin muuttuminen vähitellen yhä semanttisemmäksi asettaa kansainvälisen haasteen standardointityölle. W3C-järjestön piirissä tehdyn työn keskeisimpinä tuloksina on otettu käyttöön joukko standardeja liittyen (meta)datan, sanastojen ja ontologioiden esittämiseen sekä päättelyyn ja tiedonhakuun⁶. Näistä tärkeimmät ovat Resource Description Framework RDF ja RDF Schema⁷, Simple Knowledge Organization System SKOS⁸ (Isaac, Summers, 2009), Web Ontology Language OWL⁹ ja SPARQL-kyselykieli¹⁰. Nämä standardit määrittelevät yleisiä, yhteentoimivuuden ja päättelyn kannalta keskeisiä tietomalleja, esitysmuotoja ja semanttisia periaatteita, joita noudattamalla semanttista webiä nykyisin rakennetaan. Esimerkiksi RDF Schema -suosituksessa määritellyllä subclassOf-suhteella voidaan ilmaista käsitettä yleisempi käsite, esimerkiksi että leijonat ovat kissaeläimiä.

W3C:n standardit eivät kuitenkaan ota kantaa ontologioiden tai metadataformaattien varsinaiseen sisältöön, vaan ovat luonteeltaan sovellusriippumattomia ja perustuvat logiikkaan. Esimerkiksi subclassOf-suhde mahdollistaa leijonan ja kissaeläimen välisen suhteen ilmaisemisen, mutta varsinaiseen eläinlajien taksonomian muodostamiseen ei oteta kantaa. Sisältö- ja ontologiatyö on jätetty kunkin alan asiantuntijatahojen ja sovellusten kehittäjien tehtäväksi. FinnONTO-hanke tarttui tähän haasteeseen tavoitteenaan kansallisen, monialaisen ontologiainfrastruktuurin kehittäminen maahamme ja teknologian pilotointi hyötysovelluksissa.

⁴ <http://www.seco.tkk.fi/projects/finnonto/>

⁵ <http://www.seco.tkk.fi/>

⁶ <http://www.w3.org/standards/semanticweb/>

⁷ http://www.w3.org/standards/techs/rdf#w3c_all

⁸ http://www.w3.org/standards/techs/skos#w3c_all

⁹ http://www.w3.org/standards/techs/owl#w3c_all

¹⁰ http://www.w3.org/standards/techs/sparql#w3c_all

1.3 Ontologiainfrastruktuurin osat

Ajatusta semanttisen webin ontologiainfrastruktuurista (Hyvönen, 2008) voi verrata tie-, sähkö- ja puhelinverkkojen muodostamiin perinteisiin infrastruktuureihin. Semanttisten käsitteiden verkosto on luonnollisesti abstrakti ja näkymätön, mutta mahdollistaa hieman vastaavaan tapaan yhteyksiä kuin vaikkapa tie- tai puhelinverkko. Visiona on, että jos FinnONTO-infrastruktuurin pelisäännöillä tuotettu sisältö julkaistaan verkossa, voidaan se kytkeä ja hyödyntää automaattisesti muiden toimijoiden semanttisissa verkostoissa hieman vastaavaan tapaan, kuin uusi maantienpätkä tieverkoston osana. Kantava ajatus on, että uuden sisällön arvo rikastuu "ilmaiseksi" infrastruktuurin avulla verkon muusta sisällöstä (vastaavasti kuin uuden tieosuuden arvo syntyy sen kytkeytymisestä muihin teihin), ja toisaalta muiden verkon julkaisijoiden sisältöjen arvo rikastuu uuden tiedonsirpaleen avulla (vastaavasti kuin uusi tie parantaa muiden teiden keskinäistä saavutettavuutta). Sekä itse infrastruktuuri että siinä jo oleva tieto voidaan hyödyntää toisissa sovelluksissa jopa sellaisenaan, mikä säästää merkittävästi järjestelmien kehityskustannuksia.

FinnONTO-hankkeessa kehitetyn ratkaisumallin mukaan sisältöinfrastruktuurin keskeisiä komponentteja ovat ontologiat, niiden hyödyntämistä varten tarvittavat ontologiapalvelut sekä ontologiajärjestelmän jatkuva ylläpitoprosessi:

- Ontologiat. Joukko toisiinsa sillattuja, kansalliseen käsitteistöön perustuvia ontologioita, joilla on siltauksia myös kansainvälisiin sanastoihin ja tietosisältöihin. Näin suomalaiset kulttuurisisällöt saadaan kuvattua semanttisesti ja kytkettyä sekä toisiinsa että kansainvälisiin aineistoihin ja palveluihin, kuten Linked Open Data -pilven datajoukkoihin¹¹.
- Ontologiapalvelut. Ontologiakirjastopalveluiden ONKI¹² kautta ajantasaiset ontologiat voidaan ottaa kustannustehokkaasti käyttöön eri organisaatioissa (Tuominen et al., 2009; Viljanen et al., 2009; Suominen et al., 2012). Tällainen palvelu kannattaa toteuttaa keskitetysti perusteluina kustannussäästöt, ajantasaisuus ja yhteentoimivuus. Keskitettyä ONKI-palveluita on pilotoitu useita vuosia ONKI Living lab -ympäristössä¹³ ja sen seurauksena esimerkiksi kymmenien museoiden luettelointijärjestelmät on jo kytketty ONKIin.
- Ontologiajärjestelmän ylläpitoprosessi. Systemaattinen ja koordinoitu kansallinen organisaatio ja mekanismi, joka kantaa vastuun eri aloilla tarvittavien sanastojen yhteisöllisestä kehittämisestä jatkossa. Tässä tarvitaan keskitettyä koordinoitua, mutta ontologiatyö kannattaa hajauttaa tarpeen mukaan sinne, missä siihen on paras asiantuntemus ja resurssit käytettävissä.

¹¹ <http://linkeddata.org/>

¹² <http://www.onki.fi/>

¹³ <http://www.onki.fi/>

Kansallinen ONKI-ontologiapalvelu on ollut verkossa vuodesta 2008 ja sillä on ollut n. 14 000 ihmiskäyttäjää kuukaudessa. Palvelun rajapintoja on rekisteröitynyt käyttämään yli 400 eri tahoa. Vuoden 2014 alusta ONKI-palvelun vakiintuneimmat ja käytetyimmät osat ja toiminnallisuudet siirtyivät Kansalliskirjaston ylläpitämäksi kansalliseksi palveluksi Finto-nimisenä¹⁴.

Ontologiainfrastruktuurissa tarvittavia ontologioita voidaan ryhmitellä seuraavasti (Hyvönen, 2013):

- 1) Yleiskäsiteontologiat. Nämä ontologiat vastaavat käsitteistöltään karkeasti nykyisiä asiasanastoja, kuten YSA¹⁵ ja MASA, ilman vapaita indeksoinnin termejä (esimerkiksi paikannimet). Yleiskäsitteiden määrä on kohtuullisen pieni ja lasketaan yleensä tuhansissa tai kymmenissä tuhansissa. Esimerkiksi museoalan tiettävästi laajimmassa yleiskäsitesanastossa, Getty-säätiön Art and Architecture Thesauruksessa¹⁶, on yli 51 000 käsitettä ja 269 000 termiä, jossa mukana on hyvinkin yksityiskohtaisia käsitteitä, esimerkiksi erilaisia tuolityyppejä.
- 2) Toimijaontologiat. Toimijaontologiat (Kurki, Hyvönen, 2009) ovat henkilö- ja organisaatiorekistereitä, ja ne muistuttavat kirjastoissa käytettyjä ns. auktoriteettitietokantoja. Toimijaontologian avulla samannimiset toimijat voidaan yksilöidä ja erottaa toisistaan eri tunnisteilla ja lisätietojen avulla, esimerkiksi henkilökaimat syntymävuoden ja -paikan avulla. Toimijaontologioissa voi olla erittäin paljon yksilöitä, mutta luokkarakenne on yksinkertainen ja perustuu yleiskäsiteontologioihin. Esimerkiksi taiteilijoihin keskittyvässä Getty-säätiön Union List of Artist Names (ULAN) -sanastossa¹⁷ on n. 120 000 toimijaa (kuten Akseli Gallen-Kallela), joilla on n. 293,000 erilaista nimeä, ja toimijoita on luokiteltu kansallisuuden ja ammatin mukaan, joiden määrä lasketaan sadoissa.
- 3) Paikkaontologiat. Paikkaontologiat vastaavat kansallisten maanmittauslaitosten ylläpitämiä paikannimirekistereitä. Niiden avulla voidaan yksilöidä paikat, sijoittaa ne koordinaatistoon ja tallentaa paikkoihin liittyviä lisätietoja kuten paikkatyyppi (kylä, järvi, asema jne). Paikkaontologioissa voi olla erittäin paljon yksilöitä (miljoonia), mutta luokkarakenne on toimijaontologioiden tapaan yksinkertainen ja perustuu yleiskäsitteisiin. Esimerkiksi Getty-säätiön Thesaurus of Geographical Names (TGN) -sanastossa¹⁸ on n. 912 000 historiallista paikkaa ja Geonames-sanastossa¹⁹ yli 8 miljoonaa nykypaikkaa.
- 4) Aikaontologiat. Monella alalla keskeistä on ajan esittäminen. Lineaarisen kalenteriajan ohella voidaan viitata myös päivän ja vuodenkierron aikoihin (kevät, aamu) sekä nimettyihin aika- ja tyylikausiin (esim. rautaaika, valistuksen aika, art deco). Aikakäsitteitä ja -jaksoja ei ole kovin

¹⁴ <http://finto.fi/>

¹⁵ <http://vesa.lib.helsinki.fi/ysa/>

¹⁶ <http://www.getty.edu/research/tools/vocabularies/aat/>

¹⁷ <http://www.getty.edu/research/tools/vocabularies/ulan/>

¹⁸ <http://www.getty.edu/research/tools/vocabularies/tgn/>

¹⁹ <http://www.geonames.org/>

suuria määriä. Ajan esittämisen haasteena on mm. epätasallisen (historiallisen) aikatiedon esittäminen ja tulkinta sekä epäsäännöllisten syklisten aikojen esittäminen, kuten esimerkiksi aukioloajat.

- 5) Tapahtumaontologiat. Tapahtumat liittävät toisiinsa henkilöitä, paikkoja, aikoja ja toisia tapahtumia, ja mahdollistavat kulttuuristen historiallisten ilmiöiden sekä kokoelmien provenienssitietoon liittyvien tietojen kuvaamisen semanttisesti yhteentoimivalla tavalla. Tästä johtuen mm. CIDOC-CRM-järjestelmän²⁰ perustaksi on valittu tiedon esittäminen tapahtumina, eikä esimerkiksi Dublin Coren²¹ dokumenttiperustaista tietomallia. Jotta kaksi sisältökohdetta voidaan yhdistää toisiinsa yhteisen tapahtuman kautta, pitää olla käytössä keino viitata tapahtumiin yksikäsitteisesti samalla tavalla kuin henkilöihin tai paikkoihin. Tapahtumaontologiaan (Hyvönen, et al., 2012) tallennetaan tähän tarvittava viitetieto historiallisista tapahtumista, kuten "toinen maailmansota" tai "Napoleonin kruunajaiset".
- 6) Nimistöontologiat. Monilla tieteenaloilla on käytössä laajoja nimistöjä, joita voidaan liittää yleisontologioihin. Tällaisia ovat esimerkiksi biologian eliöiden lajilistat, geologian mineraalit, lääketieteen taudit ja lääkeaineet, kielitieteen kielet yms. Koska tällaisia yksityiskohtaisia käsitteitä saattaa olla käytössä suuria määriä ja niiden käyttö fokusoituu erikoisaloille, on tarkoituksenmukaista toteuttaa nimistöontologioita erillisinä ontologioina, jotka on kuitenkin sillattu (ripustettu) yleisempään käsitteistöön. Esimerkiksi FinnONTO:n YSO-ontologian käsite "linnut" laajenee kattavaksi, yli 10 000 maailman lintulajin taksonomiaksi AVIO-nimistöontologian kautta (Tuominen et al., 2013).

Tässä raportissa käsitellään vain FinnONTO:n yleiskäsiteontologioihin liittyvää ontologiatyötä. Laajemman kuvan hankkeesta julkaisuineen saa helpoiten hankkeen kotisivun²² kautta.

²⁰ <http://www.cidoc-crm.org/>

²¹ <http://dublicore.org/>

²² <http://www.seco.tkk.fi/projects/finnonto/>

2 Ontologiatyön tavoitteet, periaatteet ja rajoitteet

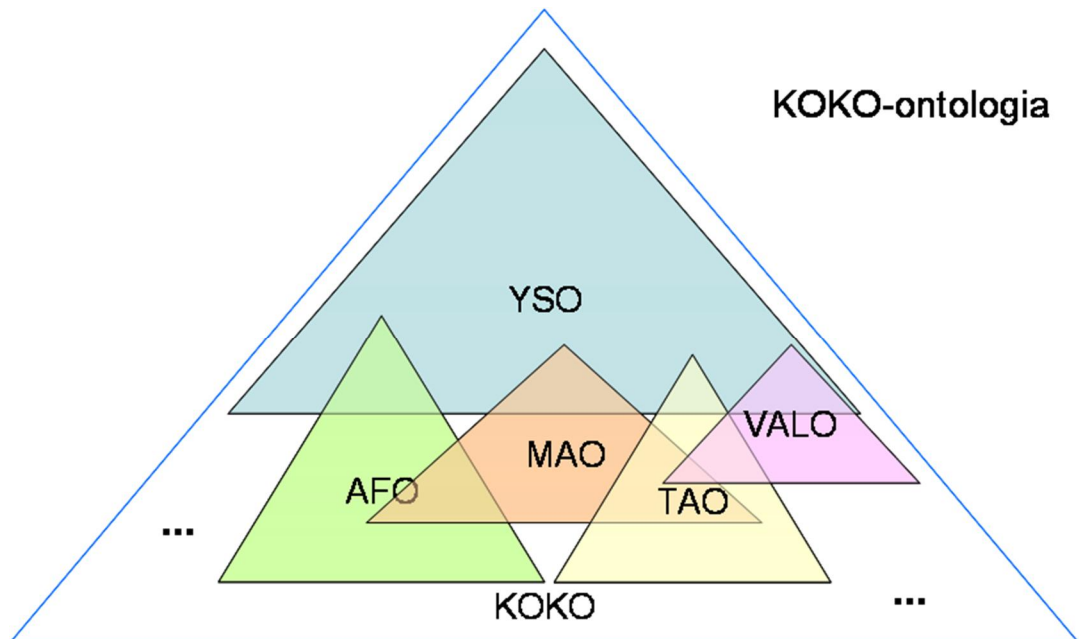
Tässä luvussa esitellään lyhyesti FinnONTO:n ontologisointityön tavoitteet, käytetyt yleiset periaatteet ja rajoitteet, joita on sovellettu KOKO-ontologiapilveen kuuluvien ontologioiden muuntamisessa asiasanastoista. Nämä on syytä arvioida ja pitää mielessä ontologioihin tutustuttaessa, kun nyt vasta alkuun saatettua laajaa työtä jatketaan kehittämällä nykyisiä ontologioita, KOKO:n rakennetta sekä lisäämällä uusia ontologioita KOKO-pilveen.

2.1 KOKO - joukko toisiinsa yhdistettyjä eri alojen ontologioita

Maamme käytetyin asiasanasto Kansalliskirjaston YSA sisälsi FinnONTO-hankkeen alussa vuonna 2003 reilusti yli 20 000 käsitettä eri aloilta. Sanaston laajentaminen oli edelleen tarpeen, mutta kaikkien eri alojen asiantuntemuksen saaminen sanastotyöhön oli haasteellista, vaikkakin ylläpitoa varten oli olemassa monialainen työryhmä. Samaan aikaan maassamme oli käytössä ja kehitettiin lukuisia YSA:n kanssa erityisesti yleisempien käsitteiden osalta päällekkäisiä sanastoja, mikä tuntui monesti tarpeettomalta moninkertaiselta työltä. Yhtenä lisähaasteena oli, että samanimisillä käsitteillä saattoi olla eri sanastoissa erilainen merkitys.

YSA:n kehitystyön jatkuessa perinteiseen tapaan FinnONTO tarttui näihin sisällöllisiin ja organisatorisiin haasteisiin esittämällä kansallisesti koordinoitua mallia, jossa sanastotyö voitaisiin yhteistuumiin jakaa eri alojen sanastojen kehittäjäryhmien kesken ja yhdistää tulokset lopulta yhdeksi laajaksi, eri alat kattavaksi ontologiaksi. Vision mukaan tämä edistäisi merkittävästi eri alojen tietosisältöjen semanttista yhteentoimivuutta ja säästäisi samalla kehityskustannuksia päällekkäisen työn vähentyessä.

Työn tuloksena syntyi prototyyppi KOKO-ontologiasta, joka on eri alojen ontologioiden muodostama Linked Data ontologiapilvi. Sen ytimenä on YSAsta kehitetty YSO-ontologia, joka muodostaa KOKO:n yläontologian (upper ontology, top ontology) sisältäen merkitysaltaan laajimmat käsittehierarkian käsitteet. Erikoisalojen tarkempi käsitteistö "ripustuu" sitten YSO:n eri haaroihin hierarkioita syventäen. Kuva 2.1 havainnollistaa mallia. FinnONTO-hankkeen aikana YSO:n perustana olleeseen YSAan tehtiin Kansalliskirjaston toimesta tuhansia muutoksia, joita on sitten päivitetty YSO-ontologiaan FinnONTO:n tutkijoiden toimesta.



Kuva 2.1. Yhteisöllinen kokonaisontologia KOKO koostuu yläontologiasta YSO ja sitä tarkentavista alaontologioista, joilla on yhteisiä käsitteitä YSO:n ja toisten alaontologioiden kanssa. Ontologioiden keskinäiset leikkaukset kuvassa ovat vain viitteellisiä ja tarkoitettu yleiskuvan havainnollistukseksi.

KOKO-ontologiapilveen kuului v. 2013 lopulla taulukossa 2.1 luetellut 15 ontologiaa, yhteensä yli 71 000 käsitettä, ja uusia ontologioita ollaan ottamassa mukaan.

KOKO-ONTOLOGIAAN KUULUVAT ONTOLOGIAT

Yleinen suomalainen ontologia	n. 26 000	Yleinen suomalainen asiasanasto
Julkishallinnon ontologia (JUHO)	n. 6 350	Valtioneuvoston asiasanasto
Kaunokirjallisuuden ontologia (KAUNO)	n. 5 100	Kaunokki-asiasanasto
Kielitieteen ontologia (KTO)	n. 950	Uralistiikan tutkimuksen bibliografian asiasanaluettelo
Kirjallisuudentutkimuksen ontologia (KITO)	n. 850	Kirjallisuudentutkimuksen asiasanasto
Kulttuurien tutkimuksen ontologia (KULO)	n. 1 500	Kulttuurien tutkimuksen asiasanasto
Liiketoimintaontologia (LIITO)	n. 3 400	Yrityssuomi.fi-portaalin käsitteistö
Merenkulkualan ontologia (MERO)	n. 1 400	Merenkulkualan asiasanasto
Musiikin ontologia (MUSO)	n. 1 000	Musiikin asiasanasto
Puolustushallinnon ontologia (PUHO)	n. 2 000	Puolustushallinnon asiasanasto
Taideteollisuusalan ontologia (TAO) / Museoalan ontologia (MAO)	n. 8 100	Muotoilun ja viestinnän asiasanasto / Museoalan asiasanasto
Terveystieteiden ja hyvinvoinnin ontologia (TERO)	n. 6 500	TESA, Stameta-asiasanasto, n. 2500 MeSH-käsitettä (Medical Subject Headings)
Valokuvausalan ontologia (VALO)	n. 2 000	Valokuvan asiasanasto
Viikin tiedekirjaston ontologia (AFO)	n. 6 000	Agriforest-asiasanasto

Taulukko 2.1. Suomalaisista asiasanastoista tuotettuja FinnONTO-ontologioita, yhteensä n. 71 150 käsitettä, jotka on sillattu (align) toisiinsa yläontologian YSO kautta KOKO-ontologiapilveksi.

Lähes jokainen KOKO-pilven ontologia perustuu suoraan maassamme jo aiemmin kehitettyyn asiasanastoon, joka on ontologisoitu tässä raportissa esitettävällä tavalla.

KOKOn jakoperuste eri aloihin ja sanastoihin on siis puhtaasti historiallinen. On ilmeistä, ettei alojen tai työryhmien jako taulukon mukaisesti ole optimaalinen, vaan esimerkiksi monia pienempiä toisiinsa liittyviä ontologioita kannattaa jatkossa yhdistää laajemmiksi kokonaisuuksiksi. FinnONTO-hankkeen puitteissa tähän ei kuitenkaan haluttu ottaa kantaa vaan ainoastaan koota eri alojen väki yhteen, jotta prosessi työn organisoinniseksi saataisiin käyntiin ensimmäisen prototyypin pohjalta.

FinnONTOssa ei myöskään kajottu KOKO-ontologioiden päällekkäisyyksien poistamiseen, joka jätettiin niin ikään sanastojen vastuutahojen asiaksi. Ontologiapilven analyysiä ja hallintaa varten kehitettiin kuitenkin erilaisia työkaluja (Pessala et al., 2011; Frosterus et al., 2013).

KOKOn osaontologiat on tätä kirjoitettaessa alustavasti sillattu vain yhteiseen YSO-ontologiaan pareittain siten, että luokkahierarkioiden toimivuus on tarkistettu yli ontologiarajojen (esimerkiksi JUHO+YSO, KAUNO+YSO jne.). Siltausta varten kullekin ontologiaparille luotiin oma Protégé-ontologiaeditorin²³ projekti, jossa oli mukana YSO ja siihen sillattava erillisontologia omissa nimiavaruuksissaan. Tällaiset yhdistelmät on julkaistu ONKI-palvelussa. Poikkeuksena kokeiltiin myös sillata toisiinsa läheisesti liittyvä ontologiakolmikko YSO, MAO ja TAO omana projektinaan. Silloin haasteeksi muodostuivat käsitteiden monimutkaistuvat riippuvuudet eri ontologioiden välillä.

Jatkokehityksen kannalta yhtenä keskeisenä kohteena tämänhetkisisissä siltauksissa ovat sellaiset osaontologioiden väliset yhteiset käsitteet, jotka eivät ole mukana YSOssa. Näiden käsitteiden hierarkiat on tarkistettu erikseen osaontologioiden ja YSON suhteen, mutta tarkistamatta on, onko ontologiarajan yli muodostuviin uusiin hierarkkisiin polkuihin voinut syntyä missä määrin virheitä. Työhypoteesina on ollut, että tällaisten riippuvuuksien eliminointi voisi toimia jatkossa metodologisena periaatteena KOKO-pilven jatkokehityksessä: osa-ontologioiden keskenään jakamat käsitteet kannattaisi luultavasti siirtää YSOon, koska

- tällaiset käsitteet ovat useimmiten merkitysalaltaan yleisiä (niitähän käytetään ainakin kahden osaontologian alalla) ja
- käsitteen hallinta kahdessa eri ontologiassa on hankalampaa kuin yhdessä.

Tavoitteena olisi tämän mukaan luoda KOKO-pilvi, jossa:

- ei olisi moniperintää (multiple inheritance) eri osa-ontologioiden kautta
- osaontologioiden väliset riippuvuudet perustuvat vain YSON kautta jaettuihin yhteisiin käsitteisiin (muuta jaettuja käsitteitä ei olisi)

Poikkeustapaukset voitaisiin käsitellä erikseen. Tämän työhypoteesin testaamista ehdotamme jatkossa KOKO-pilveä kehittäväälle ONKI-konsortiolle. Aivan selvää ei ole, millaisia haasteita ylläpidon minimointi tällä tavoin mahdollisesti aiheuttaa osaontologioiden sisältötyölle.

²³ <http://protege.stanford.edu/>

2.2 Yhteentoimivuus olemassa oleviin asiasanastoihin

Maassamme on vahva traditio käyttää sisältöjen kuvailussa ja tiedonhaussa hyväksi asiasanastoja, joten niiden avulla indeksoitua tietoa on runsaasti saatavilla. FinnONTOssa lähdettiin siksi hyödyn-tämään jo käytössä olevia asiasanastoja, ja KO-KO-ontologiapilven lähes kaikki 15 ontologiaa perus-tuvat suoraan olemassa oleviin asiasanastoihin. FinnONTO-hankkeen tavoitteena oli vain muokata niiden raken-etta ja kehittämisperiaatteita semanttisen webin aikakautta ja sovelluksia varten.

Yhteentoimivuuden tavoite tarkoitti sitä, että työn aikana kaikki aiemmat asiasanas-tojen sanat py-rittiin sijoittamaan uuteen ontologiseen struktuuriin tietoa hävittä-mättä. Mikäli työ olisi aloitettu puhtaalta pöydältä, lopputulos näyttäisi varmasti erilaiselta, sillä uusien periaatteiden soveltaminen toisilla periaatteilla muodostet-tuun sanastoon ei aina ole suoraviivaista eikä valitettavasti aina johda luontevim-paan ratkaisuun. Sama konservatiivinen periaate oli käytössä myös uusien termien lisäämisessä: ontologioita täydennettiin lähtökohtaisesti vain pakon edessä, kun se tuntui vält-tämättömältä. Tavoitteena ei siis ollut kehittää sanastoja eteenpäin sisäl-lön osalta vaan rakenteistaa olemassa olevia sanastoja hyödyllisempään muotoon. Työn aikana sanastoihin on kuitenkin tullut paljon lisäyksiä erityisesti silloin, kun monimerkityksisiä sanoja on jouduttu erittelemään saman sanan eri merkitysten perusteella. Esimerkiksi YSAssa on käsite "lapset", joka voi tarkoittaa mm. ikäluok-kaa tai perhesuhdetta – aikuisetkin ovat omien vanhempiensa lapsia. Mikäli merki-tysten erottelua ei tavalla tai toisella tietojärjestelmässä tehdä, saattaa se esimerkiksi suositella aikuisille tarpeettomasti lastenkirjoja luettavaksi tai käyntiä lastenlääkä-rillä.

Ontologioihin on lisätty myös jonkin verran ryhmitteleviä käsitteitä, joiden avulla hierarkioista on pyritty tekemään ihmisen kannalta helpommin jäsennettäviä. Näitä käsitteitä ei ole tarkoitettu käytettäväksi sisällönkuvailussa. Luontevat jaottelut li-säävät ontologian käyttökelpoisuutta esi-merkiksi sovellusten käyttöliittymissä ku-ten nykyisin paljon käytetyissä fasettihakujärjestelmissä (faceted search). Työ on tältä osin kuitenkin vasta saatettu alkuunsa. Joillain käsitteillä voi olla kymmeniä alakäsitteitä, joita kannattaisi jatkossa ryhmitellä edelleen pienemmiksi kokonai-suuk-siksi. FinnONTO-hankkeessa haluttiin kuitenkin tietoisesti vältellä tällaista sisältötyötä ja päätök-sentekoa, josta vastuu kuuluu luontevimmin kutakin sanastoa ylläpitävälle asiantuntijayhteisölle.

Eräiden ontologioiden osalta projektin käytettävissä oli alan asiantuntija ja päätet-tiin samassa yhteydessä laajentaa myös itse ontologiaa systemaattisesti. Esimerkiksi museoalan asiasanastoa MASA laajennettiin n. tuhannella uudella, erityisesti tekstii-lialan käsitteellä Museoalan ontologiaksi MAO (Kettula, 2009). Tällaisissakin tapa-uksissa uusien käsitteiden yhteentoimivuus olemassa olevaan sanastoon voitiin säi-lyttää, koska uudet termit linkittyvät semanttisessa verkossa aiempiin. Tärkeää on,

ettei vanhentuneita käsitteitä poisteta kokonaan ontologiasta, korkeintaan ne kannattaa siirtää rakenteessa syrjään indeksointikäytöstä.

Yksi haaste yhteentoimivuuden kannalta ontologisoinnissa on, että kun käsitteet esitetään uudessa struktuurissa, joka viime kädessä määrittelee niiden merkityksen, käsitteiden merkitykset väistämättä muuttuvat jonkin verran ontologisoinnin seurauksena. Tätä ongelmaa yritettiin lieventää seuraavilla ratkaisumalleilla:

1. Linkitys alkuperäisiin sanastoihin. Ontologiat on tarvittaessa linkitetty alkuperäisiin sanastoihin, jotka on muunnettu suoraan RDF(S) tai SKOS-muotoon ja ovat myös saatavilla. Näin selviää tarvittaessa termien alkuperäinen merkitys.
2. Koostekäsitteet. Monimerkityksiset sanat, joiden taustalla olevat käsitteet on eritelty ontologiassa, on esitetty ontologiassa monesti ns. koostekäsitteinä. Näiden merkitys on kuvattu linkittämällä ne käsitteisiin, joilla on yhteinen nimitys. Esimerkiksi YSAn "lapset" on edelleen YSOssa, vaikka sitä ei ole mahdollista sijoittaa luokkahierarkiaan. Sen sijaan tätä vastaava koostekäsite, joka voi joskus olla hyödyllinen, on kuvattu hierarkian ulkopuolisena käsitteenä. Näin esimerkiksi hakusanaa "lapset" käytettäessä voidaan tunnistaa sanan monimerkityksisyys, kysyä kenties käyttäjältä tarkennusta ja hakea tietokannasta sekä aiemmalla "lapset"-termillä että uudemmilla tarkemmilla merkityksillä (kuten "lapset (perheenjäsenet)") indeksoitua tietoa.

2.3 Keskittyminen ensin olennaisimpiin ontologisiin periaatteisiin

FinnONTOssa työn alla olleet asiasanastot sisälsivät yli 71 000 käsitettä laskematta mukaan käännöksiä eri kielille (vrt. taulukko 2.1). Koska Tekesin pääosin rahoittamassa teknologialähtöisessä FinnONTO-tutkimushankkeessa oli käytettävissä vain vähän resursseja sisältötyöhön, päätettiin asiasanastojen ontologisoinnissa keskittyä metodologian ja teknisten ratkaisujen kehittämiseen, joita sitten pilotoitaisiin valituissa asiasanastoissa. Ensimmäiseksi kohteeksi otettiin asiasanastoista suurin ja käytetyin eli Kansalliskirjaston Yleinen Suomalainen Asiasanasto YSA, joka muutettiin Yleiseksi Suomalaiseksi Ontologiaksi YSO. Jo aiemmin oli MuseoSuomi-hankkeessa kokeiltu Museoalan asiasanaston MASA ontologisointia. Sisältötyössä keskityttiin vain olennaisimpien ontologisten muunnosten tekoon, esimerkinomaiseen pilotointiin ja innostamaan sanastoja kehittäviä yhteisöjä jatkamaan työtä. Työn jatkuvuuden takaamiseksi - asiasanaston ja ontologian kehittäminen on jatkuva prosessi - voitaisiin suunnitelman mukaan hyödyntää asiasanastojen kehittämi-

seen jo käytössä olevia resursseja siirtämällä niitä kehittämään perinteisten asiasanastojen sijasta ontologisoituja asiasanastoja.

Suomessa käytettävien asiasanastojen rakenne on yksinkertainen, ja niiden muuttaminen syntaktisesti ja teknisesti ontologiseen W3C:n standardien mukaiseen muotoon on periaatteessa yksinkertaista: asiasanat voidaan muuttaa luokiksi (class) tai näiden yksilöiksi (instance, individual) ja asiasanojen väliset suhteet näitä yhdistäviksi ominaisuuksiksi (property). Lisäksi on sovittava käytännöistä, joilla näin syntyvän semanttisen verkon solmut ja kaaret eli resurssit (resource) nimetään yksikäsitteisesti URI-tunnisteilla (tai laajempaa merkistöä käyttävillä IRI-tunnisteilla). Nykyiset W3C:n formaatit ontologioiden esittämiseen ovat RDF(S), OWL ja SKOS. Projektin alussa vuonna 2003 RDF(S) ja OWL eivät olleet vielä lopullisesti standardeitua, ja nykyisin hyvin suosittu, RDF(S)- ja OWL-perustainen SKOS-formaatti oli täysin tuntematon. Formaattiksi valittiin tuolloin eniten käytetty, vuonna 2004 standardeitua RDF(S)-muoto, jonka käsittelemiseen oli myös saatavilla laajassa käytössä oleva ontologiaeditori, Stanfordin yliopiston Protégé. Hankkeen loppuvaiheessa käyttöön otettiin myös SKOS, kun sen standardi valmistui.

Varsinaisena haasteena ontologisoinnissa on asiasanaston semanttisen rakenteen täsmentäminen niin, että siinä toteutuvat koneellisessa päättelyssä, esimerkiksi kyselyn laajentamisessa, tarvittavat periaatteet. Keskeisimmät ontologiset muutokset ja tarkistukset, joihin työssä keskityttiin, olivat:

1. Luokkahierarkian täydentäminen. Ontologisen luokkahierarkian muodostaminen ja täydentäminen niin, että kaikilla käsitteillä (kaikkein ylin pois lukien) on ainakin yksi yläkäsite. Tämä parantaa käsitteiden linkitettävyyttä ja mahdollistaa mm. päättelyä ja kyselynlaajentamista.
2. Yksilö-luokka-suhteiden transitiivisuuden tarkistaminen. Keskeinen ontologioissa käytetty periaate luokkahierarkian muodostamisessa on, että kaikkien luokkien yksilöiden tulee olla samanaikaisesti *kaikkien* yläluokkiensa yksilöitä. Tämä ominaisuus on hyödyllinen koneellisessa päättelyssä, esimerkiksi ominaisuuksien periytyemisessä ja kyselyn laajentamisessa.
3. Monimerkityksisten asiasanojen taustalla olevien käsitteiden erottelu. Asiasanastojen monimerkityksisten käsitteiden merkitysten jakaminen eri käsitteiksi tarpeen mukaan niin, että ne voidaan sijoittaa luokkahierarkiaan.
4. Laajempi/suppeampi termi -suhteiden tarkentaminen. Asiasanastoissa käytettyjen LT/ST-suhteiden tarkentaminen yläluokkasuhteiksi (rdfs:subClassOf), osa-kokonaisuus-suhteiksi tai assosiativisiksi suhteiksi. Työssä keskityttiin vain subClassOf suhteiden kehittämiseen, mutta olemassa olevien suhteiden tarkennuksen seurauksena ontologioihin lisättiin myös

osa-kokonaisuus- ja assosiatiivisia suhteita, jotta olemassa olevaa tietoa ei olisi hävitetty.

Työ eteni niin, että ensin sanasto muutettiin automaattisesti RDF(S) muotoon, jonka jälkeen tarkistukset ja korjaukset tehtiin käsityönä Protégé-editorilla.

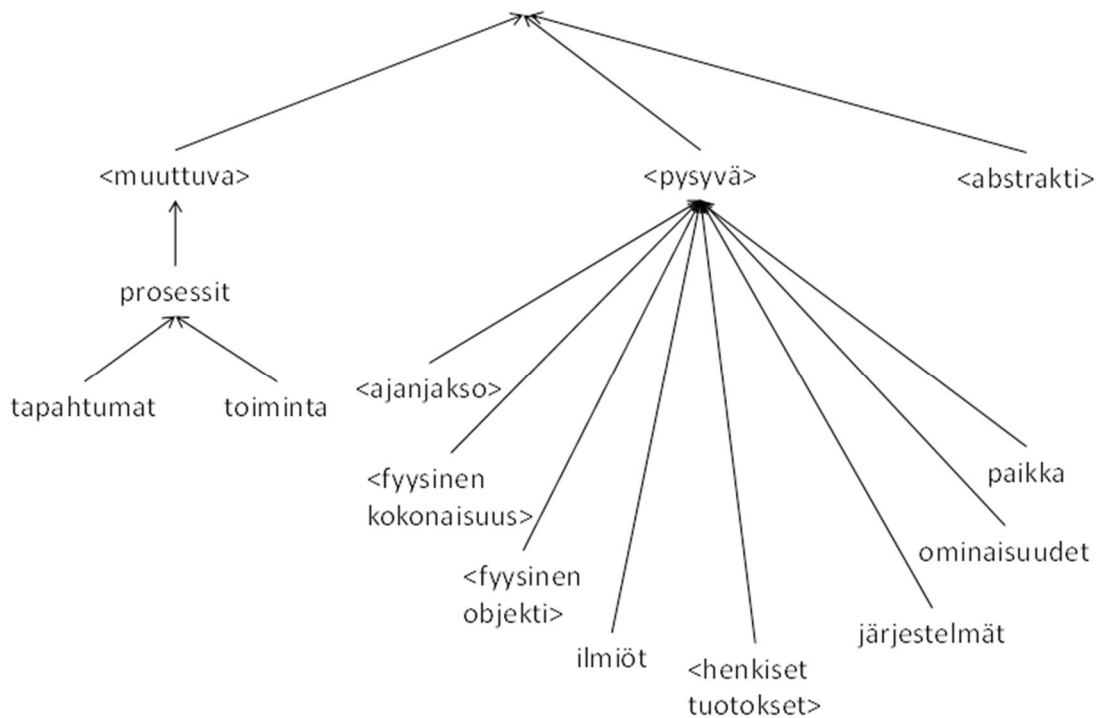
Seuraavassa kuvataan ja perustellaan hieman tarkemmin tässä työssä tehtyjä yleisiä valintoja ja käytettyjä periaatteita.

2.4 Luokkahierarkia ja sen täydentäminen

Jotta kaikille käsitteille olisi mahdollista osoittaa oma yksilöllinen paikkansa hierarkiassa, järjestelmään luotiin asiasanastoista puuttuva pääluokkarakenne, joka yhdistää osittain toisistaan irrallaan olevat käsitteet yhdeksi kokonaisuudeksi. Tämä rakenne lisättiin YSO/Allsoon, joka muodostaa KOKO-järjestelmän yläontologian (upper ontology). Tarpeen luokkarakenteen muodostamiselle on määrittänyt YSA/Allärsin käsitteistö, joka laajana ja eri erikoisalojen käsitteitä kattavana aineistona sopii hyvin hierarkian toimivuuden testaamiseen.

YSO:n luokkarakenteen ylimmän tason kolmijakomalli (kuva 2.2) on sama kuin hieman aiemmin julkaistussa alkuperäisessä DOLCE-mallissa (Gangemi et al., 2002), mutta alemmilla hierarkian tasoilla YSO/Allso muodostaa oman mallinsa, joka on muotoutunut käytännön ontologisointityön tarpeiden perusteella sijoitettaessa Yleisen suomalaisen asiasanaston käsitteitä yhtenäiseen hierarkiaan. YSO:n ei varsinaisesti toteuta DOLCE-mallia²⁴, mutta se on saanut vaikutteita siitä mm. ylärakenteen keskeisimpien luokkien osalta. Käsitteille on pyritty luomaan yhtenäinen hierarkia, jossa luokkien välillä olisi mahdollisimman vähän päällekkäisyyksiä (esim. *kulttuuriset*, *luonnontieteelliset*, *mentaaliset*, *taloudelliset* ja *yhteiskunnalliset ilmiöt*), mutta jossa luokat muodostavat kattavan kokonaisuuden myös kullakin hierarkian alemmalla tasolla niin, ettei niiden ulkopuolelle jäisi epätarkemmin luokiteltuja käsitteitä (*muut ilmiöt*).

²⁴ <http://www.loa.istc.cnr.it/old/DOLCE.html>

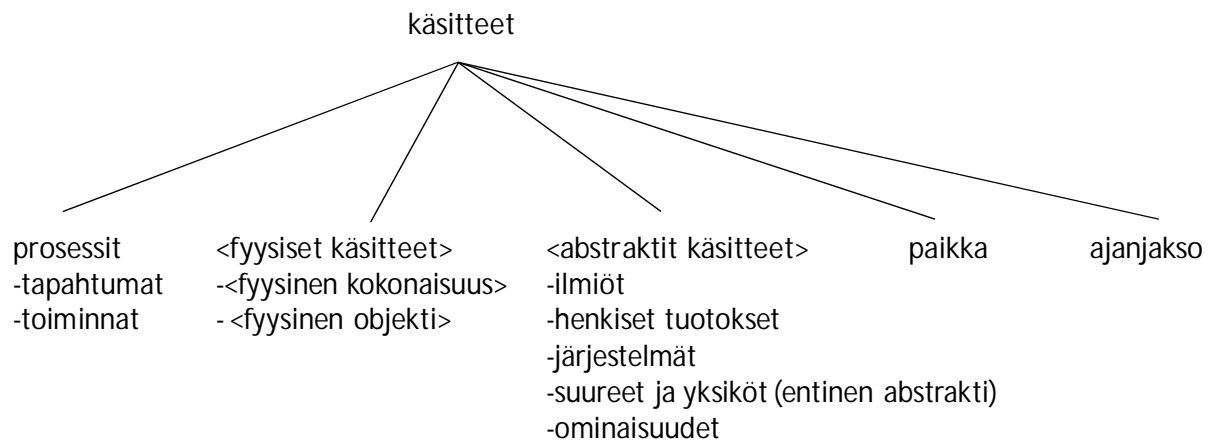


Kuva 2.2. YSO:n pääluokkajako perustuen alkuperäisen DOLCE-malliin

Ylimmällä eritellyllä tasolla pääluokkarakenteen luokat jakautuvat alkuperäisessä DOLCE-mallissa luokkiin *abstrakti*, *pysyvä* ja *muuttuva*. Näitä ylimmän tason luokkia ei ole tarkoitettu käytettäväksi aineistojen kuvailussa, sillä ne ovat liian laajoja tiedonkuvailussa ja -haussa käytettäväksi. Sama koskee myös muita pääluokkia, elleivät ne ole mukana myös YSA/Allärsissa. Kuvassa on erotettu kulmasulkeilla ne pääluokat, joita ei ole tarkoitettu käytettäväksi asiasanoina.

Vaikka pääluokkia ei ole tarkoitus käyttää asiasanoina, niillä on kuitenkin käyttöä erityisesti silloin, kun ohjataan indeksoijaa tai tiedonhakijaa ylhäältä alas (top-down) kohti merkitysalaltaan tarkempia luokkia. DOLCE-mallin ontologisiin tarkasteluihin perustuva kolmijako ei ole välttämättä ontologian loppukäyttäjän kannalta luontevin vaihtoehto. Pääluokkajaottelusta laadittiin SeCo-tutkimusryhmässä siksi myöhemmin vuonna 2013 myös toinen, käytön kannalta luontevampi versio, joka on esitetty kuvassa 2.3. Siinä on viisi pääluokkaa: *prosessit*, *fyysiset käsitteet*, *abstraktit käsitteet* sekä paikkoihin ja aikoihin liittyvät käsitteet. Aiemmalle pääluokalle *abstrakti* on annettu sen sisältöä paremmin kuvaava nimi *suureet ja yksiköt*. Intuiutiivisuutensa ohella tämä jaottelu myös heijastelee luvussa 1.3 esiteltyjä ontologioiden perustyyppisiä tapahtumaontologioita (prosessit), paikkaontologioita (paikka) ja aikaontologioita (ajanjakso). Esimerkiksi erilaiset prosessiluokat voisivat muodostaa luokkahierarkian, jonka avulla on mahdollista luoda tapahtumaontologian ydinaineistona olevia yksittäisiä tapahtumia.

Kuvan 2.3 jaottelu pitää sisällään samat luokat kuin kuvan 2.2 versio ja on luonteeltaan kosmeettinen, sillä ainoastaan ylätasen ryhmittelyä on muutettu käyttäjän kannalta intuitiivisempaan muotoon. Uutta rakennetta ei otettu SeCo-ryhmän toimesta käyttöön, koska samaan aikaan oli jo käynnistynyt ontologioiden siirtäminen Kansalliskirjaston ylläpidettäviksi alkuperäisen rakenteen pohjalta. Uskoaksemme YSO:n rakennetta tullaan jatkossa kehittämään merkittävästi hierarkian kaikilla tasoilla, monesti juuri nykyisten jaotteluiden tekemiseksi intuitiivisemmiksi loppukäyttäjille. Samassa yhteydessä kannattaa pohtia myös ylärakenteen jaotteluja.



Kuva 2.3. Käyttäjälähtöinen pääluokkajako

Pääluokkarakenteen lisäksi ontologiaan on hierarkkisen rakenteen tarkentamiseksi lisätty myös apukäsitteitä, jotka voidaan nimetä yleis- tai ammatti/erikoiskielen käytössä olevalla sanalla tai sanaliitolla, mutta joita ei ole tarvittu asiasanoituksessa, eikä siksi sisällytetty asiasanastoon. Esimerkiksi *asema (ilmiöt)* -käsite on lisätty hierarkkiseksi yläkäsitteeksi muun muassa käsitteille *ammattiasema*, *sosiaalinen asema* ja *yhteiskunnallinen asema*.

2.5 Luokkien ja yksilöiden erottaminen

Tutkimissamme suomalaisissa asiasanastossa ei tehdä eroa luokkien ja yksilöiden välillä, vaikka mahdollisuus standardeissa periaatteessa on olemassa (Aichison et al., 2000). Esimerkiksi YSAn suhteessa

Halley'n komeetta LT komeetat

Halley'n komeetta on samanlainen asiasana kuin *komeetat*, vaikka ontologisessa mielessä kyse on yhdestä erityisestä komeetta-yksilöstä. YSA-sanasto sisältää n. 6000 paikannimeä, joilla viitataan yksittäisiin paikkoihin ja jonkin verran muitakin yksilökäsitteitä, kuten *Pähkinäsaaren rauha*.

Yksilöiden ja luokkien välisen eron tekeminen ontologiassa voi olla haastavaa ja tulkinnanvaraistakin, mutkistaa rakenteita, ja erottelussa tehdyt tulkinnat saattavat haitata ontologian uudelleen käyttöä tilanteissa, joissa haluttaisiin tehdä toisenlainen luokka-yksilö-tulkinta. MuseoSuomi-hankkeen yhteydessä esimerkiksi pohdittiin, olisivatko materiaalit kuten *teräs* tai *villa* yksilöitä vai luokkia. Jokainen museokokoelman villatakki on samaan luokkaan kuuluvaa materiaalia, mutta kuitenkin materiaali ei ole fyysisesti samaa vaan esimerkiksi villa on keritty eri eläimistä. Eri villatakkien materiaalit voidaan ja itse asiassa pitääkin päättelykoneiden parhaiten hyödynnettävissä olevissa OWL-profiileissa kuvata *villa*-luokan yksilöinä, mikä kuitenkin tuntuu tarpeettoman monimutkaiselta verrattuna siihen, että materiaalikentän arvona olisi vain viittaus *villa*-luokkaan samaan tapaan kuin perinteisiä asiasanoja käytettäessä.

Erottelun tekemisestä onkin olemassa erilaisia käsityksiä ja käytäntöjä²⁵. Esimerkiksi nykyisessä SKOS-standardissa kaikki asiasanat muutetaan yleisen käsitelukan (Concept tai GroupConcept) yksilöiksi, eikä muodosteta omia luokkia kuten *villa*. RDF(S) ja erityisesti OWL-standardeissa taas luokkien ja yksilöiden käyttö päätelyssä on keskeistä ja hyödyllistä.

Koska kaikkiin tilanteisiin sopivaa patenttiratkaisua ei ole, FinnONTO-projektin alussa päätettiin jättää päätös luokkien ja yksilöiden erottamisesta ja käsittelystä sovelluksissa tehtäväksi vähäisimmän sitoutumisen ja joustavuuden periaatteiden mukaisesti (least commitment). Niinpä yksilö-luokka-jakoa ei ole tehty, vaan esimerkiksi *Halley'n komeetta* esitettiin luokkana samalla tavalla kuin *komeetat*. Ajatuksena kuitenkin on, että (yksilö)luokasta *Halley'n komeetta* voidaan tarvittaessa luoda uniikki yksilö. Näin ontologiaa voidaan tulkita tarpeen mukaan joko luokkahierarkian kautta tai vastaavaan tapaan kuin SKOS-standardissa asiasanaston sanojen rakenteen käyttökelpoisessa standardimuodossa esittävänä RDF-verkkona. Tätä voidaan sitten tulkita koneellisesti käyttötilanteen mukaan.

YSON osalta huomattava muutos YSAan nähden on kuitenkin se, että YSOon ei otettu mukaan YSAssa olevia paikannimiä. Syynä on se, että paikkaontologiat ovat luonteeltaan hyvin erilaisia kuin yleiskäsiteontologiat ja niiden kehittämistä varten FinnONTOssa oli oma tutkimushaaransa. Sen tuloksena julkaistiin mm. Maanmittauslaitoksen n. 800 000 suomalaista paikkaa sisältävä Paikannimirekisteri ensi kertaa RDF-muodossa ja otettiin koekäyttöön erillisenä ONKI Geo -palveluna²⁶ (Lindroos, 2008). Historiallisten paikkojen osalta valmistui kuntahistorian kuvaava Suomalainen ajallinen paikkaontologia SAPO (Kauppinen, Hyvönen, 2007; Hyvönen et al., 2011), joka julkaistiin ONKI-palvelussa.

²⁵ <http://www.w3.org/TR/swbp-classes-as-values/>

²⁶ <http://www.seco.tkk.fi/services/onkipaikka/>

2.6 Laajempi/suppeampi termi -suhteiden tarkentaminen

Semanttisen webin ontologioiden selkärankana on mahdollisimman hyvin määritelty luokkahierarkia. Ideana on, että luokka (class) kuten "peili" on kuvaus sen yksilöiden yhteisistä ominaisuuksista ja edustaa ilmentymiään eli yksilöitään (instance, individual), kuten vaikkapa museon kokoelmassa olevaa tiettyä peiliä. Työn alle otettujen asiasanastojen sanat viittaavat pääosin yleiskäsitteisiin, joille tällainen luokkatulkinta on pääosin luonteva ja ilmeneekin Suomessa käytetyissä monikollisissa termimuodoissa. Esimerkiksi YSAn "oopperat" viittaa yksittäisiin oopperateoksiin. Yksikössä olevien sanojen osalta tilanne ei ole niin selkeä. Esimerkiksi YSAssa on myös sana "ooppera", joka tarkoittaa oopperaa taidelajina ja saattaa vaikuttaa taidelajiluokan yksilöltä. Myös tällaiset käsitteet voidaan kuitenkin tulkita luokkina, joilla kenties on kuitenkin vain yksi ilmentymä.

Asiasanastojen termejä ja niiden suhteita alunperin luotaessa ei ole varsinaisesti ajateltu tietojenkäsittelytieteen piirissä yleisesti käytettäviä luokkia ja yksilöitä ja näiden välisiä suhteita. LT/ST suhde voi tarkoittaa yksilö-luokka-suhdetta, alaluokka-suhdetta ja lisäksi myös mm. osa-kokonaisuus-suhdetta. Esimerkiksi YSAn suhteessa

komeetat LT aurinkokunta

ei väitetä, eikä ole tarkoituskaan väittää, komeettojen olevan aurinkokuntia (luokkasuhde), vaan että komeetat esiintyvät aurinkokunnassa. Asiasanaston ajatuksena on, että tiedonhaussa komeetoista kiinnostuneen kannattaisi kenties yrittää hakea myös asiasanalla aurinkokunnat ja aurinkokunnista kiinnostuneen kenties komeetalla hakua tarkentaen. Asiasanastojen tarkoituksena on lähinnä ohjata ihmistä toisaalta tiedon indeksoinnissa ja toisaalta tiedon haussa indeksoinnissa käytettyihin hakusanoihin.

Sen sijaan seuraavassa esimerkiksi LT-suhde tarkoittaa yläluokka-suhdetta eikä osakokonaisuussuhdetta:

peilit LT huonekalut

Tiedonhaun haasteena on tulkita samalla tavalla esitetty mutta eri asiaa tarkoittava rakenne oikein, kun sen avulla tehdään päättelyä. FinnONTOssa ongelma ratkaistiin tarkentamalla LT-suhteet joko luokkasuhteiksi (geneerisyys, yleisin tapaus), osakokonaisuussuhteiksi (partitiivisuus) tai assosiatiivisiksi suhteiksi, jotka muodostavat asiasanastoissa ja ontologioissa käytettävien merkityssuhteiden kolme pääluokkaa.

2.7 Yksilö-luokka-suhteiden transitiivisuuden tarkistaminen

Keskeinen haaste luokkahierarkiaa muodostettaessa on taata se, että alaluokkien yksilöt ovat paitsi oman luokkansa niin myös transitiivisesti kaikkien yläluokkiensa yksilöitä²⁷. Tarkastellaan edellistä peili-esimerkkiä lisäämällä siihen uusi yläkäsite-suhde:

```
meikkipeilit LT peilit
peilit LT huonekalut
```

Tämän mukaan meikkipeilit ovat peilejä ja toisaalta peilit ovat huonekaluja, mikä kuulostaa järkevältä, mutta transitiivisuus molempien suhteiden ylitse ei toimi, sillä meikkipeilit eivät ole huonekaluja. Jos LT-suhteet muutetaan mekaanisesti luokkasuhteiksi (rdfs:subclassOf) seurauksena on, että haettaessa huonekaluja hakutalokseen voi tulla virheellisesti mukaan meikkipeilejä.

Transitiivisuuden tarkistaminen on erityisen haastavaa KOKO-ontologiapilvessä, kun luokkahierarkiat kulkevat eri ontologioiden kautta. Tässä ei riitä samanoloisten käsitteiden mekaaninen parittainen siltaaminen owl:sameAs-ominaisuuksilla, vaan syytä on tarkastella myös käsitteiden laajempaa hierarkiakontekstia yhdistettävissä ontologioissa, sillä transitiivisuuden pitää toimia yli ontologiarajojen.

2.8 Monimerkityksisten sanojen taustalla olevien käsitteiden erottelu

Luokkahierarkian muodostaminen edellä hahmotetulla tavalla on keskeisin ontologian rakentamisessa käytettävä semanttinen periaate. Olemassa olevaa asiasanastoa tähän muotoon muutettaessa yhtenä haasteena ovat sanastoissa esiintyvät monimerkityksiset sanat, joita ei ole monimerkityksisyytensä johdosta mahdollista sijoittaa hierarkiaan mihinkään paikkaan.

Ratkaisumalleja ongelmaan kaksi:

1. Moniperintä. Käsitteelle voidaan antaa useampia yläluokkia eri merkityksiin liittyen. Esimerkiksi valotusmittari voisi olla samanaikaisesti valokuvausväline ja mittari. Moniperintää ei ontologioissa kuitenkaan suositeta,
2. koska se ei varsinaisesti selkeytä monimerkityksiä käsitteitä ja johtaa vaikeasti ylläpidettäviin asyklisiin hierarkioihin yksinkertaisempien puumaisten hierarkioiden sijaan. Yleisenä pyrkimyksenä alalla on moniperinnän minimointi.

²⁷ Tietotekniikassa tehdään tarkka ero yksilöiden ja niiden luokkien välillä. Terminologian termistössä on alakäsite/yläkäsite ja yksilökäsite/yleiskäsite, mutta sekä yksilö- että yleiskäsitteet voivat olla alakäsitteitä.

3. Monimerkityksisten sanojen taustalla olevien käsitteiden erottelu. Merkitysten erottelu on vallalla oleva perusratkaisu ontologioissa. Käsitteet erotellaan ja pyritään sijoittamaan mahdollisimman yksinkertaisiin puumaiisiin rakenteisiin, joita sitten voidaan semanttisesti rikastaa päättelyn avulla. Haasteena tässä lähestymistavassa on, että käsitteiden määrä kasvaa, eikä ole aina selvää miten pitkälle merkitysten jaossa on syytä mennä. Esimerkiksi lukemattomia esineitä voidaan käyttää leikkikaluna (leikkiautot, leikkipyssyt, leikkiastiat, leikkihatut jne.). Pitäisikö näille kaikille luoda omat luokkansa? Ontologisessa työssä ongelman ratkaisemiseen on muitakin mahdollisuuksia kuin luokkarakenteen paisuttaminen. Esimerkiksi porkkanan käsitettä ei tarvitse jakaa kahtia porkkanaksi kasvina ja porkkanaksi ruoka-aineena, vaan käyttö ruoka-aineena voidaan ilmaista assosiatiivisella suhteella. Ontologioiden suunnittelu on sidoksissa niiden käyttötapaan metatietomalleissa, ja ongelmaa voidaan lähestyä myös sisällönkuvailun kautta. Esimerkiksi esineen kuvailutietoihin voidaan liittää tieto sen käytöstä leikkikaluna. Tämän takia kovin pitkälle meneviä jaotteluja ontologioissa on syytä välttää, ts. vähäisen sitoutumisen periaate on hyvä pitää mielessä tässäkin.

Palamme käsitteiden jakamisen problematiikkaan tarkemmin YSON esittelyn yhteydessä.

2.9 Standardiformaatin valinta: RDF(S), OWL ja SKOS

FinnONTO-ontologiat on kehitetty Protégé-editorin tukemassa RDF(S)- (ja sittemmin OWL-) formaatissa. Kun W3C:n SKOS-standardi projektin kuluessa kehitettiin ja valmistui, FinnONTO-hankkeessa kehitettiin automaattinen muunnin RDF/OWL-muodon ja SKOS-muodon välille sekä muodon tarkistin ja korjain (Skosify-työkalu) (Suominen, Hyvönen, 2012; Suominen, Mader, 2013). Kansalliskirjaston Finto-palvelussa siirryttiin tukemaan vain SKOS-standardin mukaista rakennetta. SKOS-standardi on tarkoitettu yksinkertaisten sanastojen rakenteen kuvaamiseen ja se perustuu ja integroituu yleiskäyttöisempiin RDF(S)- ja OWL-standardeihin.

2.10 Käsitteiden nimeäminen

Sanastojen kehittäjien keskuudessa on erilaisia käsityksiä siitä, pitäisikö asiasanojen olla monik-komuodossa, kuten YSAssa on pääsääntöisesti tehty, vai yksikkömuotoisia. Suomalaisissa asiasanastoissa käytetään monikkoja, mutta esimerkiksi lääketieteen sanastoissa, terminologisissa sanastoissa ja semanttisen webin ontologioissa käytetään pääasiassa yksiköllisiä muotoja. KOKO-ontologiassa tähän käytäntöön ei ole puututtu, vaan on mukauduttu alkuperäisen sanaston käytäntöön. Käytäntöjen harmonisoimiseksi sanastoon voidaan varsin helposti generoida tarpeen mukaan

yksikkö- tai monikkomuotoisia nimikkeitä, mutta tällöin on huomioitava yksikkö-monikko-jaolla mahdollisesti tehty semanttinen merkitysten erottelu.

Oma käsityksemme mukaan kannattaisi jatkossa siirtyä käyttämään yksiköllisiä muotoja kansainvälisen yleisimmän käytännön mukaisesti. Tällöin esimerkiksi YSAn termit *oppera* ja *opperat* voidaan esittää standardin mukaisesti tarkenteen (qualifier) avulla esimerkiksi muodossa *oppera (taidemuotona)* ja *oppera (teok-sena)*. Tarkenteiden käyttö nimikkeissä on hyvä käytäntö, koska se kertoo loppu-käyttäjälle selkeämmin käsitteen merkityksen.

Tavoitteellisena käytäntönä KOKO-ontologiassa on, että sen käsitteiden nimikkeet ovat kielikohtaisesti yksikäsitteisiä, ts. missään sen osaontologioissa ei käytetä samaa nimeä samalla kielellä kahdesta eri käsitteestä. Tämä vaatimus tulee suoraan yksittäisten asiasanastojen käytännöstä, mutta laajennetaan KOKOssa koskemaan myös ontologioiden yhdistelmää. Merkitykset suosittelimme erottamaan tarpeen mukaan suluissa olevilla tarkenteilla. Uuden käsitteen luominen osa-ontologiaan edellyttää siis sitä, että samalla tarkistetaan, ettei sama käsite tai edes sama nimike kuulu johonkin toiseen ontologiaan. Tämä edistää sanastojen uudelleenkäyttöä ja vähentää päällekkäisyyttä.

2.11 Moniarvoisuus

On selvää, ettei maailmaa voida kuvata riittävän tarkasti vain yhdellä ontologialla vaan eri aloilla ja sovelluksilla on omia yksilöllisiä tarpeita. Lisäksi moniin mallintamiseen liittyviin ratkaisuihin liittyy hyvin perusteltuja mielipide-eroja, eikä monissa asioissa edes ole olemassa vain yhtä totuutta. Esimerkiksi Taiwan ei ole Kiinan mukaan valtio, mistä taiwanilaiset ovat eri mieltä.

KOKO-ontologiassa tämä on huomioitu siten, että osaontologioiden kehittäjillä on mahdollisuus rakentaa oman ontologiansa rakenteet täysin itsenäisesti myös silloin, kun käsitteitä on otettu käyttöön toisesta ontologiasta, kuten YSOsta. KOKOn olemassaolo kuitenkin pakottaa järjestelmässä mukana olevan tekemään silloin tietoisesti perustellun valinnan, koska lainatut käsitteet pitää aina sillata siihen ontologian, josta ne otettu. Tämä tapahtuu RDF(S) ja OWL standardeissa yleensä `rdfs.subClassOf`- ja `owl:sameAs`-suhteilla, kun taas SKOS-sanastoissa on tarkoitukseen käytössä laajempi joukko erityisiä siltausominaisuuksia (mapping properties), kuten `skos:broadMatch`, `skos:exactMatch` ja `skos:narrowMatch`.

Ideana on, että KOKO ohjaa tutustumaan ja lähtökohtaisesti käyttämään jo olemassa olevia rakenteita ja käsitteitä toisista ontologioista, mikä vähentää päällekkäistä työtä ja parantaa ontologioiden välistä yhteentoimivuutta.

KOKOn ensimmäisessä versiossa osaontologiat on muodostettu suoraan niiden perustana olevista asiasanastoista, eikä ontologioiden päällekkäisyyden ongelmaan ole puututtu. Tutkimusten mukaan osaontologioissa on jopa 30-70% päällekkäisyyttä erityisesti YSON kanssa, mikä ei voi olla järkevää työnjaon kannalta. Tässä on kuitenkin vasta avattu työmaa eri alojen sanastojen kehittäji-en väliselle vuoropuhelulle ontologioiden työnjaosta, joka suunnitelman mukaan tulisi tapahtumaan KOKO-kokonaisuutta jatkossa ylläpitävän erityisen ONKI-konsortion toimesta. FinnONTO-hankkeen työkaluilla, kuten KOAN and MUTU, on mahdollista tutkia KOKO-ontologian rakenteita ja muutosten vaikutuksia. Jatkossa pitäisi sopia yhteisesti siitä, mitä osaontologioita oikeastaan pitäisi kehittää ja mitä sulauttaa toisiinsa, miten ontologioiden päällekkäisyyksiä kannattaisi karsia ja miten hajautettu kehitystyö jatkossa kannattaisi organisoida.

Tällä hetkellä KOKO-ontologia koostetaan algoritmisesti ja suoraviivaisesti samais-tamalla yhteiset käsitteet ja huomioimalla ontologioiden väliset hierarkiaviittaukset. Koska ontologioita ei ole alku-jaan kehitetty kokonaisuutta huomioon ottaen, ei lopputulos ole ihmislukijan kannalta aina kovin luonteva. Esimerkiksi sama käsite saattaa olla jaoteltu alaluokiksi eri ontologioissa eri periaatteella, jolloin näiden yhdistelmässä alaluokat muodostavat epäyhtenäisen joukon.

2.12 Avoin yhdistetty tieto

FinnONTO-hankkeessa neuvoteltiin lupa julkaista kaikki mukana olevat ontologiat avoimena datana. Tulosten hyödyntäminen myös kaupallisiin tarkoituksiin sekä sanastojen muokkaaminen ja sulkeminen on vapaata²⁸, ainoastaan aineiston alkuperäinen kehittäjätaho on mainittava (MIT License, CC 3.0).

²⁸ Sulkeminen tarkoittaa, että sanastosta voidaan luoda uusi versio, joka ei ole avoimesti saatavilla.

3. Asiasanastosta ontologiaksi: Case YSO

Tässä luvussa kuvataan tapaustutkimuksena asiasanaston muuttaminen ontologiaksi esimerkkinä Kansalliskirjaston YSA-sanastosta johdettu YSO-ontologia, joka muodostaa KOKO-pilven yläontologian.

3.1 Hierarkian pääluokat

Tekniikan kehityksen sekä asiasanastojen ontologisoinnin myötä on tullut mahdolliseksi laajentaa asiasanastojen käyttöä muuhunkin kuin kirjastoissa ja muissa muistiorganisaatioissa perinteisesti tehtävään aineistojen kuvailuun. Kuvailtavia kohteita voi olla monenlaisia, esimerkiksi uutiset tai sosiaalisen median aineistot ja kirjoitukset, ja ontologioita voidaan hyödyntää uusilla tavoilla verkossa, esimerkiksi puoliatomaattisen asiasanojen valinnan tukena. Ontologioiden käyttö kuvailussa tukee aineistojen keskinäistä linkitystä, mikä puolestaan mahdollistaa aineistojen entistä monipuolisemman käytön. Yksi esimerkki käyttömahdollisuuksista ovat semanttisen webin tekniikoita hyödyntävät portaalit, jotka tarjoavat mahdollisuuden hakea tietoa useista erilaisista tietolähteistä yhdellä kertaa.

Koska YSO/Allsoa voidaan käyttää monenlaisiin tarkoituksiin ja se sisältää monen erikoisalan käsitteitä, ontologialla ei ole yhtä yhtenäistä käyttäjäryhmää, jonka tarpeet ohjaisivat rakenteiden kehittämistä ja mukaan otettavan tiedon valintaa. Sen sijaan käsitetiedon esittämisessä pitää pyrkiä siihen, että tieto soveltuisi mahdollisimman moneen käyttöön.

YSO/Allson sisältö perustuu yleisesti tunnettuihin ja hyväksytyihin sanojen ja termien merkityksiin eli käsitteiden olennaisiin sisältöihin. Kielitoimiston sanakirja ja sen edeltäjät ovat olleet tärkeitä tietolähteitä, kun sanojen yleisesti hyväksytyjä merkityksiä on arvioitu YSO/Allson rakennetta kehitettäessä. Erityisesti monimerkityksisiä sanoja ja termejä käsiteltäessä on jouduttu tekemään valintoja mm. sen suhteen, mikä tieto on niin yleistä ja olennaista, että sen tulee sisältyä yleiseen ontologiaan. Osittain käsitteiden erittelyä on jo toteutettu edellä mainittujen sanakirjalähteiden perusteella, mutta osittain tietoja sanojen eri merkityksistä on vasta kirjattu ontologiaan myöhempää tarkistusta varten.

Tulevaisuudessa tällaisten valintojen tekemistä voitaisiin tukea tiiviillä yhteistyöllä ja selkeällä tehtävänjaolla yleisen ja erikoisalojen ontologioiden kehittäjien välillä. Vastuualueiden jaossa voitaisiin hyödyntää esimerkiksi nykyistä erikoisontologioiden vastuujakoa niin, että vastuu erikoisalojen käsitteistä siirrettäisiin kyseisten

erikoisontologioiden ylläpitäjille, kun taas YSO-kehittäjät jatkaisivat ylläpitoa niiden käsitteiden osalta, joiden erikoisalalta ei ole erillistä ontologiaa tai jotka ovat useille aloille yhteisiä. Järjestely edellyttäisi luonnollisesti sitä, että kaikkia yhteistyössä mukana olevia ontologioita voitaisiin ylläpitää aktiivisesti.

Alla esitetään lyhyet luonnehdinnat esitetyistä YSON pääluokista alkuperäisen jaottelun mukaisesti (vrt. kuva 2.1).

abstrakti

käsite, joka on riippumaton muihin käsiteryhmiin kuuluvista käsitteistä

Ontologian muut käsitteet voivat viitata abstrakteihin, mutta abstraktit voivat viitata vain toisiinsa. Luokka sisältää suureet (esimerkiksi pituus ja massa) ja yksiköt sekä avaruutta/euklidista tilaa tms. (esimerkiksi jatkumo) ilmentävät käsitteet. Luokan nimitys ei siis viittaa abstrakteihin käsitteisiin sanan tavanomaisessa merkityksessä.

pysyvä

käsite, jolla ei ole ajan mukaan muuttuvia olennaisia piirteitä (tai osia), vaan jonka muodostama kokonaisuus säilyy ajassa muuttumattomana, vaikka piirteiden arvot voivat muuttua

Pysyvien käsitteiden piirteiden arvojen muutoksia voidaan kuvata muuttuvien käsitteiden avulla.

muuttuva

käsite, jolla on aikasidonnaisia olennaisia piirteitä (tai osia), joiden muodostama kokonaisuus voi vaihdella ajassa

Muuttuvien käsitteiden avulla voidaan kuvata pysyvien käsitteiden piirteiden arvojen muutoksia.

Tarkastelemme seuraavassa YSON ylimpiä luokkia tämän DOLCE-järjestelmästä periytyvän jaottelun pohjalta. Käsitteiden ryhmittelyn takia mukaan on otettu ryhmitteleviä luokkia, joita ei ole tarkoitettu käytettäväksi indeksoinnissa. Esimerkiksi abstrakti, pysyvä ja muuttuva ovat tällaisia apukäsitteitä. Indeksointiin tarkoitettut käsitteet on erikseen mainittu.

3.1.1 Abstrakti (abstract)

Ontologian muut käsitteet voivat viitata abstrakteihin, mutta abstraktit voivat viitata vain toisiinsa. Luokka sisältää suureet (esimerkiksi pituus ja massa) ja yksiköt sekä avaruutta/euklidista tilaa tms. (esimerkiksi jatkumo) ilmentävät käsitteet. Luokan nimitys ei siis viittaa abstrakteihin käsitteisiin sanan tavanomaisessa merkityksessä.

3.1.2 Pysyvä (endurant)

Pysyvä-käsite jakautuu edelleen seuraavasti alaluokkiin:

ajanjakso

ajallinen kokonaisuus, jolla on tiettyjä tuntomerkkejä tai ominaisuuksia, joiden perusteella se voidaan tunnistaa, ja joiden perusteella sille voidaan määritellä vähintään summittainen alku- ja loppuaika

Ajanjaksoja ovat esimerkiksi:

- kalenteriin liittyvä ajanjakso: esim. kuukaudet, kesäaika
- luonnontieteellinen ajanjakso: esim. kaamos, vuodenaajat
- yhteiskunnallinen ajanjakso: esim. metsästysajat, teemavuodet.

fyysinen kokonaisuus

Vain sellaiset fyysisten objektien muodostamat kokonaisuudet, jotka ovat ensisijaisesti kokonaisuuksia, mutta eivät kuitenkaan toiminnallisia kokonaisuuksia eli järjestelmiä.

Fyysisiä kokonaisuuksia ovat esimerkiksi:

- elollinen fyysinen kokonaisuus: esim. puusto
- eloton fyysinen kokonaisuus: esim. astiastot.

fyysinen objekti

eloton tai elollinen aineellinen objekti, jolla on massa

Fyysisiä objekteja ovat esimerkiksi:

- aine: esim. juomat, täyteaineet, kemikaalit, laava
- konkreettinen eloton objekti: esim. helmet, muotit, pelletit

YSON 1. versiossa konkreettisten elottomien objektien alahierarkioita ei ole eritelty tarkasti. Kun hierarkioita tarkennetaan, on perusteltua rajata sellaisten käsitteiden, kuten välineet, hierarkia edelleen suppeaksi käsitteisiin, jotka selvästi ovat välineitä, sen sijaan että hierarkian perusteeksi otettaisiin "voidaan käyttää välineenä".

orgaaninen rakenne, joita ovat esimerkiksi eliöt kuten

- eläimet: esim. ihminen, koira
- kasvit: esim. puut, rikkakasvit
- kasvinosat: esim. hedelmät, kukka
- ruumiinosat: esim. kudokset, ikenet, olkapäät.

henkiset tuotokset

ajatussisältö, joka usein saa myös fyysisen muodon

Sisältö on kuitenkin fyysistä objektia tärkeämpi ja fyysisen muodon edellytys. Ai-neettomaan tuotokseen liittyy aina toimija ja toiminta, jonka tuloksena aineeton tuotos syntyy. Jos tuotokselle löytyy tarkempi luokitus kuten järjestelmä, se luokitel- laan tarkemman kuvausmahdollisuuden mukaisesti.

Henkisiä tuotoksia ovat esimerkiksi:

- kulttuuriset tuotokset, joita ovat mm. musiikilliset tuotokset: esim. pas- siot, aariat
- teokset: esim. oppimateriaali, elämäkerrat, tanssiteokset
- uskonnolliset tuotokset: esim. evankeliumi, hartauspuheet

sekä

- tekniset tuotokset: esim. kyberavaruus, tehosteet

sekä

yhteiskunnalliset tuotokset, joita ovat mm.

- menetelmät (määritelty tai säännelty tapa tehdä jotakin): esim. galvano- tekniikka, holografia
- sopimukset: esim. avioehto, kauppasopimukset
- säädökset: esim. asetukset, direktiivit
- teoriat: esim. determinismi, kaaosteoria
- tieto: esim. kaavat, muistitieto.

ilmiöt (sisältyy myös YSA/Allärsiin)

tapahtuman tai tapahtumasarjan tuloksena syntynyt havaittavissa oleva asiointila tai asiointilan vaihtelu

Ilmiöitä ovat esimerkiksi:

- kulttuuriset ilmiöt: esim. muoti, tabut

Kulttuuristen ilmiöiden aatteet-alakäsitteen hierarkiaan kuuluu joukko ismejä.

Kaikki käsitteet, joiden termi on muodostettu vastaavalla tavalla, eivät välttämättä kuitenkaan sisältönsä puolesta kuulu aatteet-hierarkiaan.

- luonnontieteelliset ilmiöt, joita ovat mm. biologiset ilmiöt: esim. ravitsemus, ruska
- fysikaaliset ilmiöt: esim. kelirikko, säteily
- kemialliset ilmiöt: esim. katalyyysi, isomeria
- mentaaliset ilmiöt: esim. tunteet, unet
- taloudelliset ilmiöt: esim. hyvitys, inflaatio
- yhteiskunnalliset ilmiöt: esim. demokratiavaje, muuttoliike; yhteiskunnallisiin ilmiöihin kuuluu erikoisala: esim. hallinto, oikeus, tieteet

Taiteen erikoisalojen hierarkiaan on sijoitettu taiteen alat, joilla on selkeä toimialamainen luonne eli esimerkiksi koulutus-, esitys-, rahoitus- tai muita institutionaalisia muotoja. Taiteen erikoisalat voivat kuulua myös taidelajien hierarkiaan, johon muut eriytyneet taiteen alat on merkitty.

- oikeudet: etu tai valta tehdä jotain; oikeus voi olla peräisin mm. perinteistä, luonnosta, hallitukselta tai yhteisön moraalikäsitteestä, esim. käyttöoikeus, ihmisoikeudet
- yhteiskunnalliset rajoitukset: esim. kiellot, velvollisuudet

järjestelmät (sis. myös YSA/Allärsiin)

tiettyjen periaatteiden mukainen toiminnallinen kokonaisuus'

Järjestelmiä ovat mm.

- kulttuuriset järjestelmät: esim. kieli, kirjoitusjärjestelmät

YSA:n asiasanapari kieli ja kielet on saanut YSO/Allsossa tarkenteet kieli (kielenkäyttö) ja kielet (asema tai alkuperä). Ensimmäisen alakäsitteiksi hierarkiassa on sijoitettu kielenkäytön perusteella eriteltävät alakäsitteet, kuten slangi, ja toisen kielen aseman tai alkuperän mukaan eriteltävät käsitteet kuten viralliset kielet ja itämaiset kielet.

- luonnontieteelliset järjestelmät: esim. ekosysteemit, ravintoketjut
- tekniset järjestelmät: esim. valaistus, laskentajärjestelmät
- yhteiskunnalliset järjestelmät: esim. aravajärjestelmä, organisaatiot.

ominaisuudet (sis. myös YSA/Allärsiin)

luonnehtiva tai kuvaava piirre

Ominaisuuksia ovat:

sisäiset ominaisuudet: "sisäänrakennettu", ilman ulkoisia vaikuttajia määrittyvä ominaisuus, joita ovat esimerkiksi:

- biologiset ominaisuudet: esim. huonokuuloisuus, kaksineuvoisuus
- fyysiset ominaisuudet: esim. korkeus, kunto

Fyysisten ominaisuuksien alakäsitteenä on YSA/Allärsistä periytyvä fyysikaalisten ominaisuuksien hierarkia.

- mentaaliset ominaisuudet: esim. avoimuus, subjektiivisuus
- sekä
- ulkoiset ominaisuudet: ominaisuus, joka määrittyy suhteessa ulkoiseen vaikuttajaan, joita ovat esimerkiksi:
 - rooli: toimintaan tai käyttöön perustuva tehtävä

Roolin tulee olla yleinen, ei esimerkiksi poikkeukselliseen käyttöön tai tiettyyn prosessiin perustuva. Rooli määritellään suhteessa yhteen tai useampiin käsitteisiin, eikä se ole käsitteen ensisijainen luokitteluperuste, ellei kyse ole roolien välisestä hierarkiasta. Esimerkiksi mies on ensisijaisesti ihminen, mutta jos miehellä on lapsi, miehellä on isän rooli. Yhteen käsitteeseen voi samanaikaisesti liittyä useita rooleja. Esimerkiksi mies voi olla isä, aviomies, vävy, ammatinharjoittaja jne. Rooli on ominaisuus-päälukassa eri merkityksessä kuin YSAn asiasana roolit, joka on sukupuoliroolien yläkäsite.

Rooli-käsitteet on ryhmitelty seuraavasti:

- aatteelliseen ryhmään liittyvä rooli: esim. kommunistit, fennofiilit
- alueelliseen identiteettiin liittyvä rooli: esim. espoolaiset
- käyttöön liittyvä rooli: esim. aineistot, korvikkeet
- oikeudelliseen asemaan liittyvä rooli: esim. asianomistajat, täysivaltaiset
- paikkaan liittyvä rooli: esim. kotipaikka, luontokohteet
- sosioekonomiseen ryhmään liittyvä rooli: esim. köyhät, tilalliset
- sukulaisuuteen (tai muuhun läheiseen ihmissuhteeseen) liittyvä rooli: esim. jälkeläiset, lesket
- terveydentilaan liittyvä rooli: esim. sairaat, vajaakuntoiset
- toimintaan liittyvä rooli: esim. jalankulkijat, ammatit (huom. YSO/Allsossa ei ole eritelty ammatteja ja tehtävänimikkeitä toisistaan)
- uskontoon liittyvä rooli: esim. munkit, profeetat
- vaikutukseen liittyvä rooli: esim. menestystekijät, riskitekijät
- väestöryhmään liittyvä rooli: esim. inuitit, romanit
- muu rooli: esim. aarteet, ihanteet

FinnONTO-mallissa on testattu yhtä tapaa luokitella ominaisuuksia: ominaisuuksien jakoa sisäisiin ja ulkoisiin. Se pohjautuu terminologian teoriassa käytettyyn vastaavaan käsitepiirteiden jakoon (Wüster, 1985). Olennaista ominaisuuksien luokittelussa on se, että luokat jakavat ominaisuuskäsitteet pienempiin ryhmiin, jotka on helpompi hahmottaa kuin suuri heterogeeninen ominaisuuskäsitteiden luokka. Jotta ryhmittely tukisi käsitteiden hahmottamista, luokitusperiaatteen tulee soveltua kaikkiin ominaisuuskäsitteisiin ilman, että merkittävä osa käsitteistä kuuluu useampaan kuin yhteen luokkaan.

paikka (sis. myös YSA/Allärsiin)
maanpinnan tai avaruuden osa

Paikkoja ovat esimerkiksi:

- alueet, kuten ihmisen määrittämät alueet: esim. hallintoalueet, matkailualueet
- luonnon muodostamat alueet: esim. aavikot, rannikot
- ihmisen määrittämä paikka: esim. kaivokset, liikuntapaikat
- luonnon muodostama paikka: esim. hiidenkirnut, koralliriutat
- muu paikka: esim. fiktiiviset paikat, paratiisi, tuonela

Paikka-sanana yksikkömuodon ja alueet-sanana monikkomuodon käyttö ontologiassa noudattaa YSA/Allärsin linjaa.

3.1.3 Muuttuva (perdurant)

Muuttuva-käsite jakautuu edelleen seuraavasti alaluokkiin:

prosessit (sis. myös YSA/Allärsiin)

Prosesseihin kuuluvat:

- tapahtumat (sis. myös YSA/Allärsiin): jokin, joka toteutuu tietyn kestoisenä; tapahtumaan voi liittyä yhden tai useammanlaisista toimintaa

Tapahtumia ovat:

- järjestetty tapahtuma: tapahtuma, joka on tarkoituksellisen toiminnan tulos
- spontaani tapahtuma: tapahtuma, jota ei ole järjestetty tarkoituksellisesti
- toiminta (sis. myös YSA/Allärsiin): jokin, joka tehdään tai toteutuu

Toisin kuin tapahtuman määrittelyssä kesto ei ole toiminnan osalta olennainen piirre.

Huomautuksia:

- Jos YSAn käsitteen nimityksenä on verbin substantivoitu muoto, joka voi viitata sekä toimintaan että toiminnan lopputulokseen, käsite on ontologiassa sijoitettu ensisijaisesti toimintahierarkiaan.
- Tutkimustoiminta-käsitteen alakäsitteiksi on hierarkiassa pyritty merkitsemään asiasanat, joiden perusosa on –tutkimus ja jotka viittaavat tutkittavaan aiheeseen. Tutkimus-käsitteen alakäsitteiksi on pyritty

merkitsemään asiasanat, joiden perusosa on –tutkimus ja jotka viittaavat tutkimuksessa käytettävään menetelmään. Käsitejaon selkeyttämiseksi olisi hyvä tarkistaa erityisesti jälkimmäisten suhdetta tutkimusmenetelmät-käsitteen alakäsitteisiin.

- Analyysin alakäsitteiksi ryhmittelyperusteen analyysi kohteen mukaan on pyritty merkitsemään asiasanat, joiden perusosa on analyysi ja jotka viittaavat analysoitavaan kohteeseen. Analyysi analyysitavan mukaan alakäsitteiksi on pyritty merkitsemään asiasanat, joiden perusosa on –analyysi ja jotka viittaavat analyysissä käytettävään menetelmään. Käsitejaon selkeyttämiseksi olisi hyvä tarkistaa erityisesti jälkimmäisen suhdetta analyysimenetelmät-käsitteen alakäsitteisiin.

Toimintaa ovat:

- havainnointi ja aistiminen: erityisesti aistien avulla tapahtuva huomiointi
- keskinäinen toiminta: elollisten osapuolten välinen toiminta, joka voi olla vastavuoroista
- kulutus ja käyttö: kohteen hyödyntäminen
- liikkuminen ja liikuttaminen: siirtyminen tai kohteen siirtäminen paikasta toiseen
- luonti: uutta luova toiminta
- mentaalinen toiminta: mielessä tapahtuva tai ajatuksiin tai tunteisiin perustuva toiminta, joka voi saada myös konkreettisia muotoja
- muutos: toiminta, jossa kohde tai tilanne voi erilaistua ilman nimetyn aktiivisen toimijan vaikutusta
- muuttaminen: toiminta, jossa aktiivinen toimija aiheuttaa kohteen tai tilanteen erilaistumisen
- oleminen: kestopensa ajan muuttumattomana säilyvä toiminta; olemassaolo, liikkumattomuus
- omistus ja omistuksen siirtäminen: kohteen pitäminen haltijalla tai kohteen haltijan vaihtaminen

Toimintojen hierarkia perustuu pitkälti WordNetissä käytettyyn toimintojen jaoteluun (Fellbaum, 1998).

3.2 Hierarkian rakentaminen

Asiasanaston ontologisoinnin yhteydessä kaikille asiasanaston käsitteille on annettava paikka hierarkiassa, vaikka asiasanastosta puuttuisikin tieto laajemmasta termistä. Näin siksi, että kokonaishierarkia, jossa hierarkiatasojen muodostama ketju ei katkea kesken, mahdollistaa koneellisen, hierarkkiseen ylä-alakäsitesuhteeseen perustuvan tietojen hyödyntämisen. Hierarkkinen yläkäsite tarvitaan myös sellaisessa tapauksessa, jossa käsite liittyy muihin käsitteisiin koostumussuhteen kautta, sillä koostumussuhteeseen ei liity käsitepiirteiden periytymistä yläkäsitteeltä alakäsitteelle kuten luokkahierarkiassa (esimerkiksi polkupyörän käsitepiirteet eivät periydy pyörän satula -käsitteelle). Hierarkkisen yläkäsitteen valinnan tulee vastata vakiintunutta kielenkäyttöä ja tukea käsitteen ymmärtämistä ja yhdenmukaista käyttöä. Nykyisissä semanttisen webin sovelluksissa hyödynnetään tyypillisesti ontologian alemman tason hierarkioita tiedonhaun rajaukseen.

YSO/Allsossa käsitteille on pyritty mieluiten löytämään yksi ensisijainen paikka edellä määritellyssä hierarkiassa. Tällöin samaan aihealueeseen liittyvät käsitteet kuten eliöt (fyysinen objekti) ja biologiset ilmiöt kuten mutaatiot (ilmiöt) sijoittuvat hierarkiassa eri pääluokkiin, koska ontologian päähierarkia ei perustu aihealuejakkoon. Joillakin hierarkian alemmilla tasoilla erikoisalojen ontologioiden vaikutus kuitenkin näkyy aihealuetta hyödyntävänä hierarkisointina (esimerkiksi kielelliset ilmiöt – äänneopilliset ilmiöt). Tällaisen hierarkisoinnin soveltuvuus rakenteiden luomiseen on kuitenkin syytä tarkistaa. Nykyisellään YSO/Allso ei sisällä tietoa YSA/Allärsin alanmukaisesta ryhmittelystä, mutta tieto alasta olisi mahdollista liittää ontologiaan yhtenä käsitteen ominaisuutena.

Silloin kun käsite sijoitetaan vain yhteen paikkaan hierarkiassa, on tarkistettava, ettei määrittely anna liian suppeaa kuvaa käsitteestä. Esimerkiksi juoksu ei aina ole kuntoliikuntaa, vaan sanalla voidaan viitata yleisemmin paikasta toiseen liikkumiseen tietyllä tavalla. Tarvittaessa käsite voidaan jakaa ja käsitteeseen liittyviä erilaisia näkökulmia voidaan esittää moniperinnän, assosiativisten suhteiden tai assosiativisten suhteiden avulla kuvattavien roolien kautta.

3.2.1 Hierarkiapalojen yhteensovittaminen

Koska osa YSO/Allson hierarkiasta periytyy YSA/Allärsista, on asiasanaston hierarkiaa pyritty tarkistamaan karkealla tasolla, jotta YSA/Allärsista periytyvä ja ontologisoinnin yhteydessä luotu hierarkia muodostaisivat yhtenäisen kokonaisuuden. Nykyisellään YSO/Allsossa on kuitenkin vielä yhdistämisen myötä syntyneitä epätarkkuuksia, jotka tulee korjata sitä mukaa kun niitä havaitaan. Lisättäessä YSA/Allärsiin uusia käsitteitä ontologiaan on esimerkiksi saattanut jäädä huomamatta, että YSA/Allärsista tuleva yläkäsitesuhde ei ole tarkka ja se olisi mahdollista tarkentaa ja yhdenmukaistaa muiden vastaavien käsitteiden hierarkian kanssa.

Hierarkian yhdenmukaistamisen tarkoituksena on parantaa mahdollisuuksia hahmottaa käsittekokonaisuuksia ja mahdollistaa johdonmukainen koneellinen käsittely. Erityisen tärkeää on estää virheet, joita syntyy ontologiaa koneellisesti käytettäessä, jos monitasoinen hierarkia ei ole kaikilta osin pätevä.

Esimerkiksi YSA/Allärsissa oli asiasanastostandardien mukaisesti määriteltynä ST-suhteiden ketju:

```
mäntyöljy LT öljy
öljy LT kaivannaiset
```

Mikäli rakenne tulkitaan luokkahierarkiaksi, tulkitseisi semanttinen hakukone mäntyöljyn tämän mukaan kaivannaiseksi, mikä ei pidä paikkaansa.

YSO/Allsossa rakenne on siksi muutettu muotoon, joka vastaa alla olevaa:

```
mäntyöljy LT kasviöljyt
kasviöljyt LT öljy
öljy RT kaivannaiset
```

3.2.2 Koostumussuhteiden käsittely

YSA/Allärsissa ei ole ylä-alakäsiteparien osalta eritelty, onko kyse hierarkkisesta vai koostumus-suhteesta. Asiasanastossa koostumussuhteita on vain vähän, mutta mikäli sellainen on ontolo-gisoinnin yhteydessä havaittu, suhde on eritelty koostumussuhteeksi, sillä koostumussuhteessa käsitepiirteet eivät periydy yläkäsitteeltä alakäsitteelle, eikä sitä siksi voida käsitellä samalla tavoin kuin hierarkkista ylä-alaluokkasuhdetta. Ontologisoinnin ensimmäisen vaiheen tavoitteena ei kuitenkaan ole ollut systemaattisesti lisätä koostumussuhteita sellaisten käsitteiden välille, joilla ei asiasanastossa ole ollut suhdetta.

3.2.3 Korvattujen termien käsittely

Ontologian hierarkiaan voi vaikuttaa se, että asiasanastoissa korvattujen termien suhde suositettavaan termiin ei aina perustu termien synonymiaan, vaan suosituksella on saatettu ohjata käyttämään suppeampaan tai laajempaan käsitteeseen viittaavaa termiä, jolloin käsitteiden välillä voi olla hierarkkinen tai koostumussuhde. Esimerkiksi:

```
alkusammutusvälineet KÄYTÄ palokalusto
aravalainat KÄYTÄ aravajärjestelmä, asuntolainat,
lainat
```

Näiltä osin ontologiaa voitaisiin kehittää erottamalla nykyisistä korvatuista termeistä todelliset synonyymit ja laajentamalla ontologia kattamaan korvaussuhteessa esitetyt käsitteet. Tällaista kehitystyötä YSO/Allsossa ei kuitenkaan vielä ole tehty.

3.2.4 Assosiativisten suhteiden käsittely

Sen lisäksi, että YSA/Allärsissa annettuja hierarkiatietoja on saatettu tarkentaa tai täydentää ontologisoinnin yhteydessä, asiasanaston assosiativinen suhde on saatettu vaihtaa YSO/Allsossa hierarkkiseksi, mikäli tällainen toimenpide on hierarkkia tarkistettaessa osoittautunut tarpeelliseksi. Esimerkiksi:

YSA/Allärs: varusmiespalvelus RT aseeton palvelus, asepalvelus

YSO/Allso: varusmiespalvelus ST aseeton palvelus, asepalvelus

Myös päinvastaiset muutokset ovat mahdollisia. Jos käsitteen sisällön tarkempi tarkastelu osoittaisi, ettei YSA/Allärsin hierarkkinen ylä-alakäsitesuhde pidä paikkaansa, se voitaisiin ontologisoinnin yhteydessä muuttaa assosiativiseksi suhteeksi.

YSO/Allson assosiativisia suhteita ei kuitenkaan vielä ole tarkistettu kattavasti. Hierarkkia on tarpeen tarkistaa ja tarkentaa myös asiasanaston (lähi)synonyymien (esimerkiksi tunnistaminen – identifiointi) osalta, jotta käsitteitä voitaisiin käyttää johdonmukaisesti.

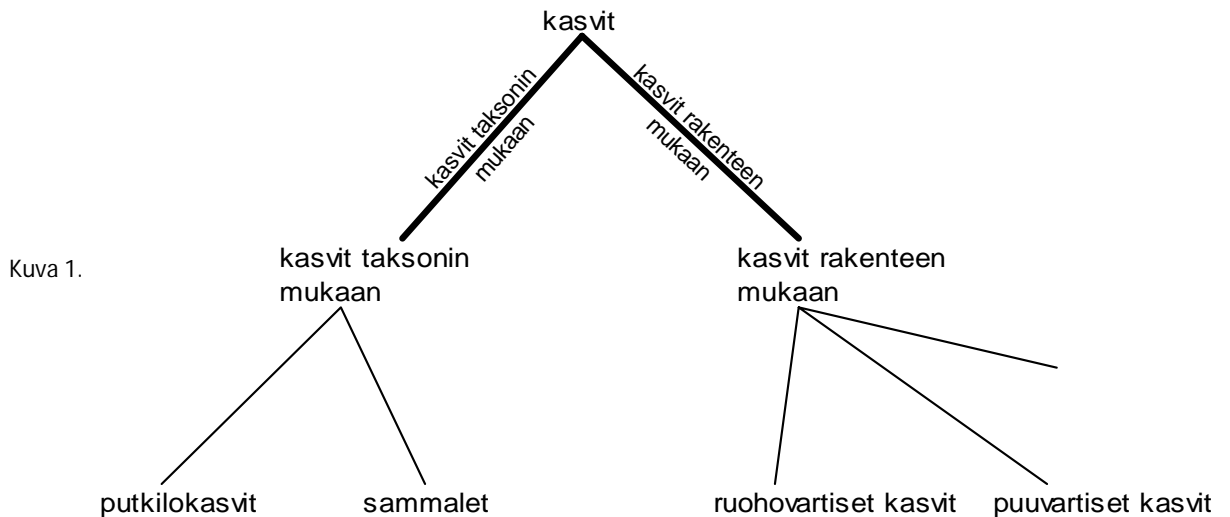
3.2.5 Asiasanaketjujen käsittely

YSA/Allärsiin sisältyviä asiasanaketjuja (esimerkiksi luonnonsuojelu – metsät, jota suositetaan käytettäväksi metsiensuojelu- ja metsien suojelu -asiasanojen sijaan) on YSO/Allsossa käsitelty asiasanojen tapaan. Mikäli useampi asiasanaketju viittaa samaan käsitteeseen (kuten luonnonsuojelu – metsät ja metsät – suojelu), ne esitetään ontologiassa yhtenä käsitteenä niin, että toinen asiasanaketju on käsitteen päänimike (skos:prefLabel) ja toinen valinnainen (skos:altLabel).

3.2.6 Hierarkkisten alakäsitteiden ryhmittely

Koska alakäsitteitä voidaan hierarkiassa usein ryhmitellä useista eri näkökulmista, YSO/Allsossa on otettu käyttöön jonkin verran ryhmitteleviä käsitteitä sellaisissa tapauksissa, joissa hierarkkia ei voida tarkentaa lisäämällä siihen vakiintuneita käsitteitä rakenteen tueksi. Ryhmittelevien käsitteiden avulla on eritelty ja nimetty toisensa poissulkevien vieruskäsitteiden muodostamia ryhmiä (esimerkiksi kasvit taksonin mukaan, kasvit rakenteen mukaan, vrt. kuva 3.1). Vieruskäsitteiden poissulkevuus tarkoittaa sitä, että em. esimerkin mukaisesti tietty kasvi ei voi olla yhtä aikaa putkilokasvi ja sammal (taksonin mukainen jako), eikä puuvartinen ja ruoho-

vartinen kasvi (rakenteen mukainen jako). Yhdellä kasvilla voi kuitenkin olla kahden eri ryhmittelyperusteen mukaisia piirteitä eli putkilokasvi voi olla puuvartinen kasvi. Eri ryhmittelyperusteiden yhdistymismahdollisuus yksilötasolla ei tarkoita, että kaikki mahdolliset yhdistelmät toteutuisivat käytännössä.



Kuva 2. Kuva 3.1. Hierarkkisten alakäsitteiden ryhmittely.

YSO/Allson nykyisessä versiossa alakäsiteryhmittelyn käyttö on kuitenkin vielä ko-keiluluontoista. Vain joidenkin ryhmittelyjen osalta on ollut mahdollista käyttää kohdealan asiantuntijoiden apua ryhmittelyyn tarkistamisessa. Myöskään kaikkia tietyn hierarkiahaaran alakäsitteistä ei ole esitetty tarkentavan ryhmittelyn mukaisesti, eivätkä ryhmittelevien käsitteiden nimitykset kaikilta osin ole viimeistellyjä. Kuvaavien ja selvästi erottavien nimitysten valinta ryhmitteleville käsitteille edellyttää, että ryhmittely tehdään kattavasti tietyn hierarkian osalta. Kattava ryhmittely taas on YSO/Allson tapauksessa usein vaikeaa, koska ontologia sisältää usein vain osan tiettyyn hierarkian osaan kuuluvista käsitteistä, eikä puuttuvia käsitteitä siis voida analysoida samanaikaisesti. Silloin kun ryhmittely on tehty tietyn hierarkiakohtaan osalta kattavasti, sen avulla voidaan tukea ja ohjeistaa aineistojen yhdenmukaista kuvailua niin, että ohjataan kuvailun tekijä käymään läpi kaikki alakäsiteryhmittelysopivien kuvailukäsitteiden löytämiseksi. Esimerkiksi kasvien osalta voitaisiin edellyttää taksonin mukaisen yläkäsitteen valintaa ja sen tarkistamista, löytyykö muista alakäsiteryhmittelysopivista kuvailuun tarvittavia käsitteitä. Ellei ryhmittelyä ole tehty, tiettyyn tarkastelunäkökulmaan perustuva käsite voi jäädä isosta käsittejoukosta helposti huomaamatta.

3.2.7 Moniperinnän käyttö

Mikäli käsite on tarpeen esittää useammasta kuin yhdestä tarkastelunäkökulmasta hierarkian kei-noin, sille voidaan antaa useampi kuin yksi hierarkkinen yläkäsite hierarkian eri haaroista. Tällaisen moniperinnän sijaan on kuitenkin suositeltavampaa käyttää näkökulmien esittämiseen as-sosiatiivisia suhteita, sillä usean rinnakkaisen hierarkian ylläpitäminen on työlästä.

Silloin kun moniperintä on sisältynyt jo asiasanastoon, sitä ei kuitenkaan ole ontologiassa muutettu assosiatiivisella suhteella kuvattavaksi, ellei käsitesuhteisiin ole sisältynyt selvää epätarkkuutta (esim. ravintokuidun yläkäsitteinä säilytetään sekä kuidut että ravintoaineet, eikä korvata yläkäsi-tesuhdetta ravintoaineisiin assosiatiivisella suhteella). Yhden hierarkian sisäinen moniperintä on ontologiassa tarpeetonta, esimerkiksi

keinutuolit LT istuimet, huonekalut

joten mikäli asiasanastossa olisi annettu käsitteelle useampi kuin yksi yläkäsite samasta hierarkiasta, jätettäisiin ontologiassa jäljelle vain tarkin yläkäsite:

keinutuolit LT istuimet

Aina kun käsitteelle annetaan useampia hierarkkisia yläkäsitteitä, on varmistettava, että kaikkien yläkäsitteiden piirteet periytyvät alakäsitteelle. Esimerkiksi

valotusmittarit LT mittarit, valokuvausvälineet

Esimerkkitapauksessa moniperintä on sallittu, koska ei ole olemassa erikseen valotusmittaria, joka olisi mittari, mutta ei valokuvausväline, vaan valotusmittari on aina molempia. Usein voidaan kuitenkin todeta mahdollisia moniperintätapauksia tarkasteltaessa, että moniperinnän sijaan on kyse useammasta samannimisestä käsitteestä, joista kullakin voi olla eri yläkäsite, sen sijaan, että yksi alakäsite perisi useamman yläkäsitteen piirteet. Jotta moniperintävirheilta vältyttäisiin, moniperintä on sallittua vain pääluokkaparien *ilmiö ja ominaisuus* sekä *tapahtuma ja toiminta* välillä. Sääntönä on, että yläkäsitettä ei saa valita kahdesta eri pääluokasta, ellei kyse ole em. pääluokkapareista. Sääntö koskee pääluokkien muodostamia hierarkioita kokonaisuudessaan riippumatta siitä, onko käsiteltävä käsite hierarkiassa suoraan pääluokan alla vai alemmalla hierarkiatasolla. Ylimmän tason luokkien (abstrakti, pysyvä, muuttuva) välillä tai sellaisten luokkien, kuten aineeton tuotos ja fyysinen objekti tai paikka ja järjestelmä, välillä moniperintä ei ole sallittua.

Erityisesti niissä tapauksissa, joissa käsitteiden merkitysten erottelutarve tuntuu syntyvän pääasiallisesti käyttöön valitun hierarkkisen pääluokkarakenteen myötä, hierarkkisen rakenteen monimutkaistuminen on pyritty estämään valitsemalla yksi ensisijainen yläkäsite. Esimerkiksi päiväkotit luokitellaan organisaatioksi (ei raken-

nukseksi) ja kirja aineettomaksi tuotokseksi (ei fyysiseksi objektiksi). Tällaisissa tapauksissa päiväkotitapahtuma ja päiväkotitapahtuma voitaisiin tarvittaessa erottaa kahdeksi eri käsitteeksi kuten myös kirja aineettomana tuotoksena (joka voidaan esittää useiden erilaisten fyysisten objektien muodossa) ja kirja fyysisenä objektina.

Luokkien välistä moniperintää rajoittamalla ei kuitenkaan voida kokonaan estää moniperintävirheiden syntymistä, sillä myöskään yhden pääluokan sisällä ei voida valita useita yläkäsitteitä ilman, että tarkasteltaisiin käsitteitä. Ei esimerkiksi ole olemassa monitoimikonetta, joka olisi yhtä aikaa kodinkone ja metsäkone, vaikka molemmat mahdolliset yläkäsitteet kuuluvat elottomiin fyysisiin objekteihin. Sen sijaan on monitoimikoneita, jotka ovat joko kodinkoneita tai metsäkoneita eli kyse on kahdesta eri käsitteestä, joilla molemmilla on oma yläkäsitteensä.

3.3 Assosiatiivisten suhteiden käyttö

Kaikki muut käsitesuhteet paitsi hierarkkiset ja koostumussuhteet ilmaistaan assosiatiivisilla suhteilla, joiden kirjo on hyvin laaja. YSO/Allsoa kehitettäessä ei ole ollut tiedossa periaatteita, joiden mukaan YSA/Allärsissa esitettävät assosiatiiviset suhteet valitaan, eikä asiansastosta ontologiaan periytyneitä assosiatiivisia suhteita ole käsitelty muussa tapauksessa kuin silloin, jos niiden muokkaaminen on liittynyt hierarkian kehittämiseen.

Vaikka assosiatiivisen suhteen käyttö korvaamassa useita hierarkkisia yläkäsitteitä (moniperintää) häivyttää osan luokitusperiaatteista, menettely on usein perusteltavissa, koska näin pystytään vähentämään hierarkiavirheiden syntymistä. Jos esimerkiksi kilpikonnat luokiteltaisiin matelijoiden lisäksi myös lemmikkieläimiksi, olisi johtopäätös se, että kaikki kilpikonnat koosta, suojelutarpeesta tms. huolimatta olisivat lemmikkieläimiä, mikä ei esimerkiksi jättäisi kilpikonnien osalta pääsääntöisesti pidä paikkaansa. Assosiatiivisella suhteella voidaan ilmaista hierarkiaa väljempi yhteys käsitteiden välillä. Sen sijaan, että kerrottaisiin kaikkien kilpikonnien olevan lemmikkieläimiä (hierarkiasuhde), todetaan että kilpikonna voidaan pitää lemmikkieläiminä (assosiatiivinen suhde). Assosiatiivisiin suhteisiin perustuvaa tietoa voidaan tuoda esiin käyttäjälle varsinaisten (hierarkkisten suhteiden hyödyntämiseen perustuvien) hakutulosten ohien mahdollisuuksina jatkaa tiedonhakuja assosiatiivisessa suhteessa esitetyn käsitteen kautta. Kun hierarkiasuhteiden merkittävistä käsitteiden välillä rajataan em. tavalla, saadaan tarkempia hakutuloksia kuin jos hierarkiaa ja assosiatiivisia suhteita ei selvästi eritellä.

Mikäli assosiatiivisia suhteita haluttaisiin kehittää ja niiden käyttöä systematisoida, olisi mahdollista tyypitellä suhteita nykyistä tarkemmin esimerkiksi erittelemällä

toiminta-käsitteiden ja pysyvän alakäsitteiden (kuten paikka, ajanjakso, henkiset tuotokset) välisiä suhteita. Mm. terminologian alan tutkimuksessa, esimerkiksi (Nuopponen, 1994), on selvitetty tyypittelymahdollisuuksia, mutta käyttöön ei ole vakiintunut yhtenäistä yleistä tyypittelyä, vaan eri käyttötarkoituksiin on muodostettu erilaisia ratkaisuja. Tekoälytutkimuksessa, tietämyksen esittämisessä (knowledge representation) ja kielitieteessä käytetään mutkikkaiden semanttisten suhteiden esittämiseen usein ns. tapahtumaperustaista mallia, jossa yhteydet käsitteiden välille muodostuvat epäsuorasti ja niihin erilaisten roolien kautta kytkeytyvien yhteyksien kautta (esimerkiksi tapahtuman tyyppi, toimija, kohde, paikka, aika jne.) (Sowa, 2000). Vaikka FinnONTO-hankkeen aikana assosiatiivisten suhteiden tarkempaa tyypittelyä ei ole sisällytetty ontologioihin, erilaisia suhteita on hyödynnetty semanttisen webin tekniikkaa ja ontologioita hyödyntävissä portaaleissa kuten Kulttuurisammossa²⁹ (Mäkelä et al., 2012) ja Kirjasammossa³⁰ (Mäkelä et al., 2011). Kulttuurisammon hakujärjestelmä voi esimerkiksi antaa suosituksia kohteista, joilla on sama tekijä tai jotka käsittelevät samaa aihetta kuin haun tuloksena löytynyt kohde.

3.3 Roolien käyttö

Rooliksi luokittelua on pyritty rajaamaan niin, että luokkaan sisällytetään vain sellaiset käsitteet, jotka eivät ole muiden kuin toisten roolien ensisijaisia yläkäsitteitä ja joiden sijoittaminen muualle hierarkiaan edellyttäisi tarpeetonta moniperintää tai käsitteiden keinotekoista erottelua, mutta ei siltikään takaisi kattavan kuvauksen aikaansaamista. Esimerkiksi aineiston roolissa voivat olla hy-vin erilaiset fyysiset objektit ja aineettomat tuotokset. Jos aineistolle annettaisiin yläkäsite sekä fyysisten objektien hierarkiasta että aineettomien tuotosten hierarkiasta, aineisto perisi molempien yläkäsitteiden piirteet. On todennäköistä, että tällaisen periaatteen noudattaminen aiheuttaisi virheitä, sillä käsitettä, jolle molempien hierarkiahaarojen piirteet periytyisivät, ei aina ole aidosti olemassa. Myöskään käsitteiden keinotekoinen eriyttäminen niin, että toinen aineisto-käsite sijoitettaisiin fyysisten objektien hierarkiaan ja toinen aineettomien tuotosten hierarkiaan ei ole perusteltua, koska tällaista jakoa ei ole suomen tai ruotsin kielessä, jotka muodostavat YSO/Allson käsittejaon alkuperäisen perustan. Vain yhteen hierarkiaan rajoittuvat käsitteet, jotka voidaan tulkita rooleiksi, on kuitenkin saatettu sijoittaa myös muuhun kuin rooli-hierarkiaan, jos kyseiselle käsitteelle on näin löytynyt luonteva paikka (esim. kotieläimet eläinten hierarkiassa).

3.5 Monimerkityksisten asiasanojen käsittely

Jos sanalla on useampia merkityksiä, kullekin ontologian kannalta tarpeelliselle merkitykselle on ontologiassa annettava oma tarkenteella varustettu nimityksensä ja

²⁹ <http://www.kulttuurisampo.fi/>

³⁰ <http://www.kirjasampo.fi/>

käsitteet sijoitettava omille paikoilleen hierarkiaan. Ontologiassa käsiteltävät aihealueet vaikuttavat siihen, mitkä sanan merkityksistä ovat tarpeellisia. Yhdellä tietyllä erikoisalalla ei esimerkiksi välttämättä tarvita kaikkia sanaan liittyviä merkityksiä (käsitteitä), kun taas useita aloja kattavan ja moneen käyttöön tarkoitettun yleisen ontologian tulee kattaa käsitteistöä monen aihealueen tarpeisiin ja erilaisiin sovel-lustarpeisiin.

Tarkenteiden käyttö helpottaa käsitteiden erottamista myös termitasolla ja tukee näin niiden johdonmukaista käyttöä. Aluksi YSO/Allsossa käytettiin tarkenteina numeroita, esim. laamannit_1 ja laamannit_2. Sittemmin numeromuotoiset tarkenteet on korvattu sanallisilla tarkenteilla, esim. laamannit (arvonimet) ja laamannit (ammatit). Jos YSO/Allsoon perustuva yhdistelmäontologia on tehty varhaisessa vaiheessa, ontologiassa saattaa vielä esiintyä numerotarkenteita sanallisten tarkenteiden sijaan. Nämä on syytä jatkossa korjata.

Sanallisen tarkenteen muodostamisessa on ensisijaisesti käytetty lähimmän yläkäsitteen nimeä. Koska tämä menettely ei kuitenkaan kaikissa tapauksissa riitä erittelemään käsitteitä (eroteltavien käsitteiden lähin yläkäsite voi olla sama tai lähin ontologiakokonaisuudessa käytettävissä oleva yläkäsite ei muuten tuo eroa esiin riittävän selvästi), on osassa tarkennusta vaativista tapauksista käytetty muunlaisia tarkenteita. Esimerkiksi vahvistus (pätevyyden antaminen) ja vahvistus (vahventaminen) tai happamuus (mausta) ja happamuus (kemialliset ominaisuudet). Pääsääntönä tarkenteiden muodostamisessa on ollut tarkenteiden erottavuus, ymmärrettävyys ja lyhyys. Aluksi asiasanojen merkitykset esitettiin em. tavalla vain niissä tapauksissa, joissa merkitysten erottelu on tehty ontologiassa, mutta myöhemmin vastaavat tarkenteet lisättiin myös niiltä osin, kun merkitysten erottelu on tehty jo asiasanas-tossa. Joissakin tapauksissa asiasanastoon on huomautukseen merkitty näkyviin asiasanan useita merkityksiä, vaikka käytössä on vain yksi asiasana, toisissa tapauksissa taas samasta sanasta on käytetty yksikkö- ja monikkomuotoa (kuten ooppera/oopperat), joilla viitataan eri käsitteisiin. Tarkenteet on lisätty myös em. yksikkö/monikkopareihin, jotta näiden asiasanojen toivotunlainen ja yhdenmukainen käyttö olisi helpompaa myös niille, joille YSA/Allärsin kehitystyössä sovitut käsitteiden erot eivät ole ennalta tuttuja.

Asiasanaston monimerkityksisten asiasanojen sumeus poistuu hierarkiaan siirrettäessä, kun on esimerkiksi tarkennettava, viitataanko lapset-asiasanalla ikäryhmään vain perheenjäsenyyteen, sillä asiasanaston monimerkityksisiin asiasanoihin liittyvää tulkinnanvaraisuutta ei voida sisällyttää käsitejakoa noudattavaan ontologiaan. Lapset-tapauksessa erittely on tarpeen tehdä: ainoaksi luokittelukriteeriksi ei kannata valita periaatteessa pysyvää perheenjäsenyyttä (lapsella on ainakin ollut vanhemmat, vaikka perheyhteys ei olisi säilynyt, ja henkilö säilyy vanhempiensa lapsena myös aikuistuessaan), sillä yleisempänä kiinnostuksen kohteena saattaa olla lapset ikäryhmänä. Kun käsitteet on eritelty, niiden alakäsitteet saadaan oikeille paikoilleen hierarkiaan. Jos kuvailua varten tarjotaan käyttöön tarkkoja käsitteitä, voidaan

käsitevalinnalla tuoda esiin vastaava tarkennus, jonka konteksti luo juoksevassa tekstissä. Konteksti esimerkiksi yleensä paljastaa, onko kyse aluksen, astian vai kirjan kannesta, ja mikäli nämä kolme eri käsitettä erotellaan ontologiassa, voidaan kuvailu tehdä tarkemmin kuin monimerkityksisen kannet-sanavan avulla. Tarkempien kuvailukäsitteiden erittelyn myötä hakutulokset voidaan ryhmitellä/rajata entistä tarkemmin ja näin helpottaa relevanttien tietolähteiden löytymistä. Jos sanan merkitys tietyssä yhteydessä on kuitenkin tahallisesti tai tahattomasti jätetty tarkentamatta, ei kuvailukäsitteen valinnallakaan voida tilannetta muuttaa.

Joissakin tapauksissa kuvailussa ei käytetä kaikkia asiasanojen yleiskielisiä merkityksiä, esimerkkinä asiasana sijoitus (johdettu verbistä sijoittaa; määritelmien lähde Kielitoimiston sanakirja):

1. paikkaan liittyen
2. aikaan, tilanteeseen, olosuhteisiin tms. liittyen
3. jhk toimipaikkaan, työhön, palvelukseen asettamisesta t. ottamisesta, työn järjestämisestä jkille
4. käyttää rahaa taloudellisen hyötymisen toivossa esim. arvopaperien t. arvoesineiden ostamiseen, osuuden t. omistusoikeuden hankkimiseen jssak yrityksessä tms., investoida.

Kaikki yleiskieliset merkitykset eivät välttämättä ole tarpeen ontologiassakaan, mutta tässä tapauksessa ontologiassa on eritelty kolme Kielitoimiston sanakirjan merkityksistä. Neljäs merkitys on eritelty jo asiasanastossa valitsemalla tähän käyttöön asiasana sijoitustoiminta. Menettely, jossa jo-takin asiasanan merkitystä päätetään kuvata erillisellä asiasanalla, vaikeuttaa kuitenkin asiasanojen systemaattista käyttöä ja tiedonhakua. Tällaisissa tapauksissa käyttäjän olisi tunnettava asiasanoituskäytäntö tarkasti osatakseen ottaa käyttöön oikean hakusanan ja esimerkiksi huomatakseen, että kaikkiin sijoitus-sanavan merkityksiin ei tällaisessa tapauksessa pidä viitata sijoitus-sanalla.

Vaikka tarkkojen asiasanojen käyttö ei aina ole mahdollista tai välttämätöntä, niiden käyttömahdollisuuden tarjoaminen ontologian avulla on kuitenkin perusteltua. Silloinkin kun valitaan tiettyä teosta kuvaavat asiasanat, tarkemman asiasanan käytöstä on hyötyä, sillä se mahdollistaa esimerkiksi hakutulosten täsmällisen rajauksen tiedonhakuovelluksissa, vaikka käyttäjä ei tuntisi monipuolisesti asiasanoitusta tai eri hakumenetelmiä. Kun taas halutaan kuvailla kokonaisen teoksen sijaan esimerkiksi yhtä, sisällöltään suppeaan aihealueeseen rajautuvaa verkkosivua, ei kuvailu ole mielekäästä, ellei käytössä olevia käsitteitä ole eritelty riittävän tarkasti. Jos käytössä on vain yleisen tason kuvailukäsitteitä, ne eivät riitä erottamaan haku kohteita toisistaan, esimerkiksi toisiinsa liittyviä aiheita käsitteleviä verkkosivuja. Tarkasti määritellyt käsitteet helpottavat myös mahdollisesti tarvittavien erikielisten vastineiden valintaa. Kielten välisten käsite-erojen takia hyvän vastineen valinta ei aina ole mahdollista ilman asiasanan merkityksen selkeää rajausta.

3.5.1 Koostekäsitteet

Sellaisia kuvailutilanteita varten, joissa tarkan asiasanan käyttö ei ole aineiston luonteen tai käytettävissä olevien resurssien takia mahdollista, YSO/Allson oheen on sijoitettu hierarkiaan kuulu-mattomia koostekäsitteitä, jotka ilmentävät asiasanaston tarkentamattomien asiasanojen tapaan useampaa kuin yhtä käsitettä. Koostekäsitteet on luotu aina sen kielen näkökulmasta, jossa moni-merkityksisyys ilmenee. Jos monimerkityksisyyttä esiintyy monikielisessä ontologiassa saman mallin mukaisesti useammassa kielessä, luodaan yksi koostekäsite, jolle annetaan tarvittavat nimi-tykset kullakin kielellä. Jos monimerkityksisyyttä taas esiintyy vain yhdellä ontologian kielistä, luodaan koostekäsite tuon kielen näkökulmasta ja annetaan sille nimitykset vain yhdellä kielellä. Koostekäsitteiden käyttö aineistojen kuvailussa on tarkoitettu rajattavaksi niihin tapauksiin, jossa tarkan kuvailukäsitteen valinta ei ole mahdollista.

Toistaiseksi YSO/Allsoon on luotu koostekäsitteitä vain suomen ja ruotsin tarpeiden perusteella.

3.6 Monikielisen ontologian rakentaminen

Yksi ontologian tärkeä etu asianastoihin verrattuna on monikielisyys tai pikemminkin kieliriippu-mattomuus. Semanttisessa webissä sekä tiedon indeksointi että haku perustuvat kieliriippumat-tomiin URI-tunnisteisiin samalla kun ihmiskäyttäjälle voidaan näyttää näihin liittyvät nimikkeet tarpeen mukaan eri kielillä. KOKOn suomalaisissa asiasanastoissa ei tunnisteita ollut käytettävissä ja ne olivat yleensä yksikielisiä.

YSO-ontologiaan on yhdistetty ruotsinkielinen ontologiamuunnos Allso, joka on muokattu YSAa vastaavasta Allärs-asiasanastosta. FinnONTO-projektin yksi tulos oli myös ensimmäinen YSA/YSOn käänös englanniksi.

YSO/Allsoa on aluksi kehitetty yksikielisenä, suomen kielen näkökulmasta. Koska YSAa vastaavat ruotsinkieliset asiasanat ovat kuitenkin saatavilla Allärsistä ja ontologiaan on lisätty Helsingin kaupunginkirjaston luokitusjärjestelmästä ³¹ (HKLJ) saadut YSO/Allson käsitteitä vastaavat englanninkieliset vastineet, voidaan YSO/Allsosta kehittää monikielinen ontologia. Åbo Akademi onkin tarkistanut ja täydentänyt Allärsistä peräisin olevia ruotsinkielisiä vastineita ontologian vaatimalla tavalla. FinnONTO-hankkeen aikana ontologiaan on lisätty myös englanninkielisiä vastineita HKLJ:stä tulevan aineiston täydentämiseksi.

³¹ <http://hklj.kirjastot.fi/>

Monikielistä ontologiaa kehitettäessä ja ylläpidettäessä on huomioitava, että lähtökielenä toimi-vaan suomen osuuteen tehtävät muutokset saattavat edellyttää muutoksia myös muihin kieliin. Vaikka Allärsia on kehitetty aktiivisesti YSA:n myötä, suomen ja ruotsin vastaavuus YSO/Allsossa vaatii vielä tarkistuksia, sillä ruotsin osuuden kehitystä ei pystytty ohjelmistoteknisistä syistä täysin yhdistämään suomen osuuden kehitykseen FinnONTO-hankkeen aikana.

Monikielisen ontologian rakentaminen edellyttää käytettävien kielten, kääntämisen ja ontologioiden asiantuntemusta. Kun etsitään käsitettä vastaavaa vieraskielistä nimitystä, on tarkistettava tai selvítettävä, mihin käsitteeseen lähtökielen asiasana viittaa ja missä konteksteissa sitä käytetään. Erikoisalojen termit vaativat jo itsessään asiantuntemusta, mutta yleistajuisempienkaan asiasanojen merkitys ei aina ole ilmiselvä pelkän sanan kieliasun perusteella. Näin on esimerkiksi monimerkityksisten sanojen kohdalla, ellei sanojen taustalla olevia käsitteitä ole eritelty. YSO/Allsoa varten käsitteiden merkitystä ja käyttökontekstia koskevaa tietoa löytyy YSA/Allärsista ja esim. Fennica-tietokannasta: YSA/Allärsista voidaan tarkistaa, mitä alaa tai aloja asiasana koskee ja Fennicasta millä tavoin asiasanaa on siellä käytetty sisällönkuvailussa. Erikoisalojen asiasanojen merkityksen ja käyttökontekstit tuntevat varmimmin kyseisen alan asiantuntijat ja asiasanastoja hallinnoivien organisaatioiden asiasanastovastaavat, jotka ovat päättäneet asiasanan käyttöönnotosta.

Kun lähtökielen käsite on selvillä, on varmistettava luotettavan lähteen avulla, että kohdekielen vastinetta käytetään mahdollisimman tarkasti samassa merkityksessä kuin lähtökielen sanaa ja et-tä vastine on vakiintuneessa käytössä. Vieraskielisiä vastineita valittaessa on käytettävä parhaita mahdollisia tietolähteitä, erityisesti erikoissanastoja ja -sanakirjoja sekä asiantuntijoita niiltä aloilta, joihin käsitteet liittyvät. Mikäli luotettavista lähteistä löytyy useita vastinevaihtoehtoja, on selvitettävä, mikä vaihtoehdoista vastaa tarkimmin lähtökielen käsitettä. Usein vaihtoehdoilla saattaa olla käyttökontekstiin liittyviä eroja tai vivahde-eroja, jotka selviävät termien käyttöä vertailemalla. Jollei merkityseroja ole, valitaan kielenhuoltajien suosittelema tai eniten käytetty vastine. Mikäli vastine löytyy epäsuorasti esimerkiksi saman aihepiirin erikielisiä tekstejä vertailemalla, on varmistettava, että lähtö- ja kohdekielen sanat viittaavat samaan käsitteeseen. Tämä selviää yleensä perehtymällä kohdekielen termin käyttöön laajemmin.

Vieraskielisiä vastineita valittaessa on huomioitava myös sanojen mahdollinen monimerkityksisyys. Asiasanan ollessa monimerkityksinen, sen taustalla olevat käsitteet on voitu eritellä jo muokattaessa ontologiaa lähtökielen näkökulmasta. Osa sanojen monimerkityksisyydestä tulee kuitenkin esiin vasta vastinetyötä tehtäessä. Mikäli kohdekielestä kuitenkin löytyy kullekin käsitteelle oma erillinen nimityksensä, on suositeltavaa valita vastineeksi tällainen nimitys. Jos näin ei ole, on vastine varustettava tarkenteella kuten lähtökielessäkin. Jotta ei syntyisi tilanteita, joissa sama tarkentamaton vastine valitaan useamman käsitteen vieraskieliseksi nimityk-

seksi, on hyvä tarkistaa ennen vastineen lisäämistä, ettei se ole käytössä jo toisen käsitteen nimityksenä. Jotta käsitteiden ero näkyisi myös nimitysten tasolla, ontologiassa ei saa olla kahta erillistä käsitettä, joista on suositettu käytettäväksi samaa nimitystä (skos:prefLabel) samalla kielellä.

3.6.1 Monikielisen ontologian haasteet

Kulttuurierojen vuoksi kaikille käsitteille ei ole olemassa vakiintuneita nimityksiä eri kielissä. Jos kulttuurit ovat toisilleen läheisiä ja ovat saaneet paljon vaikutteita toisistaan, eroja on vähemmän, mutta esimerkiksi yhteiskunnan rakenteisiin ja toimintaan liittyvissä käsitteissä eroja on silti usein. Joissakin tapauksissa suositettavat nimitykset tuovat esiin eri kielten käsitteiden välisiä eroja, joissakin taas hämärtävät niitä. Suositettavien nimitysten valintaan kannattaa panostaa, sillä ne ovat viestinnän ja tiedonhaun lähtökohta, vaikka käsitteen tunnisteena käytettäisiinkin URLia.

YSO/Allso sisältää käsitteitä, jotka rajoittuvat lähinnä suomalaiseen kulttuurialueeseen, eikä niille ole tarkkoja vakiintuneita vastineita kielialueilla, joihin käsitteillä ei ole liittymäkohtaa. Tällaisia ovat esimerkiksi Suomen historiaan liittyvät käsitteet kuten kalterijääkärit ja Viipurin pamaus, yhteiskuntajärjestelmään liittyvät käsitteet kuten perhekeskus ja erikoisnuorisotyö sekä paikalliseen kulttuuriin liittyvät käsitteet kuten rintamamiestalo. Yhteiskunnallinen kehitys ja kulttuuri ovat vaikuttaneet myös ammattinimikkeiden osittaiseen eriytymiseen (esimerkiksi kunnanlääkäri ja lähikasvattaja). Oman haasteensa aiheuttavat myös käytöstä poistuneet käsitteet kuten asemapäällikkö ja suostuntavero, sillä niille mahdollisia vastine-ehdokkaita on usein vähemmän kuin edelleen käytössä oleville käsitteille, eikä käsitteiden sisällystään aina löydy riittävästi vastineen valintaa tukevaa tietoa.

Lähinnä suomalaiseen kulttuurialueeseen liittyvien käsitteiden lisäksi vastineiden valinta voi olla vaikeaa myös sellaisissa tapauksissa, joissa eri kielten käsitteet vastaavat toisiaan vain osittain. Tällöin vastinetta valittaessa joudutaan miettimään, sopiiko vastineeksi kohdekielen sana, joka viittaa erilailla rajautuvaan käsitteeseen kuin lähdekielen sana vai pitäisikö käsitteiden ero tuoda esiin myös sanatasolla luomalla uusi vastine.

Käsite-erot ja terminmuodostuksen erot vaikuttavat myös hierarkian esittämiseen. Esimerkiksi YSO/Allson tarvikkeet-käsitteellä on erilaisia alakäsitteitä, joissa suomen sanan perusosalle (tarvike) löytyy yhteensä kahdeksan erilaista vastinetta englanniksi: atk-tarvikkeet = computing accessories, pakkaustarvikkeet = packaging materials, rakennustarvikkeet = building components, sairaanhoitotarvikkeet = nursing equipment jne. Joissakin tilanteissa vastaavaa variaatiota voi karsia tarkistamalla, olisiko yhtenäisen terminmuodostusperiaatteen käyttö mahdollista, mutta toisaalta yhtenäistämistä ei pidä tehdä kielen käytänteiden vastaisesti, vaan valinnoilla tulisi aina olla vakiintuneeseen kielenkäyttöön liittyvä peruste. Se, ettei käsit-

teiden nimityksiä eri kielissä tai yh-den kielen hierarkiakokonaisuudessa ole muodostettu yhtenevällä tavalla, ei välttämättä tarkoita, etteivätkö käsitteet kuuluisi samaan hierarkiaan. Käytännössä terminmuodostusperiaatteiden ja käsitteiden väliset erot vaikeuttavat kuitenkin kaikille alakäsitteille sopivan yläkäsitteen nimityksen valitsemista.

Jos halutaan, että kaikilla ontologian käsitteillä on nimitys kaikilla ontologian kielillä, voi vastinetyön tekijä muodostaa käännösvastineen sellaisessa tapauksessa, jossa vakiintunutta vastinetta ei ole olemassa. Tällaisissa tilanteissa muodostettujen käännösvastineiden ongelmana on kuitenkin se, että ne eivät ole käyttäjille ennalta tuttuja, eikä niitä siksi välttämättä osata hyödyntää. Jos em. periaatteen noudattaminen ei ole välttämätöntä, kannattaa vastine jättää lisäämättä, ellei vakiintunutta vastinetta ole olemassa tai vastine-ehdotuksen sopivuutta ei voida tarkistaa kohdekielen hyvin hallitsevilta asiantuntijoilta. Varsinkaan erisnimien kääntäminen ei ole suositeltavaa, ellei nimen haltija itse ole ottanut käyttöön nimeä eri kielillä.

3.6.2 Ruotsinkieliset asiasanat

YSO/Allson suomen käsitteitä vastaavat ruotsinkieliset asiasanat tulevat Allärsista. Vaikka ruot-sinkieliseltä käsitteeltä puuttuisi suomenkielinen vastine, käsite voidaan sijoittaa ontologian hierarkiaan. Allärsin mallin mukaan ruotsin käsitteelle valitaan yläkäsitteeksi aina toinen ruotsin käsite, vaikka joissain tapauksissa suomenkielisestä hierarkiasta voisi löytyä tarkempi yläkäsite. Esimerkiksi floder-käsitteen yläkäsitteeksi tulee strömmande vatten – ei suomen joet-käsite.

Jotta asiasanastosta voitaisiin kehittää aidosti monikielinen ontologia, on tarkistettava kaikkien ontologiassa tarkennettujen käsitteiden osalta, että asiasanan ruotsinkielinen vastine pätee myös ontologiassa tehtyjen hierarkian tarkennusten jälkeen. Tarkistuksen apuna käytetään ensisijaisesti hierarkista yläkäsitettä ja sen lisäksi käsitteestä lisätietoja antavia assosiatiiivisia suhteita tai huomautustietoja. Tämä on tarpeen siksi, että asiasanastossa tarkentamattoman käsitteen vastine saattaa olla valittu toisin perustein kuin mikä on myöhemmin valittu ontologiassa esitettäväksi käsitteeksi.

Tarkistusta vaativat erityisesti sellaiset tapaukset, joissa YSO/Allsossa on otettu käyttöön tietty asiasana useammassa kuin yhdessä merkityksessä – tällöin Allärsissa annettu ruotsinkielinen vastine ei välttämättä sovi kaikkiin merkityksiin. Esimerkiksi kirjamuotoisen oppaan vastine han-dledning ei sovi käytettäväksi viitattaessa oppaana toimivaan henkilöön. Monimerkityksisyys saattaa olla implisiittisesti mukana myös YSAn aineistossa niin, että suomenkieliselle asiasanalle on annettu useampi ruotsinkielinen vastine, jotka eivät ole toistensa synonyymeja – tällaisessa tapauksessa vastineet on kohdistettava eri käsitteisiin. Esimerkiksi:

YSA/Allärs:

monitoimikoneet LT kodinkoneet, metsäkoneet / flerprocessmaskiner, matberedare

YSO/Allso:

monitoimikoneet (metsäkoneet) - flerprocessmaskiner
monitoimikoneet (kodinkoneet) - matberedare

Lisäksi on tunnistettava tapaukset, joissa suomen ja ruotsin välillä on todellisuudessa (eikä vain puuttuvan käsite-erittelyn vuoksi) yhden suhde useampaan - käsitesuhde. Näiden tapauksien tunto-merkinä ei aina ole se, että asiasana puuttuisi toisella kielellä, vaan 1) samaa vastineena olevaa asiasanaa on voitu asiasanastossa käyttää useamman käsitteen kohdalla tai 2) vain yksi asiasanan merkityksistä näkyy asiasanastossa.

Esimerkiksi:

YSA/Allärs:

ruotsinkielinen sabotage-sana viittaa sekä sabotaasi että tuhotyö-käsitteisiin

YSO/Allso:

sabotage-sanaa on tarkennettu sulkumerkinnän avulla sen mukaan, mihin käsitteeseen se viittaa

sabotaasi - sabotage (målinriktad skadegörelse)

tuhotyö - sabotage (juridisk kategori)

Suomen kielen sivuvaikutukset-sana:

YSA/Allärs:

sidoeffekter – sivuvaikutukset (biverkningar-merkitys puuttuu)

YSO/Allso:

sidoeffekter - sivuvaikutukset (ennalta-arvaamattomat, neutraalit)

biverkningar - sivuvaikutukset (haitalliset)

YSO/Allsossa siis lisätään sulkutarkenne siihen asiasanaan, joka on monimerkityksinen, olipa ky-seessä sitten suomen tai ruotsin sana. Kun kieliä käsitellään tasaveroisina, voi molempien kielten lähtökohdista syntyä tilanteita, joissa kielten välinen käsite-ero aiheuttaa toisen kielen kannalta keinotekoisesti erottelun. Tarkenteiden luominen on haasteellista, sillä käsitteiden väliset erot eivät välttämättä ole yksiselitteisiä, eikä aina ole käytettävissä lähteitä, jotka selvittäisivät riittävällä tarkkuudella kahden eri kielen käsitteiden välisiä eroja.

Koska YSAa ja Allärsia on kehitetty yhdenmukaisesti, riittää että ruotsin kielen osalta tarkistetaan hierarkia. Ruotsin käsitteiden assosiatiivisia suhteita ei tarvitse erikseen käsitellä, koska ne vastaavat YSON assosiatiivisia suhteita ja tulevat käsitellyksi yhteisestä näkökulmasta sitten, kun niitä käsitellään suomenkielistä osiota kehitettäessä.

3.6.3. Asiasanojen englanninkieliset vastineet

YSO/Allson linkittäminen kansainvälisiin ontologioihin edellyttää englanninkielisten vastineiden lisäämistä ontologiaan. Yli 2/3 asiasanojen englanninkielisistä vastineista on tuotu YSO/Allsoon Helsingin kaupunginkirjaston luokitusjärjestelmän (HKLJ) englanninkielisestä versiosta, jonka käännöstoimisto MarTina teki vuonna 2007. Koska englanninkieliset vastineet on kerätty luokitusjärjestelmän perusteella ennen ontologian rakentamista, työssä ei ole voitu hyödyntää vastineiden valinnan tukena ontologian rakenteita, eikä käytettävissä ole ollut ohjeistusta esimerkiksi ontologiassa käytettävien asiasanojen muodosta. Näin ollen käännökseen liittyy sanojen monimerkityksisyyden kautta syntyneitä virheitä, eikä monimerkityksiin, useiden eri käsitteiden nimityksenä esiintyviin sanoihin ei ole liitetty tarkenteita. Joissakin tapauksissa myös asiasanan muoto poikkeaa suomessa ja ruotsissa käytetyistä muodoista niin, että asiasanan perusosa on erotettu määreestään pilkulla.

Englanninkielisen HKLJ-luokituksen lisäksi puuttuvia englannin vastineita on lisätty YSO/Allsoon FinnONTO-hankkeessa. Vuonna 2009 ontologiaa täydennettiin noin 6 000 HKLJ:stä puuttuneella vastineella ja myöhemmin vastine lisättiin vielä noin 2 000 käsitteelle, jotka ovat tulleet YSO/Allsoon YSA/Allärs-päivityksen kautta. FinnONTO-hankkeen aikana ei kuitenkaan systemaattisesti tarkistettu HKLJ:stä tulleita vastineita, eikä HKLJ:n ja FinnONTO-työssä lisättyjen vastineiden yhteentoimivuutta hierarkiassa.

Kun YSO/Allsoon on lisätty englannin vastineita, työn tarkoituksena ei ole ollut laajentaa YSO/Allsoa englannin kielen käsitteillä, vaan antaa englanninkieliset vastineet olemassa oleville YSO/Allson käsitteille, jotta ontologian siltaus muihin englantiä sisältäviin ontologioihin olisi mahdollista.

Yksi YSO/Allson mahdollinen siltauksen kohde on laaja englanninkielinen WordNet, joka on kään-netty monille kielille, myös suomeksi. Kun asiasanoilla on englanninkieliset vastineet, suomen käsitettä vastaavat englannin käsitteet voidaan etsiä englannin vastineen avulla WordNetistä ja sillata YSO/Allsoon. Siltaus voidaan tehdä alustavasti koneellisesti ja tarkistaa/vahvistaa ihmistyönä. Samalla kun tarkistetaan koneellisesti tehdyn ekvivalenssilinkityksen paikkansapitävyys, voidaan tarkistaa myös onko YSO/Allsoon alun perin valittu vastine oikea suhteessa siihen, missä ontologi-an hierarkian kohdassa nimettävä käsite sijaitsee.

Siltausta varten on luotava säännöt esimerkiksi sille, millaisia vastaavuusvaatimuksia käsitteille asetetaan, jotta ne voitaisiin tarkistuksen yhteydessä vahvistaa ekvivalenteiksi eli kuinka tarkkaan eri kielten käsitteiden sisältöjen tulee vastata toisiaan. On myös päätettävä pidetäänkö WordNetiä ensisijaisena englanninkielisten vastineiden lähteenä eli korjataanko aiemmin poimittuja vastineita WordNetin mallin mukaan ja toisaalta miten toimitaan niissä tapauksissa, joissa käsitettä ei löydy WordNetistä (esimerkiksi erityisesti suomalaiseseen kulttuuriin ja yhteiskuntaan liittyvät käsitteet). Erityisen haasteen muodostavat käsitteet, joilla on kielten välillä yhden suhde useampaan -käsitesuhde eli esimerkiksi yhtä englannin käsitettä vastaa kahden suomen käsitteen osittainen yhdistelmä. Näissä tapauksissa sopivimman vastineen valinnalle ja ekvivalenssin epätäydellisyyden arvioinnille tarvitaan peruseriaatteet, joita voidaan soveltaa eri tapauksissa.

3.7 Ontologioiden päivitys

Jotta ontologia pysyisi ajantasaisena, sitä on päivitettävä esiin tulevien tarpeiden perusteella jat-kuvaluonteisesti. Muutostarpeita tulee väistämättä käsitteistön kehityksen myötä, mutta muutoksia voidaan tarvita myös ontologian rinnalla mahdollisesti kehitettävän asiasanaston muutosten myötä tai muissa ontologioissa tehtävien muutosten vuoksi. Ontologiaan tarvittavat muutokset voivat liittyä suositettiin (erikielisiin) asiasanoihin tai käsitteiden välisiin suhteisiin sekä käsitteiden lisäyksiin.

Jotta eri erikoisalojen käsitteistöön ja ontologioiden erilaisiin käyttötarkoituksiin perustuvat tarpeet voitaisiin ottaa huomioon ja välttää päällekkäistä työtä ontologiakehityksessä, tarvitaan tiivistä yhteistyötä YSO/Allson ja sen pohjalle rakennettujen erikoisalojen ontologioiden ylläpitäjien kesken.

3.7.1 Asiasanaa koskevat muutokset

Vaikka ontologian käsite muuten säilyisi ennallaan, voi olla tarpeen vaihtaa nimitystä, jota suositetaan käytettäväksi käsitteeseen viitattaessa tai lisätä käytössä olevalle nimitykselle synonyymi. Jos päivitystieto tulee asiasanastosta tai ontologiasta, joka ei sisällä koneluettavaa tietoa aiemmin käytössä olleen ja uuden nimityksen välisestä

suhteesta, uusi nimitys näyttäytyy päivityksen alkuvaiheessa uutena käsitteenä. Tällaisessa tilanteessa ontologian kehittäjän tulisi tunnistaa käsite ja yhdistää uusi nimitystä koskeva tieto ontologiassa jo olevan käsitteen tietoihin eli merkitä uusi nimitys joko käsitteen suositettavaksi nimitykseksi (skos:prefLabel) tai synonyymiksi (skos:altLabel) päivitystiedon sisällön mukaan. Samalla kun tiedot yhdistetään yhden käsitteen tiedoiksi, poistetaan uuteen nimitykseen mahdollisesti liitetty käsitekohtainen URI tarpeettomana. URIn poistaminen tässä vaiheessa on vielä mahdollista, koska URIa ei ole julkaistu, vaan se on otettu käyttöön ainoastaan kehitystyössä. Julkaisun jälkeen käsitteen URIa ei koskaan muuteta eikä poisteta.

Silloin kun suositettavaa käsitteen nimitystä vaihdetaan, tulisi myös varmistaa, onko edelleen kyse yhdestä ja samasta käsitteestä vai onko nimityksen muutoksen syynä se, että käsitteen sisältö on muuttunut ja itse asiassa vanha ja uusi käsite pitää säilyttää erillisenä. Tämä päätös pitää tehdä tapauskohtaisesti. Esimerkiksi sanat inhibiitti ja inhibiittori viittaavat samaan käsitteeseen, eikä niitä pidä säilyttää ontologiassa erillisenä, vaan suositettava asiasana + korvattu asiasana parina. Sen sijaan mielisairaala ja psykiatrisen sairaala viittaavat kahteen eri käsitteeseen, joista ensimmäinen sopii käytettäväksi vain historiallisessa yhteydessä.

Jos päivitystiedon lähde ei sisällä tietoa käsitteen vanhan ja uuden nimityksen yhteydestä, ontologian kehittäjän voi olla vaikea tunnistaa tilanteita, joissa kyse on vain nimityksen vaihtumisesta tai lisäyksestä, eikä uudesta käsitteestä. Käsitteen tunnistamiseksi pitäisi tuntee aineisto erittäin hyvin tai huomata uudeksi oletettua käsitettä hierarkiaan sijoitettaessa, että käsite on jo mukana ontologiassa vaikkakin eri nimityksellä varustettuna. Tällainen tilanne voi syntyä yhden ontologian sisällä silloin, kun ontologiaan on lisätty tietty käsite rakenteellisista syistä ja sama käsite lisätään (mahdollisesti eri nimityksellä varustettuna) myöhemmin asiasanastoon, jonka perusteella ontologiaa päivitetään. Yhdistelmäontologioiden kehityksessä vastaavia tilanteita syntyy helposti, jos samasta käsitteestä on käytetty eri nimitystä YSO/Allsossa ja erikoisalan ontologiassa.

Kun on kyse monikielisestä ontologiasta, jossa tarkistuksen yhteydessä vaihdetaan asiasanan vastinetta jollakin kielellä, voi olla tarpeen tarkistaa, ettei vastinevalinta muuta tulkintaa käsitteestä ja siten vaikuta ontologian hierarkiaan. Tällaiset muutostarpeet ovat todennäköisesti harvinaisia, mutta myöhemmin virheelliseksi osoittautunut vastinevalinta on saattanut vaikuttaa käsitteen hierarkian rakentamiseen lähtökielen käsitteen rajauksen ollessa epävarma, ja siksi tällainen varmistus on tarpeen. Vastinetta valittaessa on otettava huomioon soveltuvuus sekä yleiskielen että erikoiskielen näkökulmasta, mikäli käsite on käytössä molemmista.

3.7.2 Käsitteen hierarkiaa koskevat muutokset

Koska asiasanastosta ontologiaan siirryttäessä ontologisointi vaatii yleensä runsaasti hierarkiaan kohdistuvia muutoksia, ei ole mielekäs käydä jokaisen päivityksen yhteydessä läpi samoja asiasanaston ja ontologian välisiä hierarkiaeroja, jotka on jo aiemmin käsitelty, vaan huomio tulee kohdistaa asiasanaston sisällä tapahtuneisiin hierarkiamuutoksiin ja niiden mahdollisiin ontologian hierarkiaan kohdistuviin vaikutuksiin. Muutokset voivat koskea joko ylä- tai alakäsitteen lisäystä tai tällaisen suhteen poistoa tai näiden yhdistelmää.

3.7.3 Käsitteen assosiatiivisissa suhteissa tapahtuneet muutokset

Käsiteltäessä asiasanaston ja ontologian välisiä assosiatiivisten suhteiden muutoksia, uudet assosiatiiviset suhteet tuodaan ontologian kehitysversioon omana tietotyypinään, jotta niiden vertailu olemassa oleviin assosiatiivisiin suhteisiin olisi mahdollisimman helppoa. Vertailun tuloksena uudet hyväksyttävät assosiatiiviset suhteet siirretään varsinaisiksi julkaistaviksi assosiatiivisiksi suhteiksi ja tarpeettomat suhteet poistetaan. Osa asiasanaston assosiatiivisista suhteista saattaa olla ontologiassa tarpeettomia, jos käsitteiden hierarkiaa on ontologisoitaessa muutettu, sillä samojen käsitteiden välillä ei pidä olla sekä hierarkkista että assosiatiivista suhdetta. Assosiatiivista suhdetta ei myöskään tarvitse toistaa hierarkian alemmilla tasoilla, jos se on yläkäsitteen olennainen piirre ja periytyy alakäsitteille. Koska em. muutokset voivat olla tarpeen, ei assosiatiivisia suhteita voida siirtää päivityksen yhteydessä asiasanastosta ontologiaan automaattisesti, vaan ne tulee tarkistaa erikseen.

3.7.4 Ontologian käsitekokonaisuuteen liittyvät muutokset

Pääsääntöisesti YSO/Allsoon tehtävät käsitteiden lisäykset ovat perustuneet YSA/Allsiin tehtyihin lisäyksiin. Tämän lisäksi YSO/Allsoon voidaan lisätä apukäsitteitä hierarkian selkeyttämiseksi tai eritellä käsitteitä, joihin viitataan monimerkityksisillä asiasanoilla.

Jos asiasanastoa ontologisoitaessa eritellään useampia käsitteitä, joihin voidaan viitata samalla asiasanalla, on tarkistettava, että tällaisten käsitteiden mahdolliset alakäsitteet sijoittuvat oikeille paikoilleen hierarkiaan. Kun eritelyihin käsitteisiin liittyy asiasanaston myöhemmissä versioissa muutoksia, on tarkistettava, mihin ontologian eritellyistä käsitteistä muutokset liittyvät ja tehtävät muutokset kyseisen käsitteen tietoihin.

Edellä kuvattu käsitteiden erittely saattaa myös myöhemmin muuttaa muotoaan. Muutos tapahtuu esimerkiksi jos ontologian rakenteessa otetaan huomioon asiasanastossa tehty muutos, jossa aiemmin ontologiassa mukana ollut, mutta

asiasanastosta puuttunut käsite otetaan mukaan, ja siitä suositetaan käytettäväksi eri nimitystä kuin ontologiassa. Esimerkiksi YSO/Allsoon oli otettu jäljittely (matkiminen) -käsite ennen kuin matkiminen tuli mukaan YSA/Allärsiin. Tällaiset tilanteet pitäisi kuitenkin pyrkiä estämään asiasanasto- ja ontologiatyön koordinoinnilla.

Mikäli käsitteiden erittelyssä on tapahtunut virhe ja se joudutaan tästä syystä purkamaan, on tarkistettava, että mahdolliset alakäsitteet sijoittuvat hierarkiaan oikeille paikoilleen.

Käsitteitä saatetaan poistaa käyttöön suositettavasta ontologian käsitekokonaisuudesta, mikäli niiden ontologisointiin liittyy virhe. Sellaisia käsitteitä, joille on annettu ja julkaistu pysyvä URI, ei kuitenkaan koskaan poisteta kokonaan, vaan niillä säilytetään viittaussuhde käytössä olevaan ontologian versioon. Näin voidaan taata eri yhteyksissä tehtyjen URI-perustaisten merkintöjen toimivuus.

3.8. Ehdotuksia jatkotoimenpiteiksi

Tässä ontologiatyön kuvauksessa on kerrottu FinnONTO-hankkeen aikana tehdyn ontologiatyön perusteista. Kuten kuvauksesta käy ilmi, hankkeen aikaisen työn tuloksiin liittyy kokeiluja ja keskeneräisyyttä, joiden osalta on tarpeen tehdä pysyviä linjauksia YSO/Allson jatkokehitystä suunniteltaessa.

Jotta päätökset voitaisiin tehdä kestävältä pohjalta, niiden tulee perustua ontologian sisällön tuntemiseen kokonaisuutena. Paras tapa saada kokonaiskäsitteily ontologiasta on tehdä ontologian kehitystyötä, käydä läpi hierarkiaa ja samalla korjata nykyisen rakenteen sisäisiä epä johdonmukaisuuksia ja virheitä. Tämän jälkeen on mahdollista arvioida, mitä vaikutuksia mahdollisilla kokonaisrakenteen muutoksilla olisi. Koska rakenteen muutokset voivat vaikuttaa YSO/Allson lisäksi kaikkiin sen ympärille rakennettuihin yhdistelmäontologioihin, on mietittävä tarkkaan, kuinka merkittävää hyötyä muutoksilla saataisiin aikaan ja kuinka paljon työtä muutosten toteuttaminen vaatisi.

Kokonaishierarkiaa koskevien linjausten lisäksi tarvitaan päätöksiä ja jatkotyön ohjeistusta esimerkiksi seuraavista:

- monimerkityksisten asiasanojen taustalla olevien käsitteiden erittely ja merkitseminen sekä näihin liittyvä ruotsin- ja englanninkielisten vastineiden esittäminen
- hierarkiaa ryhmittelevien käsitteiden käyttö
- koostekäsitteiden käyttö
- roolikäsitteiden hierarkisointi

-
-
- yksilökäsitteiden/erisnimien käsittely
 - o vaikka maantieteellisten nimien ryhmä on jätetty pois, ontologiassa on vieläkin paikkoihin viittaavia nimiä kuten Helsinki-Vantaan lentokenttä sekä muita erisnimiä
 - synonyymisten ja ei-synonyymisten ohjaustermien käsittely
 - asiasanasuosituksen muutosten esittäminen suhteessa käsitejakoon eli miten esitetään saman käsitteen muuttunut nimitys ja miten eri aikoina käytetyt, mutta toisiaan vastaavat erinimiset käsitteet
 - o käsitteen muutokset suhteessa URIn muutokseen: milloin otetaan käyttöön uusi URI.

Sääntöjen luomisen jälkeen sisältötyön tueksi tarvitaan joukko aineiston koneelliseen tarkistukseen sopivia menetelmiä, jotta laajan aineiston hallinnointi olisi mahdollista.

4 Yhteenveto

Tässä raportissa esitettiin yhteenveto FinnONTOssa kehitetystä ja pilotoidusta menetelmästä maas-samme käytössä olevien asiasanastojen muuttamiseksi joukoksi toisiinsa yhdistettyjä ontologioita. Tarkemmin käytiin läpi tapaustutkimuksena YSA/Allärsin muuttaminen kolmikieliseksi YSO-ontologiaksi. Tavoitteena on, että tämä kokemusperäinen esitys voisi toimia opastuksena KOKO-järjestelmän eri osa-ontologioita jatkossa kehitettäessä. Käsityksemme mukaan asiasanastoista kannattaa siirtyä ontologioihin, mutta asiasanastoja voidaan hyödyntää myös sellaisenaan. ONKI-palvelussa on julkaistu lukuisia asiasanastoja sovellusten kannalta näppärässä SKOS-muodossa ilman, että niille olisi tehty mitään semanttisia muutoksia.

Raportin toisena tavoitteena oli antaa lukijalla tarkempi käsitys siitä, miksi KOKO-ontologiat ovat rakenteeltaan sellaisia kuin ne ovat ja mihin tehdyt ratkaisut perustuvat. Kun työtä jatketaan ja käyttöön otetaan kenties uusia menetelmiä, voidaan uusia päätöksiä toivon mukaan tehdä paremmin tähän raporttiin koottujen aiempien kokemusten ja valintojen perusteella.

Kolmantena tavoitteenamme on ollut kuvata ontologiatyön nykytila FinnONTO-hankkeen jäljiltä, jotta alalla jatkossa tarvittavan työn sisältöä ja määrää voitaisiin paremmin arvioida. Tutkimushanke on luonut aiemman asiasanastotyön perustalle uudet puitteet ja koonnut yhteisön, jonka toimesta eri alojen ontologiatyötä voidaan jatkossa viedä toivon mukaan yhteisöllisesti eteenpäin. Olemme kuitenkin astuneet vasta ensimmäisen askeleen tälle tielle: esimerkiksi KOKO-pilven jäsentely ja työn koor-dinointi pitää vielä tehdä, pilottivaiheessa tehdyt rakenteet kaipaavat täsmennyttä, käännöksiä ei ole tarkistettu kuin osittain ja kotimaisten ontologioiden siltaaminen kansainvälisiin ontologioihin on tekemättä. KOKO-ontologioiden ensimmäisten versioiden pohjalta voidaan tähän kuitenkin nyt ryhtyä. Tärkeä FinnONTO-hankkeen tulos on, että ontologiatyö on sen seurauksena hyväksytty osaksi kansallista tietoinfrastruktuuria, joten alalla on toivon mukaan jatkossa käytettävissä enemmän resursseja.

Ontologiatyö, jo pelkästään insinöörien filosofiasta uusiokäyttöön ottama termi ontologia, on herättänyt monissa semanttisia intohimoja ja kritiikkiä. Myös FinnONTOlla on kriitikkonsa ja ontologisoinnin hyödyllisyyttä suhteessa työssä tarvittavaan lisäpanostukseen on epäilty. Kritiikkiä on herättänyt ontologisoinnin yhteydessä syntyvä yhä tarkempi jaottelu merkityksiin, mikä johtaa yhä isompaan ja mutkikkaampaan käsittehierarkiaan. Tämä lisää paitsi ylläpitotyötä myös työtä indeksoinnissa ja saattaa rajata termeihin liittyviä vivahteita liikaa. Jos esimerkiksi kuvailaan Albert Edelfeltin teos "Pariisin Luxembourgin puistossa", pitääkö siinä leikkivät lapset kuvata erikseen sekä ikäryhmänä, perhesuhteena että sosiaalisena ryhmä-

nä, ja katetaanko näillä lapsiin liittyvä merkitysten koko kirjo? Eikö olisi yksinkertaista vain lisätä asiasana lapset maalauksen metatietoihin? Eikö maailman kuvaamista loogisena struktuurina ole jo kokeiltu 80-luvulla tekoälytutkimuksen piirissä asiantuntijajärjestelmien yhteydessä ja opittu silloin, ettei se ole mahdollista?

Näitä huolia on pyritty huomioimaan FinnONTO:n piirissä. Tavoitteeksi asetettiin alun perin minimaalinen ja mahdollisimman yleiskäyttöinen sanastojen ontologisointi, joka kuitenkin ratkaisisi eräitä keskeisimpiä asiasanastoihin ja asiasanastotyöhön liittyviä haasteita tietotekniikan sovellusten kannalta. Merkitysten pirstoutumiseen liittyvää problematiikkaa pyrittiin huomioimaan mm. koostekäsitteillä, joita tarvitaan myös yhteentoimivuuden rakentamiseksi aiempiin asiasanastoihin. Ontologian laajenemisen ongelma liittyy samalla metatietomallien kehittämiseen, mikä kannattaa ottaa jatkossa myös huomioon. Ontologioitahan käytetään metatiedon esittämiseen eikä yksinään. Esimerkiksi Finn-ONTO-hankkeen piirissä kehitetty JHS-suositus 183 ”Julkisen hallinnon palvelujen tietomalli ja ryhmityt verkkopalveluissa” päätyi luomaan yhden laajan palveluontologian sijasta yleisemmän metatietomallin, jonka metatietokentille voidaan valita arvoja eri ontologioista. Tekoälytutkimuksen kokemuksista kumpuavaan kritiikkiin on todettava, että semanttisen webin lähtökohdat ovat logiikan soveltamisen suhteen hyvin erilaiset verrattuna tekoälyyn. Semanttisen webin ontologioiden ei tarvitse olla virheettömiä ja loogisesti eheitä voidakseen olla silti hyödyllisiä, päinvastoin kuin vaikkapa sytostaattihoitoa annostelevan asiantuntijajärjestelmän. Web ylipäätyäkään ei ole virheetön, mutta silti hyödyllinen. On kuitenkin selvää, että täsmälliseen tietoon perustuva logiikka asettaa rajoitteita luonteeltaan epätäsmällisen maailman kuvaamisella, ja että yhä automaattisemmin menetelmin ja yhteisöllisemmin tuotetun datan laatu tulee olemaan keskeisiä haasteita semanttisessa webissä.

Uusista yhä syvällisemmistä haasteista huolimatta webin kehitys on kuitenkin astumassa uudelle semanttiselle tasolle eikä paluuta entiseen ole. Web of Data luo WWW:lle uuden, W3C:n standardeihin perustuvan rakenteellisen sisällöllisen kerroksen, joka on hyödyllinen ja tullut jäädäkseen.

5 KIRJALLISUUTTA

1. G. Antoniou, F. van Harmelen: A Semantic Web Primer (Third Edition). The MIT Press, 2012.
2. Ricardo Baeza-Yates, Berthier Ribeiro-Neto: Modern Information Retrieval: The Concepts and Technology behind Search (2nd Edition). ACM Press Books, 2011.
3. Jean Aitchison, David Bawden, Alan Gilchrist: Thesaurus Construction and Use: A Practical Manual. Europa Publications, London, 2000.
4. Christiane Fellbaum (ed.): WordNet: An Electronic Lexical Database. The MIT Press, 1998.
5. Aldo Gangemi, Nicola Guarino, Claudio Masolo, Alessandro Otramari, Schneider Luc: Sweetening Ontologies with DOLCE. Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web Springer-Verlag, 2002, pp. 166-181.
6. Tom Heath, Christian Bizer: Linked Data: Evolving the Web into a Global Data Space. Claypool & Morgan, CA, 2011.
7. Matias Frosterus, Jouni Tuominen, Sini Pessala, Katri Seppälä, Eero Hyvönen: Linked Open Ontology Cloud KOKO--Managing a System of Cross-domain Lightweight Ontologies. The Semantic Web: ESWC 2013 Satellite Events, pp. 296-297, Springer-Verlag, Berlin Heidelberg, Montpellier, France, May 26-30, 2013.
8. Eero Hyvönen, Eetu Mäkelä, Mirva Salminen, Arttu Valo, Kim Viljanen, Samppa Saarela, Miikka Junnila and Suvi Kettula: MuseumFinland - Finnish Museums on the Semantic Web. Journal of Web Semantics, vol. 3, no. 2, pp. 25, 2005.
<http://www.seco.tkk.fi/publications/2005/hyvonen-makela-et-al-museumfinland-finnish-2005.pdf>
9. Eero Hyvönen: Miksi asiansastot eivät riitä vaan tarvitaan ontologioita? Tietolinja, 2005.
<http://urn.fi/URN:NBN:fi:ELE-1070613>
10. Eero Hyvönen: FinnONTO-malli kansallisen semanttisen webin sisältöinfrastruktuurin perustaksi - visio ja sen toteutus. Raportti, Teknillinen korkeakoulu, Espoo, 12.9.2008.
<http://www.seco.tkk.fi/publications/2008/hyvonen-ONKI-yleisesitys-2008.pdf>
11. Eero Hyvönen, Kim Viljanen, Jouni Tuominen, Katri Seppälä: Building a National Semantic Web Ontology and Ontology Service Infrastructure--The FinnONTO Approach. Proceedings of the European Semantic Web Conference ESWC 2008, Springer-Verlag, 2008.

-
-
12. <http://www.seco.tkk.fi/publications/2008/hyvonen-et-al-building-2008.pdf>
Eero Hyvönen: Kulttuurisampo - suomalainen kulttuuri semanttisessa webissä. Muistiorganisaati-oiden ja kansalaisten yhteisöllinen kansallinen julkaisujärjestelmä. Raportti, Teknillinen korkeakoulu, Espoo, 25.9.2008b.
<http://www.seco.tkk.fi/publications/2008/hyvonen-Kulttuurisampo-2008.pdf>
 13. Eero Hyvönen: Preventing Interoperability Problems Instead of Solving Them. Semantic Web Journal, IOS Press, no. 2-3, December, 2010.
<http://www.semantic-web-journal.net/content/new-submission-preventing-interoperability-problems-instead-solving-them>
 14. Eero Hyvönen, Jouni Tuominen, Tomi Kauppinen, Jari Väätäinen: Representing and Utilizing Changing Historical Places as an Ontology Time Series. Geospatial Semantics and Semantic Web: Foundations, Algorithms, and Applications (Naveen Ashish and Amit Sheth (eds.)), Springer-Verlag, 2011.
<http://www.seco.tkk.fi/publications/2011/hyvonen-et-al-sapo-2011.pdf>
 15. Eero Hyvönen, Thea Lindquist, Juha Törnroos, Eetu Mäkelä: History on the Semantic Web as Linked Data - An Event Gazetteer and Timeline for World War I. Proceedings of CIDOC 2012 - Enriching Cultural Heritage, Helsinki, Finland, CIDOC,
<http://www.cidoc2012.fi/en/cidoc2012/programme>, June, 2012.
 16. Tomi Kauppinen, Eero Hyvönen: Modeling and Reasoning about Changes in Ontology Time Series. Ontologies: A Handbook of Principles, Concepts and Applications in Information Systems (Rajiv Kishore, Ram Ramesh and Raj Sharman (eds.)), Integrated Series in Information Systems, pp. 319-338, Springer-Verlag, New York (NY), 2007.
 17. Suvi Kettula: Semanttisen webin ontologisen tekstiilikäsittelön kehittäminen ja liittäminen mu-seon luettelointijärjestelmiin. Väitöskirja, Helsingin yliopisto, käsityötieteen laitos, 2009.
<http://www.doria.fi/handle/10024/50313>
 18. Jussi Kurki: Toimijaontologiat ja niiden käyttö semanttisessa webissä. Pro Gradu, Helsingin yliopisto, 2011.
<http://www.seco.tkk.fi/publications/2011/kurki-gradu-2011.pdf>
 19. Jussi Kurki, Eero Hyvönen: Authority Control of People and Organizations on the Semantic Web. Proceedings of the International Conferences on Digital Libraries and the Semantic Web 2009 (ICSD2009), September, 2009. Trento, Italy.
<http://www.seco.tkk.fi/publications/2009/kurki-hyvonen-onki-people-2009.pdf>
 20. Robin Lindroos: Paikkatiedon ontologiapalvelu. Diplomityö, Teknillinen korkeakoulu, 2008.
<http://www.seco.tkk.fi/publications/2008/onkipaikka-diplomityo-2008.pdf>
 21. Riitta-Liisa Leskinen (toim.): Museoalan asiasanasto (MASA). Museovirasto, 1998.

22. Eetu Mäkelä, Eero Hyvönen, Tuukka Ruotsalo: How to deal with massively heterogeneous cultural heritage data – lessons learned in CultureSampo. Semantic Web Journal, IOS Press, 2012.
<http://www.semantic-web-journal.net/content/how-deal-massively-heterogeneous-cultural-heritage-data-%E2%80%93-lessons-learned-culturesampo>
23. Eetu Mäkelä, Kaisa Hypén, Eero Hyvönen: BookSampo--Lessons Learned in Creating a Semantic Portal for Fiction Literature. Proceedings of ISWC-2011, Bonn, Germany, Springer-Verlag, 2011.
<http://www.seco.tkk.fi/publications/2012/tuominen-et-al-ontologiapalvelut-2012.pdf>
24. Anita Nuopponen: Begreppssystem för terminologisk analys. Vaasan yliopisto. 1994.
25. Sini Pessala, Katri Seppälä, Osma Suominen, Matias Frosterus, Jouni Tuominen, Eero Hyvönen: MUTU: An Analysis Tool for Maintaining a System of Hierarchically Linked Ontologies. Proceedings of the ISWC 2011 Workshop Ontologies Come of Age in the Semantic Web (OCAS). CEUR Works-shop Proceedings, Vol-809, 2011.
26. Sanastokeskus TSK:n TEPA-termipankki. <http://www.tsk.fi/tepa>.
27. Antoine Isaac, Ed Summers (Eds): SKOS Simple Knowledge Organization System Primer: The skos:altLabel property makes it possible to assign an alternative lexical label to a concept. W3C Working Group Note 18, August, 2009. <http://www.w3.org/TR/skos-primer/>.
28. Steffen Staab, Rudi Studer (Eds): Handbook on Ontologies. Springer-Verlag, 2009
29. Osma Suominen, Alex Johansson, Henri Ylikotila, Jouni Tuominen, Eero Hyvönen: Vocabulary Services Based on SPARQL Endpoints: ONKI Light on SPARQL. Poster proceedings of the 18th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2012), Galway, Ireland, October, 2012.
30. Jouni Tuominen, Matias Frosterus, Kim Viljanen, Eero Hyvönen: ONKI SKOS Server for Publishing and Utilizing SKOS Vocabularies and Ontologies as Services. Proceedings of the 6th European Semantic Web Conference (ESWC 2009), Heraklion, Greece, May 31 - June 4, 2009. Springer-Verlag.
<http://www.seco.tkk.fi/publications/2009/tuominen-et-al-onki-skos-2009.pdf>
31. Jouni Tuominen, Nina Laurenne, Mikko Koho, Eero Hyvönen. The Birds of the World Ontology AVIO. The Semantic Web: ESWC 2013 Satellite Events, Lecture Notes in Computer Science Volume 7955, 2013.
32. Jouni Tuominen, Kim Viljanen, Eero Hyvönen: Ontologiapalvelut semanttisessa webissä, Tietojenkäsittelytiede, Tietojenkäsittelytieteen Seura ry, 2012.
<http://www.seco.tkk.fi/publications/2012/tuominen-et-al-ontologiapalvelut-2012.pdf>

-
-
33. Kim Viljanen, Jouni Tuominen, Eero Hyvönen: Ontology Libraries for Production Use: The Finnish Ontology Library Service ONKI. Proceedings of the 6th European Semantic Web Conference (ESWC 2009), Heraklion, Greece, May 31 - June 4, 2009. Springer-Verlag.
<http://www.seco.tkk.fi/publications/2009/viljanen-et-al-www.yso.fi-2009.pdf>
 34. Eugen Wüster: Einführung in die allgemeine Terminologielehre und terminologische Lexikographie. Fachsprachliches Zentrum, Handelshochschule Kopenhagen, 1985.

6 LIITE: Keskeisiä käsitteitä

Tässä liitteessä esitetään joitain ontologiatyön keskeisten käsitteiden määritelmiä käsitejärjestelmän systematiikkaan perustuvassa järjestyksessä, jotta järjestys tukisi käsitteiden vertailua. Määritelmien lähdeaineistot ovat käytettävissä Sanastokeskus TSK:n TEPA-termipankin kautta osoitteessa <http://www.tsk.fi/tepa>.

asiasanasto

ihmisen tulkittavaksi tarkoitettu asiasanojen ja ohjaustermien luettelo, jota käytetään aineistojen sisällönkuvailuun

[lähde: JHS 183 Julkisen hallinnon palvelujen tietomalli ja ryhmittely verkkopalveluissa]

Asiasanastoa, jossa asiasanojen yhteydessä ilmaistaan niiden suhteet muihin asiasanoihin ja ohjaustermeihin, kutsutaan tesaurukseksi. Osa FinnONTO-mallin mukaan ontologisoiduista asiasanastoista on tesauruksia (esimerkiksi Yleinen suomalainen asiasanasto YSA).

ontologia

koneen ja ihmisen tulkittavissa oleva, yhteisesti sovittu täsmällinen kuvaus sovellusalueen käsitteistä ja näiden välisistä suhteista

[lähde: JHS 183 Julkisen hallinnon palvelujen tietomalli ja ryhmittely verkkopalveluissa]

Ontologian sovellusalueeseen voi sisältyä useita erikoisaloja, mikäli ontologiaan kuuluu riittävän kattavasti kyseisten alojen käsitteitä.

linkittäminen <ontologiat>

minkä tahansa suhteen luominen eri ontologioiden käsitteiden välille

siltaaminen, peilaaminen, mäppääminen <ontologiat>

ekvivalenssi- tai hierarkiasuhteen selvittäminen ja merkitseminen eri ontologioissa olevien käsitteiden välille

asiasana

sisällönkuvailussa käytettävä sana tai sanaliitto, joka kuvaa tiettyä käsitettä

[lähde: muokattu Tietohuollon sanaston, TSK 20, pohjalta]

termi

erikoisalalla käytettävä yleiskäsitteen nimitys

[lähde: Terminologian sanasto, TSK 36]

käsite

tiedon yksikkö, joka muodostuu käsitepiirteiden ainutkertaisesta yhdistelmästä

[lähde: Terminologian sanasto, TSK 36]

yleiskäsite

käsite, joka vastaa kahta tai useampaa tarkoitetta, joilla on yhteisiä ominaisuuksia

Esimerkiksi: 'moottori', 'meri' ja 'lahjakkuus'.

[lähde: Terminologian sanasto, TSK 36]

yksilökäsite

käsite, joka vastaa yhtä ainutkertaista tarkoitetta

Yksilökäsitteisiin viitataan erisnimillä kuten 'Jupiter' ja 'Itämeri'.

[lähde: Terminologian sanasto, TSK 36]

Ontologioissa instanssit eli käsitteen ilmentymät vastaavat lähinnä yksilökäsitteitä.

vieruskäsite

käsite, jonka lähin hierarkkinen yläkäsite on sama kuin siihen verrattavalla toisella

käsitteellä ja joka sijoittuu tämän kanssa samaan käsitejärjestelmän ulottuvuuteen tai jonka koostumussuhteinen yläkäsite on sama kuin siihen verrattavalla käsitteellä

[lähde: Terminologian sanasto, TSK 36]

yläkäsite

käsitejärjestelmän käsite, jolla on vähintään yksi alakäsite

hierarkkinen yläkäsite

käsite, joka on hierarkkisessa suhteessa toiseen käsitteeseen ja jonka sisältö on osa tämän toisen käsitteen sisältöä

[lähde: Terminologian sanasto, TSK 36]

koostumussuhteinen yläkäsite; koostumussuhteen yläkäsite; kokonaisuuskäsite

koostumussuhteinen käsite, joka vastaa kokonaisuutta

[lähde: Terminologian sanasto, TSK 36]

hierarkia

hierarkkisiin käsitesuhteisiin perustuva käsitejärjestelmä

[lähde: JHS 183 Julkisen hallinnon palvelujen tietomalli ja ryhmittely verkkopalveluissa]

hierarkkinen suhde; hierarkkinen käsitesuhde; hierarkiasuhde
käsitesuhde, jossa yhden käsitteen sisältö muodostuu toisen käsitteen sisällöstä sekä lisäksi yhdestä tai useammasta erottavasta piirteestä

[lähde: Terminologian sanasto, TSK 36]

koostumussuhde; osa-kokonaisuussuhde
käsitesuhde, jossa yksi käsite vastaa tiettyä kokonaisuutta ja toinen kokonaisuuden osaa

Esimerkiksi 'polkupyörän' ja 'polkimen' välisessä koostumussuhteessa 'polkupyörä' vastaa kokonaisuutta ja 'poljin' kokonaisuuden osaa.

[lähde: Terminologian sanasto, TSK 36]

assosiatiivinen suhde; funktiosuhde
käsitesuhde, joka perustuu muuhun kuin hierarkkiseen tai koostumussuhteeseen liittyvään aiheytteen käsitteen tarkoitteiden välillä

[lähde: Terminologian sanasto, TSK 36]

käsitepiirre
tarkoitteen tai tarkoitejoukon ominaisuuden abstraktio

[lähde: Terminologian sanasto, TSK 36]

olennainen käsitepiirre
käsitteen selvittämisen kannalta välttämätön käsitepiirre

[lähde: Terminologian sanasto, TSK 36]

tietotyyppi
tiedon laji

Esimerkiksi ontologiassa voidaan eritellä käsitteisiin liittyviä tietotyyppisiä kuten erilaisia nimityksiä ja käsitteiden välisiä suhteita.