

Aalto-yliopisto
Perustieteiden korkeakoulu
Tietotekniikan tutkinto-ohjelma

Sini Pessala

Linkitettyjen kevytontologioiden muutosten kuvaaminen

Diplomityö
Espoo, 9. huhtikuuta 2015

Valvoja: Professori Eero Hyvönen
Ohjaaja: Diplomi-insinööri Matias Frosterus
Filosofian maisteri Jouni Tuominen

Tekijä:	Sini Pessala		
Työn nimi:	Linkitettyjen kevytontologioiden muutosten kuvaaminen		
Päiväys:	9. huhtikuuta 2015	Sivumäärä:	vi + 76
Professori:	Mediatekniikka	Koodi:	T-75
Valvoja:	Professori Eero Hyvönen		
Ohjaaja:	Diplomi-insinööri Matias Frosterus Filosofian maisteri Jouni Tuominen		
<p>Ontologioiden avulla pyritään mallintamaan tiettyä alaa määrittelemällä kyseiseen alaan kuuluvia käsitteitä ja näiden välisiä yhteyksiä. Näitä käsitteitä voidaan käyttää alaa käsittelevän aineiston annotointiin. Lisäksi ontologioita voidaan linkittää toisiin ontologioihin, jolloin päästään hyödyntämään muualla tehtyä työtä. Ontologioita joudutaan luonnollisesti päivittämään, kun uusia käsitteitä tai muita muutoksia tulee alalle. Tällöin myös kyseiseen ontologiaan linkittyneitä ontologioita voi joutua muuttamaan, jotta linkittyneet ontologiat pysyvät keskenään konsistentteina.</p> <p>Tässä työssä luotiin MUTU-työkalu auttamaan ontologiakehittäjää edellä kuvatussa muutostilanteessa. Työkalu välittää muutokset ontologiakehittäjälle, lajittelee muutokset muutoksen tyyppin mukaan sekä esittää nämä muutokset linkitetyn ontologian kehittäjälle. Tieteellisenä tuloksena työkalu sisältää algoritmin, joka pyrkii ennustamaan, milloin toisessa ontologiassa tapahtunut muutos aiheuttaa muutoksen kehittäjän ontologiaan. Työkalun tavoitteena on helpottaa linkitetyn ontologian päivitystä varsinkin suurilla muutosmäärillä.</p> <p>Työkalun luomia muutoslistoja sekä muutosten lajittelun toimivuutta testattiin ontologiakehittäjillä todellisessa käyttötilanteessa. Työn tuloksena oli, että lajitteluheuristiikka toimi aineistossa hyvin käsitteiden ominaisuuksien muutoksissa ja auttaisi täten priorisoimaan läpikäytäviä muutoksia ontologiaa päivitettäessä.</p>			
Asiasanat:	linkitetyt ontologiat, erikoisontologiat, muutosten visualisointi, muutosten luokittelu, ontologioiden kehitys, ontologioiden ylläpito, tesaurokset, semanttinen web		
Kieli:	suomi		

Author:	Sini Pessala	
Title:	Visualising Changes in a System of Linked Lightweight Ontologies	
Date:	April 9, 2015	Pages: vi + 76
Professorship:	Media Technology	Code: T-75
Supervisor:	Professor Eero Hyvönen	
Instructor:	Matias Frosterus M.Sc. (Tech.) Jouni Tuominen M.Sc.	
<p>Ontologies aim to model a field by listing concepts and their relationships in the field. These concepts are then used in annotating data regarding the field. In addition, ontologies can be linked to other ontologies, thus utilising the work and knowledge of other ontology developers. Naturally, ontologies need to evolve due to the changes and advances of the field. In this case, also other ontologies linked to the ontology might require updates to ensure the consistency of the ontologies.</p> <p>In this work, I created a tool for assisting the ontology developer in the above described update. The tool was named MUTU and it conveys the changes to the ontology developer, classifies the changes by their type and visualises the changes for the developer. As a scientific contribution, this work introduces an algorithm for classifying changes by their priority. The algorithm aims to predict which changes in the related ontology will cause changes in the ontology of the developer. The aim of the tool is to aid in updates with a large amount of changes.</p> <p>The created change lists and the performance of the change classification were tested in a real-life update task. The outcome of the evaluation was that the priority classification performance was good with the changes of property values, thus being an aid in prioritising changes in updating ontologies.</p>		
Keywords:	linked ontologies, domain ontologies, change visualisation, change classification, ontology maintenance, ontology update, thesauri, Semantic Web	
Language:	Finnish	

Kiitokset

Tämä diplomityö on tehty suurimmalta osin Aalto-yliopiston ja Helsingin yliopiston yhteisessä Semanttisen laskennan tutkimusryhmässä osana FinnONTO2-hanketta.

Haluan kiittää professori Eero Hyvöstä työni valvomisesta. Kiitokset kuuluvat myös työni ohjaajille Matias Frosterukselle ja Jouni Tuomiselle aina käytettävänä olemisesta työni ohjaamiseen ja tukemiseen mitä mielikuvituksellisimmin tavoin.

Kiitokset ansaitsee myös Osma Suominen, joka ohjelmoi ontorip-työkalun työtä varten, sekä Katri Seppälä, jonka kanssa hahmottelimme MUTUun tarvittavia ominaisuuksia.

Suurkiitokset kuuluvat MUTU-työkalun testikäyttäjille. Samaan soppaan pääsee jälleen käsiksi, kun MUTUsta ilmestyy seuraava versio!

Lisäksi tahdon kiittää muita Semanttisen laskennan tutkimusryhmäläisiä hetkistä, jotka liittyivät diplomityöhöni enemmän tai usein vähemmän. Kiitokset lähtevät myös Finto-palvelun tiimille, joidenka kanssa työskennellessäni työni aihe on syventynyt ja pölyttynyt kirjastomaailman ajatuksilla. Lopuksi kiitän vielä Matti Nelimarkkaa diplomityöni loppurutistuksessa tukemisesta ja motivoinnista.

Nummela, 9. huhtikuuta 2015

Sini Pessala

Sisältö

1	Johdanto	1
2	Linkitetyt ontologiat	4
2.1	Ontologioiden eroja	5
2.2	Ontologioiden tekninen toteutus	9
2.3	Ontologioiden linkittäminen ja yhdistely	10
3	Linkitettyjen ontologioiden päivittäminen	15
3.1	Ontologioiden muutokset	15
3.2	Ontologioiden kehittäminen	18
4	Menetelmä	25
4.1	Aikaisemmassa tutkimuksessa sivuun jääneitä aiheita	25
4.2	Menetelmä ontologioiden suurten muutosmäärien esittämiseen . . .	27
5	MUTU-muutostentunnistustyökalu	30
5.1	MUTU-työkalun käyttäminen	30
5.2	Muutostyypit ja muutosten tärkeys	34
5.3	Muutosten visualisointitavat	36
5.4	MUTU päivityksen osana	41
6	MUTUn suoriutuminen	44
6.1	Testikäyttäjät ja erikoisontologiat	44
6.2	Tulokset	47

7	Yhteenveto ja lopputulokset	56
7.1	Tutkimuskysymyksiin vastaaminen	56
7.2	MUTU-työkalun kehityskohteita	57
7.3	Jatkotutkimusaiheita	58
	LIITTEET	66
A	Loppukysely	67
B	Loppuhaastattelun runko	72
C	Kiinnostavat muutokset -taulukko	74
D	Muutosmäärät-taulukko	75
E	Erikoisontologiakehittäjien tekemät muutokset -taulukko	76

Luku 1

Johdanto

Luovimme päivittäin suurissa tietomassoissa etsien ja selaillen aineistoja. Modernit hakukoneet löytävät helposti hakua vastaavia tekstimuotoisia asiadokumentteja, mutta ei-tekstipohjaisten aineistojen, kuten multimedian, äänitteiden tai museoesineiden, hakujärjestelmät eivät ole yhtä kehittyneitä [53].

Näiden hankalasti haettavien aineistojen löytämistä voi parantaa lisäämällä metadataa hakukoneiden hyödynnettävässä muodossa. Esimerkiksi kirjaan voidaan liittää sen käsittelemät teemat tai museoesineestä voidaan ilmoittaa käytetty materiaali. Jotta metadata olisi yhtenäistä, käytetään aineiston annotointiin alan käsitteitä kuvaavia sanaston kaltaisia malleja, kuten asiasanastoja [17], tesauroksia [17] tai kevytontologioita (engl. *lightweight ontology*) [15]. Käsitte pohjaisia kevytontologioita on taas alettu ottaa käyttöön Suomen muistiorganisaatioissa vasta viimeisen kymmenen vuoden aikana.

Jotta ontologiat tukisivat aineiston hakemista mahdollisimman hyvin, tulee niiden vastata aineiston annotoijan tarpeita sekä olla ajantasaisia. Tämä tarkoittaa, että ontologioita ei luoda kerralla valmiiksi, vaan niitä täytyy kehittää ja päivittää. Ontologian kehittäjänä tulisi olla kyseisen aihealueen asiantuntija, jotta ontologian rakenne ja käsitteiden väliset suhteet ovat muodostettu erikoisalan tiedon perusteella.

Annotointiin tarkoitettujen ontologioiden koko vaihtelee tuhansista kymmeneen tuhansiin käsitteisiin riippuen ontologian aihepiiristä. Laajimpia näistä ontologioista kehitetään useamman henkilön voimin, joko samassa tai erillisissä ontologioissa. Erillisissä ontologioissa kehitettäessä on mahdollista olla yksi laaja, yleisiä käsitteitä sisältävä yleisontologia, jota laajentaa siihen linkitetty erikoisontologioiden joukko. Tällaisia yleisontologioita ovat esimerkiksi SUMO [46] ja KOKO [16].

Ontologiaa päivitettäessä voivat muiden alojen tai yleiskäsitteiden muutokset vai-

kuttaa erikoisalan käsitteistöön. Läheisesti yhteistyötä tekevät ontologiakehittäjät voivat tiedottaa muutoksista toisilleen suullisesti, kirjallisesti tai selaamalla ontologian muutoslokia. Mikäli kehitystä tehdään fyysisesti eri paikoissa tai organisaatioissa, voi tiedonkulku olla hankalampaa, varsinkin mikäli erikoisalan ontologian päivittäminen ei ole säännöllistä. Organisaation resurssien mukaan vaihteleva kehitys voi johtaa tilanteeseen, jossa katkonaisen kehityksen takia myös muuttuneiden käsitteiden määrä voi olla suuri, eikä kunnollisia välineitä muutoksista tiedottamiseen ole.

Tämä diplomityö keskittyy tarkastelemaan, miten ontologioiden muutoksia voi välittää ontologiakehittäjille, joiden ontologia on linkittynyt muuttuneeseen ontologiaan. Työssä keskitytään erityisesti yleisontologian muutosten välittämiseen erikoisontologian kehittäjille, mutta menetelmä on sovellettavissa myös suuriin yksittäisiin ontologioihin, joiden kehittäjillä on vastuullaan tietyn alan käsitteet.

Tässä diplomityössä keskitytään seuraavaan tutkimuskysymykseen ja sen osaongelmiin:

Miten kevytontologioita kehittävien erikoisontologiakehittäjien päivitystyötä voidaan tukea, kun yleisontologian muutosmäärät ovat suuria?

- Voidaanko yleisontologian muutoksista erotella automaattisesti sellaiset, jotka aiheuttavat muutoksia erikoisontologiassa?
- Onko mahdollista käydä läpi nämä muutoskategoriat jossakin tietyssä järjestyksessä?

Työssä keskitytään vain RDFS- ja SKOS-mallinnuskielillä esitettävissä oleviin kevytontologioihin. Työn ulkopuolelle rajautuvat raskasontologiat, ontologioiden validointi [10, 69], muutosten levittäminen (engl. *change propagation*) [6] sekä instanssidatan muutoksissa validina säilyttäminen [48].

Tutkimuskysymyksiin vastattiin luomalla prototyypityökalu muutosten esittämiseen. Tämän prototyypin käyttöä arvioitiin todellisessa ontologiapäivitystilanteessa.

Diplomityön sisältö on seuraavanlainen: Luvussa 2 tutustutaan tarkemmin ontologioihin sekä näiden välisiin linkkeihin. Muutoksiin ja kehitykseen linkitetyissä ontologioissa syvennyttään luvussa 3. Luku 4 listaa aikaisemmassa tutkimuksessa sivuun jääneitä aiheita sekä esittelee menetelmän, jota on käytetty luvussa 5 esiteltävän muutostyökalun pohjana. Luvussa 6 esitellään arviointi työkalun käytöstä päivitystilanteessa. Luvussa 7 tarkastellaan vastauksia tutkimuskysymyksiin sekä hahmotellaan tulevaisuuden tutkimus- ja kehityssuuntia.

Tämä diplomityö perustuu osittain MUTU-työkalusta aikaisemmin julkaistuun artikkeliin [51].

Luku 2

Linkitetyt ontologiat

Termi *ontologia* on alun perin ollut käytössä filosofiassa merkitsemässä oppia todellisuudesta. Ontologiatutkimus voimistui erityisesti 80- ja 90-luvuilla, jolloin tekoäly-, ohjelmistokehitys- ja tietokantatutkimus huomasivat toisistaan tietämättään tarvitsevansa tapoja korkeamman tason konseptuaalisia malleja. Nämä alat omaksuivat ontologia-termin merkitsemään esityksiä käsitteistä ja niiden välisistä suhteista. Gruber [21] määritteli ontologian olevan ”tietyn aihealueen käsitteistön eksplisiittinen malli” (suom. [63]). Tämä määritelmä on hyvin laaja, eikä juurikaan ota kantaa toteutustapaan, jotta se olisi yhteensopiva näiden kolmen alan näkemysten kanssa. [57]

Muita määritelmiä ja selityksiä ovat tarjonneet muun muassa Struder [61] sekä Guarino ja Giaretta [18]. Uschold ja Jasper [27] määrittelivät ontologian yleisemmin näin:

An ontology may take a variety of forms, but it will necessarily include a vocabulary of terms and some specification of their meaning. This includes definitions and an indication of how concepts are inter-related which collectively impose a structure on the domain and constrain the possible interpretations of terms.

Heidän määritelmänsä pyrkii kokoamaan yhteen useilla tieteenaloilla käytetyt ontologiat tai ontologian kaltaiset tiedonesitystavat. Yksi keskeisistä ontologioiden tarkoituksista on toimia yhteisenä ”sanastona” useamman toimijan kesken. Tämä tarkoittaa, että hyödynnettävyyden takia ontologioiden tulisi olla yhteisesti sovit-tuja, koneluettavia formaaleja esityksiä. Tavoitteena on ontologioiden käyttäminen kommunikaation tehostamisessa sekä koneiden että ihmisten välillä toteuttaen täten Semanttisen Webin periaatteita [5].

Sovelluskohde vaikuttaa vahvasti siihen millaisia tarpeita ontologiaa kohtaan on. Käyttötapa määrää paljonko käsitteitä tarvitaan ja kuinka paljon rakennetta käsitteisiin ja niiden välille tarvitaan. Gómez-Pérez ja kumppanit listaavat seuraavia käyttökohteita ontologioille [20, s. 8]

- luonnollisen kielen käsittely
- tietämyksen hallinta
- verkkokauppa
- älykäs tietojen yhdistely
- semanttinen web

Tyypillisesti semanttisessa webissä toimivat agentit tarvitsevat päättelyyn paljon aksiomia ja rajoituksia sisältäviä ontologioita [24] kun taas tietämyksenhallinnassa on käytetty pääosin rakenteeltaan kevyempiä tesauruksia [25]. Seuraava aliluku esittelee erilaisia ontologioita sekä näiden luokittelutapoja.

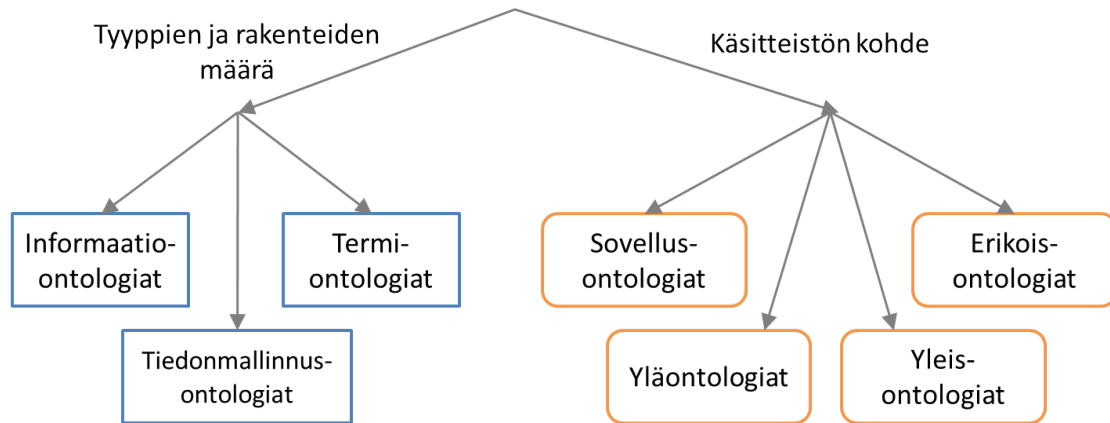
2.1 Ontologioiden eroja

Ontologioita tehdään erilaisilla aloilla, kuten biologiassa, kirjastoissa ja tekoälytutkimuksessa, joten ontologioiden näkökulmat että käyttökohteet voivat erota toisistaan huomattavasti. Aikaisemmassa tutkimuksessa ontologioita onkin luokiteltu monin eri tavoin [37, 45, 65]. Seuraavaksi esittelen kaksi erilaista tapaa jaotella ontologioita: Van Heijstin ja kumppanien [65] kahden näkökulman jaottelu käsitteistön rakenteiden sekä kohteen mukaan. Tämän jälkeen esittelen Lassilan ja McGuinnessin [37] ontologian sisäisen rakenteen rikkauteen perustuvan jatkumon. On hyvä huomata, että Van Heijstin sekä Lassilan ja McGuinnessin tavat jaotella ontologioita eivät ole toisiaan poissulkevia, vaan ne täydentävät toisiaan. Van Heijstin esitystapa keskittyy enemmän ontologioihin käyttötarkoituksen perusteella ja Lassila ja McGuinness tarkastelevat ontologioiden rakennetta.

Van Heijstin ja kumppanien ontologiakategoriat

Jaottelu käsitteistön tyyppin ja rakenteen mukaan kuvan 2.1 mukaisesti:

- **Termiontologiat** (engl. *Terminological Ontologies*) ovat sanaston tyyppisiä malleja, jotka määrittävät tietyn alan termit.

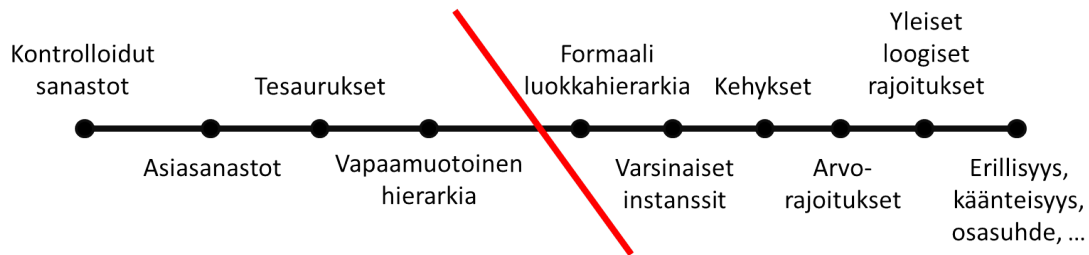


Kuva 2.1: Van Heijstin ja kumppaneiden ontologiakategoriat [65] mukailtuna Gómez-Pérezin ja kumppanien kuvasta [20]

- **Informaatio-ontologiat** (engl. *Information Ontologies*) määrittävät skeemamaisesti mallin tiedon tallennukseen tietokantaan.
- **Tiedonmallinnusontologiat** (engl. *Knowledge modelling Ontologies*) pyrkivät mallintamaan aihealuettaan. Tämän johdosta tiedonmallinnusontologiat ovat yleensä rakenteeltaan rikkaampia kuin informaatio-ontologiat.

Jaottelu käsitteistön kohteen mukaan:

- **Sovellusontologiat** (engl. *Application Ontologies*) ovat tiettyä sovellusta varten luotuja ontologioita. Nämä ovat yleensä kokoelmia erikois- ja yleisontologioiden käsitteistä sekä sovelluskohtaisista laajennuksista. Sovellusontologiat eivät ole pääsääntöisesti uudelleenkäytettävissä muissa sovelluksissa.
- **Erikoisontologiat** (engl. *Domain Ontologies*) sisältävät tietyn alan käsitteistön.
- **Yleisontologiat** (engl. *Generic Ontologies*) sisältävät yleisiä käsitteitä, joita halutaan hyödyntää useammilla aloilla [65]. Yleisontologiat voivat sisältää jopa kymmeniä tuhansia käsitteitä ja erikoisontologiat voivat osaltaan laajentaa yleisontologioiden käsitteistöä.



Kuva 2.2: Lassilan ja McGuinnessin ontologiajatkumo [37] mukailtuna Gómez-Pérezin ja kumppanien kuvasta [20]

- **Yläontologiat** (engl. *Upper Ontology*), joita aiemmin kutsuttiin englanniksi termillä *Representation Ontologies*, kuvaavat neutraalilla tavalla yleisimmän tason käsitteistöä. Yleis- ja erikoisontologiat pohjautuvat yläontologioiden luomille käsitejaoille [65]. Yläontologiat koostuvat tyypillisesti muutamista kymmenistä käsitteistä, joiden tavoitteena on kategorisoida muiden ontologioiden käsitteet erillisiin kategorioihin.

Lassilan ja McGuinnessin ontologiajatkumo

Lassila ja McGuinness [37] jaottelevat ontologiat niiden sisäisten rakenteen rikkauksen mukaan ilmenevänä jatkumona (kuvassa 2.2).

Heidän jatkumonsa sisältää seuraavat ontologioiden rikkaus- ja tarkkuustasot:

- **Kontrolloidut sanastot** (engl. *Controlled vocabularies*) eli termilistat
- **Asiasanastot** (engl. *Glossary*), joissa on termien listaamisen lisäksi niiden määrittelyt luonnollisella kielellä.
- **Tesaurukset** (engl. *Thesauri*) sisältävät termien välisiä yhteyksiä, mutta eivät välttämättä sisällä eksplisiittistä hierarkiaa.
- **Vapaamuotoinen hierarkia** (engl. *Informal is-a*), jossa käsitteillä on ylä-käsite, mutta hierarkia ei noudattele formaalia luokkahierarkiaa.
- **Formaali luokkahierarkia** (engl. *Formal is-a*), jossa käsitteet ovat ylä-käsitteensä aitoja alakäsitteitä, mikä mahdollistaa ominaisuuksien periyttämisen yläluokilta.
- **Varsinaiset instanssit** (engl. *Formal instance*), jotka formaalin luokkahierarkian lisäksi sisältävät aihealueen instansseja.

- **Kehykset (ominaisuudet)** (engl. *Frames (properties)*), jonka tyyppisessä ontologiassa käsitteille voi määritellä ominaisuuksia, jotka periytyvät käsitteen alakäsitteille ja instansseille.
- **Arvorajoitukset** (engl. *Value Restrictions*), jotka voivat rajoittaa instanssien ominaisuuksien arvoja.
- **Yleiset loogiset rajoitukset** (engl. *General Logical Constraints*), jossa ontologiakehittäjä voi määritellä rajoitteita ensimmäisen kertaluvun logiikkaa käyttäen.
- **Erillisuus, käänteisyys, osasuhte** (engl. *Disjointness, Inverse, Part-Of*) ja muut ilmaisuvoimaisemmat ominaisuudet ja menetelmät.

Lassilan ja McGuinnessin jatkumo perusteella viivan vasemmalla puolella olevia taksonomian omaisia ontologioita voidaan kutsua kevytontologioiksi (engl. *lightweight ontology*)[15] ja oikealla puolella olevia aksiomia ja rajoituksia sisältäviä ontologioita raskasontologioiksi (engl. *heavyweight ontology*)[20, s. 8]. Giunchiglia [19] lisää vielä Lassilan ja McGuinnessin kevytontologian määritelmään vaatimuksen yhtenäisestä, puumuotoisesta käsittehierarkiasta.

Kun Lassilan ja McGuinnessin vuonna 2001 luomaa jatkumoa tarkastellaan nykyontologioiden ja ontologiakielien valossa, voidaan huomata että raskasontologioiden ilmaisuskaala loppuu kesken. Esimerkiksi vuonna 2004 julkaistun OWL-ontologiakielen avulla pystyy ilmaisemaan sekä erillisyyttä että käänteisyyttä. Jatkumoa tulisikin päivittää erittelemään myös nykyistä raskasontologioiden kirjoja.

Toisin kuin Lassilan ja McGuinnessin jatkumo, ISO 25964-2 -standardi [26] näkee, että ontologian ja tesauksuksen eroja ovat:

- Ontologioissa instanssit ja luokat on erotettu toisistaan selvästi, kun taas tesauksissa tämä ei ole välttämätöntä
- Tesauksukset on tarkoitettu ihmisten käytettäväksi, kun taas ontologiat on tehty koneluettaviksi. Tämän takia tesauksissa on tärkeää olla listattuna saman termin muita muotoja, kun taas ontologioissa voi usein riittää vain oletusnimike (engl. *prefLabel*), sillä keskitytään päättelyyn.
- Tesauksissa on tärkeää erotella termin merkitykset toisistaan, esim. ”aura (maatalousväline)” ja ”aura (oire)”. Ontologioissa termitason disambiguoiti ei ole niin tärkeää, sillä rakenne ei ole tarkoitettu käyttäjien nähtäväksi.
- Ontologiassa on tärkeää esittää käsitteiden väliset suhteet eksplisiittisesti, kun taas tesauksessa tarkoitus on tukea ihmistä sopivan termin etsimisessä.

Ontologioissa halutaan esimerkiksi erotella hierarkkiset ja osa-kokonaisuussuhteet toisistaan, kun tesauksissa riittää laajempi termi -suhde.

Kuten ylläolevia ontologioiden luokitteluja tarkastellessa voi huomata, että ontologioita voi jaotella useilla eri tavoilla näkökannasta riippuen. Lisäksi ontologia-käsitteen merkitys vaihtelee tutkimuksesta toiseen, kuten huomattiin verratessa Lassilan ja McGuinnesin jatkumoa ISO 25964-2 -standardiin. Tämän merkityksen vaihtelevuuden takia esitän seuraavaksi, miten tässä työssä käytetään ontologia-termiä.

Ontologia-termin määritelmä tässä työssä

Tässä työssä:

1. Ontologia on tesauksenkaltainen kevytontologia sisältäen tuhansia käsitteitä, mutta suhteessa vähän ominaisuustyyppettä. Ontologian sovelluskohde on tiedonhaun tukeminen.
2. Ontologian käsitteellä tulee aina olla määriteltynä yläkäsite ja hierarkia perustuu luokkahierarkiaan.
3. Käsitteellä on yksilöivä ja päivityksissä muuttumaton URI-tunniste ja mahdollisuus olla nimikkeitä (engl. *label*) useammalla kielellä.

On huomattava, että myös tesaurus voi täyttää yllä olevat vaatimukset, mutta ontologialla näiden täytyy aina täytyä.

2.2 Ontologioiden tekninen toteutus

Ontologioiden teknisen toteutuksen pohjana on Resource Description Framework -tietomalli (RDF) [38]. RDF perustuu subjekti-predikaatti-objekti-kolmikoihin (engl. *triple*). Subjekti on kuvailtava resurssi, predikaatti kertoo millainen ominaisuus subjektiin on liittynään ja objekti voi olla joko toinen resurssi tai literaaliarvo. Nämä kolmikot muodostavat suunnatun verkon kuvaten sen solmuja ja kaaria. Keskeinen asia käsitteeresurssien verkon luomisessa on käsitteiden yksilölliset URI-tunnisteet (engl. *Uniform Resource Identifier*) [4]. URI-tunnisteet mahdollistavat tietojen yhdistelyn, sillä yksilöllisiä tunnisteita tarvitaan käsitteeseen käsitteeseen viittaamisen, linkittämisen ja kuvauksen laajentamisen muissa tietolähteissä.

RDF itsessään sisältää vain tarvittavat elementit verkkomaisen datan esittämiseen. RDF Schema (eng. *RDFS*) [42] laajentaa RDF:ää lisäämällä siihen luokkia ja ominaisuuksia, jotka mahdollistavat käsite- ja ominaisuushierarkioiden luomisen sekä rajoitteet. Koska RDF ja RDFS antavat vain mallin verkkomuotoisen datan esittämiseen, tarvitaan lisäksi RDF-muotoiseen dataan pohjautuvia tiedonesityskieliä ja -standardeja. Esimerkkinä näistä ovat SKOS (engl. *Simple Knowledge Organization System*) [43] sekä OWL (engl. *Web Ontology Language*) [66]. SKOS on suunniteltu sanastojen, luokitusten sekä kevytontologioiden esittämiseen, minkä takia siinä on valmiiksi ominaisuuksia ja luokkia käsitteiden ja nimikkeiden välisten suhteiden ilmaisemiseen, kuten `skos:prefLabel`- tai `skos:closeMatch`-ominaisuudet. OWL puolestaan on tarkoitettu raskaammille ontologioille. Sillä voi ilmaista rajoituksia ja sääntöjä, mikä mahdollistaa logiikkaa hyödyntävän päättelyn.

RDF itsessään ei ole tiedostomuoto, vaan tietomalli verkkomuotoisen datan esittämiseen. RDF-muotoisen datan tallennukseen onkin olemassa useita serialisointimuotoja. Aikaisemmin yleisin RDF-muotoisen datan serialisointimuoto oli RFD/XML [68], jossa RDF-graafi tallennettiin yleisesti käytössä olleessa XML-formaatissa. XML on yleisformaatti, joten se ei ole optimoitu RDF-muotoisen datan esittämiseen. Tämän takia RDF/XML-muotoisen datan lukeminen ja kirjoittaminen on ihmiselle työlästä sen toisteisuuden vuoksi. Turtle-formaatti [3] sitä vastoin suunniteltiin vartavasten RDF-muotoista dataa varten. Se on helposti ihmisluettavaa ja se on myös tiedostokooltaan vastaava RDF/XML-tiedostoa huomattavasti pienempi. Lisäksi tuoreimpana serialisointimuotona on JSON-LD [36], missä RDF-data serialisoidaan nykyisin web-kehityksessä suosittuun JSON-formaattiin.

2.3 Ontologioiden linkittäminen ja yhdistely

Ontologioiden luominen on työlästä ja aikaa vievää. Usein onkin helpompaa hyödyntää tai laajentaa jo olemassa olevia ontologioita kuin luoda täysin uusi ontologia. Olemassa olevia ontologioita ja muita tietolähteitä ei tarvitse kopioida omaan ontologiaansa, vaan niihin voi viitata linkkien avulla. Tämä mahdollistaa toisilla ontologioilla annotoidun datan hyödyntämisen muissa sovelluksissa. Lisäksi linkittämisen etuna on se, että kehittäjät eivät tee päällekkäistä työtä käsitteiden päivittämiseen, vaan kehitysresurssit voidaan käyttää omien erikoiskäsitteiden ylläpitoon.

Ontologioita voidaan yhdistää tai liittää toisiinsa RDF-tietomallin avulla. RDF-tietomalli perustuu kolmikielisiin, joten ontologioiden liityntäkohta on sellainen kolmikko, jossa subjekti ja objekti ovat eri ontologioista. Tällöin subjektin ja objektin nimiavaruudet eivät ole samat [7]. Tavoitteena on saada mahdollisimman paljon

yhtymäkohtia ontologioiden välille. Mitä enemmän näitä yhteyksiä eri käsitteiden välillä on, sitä paremmin pystytään ymmärtämään miten ontologioiden käsitteet liittyvät toisiinsa. Periaatteessa ontologioita yhdistävä ominaisuus voi olla mikä tahansa ja ontologioiden välillä voi olla useita erityyppisiä linkitettäviä ominaisuuksia. Tästä yleisimpänä esimerkkinä ontologioiden välisiin linkkeihin on on `rdf:type`-ominaisuus, jonka avulla usein määritellään käsitteelle yleinen tyyppi, kuten `skos:Concept`. Tässä työssä ontologioiden linkittämisellä tarkoitetaan kuitenkin kahden ontologian välisiä yhteyksiä, jossa yhdistävä ominaisuus on hierarkkinen tai ekvivalenssiyhteys.

Myös ontologioiden modularisoinnissa alalla keskitytään ontologioiden välisiin yhteyksiin. Modularisoinnissa on kaksi päätapaa: suuresta ontologiasta pienempien moduleiden eristäminen sekä laajan ontologian muodostaminen pienemmistä ontologiamoduuleista.[55] Ontologioiden modularisointi kuitenkin on pääasiassa keskittynyt automaattisiin menetelmiin moduleiden eristämässä, yhdistelyssä sekä validoinnissa usein logiikkaa sisältävien ontologioiden kanssa. Tässä työssä keskitytään kevytontologioiden kehittäjiin, joten modularisointi alana rajautuu työn ulkopuolelle.

Linkitettyjen ontologioiden väliset yhteydet

Ontologioiden välisten yhteyksien sanallinen kuvaaminen on rikasta sekä englanniksi että suomeksi. Englanniksi voidaan puhua termeistä *ontology mapping* ja *ontology alignment*, joissa ontologiat itsessään eivät muutu linkityksiä lisätessä, sekä *ontology reconciliation*, jossa ontologioita voidaan muokata linkityksiä etsittäessä ja lisätessä [11]. Lisäksi myös alignment-termiä vastaava *ontology linking* -termiä on esiintynyt kirjallisuudessa. Suomessa termejä *peilaaminen*, *siltaaminen* sekä *linkittäminen* voidaan käyttää sekä samaa aihealuetta olevien ontologioiden että ylä-erikoisontologiaparien välisiin suhteisiin. *Ripustaminen*-termiä voi puolestaan käyttää vain kuvaillaessa erikoisontologian suhdetta yläontologiaansa ajateltaessa ontologiaa tietojenkäsittelyn puumaisena tietorakenteena.

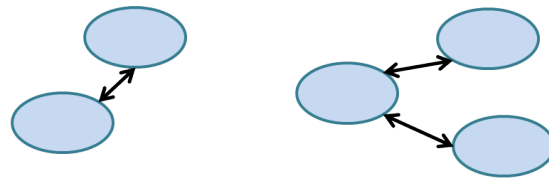
Hameed ja kumppanit [23] ovat listanneet kolme rakennemallia ontologioiden linkittämiseen (kuvassa 2.3).

Ontologiaparin väliset yhteydet

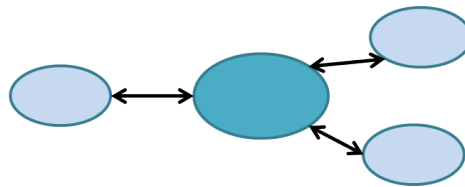
Linkattavien ontologioiden välille on luotu suora linkitys ja ne ovat ontologiakehittäjän hallittavissa. Menetelmä on joustava, mutta ylläpidettävien linkitysten määrä kasvaa ja voi viedä ylimääräisiä resursseja.

Ontologioiden yhteydet yhteiseen ontologiaan

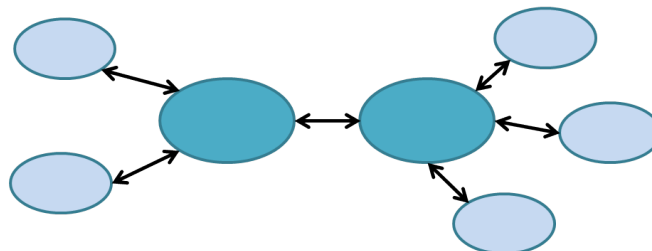
Ontologiat on linkitetty yhteen yhteiseen ontologiaan, minkä ansiosta eri on-



a) Yksittäisten ontologioiden välisiä linkityksiä



b) Linkityksiä yhteen yhteiseen ontologiaan



c) Useita linkitettyjen ontologioiden ryhmiä ja näiden välisiä linkityksiä

Kuva 2.3: Hameedin ja kumppanien [23] rakennemallit ontologioiden linkittämiseen

tologioissa olevien käsitteiden välinen yhteys on tiedossa ilman erillisen linkityksen tekemistä. Varjopuolena menetelmässä on yhteisen ontologian ylläpito ja se, että linkit muihin ontologioihin eivät ole kehittäjän hallinnassa.

Yhteisten ontologioiden yhteydet toisiinsa

Yhteiset ontologiat voi myös linkata toisiinsa, jolloin yhteydellisten ontologioiden määrä kasvaa entisestään yhdellä ontologiaparin välisellä linkityksellä.

Vastaava jako on myös esillä ISO 25964-2 -standardissa [26], jossa lisäksi mainitaan, että linkkauksilla on suunta. Tämä ei niinkään tule esiin tehdessä linkityksiä yhdestä käsitteestä toiseen, vaan tilanteissa, joissa yhtäläisyyden esittämiseen tarvitaan useampia käsitteitä sekä logiikkaa. Tässä työssä kuitenkin pitäydytään lähtökohdassa, että ontologioiden väliset linkit ovat kahden yksittäisen käsitteen välillä.

On hyvä huomata, että Hameedin ja kumppanien tai ISO 25964-2 -standardin rakennemallit eivät ota kantaa siihen, mikä ontologioiden hierarkioiden suhde toisiinsa on. Sunagawa ja kumppanit [62] esittelevät ontologioiden mahdolliset suhteet toisiinsa:

Hierarkkinen suhde

Ontologioilla on hierarkkinen suhde, jos on olemassa kolmikko, jossa predikaatti on hierarkkinen suhde ja kolmikron subjekti ja objekti ovat eri ontologioista. Tällöin ontologioita kutsutaan *yläontologiaksi* ja *alaontologiaksi*.

Viittaussuhde

Ontologioilla on viittaussuhde, jos on olemassa kolmikko, jossa subjekti ja objekti ovat eri ontologioista. Tällöin puhutaan subjektin ontologiaa kutsutaan *lähdeontologiaksi* ja objektin *kohdeontologiasta*.

On hyvä huomata, että hierarkkinen suhde on erikoistapaus geneerisemmästä viittaussuhteesta. Teoriassa ontologioiden väliset linkit voi toteuttaa käyttäen mitä tahansa ominaisuutta, mutta kuitenkin käytännössä linkkaukseen käytetyt ominaisuudet, varsinkin tesauroyksissä, ovat ekvivalenssiyhteys, hierarkkinen sekä assosiatiivinen suhde [26]. Usein selkeyden ja teknisen toteutuksen yksinkertaisuuden vuoksi linkitykset saatetaan tehdä vain ekvivalenssiyhteyttä käyttäen. Esimerkkinä hierarkkisesta ontologiasuhteesta on muun muassa Yleinen suomalainen ontologia YSO¹ [54] erikoisontologioineen² sekä SUMO [46] erikoisontologioineen,

¹<http://www.seco.tkk.fi/ontologies/yso/>

²<http://www.seco.tkk.fi/ontologies/>

kuten [46, 56, 58]. owl:sameA-suhde on käytössä esimerkiksi Freebasen³ ja DBpedia⁴ [39] välillä kun taas closeMatch-suhdetta on käytetty Kongressin kirjaston⁵, Saksan⁶ sekä Ranskan⁷ kansalliskirjastojen auktoriteettitietokantojen väliseen linkitykseen.

Lopuksi esittelen ontologioiden välisten linkkien tallentamisvaihtoehtoja. Tesaurusstandardissa ISO 25964-2 [26] mainitaan, että ontologioiden linkitysten vaihtoehtoja tallennuspaikaksi ovat säilyttäminen ja ylläpito linkityksen lähde- tai kohdeontologiassa, säilytys erillään kummastakin ontologiasta tai samassa järjestelmässä kuin missä lähde- ja kohdeontologiat ovat. Tämä tarkoittaa, että ontologioiden välisiä linkityksiä voivat tehdä muutkin kuin linkitettävien ontologioiden kehittäjät, joten kyseisten ontologioiden kehittäjät ole välttämättä tietoisia linkitysten olemassaolosta. Mikäli linkityksiä ei ole julkaistu avoimesti, ei linkityksen kohteena oleva ontologia välttämättä tiedä linkityksistä, jolloin ei ole mahdollista huomioida linkityksiä ontologioiden päivityksessä.

³<https://www.freebase.com/>

⁴<http://dbpedia.org/>

⁵<http://id.loc.gov/authorities/subjects.html>

⁶<http://www.dnb.de/gnd>

⁷<http://rameau.bnf.fr/>

Luku 3

Linkitettyjen ontologioiden päivittäminen

Tässä luvussa keskitytään linkitettyjen ontologioiden kehittämiseen ja kehittämisen aikana tehtäviin muutoksiin. Tarkastelun kohteena ovat vain olemassa olevien ontologioiden muutokset, joten ontologioiden luominen tyhjästä ja aikaisempia lähteitä tai ontologioita hyödyntäen rajautuu työn ulkopuolelle. Näihin voi tutustua muun muassa viitteissä [22, 50, 54, 67]. Luvussa tutustutaan ensin ontologioiden päivityksiin ja muutoksiin yleisellä tasolla. Tämän jälkeen tarkastellaan muutoksia konkreettisemmalla tasolla ja lopuksi perehdytään kirjallisuuteen ontologioiden kehittämisestä sekä muutostyökaluista.

3.1 Ontologioiden muutokset

Ontologia voidaan luoda kerralla valmiiksi tiettyä tarkoitusta varten. Usein ontologiaa kuitenkin halutaan päivittää vastaamaan muuttuviin tilanteisiin tai tarpeisiin. Klein ja Fensel [35] ovat määritelleet, että ontologian kehitystarpeet tulevat joko ontologian aiheen, käsitteellistämisen tai spesifikaation muutoksista. Flouris ja kumppanit [14] ovat lähestyneet päivitystarvetta käytännöllisemmästä näkökulmasta listaten muutokseen alla olevat syyt:

- maailma on muuttunut
- muutos näkökulmaan, jolla ontologian aihepiiriä tarkastellaan
- löydetty ongelma alunperäisessä käsitteellistämisessä
- halu lisätä uusia ominaisuuksia käyttäjätoiveiden pohjalta

- aihepiiristä aikaisemmin saavuttamattomissa olleen tiedon saaminen

Mädche ja kumppanit [41] taas listaavat syitä ontologioiden muutokseen ohjelmistotuotannon näkökulmasta. Ontologian muutoksia aiheuttavat heidän mukaansa:

- käyttöympäristön muutos, esimerkiksi ontologiaan voidaan haluta sisällyttää uuden alan käsitteistöä
- käyttäjävaatimusten muutos, esimerkiksi ontologiaa käyttävälle sovellukselle voi tulla uusi käyttäjäryhmä
- ontologian käyttöä tarkkailtaessa huomatu uudet vaatimukset

Näissä kaikissa muutoslistoissa on yhteistä se, että muutospaineen tuo joko aihepiirin tai käyttöympäristön muutos. Ontologiat eivät siis eivät ole pelkästään aihepiirinsä kuvauksia, vaan niistä heijastuu myös niiden käyttötapa.

Koska ontologioiden muutos- ja päivitystarpeet ovat erilaisia, ne voivat lähteä kehittymään erilaisiin suuntiin. Flouris ja kumppanit [14] listaavat eri näkökulmat, joista ontologioiden kehitystä ja muutoksia tutkimuksessa on tarkasteltu:

- ontologioiden evoluutio, versiointi (engl. *ontology evolution, versioning*)
- siltaaminen, morfismi, linjaus, artikulaatio, kääntäminen (engl. *mapping, morphism, alignment, articulation, translation*)
- integraatio, yhdistely (engl. *integration, merging*)

Ontologioiden evoluutio ja versiointi keskittyvät ratkaisemaan ontologioiden muutoksista ja päivityksistä johtuvia yhteensopivuuden haasteita. Ala on kehittynyt tietokantatutkimuksen skeemaevoluutiosta ja versioinnista (engl. *schema evolution and versioning*) [48]. Tietokannoissa evoluutiolla tarkoitetaan skeeman muutosta siten, että muutoksen aikana dataa ei häviä. Versioinnilla taas tarkoitetaan mahdollisuutta kysellä dataa skeeman eri versioiden avulla. Ontologioiden kanssa evoluutiolla tarkoitetaan usein ylipäättänsä muutoksia ontologiassa, kun taas versioinnissa ollaan kiinnostuttu alaspäin yhteensopivuudesta sekä instanssidatan validina ja käytettävänä säilymisestä.

Ontologioiden yhtäläisyyksien kartoittamiseen on useita samankaltaisia termejä ja aloja. Näitä ovat ontologioiden siltaaminen, morfismi, linjaus, artikulaatio ja käännökset, jotka keskittyvät vaihtelevan sisältöisten ja esitystapaisten ontologioiden välisten yhtäläisyyksien löytämiseen. Kyseessä voi olla esimerkiksi saman alan

kaksi tai useampi ontologiaa, joiden samankaltaisuutta yritetään kartoittaa tai erikoisontologia, joka halutaan linkittää yleisontologiaan. Suurin osa näistä aloista on kiinnostunut etsimään käsitteiden välisiä yhteyksiä ja esittämään tulokset halutussa muodossa, esimerkiksi ontologiaan lisättyinä linkkeinä ontologioiden välillä tai ontologiana yhteisistä käsitteistä. Ontologioiden integraatiossa ja yhdistelyssä tavoitteena on luoda uusi yhtenäinen ontologia hyödyntäen useampia ontologioita ja näiden välisiä linkityksiä.

On hyvä huomata, että yllä esiteltyt ontologioiden yhtäläisyyksien alat keskittyvät lähes pelkästään menetelmien kehittämiseen tai parantamiseen. Ontologiakehittäjien tarpeet eivät ole juurikaan esillä ontologioiden evoluutiota käsittelevässä kirjallisuudessa.

Muutostyypit

Ontologian päivityksessä ontologiaan tehtäviä muutoksia voidaan kategorisoida usein eri tavoin. Klein ja kumppanit [33] ovat jakaneet ontologian muutokset kolmeen kategoriaan.

- **ei-loogiset muutokset**, kuten muutokset luonnollisella kielellä esitettyihin arvoihin
- **loogisten määritelmien muutokset**, eli muutokset käsitteen määrittelyyn, kuten subClassOf-ominaisuuden arvojen muutokset. Termi on harhaanjohtava, koska sillä tarkoitetaan käsitteen muita kuin merkkijonoarvoihin tulevia muutoksia eikä suinkaan logiikkaan liittyviä muutoksia.
- **tunnistemuutokset**, kuten resurssien lisäykset tai poistot

Yleisempi jakoperiaate on kuitenkin Stuckenschmidtin ja Kleinin [60] muutosten jako atomisiin (engl. *atomic*) sekä kompleksisiin (engl. *complex change*) muutoksiin. Atomiset muutokset tunnistetaan hyödyntämällä käytetyn ontologiakielen ominaisuuksia, eikä tietoa ontologiasta itsestään käytetä. Kompleksiset muutokset taas voivat hyödyntää tietoa ontologian rakenteesta tai ne olla yhdistelmä atomisia ja kompleksisia muutoksia. Lisäksi kompleksisia muutoksia voidaan käyttää muutosten syiden tai vaikutusten esittämiseen.

Klein ja Noy ovat useampaan otteeseen listanneet ontologioiden muutostyyppejä. Vuonna 2003 he listasivat yli 80 muutostyyppiä [34]. Muutostyyppejä ei oltu eritelty tarkemmin, vaan mainittiin näiden pohjautuvat OWL-ontologiakielen ominaisuuksiin. Oletettavasti muutostyypit ovat atomisia, sillä niiden tavoitteena on mallintaa ontologioiden muutoksia. Seuraavana vuonna he listasivat ontologioille

22 muutosoperaatiota [48], jolloin tarkastelu kohdistui siihen, millaisissa muutoksissa instanssidata säilyy validina. Lähes kaikki muutostyypeistä ovat kompleksisia yhdistäen useampia atomisia muutoksia ja tietoa ontologian rakenteesta. Lisäksi monet muutostyypeistä ovat ominaisuuksien määritelmien muutoksia tai muutoksia millä käsitteillä kyseistä ominaisuutta voi käyttää. Kumpikin näistä muutostyyppilistoista oli siis tarkoitettu mallintamaan muutoksia raskasontologioissa, joiden skeemaa tyypillisesti muutetaan ontologiaa kehittäessä. Skeemamuutoksia ei juurikaan tehdä kevytontologioille, joten suurin osa muutuskategorioista jäisi käyttämättä kevytontologian muutoksia mallintaessa.

Kompleksisia muutostyyppisiä ovat listanneet myös muun muassa Stojanovic ja kumppanit [59] sekä Stuckenschmidt ja Klein [60]. Stuckenschmidt ja Klein määrittelivät 120 kompleksista muutosta, mutta mainitsevat määrän kasvavan koko ajan, sillä kompleksisia muutoksia voi määrittellä lisää. Lisäksi Klein [34] huomauttaa, että kompleksisia muutoksia ei voi täydellisesti tunnistaa, sillä käsitteisiin voidaan tehdä muitakin muutoksia. Tämän takia kompleksisten muutosten tunnistaminen vaatii heuristiikkaa.

Poiketen ontologioiden muutoksen näkökulmasta, tesaurusstandardi ISO 25964-1 [25] listaa tesauruksen muutoksia kehittäjänäkökulmasta. Muutostyyppit ovat selvästi tarkoitettu ihmisten luettavaksi, sillä osa kuvauksista on epämääräisiä, kuten kohta ”k) hierarkian muuttaminen”. Monet muutoksista ovat kompleksisia muutoksia yhdistellen useita muutoksia laajemmaksi käsitteelliseksi muutokseksi.

3.2 Ontologioiden kehittäminen

Tässä aliluvussa perehdytään ontologioiden kehittämiseen sekä linkitettyjen ontologioiden kehityksen haasteisiin. Usein ontologia tai ontologiakokonaisuus ei ole täysin yhden kehittäjän vastuulla. Yleisempi tilanne on, että useampi kehittäjä kehittää samaa ontologiaa tai toisiinsa linkittyneitä ontologioita. Noy ja kumppanit [47] ovat listanneet seuraavia ontologioiden yhteiskehityksen ulottuvuuksia:

Synkroninen vs. asynkroninen kehittäminen

Synkronisessa kehityksessä ontologiakehittäjät pystyvät välittömästi näkemään ontologiassaan muiden tekemät muutokset. Muutosten välitön näkeminen vähentää huomattavasti päällekkäisten muutosten tekemistä. Asynkronisessa kehittämisessä taas kehittäjät tekevät itsenäisesti muutoksia ontologian kopioon ja nämä tehdyt muutokset lopulta yhdistetään. Koska kehittäjät eivät voi nähdä toistensa tekemiä muutoksia, voi yhdistelyssä tulla päällekkäisiä tai ristiriitaisia muutoksia. Nämä ristiriidat tulee ratkaista ontologioiden yhdistelyn aikana.

Jatkuva julkaisu vs. ajoittainen arkistointi Jatkuvassa julkaisussa ontologian sen hetkinen versio on julkisesti nähtävillä. Täten ontologiasta ei ole erikseen talletettuja ja nimettyjä versioita. Ajoittaisessa arkistoinnissa ontologiasta laitetaan esille ja tallennetaan julkaistava versio. Ero on siis versioinnissa, ei julkaisutahdissa.

Kuratoitu vs. kuratoimaton Kuratoidussa ontologiassa ennen ontologian julkaisua ontologian kehityksen vastuuhenkilö käy läpi ontologian muutokset yksitellen joko hyväksyen tai hyläten ne. Kuratointi tapahtuu lähes aina ajoittaisen arkistoinnin ontologioissa ennen seuraavan version julkaisua.

Valvottu vs. valvomaton Valvotussa ontologiakehityksessä ontologiassa tehdyt muutokset sekä niihin liittyvä mahdollinen metadata tallennetaan. Valvomattomassa ontologiakehityksessä taas muutoksia ei tallenneta ollenkaan tai ne eivät ole hyödynnettävissä muissa työkaluissa.

Ontologioiden yhteiskehityksessä on tyypillistä, että kehittäjillä on erilaisia rooleja. Esimerkiksi kaikki kehittäjät eivät välttämättä osallistu ontologian kuratointiin. Varsinkin raskaiden ontologioiden kehityksessä on tyypillistä, että kehityksessä on erikseen aihealueen asiantuntijat sekä ontologiakehittäjät. Asiantuntijat tietävät millaisia käsitteitä heidän alallaan on, mutta heillä ei ole konkreettista tietoa ontologian rakentamisesta. Ontologiakehittäjät ilmaisevat ontologiamuodossa asiantuntijoiden mallit alastaan. Samoin tesauruskeittäjä voi kysyä asiantuntijoilta apua termien määritelmiin ja käsitteiden suhteisiin, mikäli ala on hänen osaamisalueensa ulkopuolella.

Mikäli ontologiakehittäjillä on omat vastuualansa ontologioiden kehittämisessä, eli he toimivat asiantuntijoina ja ontologiakehittäjinä samanaikaisesti, voidaan roolit jakaa käsitteiden aihealueiden mukaan. Yksi kehittäjä voi olla vastuussa esimerkiksi maatalouteen liittyvistä käsitteistä ja toinen kalan kasvatusta sivuavista. Ontologioiden kehittäminen pienempinä ontologioina ja näistä ontologiapalasista isomman kokonaisuuden koostaminen helpottaa kehitystä. Yhdellä kehittäjällä on vain tietty aihealue työn alla, jolloin ontologia pysyy hallittavan kokoisena. Ontologioiden jakaminen pienempiin aihealueisiin mahdollistaa sen, että ontologiaa työstää kyseisen aihealueen asiantuntija, jolloin ontologian sisällöstä tulee eksaktimpi.

Yleensä useamman kehittäjän ontologiassa tavoite on se, että kaikki kehittäjät pystyisivät kehittämään ontologiaa samanaikaisesti ja näkemään toistensa tekemät muutokset. Valitettavasti yksittäisen ontologian yhteiskehitys ei ole aina mahdollista, jolloin organisaatiossa saatetaan haluta lähteä kehittämään omaa ontologiaa. Syitä voivat olla muun muassa:

- Ontologian kehittäjäryhmään liittyminen voi olla rajoitettua. Kehittäjäksi pääseminen voi vaatia tietyssä organisaatiossa olemista tai mahdollisuuden sitoutua ontologian kehitykseen.
- Ontologiaeditori ei välttämättä tue tiedostomuotoa, yhteiskehitystä tai se on liian kallis kehitysorganisaatioille.
- Kehittäjien tai -organisaatioiden näkemykset ontologian tavoitteista ja kehitystahdistista eivät vastaa oman organisaation tavoitteita. Ontologia voidaan haluta tehdä muutoksia nopealla aikataululla, tehdä hyvin spesifejä muutoksia tai pitää ontologia oman organisaation sisällä.

Tilanteessa, jossa ontologiaa ei pystytä kehittämään yhtenä ryhmänä, vaihtoehtona on jakaa ontologian vastuualueet erillisiin ontologioihin, jolloin muodostuu yleisontologian ja sen erikoisontologioiden kokonaisuus.

Linkitettyjen ontologioiden kehittämisen haasteet

Verrattuna täysin erillään kehitettäviin ontologioihin, linkitetyissä ontologioissa hyödynnetään muualla tehtyä työtä linkittämällä erillisiä ontologioita toisiinsa. Linkitettyjen ontologioiden kehityksen haasteena ovatkin nämä yhteydet muihin ontologioihin. Luvussa 2.3 esiteltiin kaksi erilaista motiivia ontologioiden linkittämiseen. Nämä vaikuttavat myös linkityksen ylläpitoon.

Kun ontologia on linkitetty toiseen ekvivalenssiyhteyksien avulla, on tavoitteena yleensä toisaalla tehdyn ontologian tai annotoidun aineiston hyödyntäminen. Ekvivalenssiyhteydellä ilmaistaan, että kaksi käsitettä ovat samat, eikä käsitteiden merkitys useinkaan muutu radikaalisti ontologiaa päivitettäessä. Tämän takia käsitteiden välisten linkkien ajantasaisuutta ei tarvitse kriittisesti ylläpitää, sillä linkitys vain lisää ylimääräistä tietoa itsenäisesti toimivaan ontologiaan.

Tilanne on toinen, mikäli käsitteitä on yhdistelty toisiinsa myös hierarkkisilla ominaisuuksilla. Hierarkkiset ominaisuudet ovat transitiivisia ominaisuuksia vaatien koko hierarkiaketjun olevan validi. Hierarkkisia suhteita voi hyödyntää päättelyyn tai kyselynlaajentamiseen, joten näiden tulisi olla oikein, jotta haut tuottavat haluttuja tuloksia. Tällainen tilanne on esimerkiksi erikoisontologioilla, jotka ovat linkittyneet yleisontologiaan saadakseen täydellisen hierarkian. Hierarkkisesti linkittyneessä ontologiassa on tärkeää tietää, mitä muutoksia linkityksen kohdeontologiassa on tapahtunut ja voivatko ne rikkoa hierarkian konsistenttiuden tai muuten luoda tarvetta lähdeontologian muuttamiseen.

Luvussa 2.3 mainittiin myös, että riippuen ontologioiden välisen linkityksen säilytys- ja ylläpitopaikasta, ei linkityksen kohteena oleva ontologia ole välttämättä tie-

toinen koko linkityksen olemassaolosta. Jos kohdeontologia ei tiedä linkityksen olemassaolosta, voidaan kohdeontologian kehityksessä tehdä ratkaisuja, jotka aiheuttaisivat epätoivottuja muutoksia linkityksen lähdeontologiaan. Vaikka linkitykset olisivat tiedossa, voivat ontologian kehittäjät päättää, että ontologiaa kehitetään haluttuun suuntaan, ja linkityksen lähdeontologiat joutuvat mukautumaan tämän kohdeontologian muutoksiin. Tällainen tilanne on usein yleisontologioilla, joihin on sitoutunut useita erikoisontologioita. Erikoisontologioiden tulisi siis reagoida yleisontologian muutoksiin ja täten varmistaa erikoisontologian konsistenttius. Täten yleisontologian muutokset aiheuttavat muutospainetta erikoisontologiaan, mutta erikoisontologian muutoksia ei oteta huomioon yleisontologian kehityksessä.

Mikäli linkittyneiden ontologioiden kehittäjien välinen yhteydenpito ei ole tiivistä, syntyy ontologioihin usein päällekkäisiä käsitteitä. Ontologioita voidaan kehittää samanaikaisesti, mutta vasta julkaistusta versiosta saatetaan nähdä, että sama käsite on lisätty kumpaankin ontologiaan. Päällekkäisyyteen reagoiminen riippuu erikoisontologian kehityksiperiaatteista. Mikäli erikoisontologian tarkoitus on kuvata alansa käsitteistö kattavasti, erikoisontologia tyypillisesti sisältää päällekkäisiä käsitteitä yleisontologiansa kanssa. Tyypillisesti nämä päällekkäiset käsitteet on ilmaistu ekvivalenssiyhteydellä, joten tässäkin tapauksessa voidaan lisätä ekvivalenssi yleisontologian ja erikoisontologian saman käsitteiden välille. Erikoisontologian tavoite voi myös olla listata vain yleisontologiasta löytymättömät erikoisalan käsitteet, ja siten minimoida päivitystyötä. Tällaisessa tilanteessa tulee tarkastella onko päällekkäinen käsite sama, eli tulisiko erikoisontologian käsite poistaa. Varsinkin samaan yleisontologiaan linkittyneiden ontologioiden välillä on usein paljon päällekkäisyyksiä, vaikka alat tuntuisivat intuitiivisesti olevan sanastoltaan kaukana toisistaan. Esimerkkinä työkalusta erikoisontologioiden päällekkäisyyksien tarkasteluun on KOAN-työkalu¹.

Ontologian muutosten tunnistaminen ja esittäminen

Aikaisemmassa tutkimuksessa ei ole eksplisiittisesti määritelty, mitä ontologioiden muutokset ovat. Tässä työssä ontologian muutoksella tarkoitetaan sitä, että jonkin URI-resurssin jonkin ominaisuuden arvoon tulee muutos. Käsitteen muutos voi olla arvon lisäys, poisto tai muuttuminen. Tähän määritelmään mahtuu myös käsitteiden lisääminen ja poistaminen, sillä käsite lisätään ontologiaan, jos ontologiaan ilmestyy uusi subjektina toimiva URI-tunniste. Koska RDF-tiedostoissa kolmikoiden järjestyksellä ei ole väliä, ei ontologian muutoksia pysty luotettavasti tunnistamaan ilman tiedoston lataamista muistiin ja muuttamista verkkomuotoon. Tässä työssä ei tarkastella syntaktisen tason muutoksia, joita tulee seriali-

¹<http://www.seco.tkk.fi/tools/koan/>

sointiformaattia vaihdettaessa, eikä tyhjiä solmujen (engl. *blank node*) muutosten tunnistamista.

Ontologioiden tulee reagoida alansa sekä käsitteellistämisen muutoksiin. Tätä kutsutaan ontologioiden kehittymiseksi (engl. *ontology evolution*) [14]. Noy ja Klein [48] listaavat kaksi tapaa ontologian kehittymiseen ja muutosten tunnistamiseen. Jäljitettävässä (engl. *traced*) ontologiakehityksessä muutokset ovat operaatioita ja niihin voi reagoida heti muutoksen tapahduttua. Jäljittämättömässä (engl. *un-traced*) kehityksessä on vain kaksi ontologian versiota, eikä tietoa siitä, mitä operaatioita ontologioiden välillä on tehty. Nämä vastaavat valvottua ja valvomaton- ta ontologiakehitystä. Kummallakin muutosten tunnistustavalla on mahdollista löytää ja listata ontologiaan tehdyt muutokset kahden version välillä.

Ontologian muutosloki tai kahta ontologiaversiota hyödyntämällä voidaan tunnistaa ontologiassa tapahtuneet muutokset. Näitä muutoksia on mahdollista esittää eri muodoissa. Klein ja Noy [34] ovat esitelleet seuraavat muutosten esitysmuodot:

Muutosloki (engl. *change log*) listaa ontologian aikaisemman ja myöhemmän version välillä tehdyt muutokset aikajärjestyksessä [34]. Muutosloki on helppo saada omasta kehitettävästä ontologiasta, mikäli ontologiaeditori tallentaa muutokset, mutta muualla kehitettävistä ontologioista ei välttämättä ole muutoslokiä saatavilla. Ontologioiden yhteiskehityksessä tulee lisäksi haasteita kehittäjiä muutoslokien yhdistämisestä [34]. Muutosloki on esitystavoista ainoa, joka mahdollistaa ajallisen tallentamisen ja täten muutosten kumoamisen tai uudelleen tekemisen.

Rakennevertailussa (engl. *structural diff*) listataan, mitkä käsitteet löytyvät mistäkin ontologiaversiosta sekä onko käsite uudelleennimetty [34]. Moderneissa ontologioissa käsitteillä on pääsääntöisesti pysyvät URI-tunnisteet, joten rakennevertailutiedostosta ei ole juurikaan hyötyä.

Käsitteellisillä muutoksilla (engl. *set of conceptual changes*) esitetään, miten itse käsitteet ovat muuttuneet muutoksissa [34]. Esimerkiksi, uuden käsitteen lisäyksen myötä alkuperäisen käsitteen merkitys on voinut kaventua.

Muutosjoukko (engl. *transformation set*) sisältää operaatiot, joita suorittamalla ontologian vanhemmasta versiosta saadaan uudempi versio. Muutosjoukko ja muutosloki ovat samankaltaisia, mutta ne eroavat toisistaan muutamalla tavalla. Ensinnäkin, muutosjoukko sisältää vain muutokset, jolla version saa muutettua toiseksi, kun taas muutoslokissa ovat kaikki käyttäjän tekemät muutokset. Toisekseen, muutosjoukossa operaatioiden järjestyksellä ei ole väliä, kun taas muutoslokissa muutokset ovat aikajärjestyksessä. Tosin hyödynnettäessä muutosjoukkoa versiosta toiseen siirtymiseen, käsitteiden

luominen tulee tehdä ennen muiden muutosten toimeenpanoa. Lopuksi, muutosjoukkoja voi olla useita mahdollisia kahden version välillä, kun taas muutosloki on yksi toteutettujen muutosoperaatioiden ilmentymä. Mikäli muutoslokissa ei ole kumoamisoperaatioita, on muutosloki määrityksellisesti muutosjoukko.

Ylläolevista esitysmuodoista vain käsitteellisten muutosten tunnistaminen vaatii ymmärrystä ontologian rakenteesta. Muutosloki, muutosjoukko sekä rakennevertailu pystytään luomaan ilman ontologian rakenteeseen perehtymistä. Tosin mikäli rakennevertailussa halutaan tietoa uudelleennimetyistä resursseista, vaatii tämä sekä tietoa ontologian rakenteesta että heuristista päättelyä määrittämään uudelleennimeämiset.

Muutosten välittäminen linkitettyihin ontologioihin

Edellisessä aliluvussa esiteltiin esitystapoja yksittäisen ontologian muutosten listaukseen. Linkitettyjen ontologioiden tapauksessa ontologian muutosten tunnistamisen jälkeen muutokset täytyy välittää muille ontologiaan linkittyneille ontologioille. Ontologioiden muutosten välittämiseen (engl. *change propagation*) on kaksi tapaa: työntävä välittäminen (engl. *push-based*) sekä vetävä välittäminen (engl. *pull-based*) [6], jotka vastaavat löyhästi synkronista ja asynkronista ontologiakehitystä.

Vetävässä välittämisessä ontologiankehittäjä pyytää halutessaan muutokset ontologioista, joihin hänen ontologiansa on linkittynyt. Työntävässä muutosten välittämisessä linkittynyt ontologia ei voi itse päättää, milloin ottaa muutoksia vastaan, vaan muutokset välitetään automaattisesti kehittäjän ontologiaan. Välystiheyksiä on useampia [40]: Jaksottaisessa muutosten välittämisessä (engl. *periodic delivery*) muutokset työnnetään linkittyneisiin tietyn ajanjakson välein, kun taas vaihtuvajaksaisessa välittämisessä (engl. *ad-hoc delivery*) ne työnnetään määrittelemättömien ajanjaksojen välein. Välittömässä muutokset työnnetään heti niiden ontologiaan suorittamisen jälkeen. Työntävä välitön muutostenvälittäminen soveltuu tilanteisiin, joissa linkitettyjen ontologioiden konsistenttius on tärkeää, sillä muissa välitysvaihtoehdoissa ontologiat voivat olla keskenään epäkonsistentteissa tiloissa. Hyväksyttäessä ontologioiden väliaikaiset epäkonsistenttiudet, on vetävä välittäminen parempi vaihtoehto linkitettyyn kehittämiseen.[40]

Työkaluja ontologioiden muutosten hallitsemiseen

Ontologioiden muutosten alalla ei ole syntynyt yhtenäistä muotoa ontologioiden muutosten esittämiseen tai tallennukseen. Tämän takia on olemassa monia erillisiä työkaluja ontologioiden muutosten tunnistamiseen ja esittämiseen. Nämä muutos työkalut ovat jaettavissa kahteen päätyyppiin: muutoslokiin pohjautuviin ja muutosjoukkoon pohjautuviin.

Muutoslokipohjaisia työkaluja ovat muun muassa Protégé-editorin Change-management-lisäosa [47] sekä Khattakin ja kumppanien nimetön Protégé-lisäosa [31], joista kumpikin olettaa, että ontologiaa kehitetään kyseisellä editorilla. Kumpikin työkaluista pohjautuu omaan muutosontologiaansa; ensimmäisessä ontologiana on Change and Annotation Ontology (ChAO) ja jälkimmäisessä Change History Ontology (CHO).

Muutosjoukkopohjaisissa työkaluissa taas ei ole vaatimusta, että ontologiaa täytyisi kehittää sillä editorilla, jolle muutostyökalu on ohjelmoitu. Esimerkkeinä tällaisista työkaluista on ContentCSV-työkalu [28] sekä PROMT-lisäosa Protégé-editoriin [47], joka myöskin perustuu ChAO-ontologiaan muutosten esittämisessä.

Muutoslokipohjaisten editorien etu muutosjoukkopohjaisiin on se, että saman editorin sisällä kehitettäessä muutokset tallennetaan aikajärjestyksessä, minkä lisäksi niihin voi lisätä ylimääräistä tietoa annotoimalla. Luonnollisesti lokipohjainen muutosten esittäminen vaatii, että ontologian kaikki kehittäjät käyttävät samaa editoria. Muutosjoukkopohjaisessa muutoksien esittämisessä muutokset voi tunnistaa ontologiasta sen kahta versiota vertailemalla, jolloin kehittäjät eivät ole sidottuja tietyn editorin käyttöön. Haittapuolena tällöin tosin on se, että mahdollisesti menetetään editorin tuottamat muutostiedot, kuten milloin muutos on tehty ja kuka sen tekijä oli. Myöskään muutostietoja ei välttämättä saada näkyviin kehittäjän omassa editorissa, vaan hän voi joutua käyttämään erillistä ohjelmaa muutoksien tarkasteluun.

Luku 4

Menetelmä

Tässä luvussa tutustutaan ensin ontologioiden muutosten osa-alueisiin, jotka ovat jääneet vähälle huomiolle aikaisemmissa tutkimuksissa. Tämän jälkeen esittelen menetelmän auttamaan erikoisontologioiden kehittäjiä yläontologioiden suurten muutosmäärien kanssa.

4.1 Aikaisemmassa tutkimuksessa sivuun jääneitä aiheita

Ontologioiden hyödyt tulevat parhaiten esille, kun ontologiaa käyttää ja kehittää suuri määrä ihmisiä eri organisaatioista. Tämä kuitenkin johtaa erinäisiin haasteisiin ontologioiden kehityksessä. Pinto ja kumppanit [52] ovat huomanneet, että aikaisemmat tutkimukset eivät ole ottaneet huomioon, että käytännössä ontologiakehitys on usein hajautettua, osittain autonomista, iteratiivista ja sitä tekevät muutkin kuin ontologiaekspertit. Avaan tarkemmin seuraavaa kolmea vähän käsitellyllä haasteella:

- Katkonaisesta ontologiakehityksestä johtuvat suuret muutosmäärät.
- Linkitettyjä ontologioita kehittävien organisaatioiden eroavat teknologiset valinnat.
- Ontologiakehittäjien näkökulma ja tarpeet ontologioiden muutoksista tiedottaessa.

Katkonainen ontologiakehitys

Kuten Pinto ja kumppanit [52] toivat esiin, ontologiakehitys on harvoin jatkuva. Vähäiset ajalliset tai taloudelliset resurssit johtavat siihen, että ontologioiden kehitys on usein jaksottaista. Linkitetyillä ontologioilla katkonainen kehitys voi johtaa siihen, että linkityksen kohdeontologiassa on voinut kehitystauon aikana tapahtua suuri määrä muutoksia. Nämä muutokset tulisi käydä läpi ennen lähdeontologian omien päivitysten tekemistä, jotta ontologia pysyisi ajantasaisena ja konsistenttina. Vaikka linkityksen lähdeontologiaa kehitettäisiinkin jatkuvasti, on myös mahdollista, että kohdeontologia ei ole jatkuvan julkaisun mallissa. Tällöin ontologian uusi versio julkaistaan vain tietyn ajanjakson välein, jolloin muutokset tulevat yhdellä kertaa suurena massana.

Kummassakin ylläkuvaillussa tapauksessa haasteena on kehittäjälle tuleva suuri määrä muutoksia. Tällaisessa tilanteessa on vaikeaa saada käsitystä ontologian ja käsitteen muutoksista sekä ymmärtää, miten muutoksiin tulisi reagoida linkityksen lähdeontologiassa.

Suurten muutosmäärien huomioimisesta ei ole ollut puhetta aikaisemmassa tutkimuksessa. Työkalujen arvioinnit on tehty pienillä ontologioilla, kuten tutkimuksissa [2, 32], ja kuvankaappauksissa on muutoksia kerrallaan nähtävissä vain muutamia, kuten tutkimuksissa [29, 47].

Linkitetty ontologiakehitys eri organisaatioissa

Luvussa 3.2 listattiin syitä, miksi erillisiä linkitettyjä ontologioita kehitettäisiin yksittäisen ontologian yhteiskehityksen sijaan. Koska linkitettyjä ontologioita kehitetään yleensä eri organisaatioissa, voi toisiinsa linkittyneiden ontologioiden kehittämiseen olla käytössä eri ontologiaeditoreita. Lisäksi editorien muutostyökaluilla ei ole yhteistä muutostentallennusmuotoa, joten editoreiden väliset muutosmuodot ovat keskenään yhteentoimimattomia, eikä muutoksia voi siirtää editorista toiseen ilman erillistä muunnosta. Ei ole myöskään itsestäänselvyys, että ontologioista ylipäätään julkaistaisiin muutoslistoja julkisesti.

Aiemmassa tutkimuksessa ei ole otettu huomioon linkitetyn ontologiakehityksen haasteita eri editorien ja teknologioiden käytössä, vaan on oletettu, että hajautettu kehitys tapahtuu samalla editorilla, kuten esimerkiksi Sunagawa ja kumppanit [62] ovat ajatelleet.

Ontologiakehittäjien näkökulma

Aikaisemmassa tutkimuksessa ollaan keskitytty uusien menetelmien ja työkalujen luomiseen ontologioiden muutosten tunnistamiseen ja välittämiseen. Näissä töissä muutosten esitysmuotojen ja optimaalisten käyttöliittymäratkaisujen tutkiminen sekä ontologiakehittäjien tarpeet ovat olleet taka-alalla.

Muutosten esittämistapoja on esitelty lyhyesti sanallisesti ja muutosten esittämisestä on ollut muutamia kuvankaappauksia, kuten [28, 47]. Kuitenkaan keskustelua käyttöliittymätoteutusten hyvistä ja huonoista puolista ei ole.

Valitettavan usein työkalujen arvioinneissa testikäyttäjät eivät ole itse ontologiakehittäjiä, tai itse testit kestävät vain muutamia tunteja. Näillä saadaan kyllä testattua toimiiko työkalu odotetusti, mutta arvokasta lisätietoa ontologiakehittäjien tarpeista ja toiveista ei tule kentältä kuultavaksi. Poikkeuksena tälle ovat olleet Millerin ja kumppanien [44] suunnitteluperiaatteet, Noyn ja kumppanien [47] listatamat käyttäjätarpeet ontologioiden yhteiskehityksessä sekä Braunin ja kumppanien [8] SOBOLEO-editorin käytön arviointi.

4.2 Menetelmä ontologioiden suurten muutosmäärien esittämiseen

Tässä luvussa esittelen oman menetelmäni siihen, miten erikoisontologiakehittäjille tulevia yleisontologian suuria muutosmääriä pystyy muuttamaan hallittaviksi ja havainnoitaviksi kokonaisuuksiksi. Alla käytän termejä yleisontologia ja erikoisontologia selvyuden vuoksi, mutta menetelmä on sovellettavissa mille tahansa linkitetuille lähde- ja kohdeontologioille. Lisäksi menetelmää voi hyödyntää myös yhteiskehittäviin kevytontologioihin, joissa kehittäjillä on vastuullaan tietyt merkityt käsitteet.

Menetelmä perustuu neljään pääajatukseseen:

- Yleisontologian muutosten läpikäyntijärjestystä voi priorisoida, koska kaikki yleisontologian muutokset eivät välttämättä aiheuta muutoksia erikoisontologiaan.
- Erikoisontologiakehittäjä voi hyötyä muistakin tiedoista kuin pelkästään listata yleisontologian muutoksia.
- Kaikki yleisontologian yhteen käsitteeseen tehdyt muutokset tulee nähdä kerralla.

- Yleisontologian muutostietojen tulee olla editoririippumattomassa muodossa.

Muutosten priorisointi

Laajojen kevytontologioiden suurten muutosmäärien alla on tärkeää priorisoida muutosten teon järjestystä. Kevytontologioita käytetään pääasiassa aineiston annotoinnissa ja hakujärjestelmissä, joten ontologian konsistenttius ei ole niin kriittistä kuin päättelyyn käytettävillä raskasontologioilla. Tämän johdosta kevytontologioiden muutoksia voi priorisoida siten, että muutosten tarkastelu aloitetaan niistä yleisontologian muutoksista, jotka mitä luultavimmin vaativat muutoksia yleisontologiaan liittyneessä erikoisontologiassa. Tämä vastaa Falconerin ja kumppanien ehdottamaa muutosten suodattamista [12] sekä Braunin ja kumppanien raportoimaa tarvetta käyttäjäkohtaiselle personoinnille [8].

Muutokset, jotka ovat tapahtuneet lähellä erikois- ja yleisontologian linkityskohdita voivat suuremmalla todennäköisyydellä aiheuttaa muutoksia erikoisontologiassa kuin kauempana yhtymäkohdasta tapahtuneet muutokset. Esimerkiksi jos erikoisontologia käsittelee kielitiedettä, on foneemi-käsitteen muutos todennäköisesti kiinnostavampi kuin kaulahuivit-käsitteen, sillä erikoisontologialla on luultavasti yhteyksiä muihin kielitieteen käsitteisiin, muttei vaatetukseen liittyviin käsitteisiin. Kun tärkeimmät muutokset on käyty läpi, voidaan myös epätodennäköisemmät muutospaineen aiheuttajat käydä läpi, mikäli resursseja on yhä käytössä.

Vaikka yleisesti voi sanoa, että muutos on tärkeä ontologioiden linkityskohdan lähellä, vaikuttaa tärkeyteen lisäksi muita seikkoja. Myös tehdyn muutoksen tyyppi, ontologioiden väliseen linkitykseen käytetyt ominaisuudet sekä erikoisontologiassa käytössä olevat ominaisuudet voivat vaikuttaa siihen, mitkä muutokset ovat erikoisontologiakehittäjälle tärkeitä.

Muitakin tietoja kuin lista muutoksista

Yleisontologian muuttuessa voi erikoisontologian kehittäjälle olla hyödyllistä nähdä myös muutoksista johdettuja tietoja, esimerkiksi tieto siitä, mitkä yleisontologiaan lisätyistä käsitteistä jo ovat erikoisontologiassa. Tähän informaation tarvitaan yleisontologian muutostiedon yhdistäminen erikoisontologian tietoon vastaavasti kuin Stuckenschmidtin ja Kleinin kompleksiset muutokset [60]. Muutosten esittämisen ei tulisi välttämättä rajautua pelkkiin muutosten esittelemisiin vaan tarjota sel-laisia listoja ja näkymiä, mitä erikoisontologiakehittäjät tarvitsevat päivityksensä tueksi.

Käsitteen muutokset nähtävissä kerralla

Ontologian käsitteen kaikki muutokset tulee olla kerralla nähtävissä, kuten Prompt-Diff-työkalussa [49]. Tämä mahdollistaa yleiskuvan saamisen käsitteestä, toisin kuin ISO 25964-1 -standardin pelkkä muutostyypeittäin listaaminen.

Editoririippumattomuus

Ontologioiden muutosten tunnistamiseen ja esittämiseen ei ole standardeja. Editorien tarjoamat muutosten tarkastelujärjestelmät poikkeavat toisistaan ja tuottavat erilaista muutosdataa. Välillä myös ontologiaeditoria voidaan haluta vaihtaa, jolloin haasteeksi tulee muutostietojen muuttuminen. Lisäksi yleisontologiaa ja sen erikoisontologioita ei välttämättä kehitetä samaa editoria käyttäen. Näiden syiden takia on tärkeää, että muutosten esittäminen ei vaadi editorispesifejä ratkaisuja, vaan että muutokset voidaan tunnistaa ja esittää mahdollisimman samankaltaisesti editorista riippumatta.

Luku 5

MUTU-muutostentunnistustyökalu

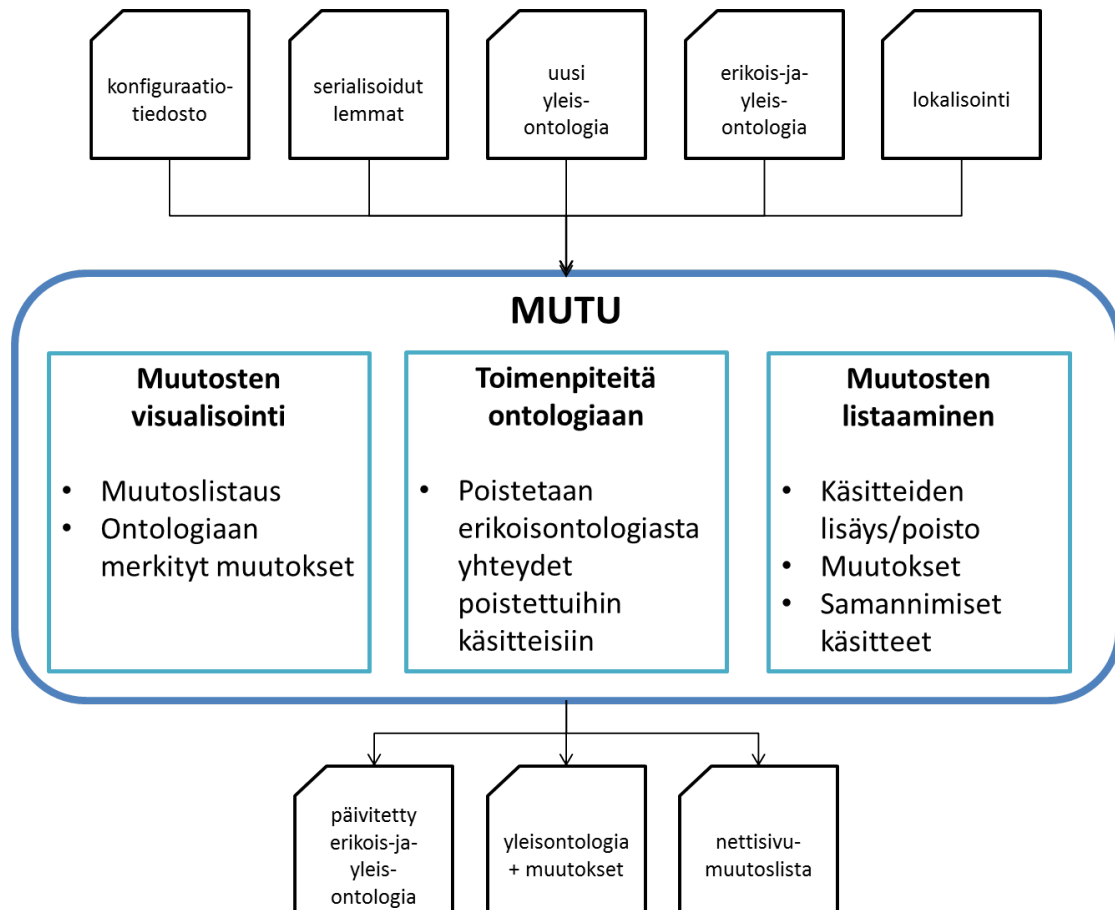
Tässä diplomityössä kehitettiin prototyypityökalu yleisontologian muutosten tunnistamiseen ja välittämiseen erikoisontologiakehittäjille. Työkalu sai nimekseen MUTU ja sen ominaisuudet perustuvat luvun 4 periaatteille.

MUTU-työkalun kehityksen pohjana olivat yleisimmät ominaisuudet, joita kevytontologioissa käytetään. Lisäksi kehityksen alussa keskusteltiin tarpeista ontologiakehittäjän kanssa, minkä johdosta työkaluun sisällytettiin samannimisten käsitteiden tunnistaminen yleis- ja erikoisontologian välillä sekä kehittäjälle turhien ominaisuuksien sivuuttaminen muutoslistoja luodessa. Lisäksi assosiattiivisen ominaisuuden muutoksille lisättiin erilliset alakategoriat suhteen lisäykselle, poistolle sekä korvaukselle.

Tässä luvussa esitellään ensin, miten MUTU-työkalua käytetään. Tämän jälkeen perehdytään muutostyyppeihin ja tärkeisiin muutoksiin. Lisäksi esittelen miten MUTU esittää muutokset sekä mitä muita skriptejä MUTUn käyttämiseen saatetaan tarvita.

5.1 MUTU-työkalun käyttäminen

Kuvassa 5.1 on kaavio siitä, mitä tiedostoja tarvitaan MUTU-muutoslistausten luomiseen sekä mitkä ovat MUTU-ohjelman pääkomponentit ja lopputulostiedostot. MUTUn itsensä toiminta jakautuu kolmeen osaan: muutosten listaamiseen, erikoisontologian muokkaamiseen sekä muutosten esittämiseen.



Kuva 5.1: MUTUn pääkomponentit syöte- ja tulostetiedostoiin

Syötetiedostot

MUTU-ohjelma vaatii ajoonsa seuraavat tiedostot:

- **erikois-ja-yleisontologia-tiedosto**
RDF/XML-muotoinen tiedosto, joka sisältää erikoisontologian ja vanhan version yleisontologiasta
- **uusi versio yleisontologiasta**
RDF/XML-muotoinen tiedosto, joka sisältää uusimman version yleisontologiasta
- **serialisoidut lemmat**
yleisontologian etukäteen perusmuotoistetut nimikkeet
- **konfiguraatiotiedosto**
tekstitiedosto, joka sisältää ontologioiden tietoja: ontologioiden käsitteiden käytössä olevat nimiavaruudet ja erityyppisten ominaisuuksien URI-tunnisteet. Lisäksi tiedosto sisältää listan ominaisuuksista, joiden muutokset tulisi sivuuttaa.
- **lokalisointitiedosto**
tekstitiedosto, joka sisältää muutoslistauksen otsikot ja muut tekstit yhdellä kielellä

Muutosten listaaminen

Muutosten listaamisessa edetään siten, että ensin selvitetään mitkä käsitteet on lisätty ja poistettu yleisontologiasta versioiden välillä. Lisätyt käsitteet löytyvät vain uudesta versiosta ja poistetut vain vanhasta versiosta. Kun lisättyjen ja poistettujen käsitteiden listat on luotu, tarkastellaan kummassakin ontologian versiossa olleita käsitteitä.

Ontologian muutoksien selvittäminen tehdään käsite kerrallaan, verraten mitä ominaisuuksia ja niiden arvoja käsitteellä on ollut yleisontologian uudessa ja vanhassa versiossa diff-algoritmin avulla. Diff-algoritmin tuottamien lisäyksiä ja poistoja yhdistelemällä voidaan esittää myös korvatut arvot, eli ominaisuudet, joille on sekä lisätty että poistettu arvo. Tämän jälkeen muutokset lajitellaan muuttuneen ominaisuuden mukaan muutostyyppeihin.

Samannimisten käsitteiden tunnistaminen perustuu yleis- ja erikoisontologian nimikkeiden perusmuotojen vertailuun. Tavoitteena on löytää mahdollisimman paljon ehdotuksia samoiksi käsitteiksi, joten vertailuun on sisällytetty sekä oletusnimikkeet (`prefLabel`) että vaihtoehtoiset nimikkeet (`altLabel`). Kun on tunnistettu

käsitteet, joilla on samannimisiä nimikkeitä, hylätään näistä käsiteparit, joilla on jo ekvivalenssiyhteys toisiinsa.

Erikoisontologian muokkaaminen

On mahdollista, että yleisontologiasta poistuneisiin käsitteisiin on luotu linkkejä erikoisontologiasta. Tämän takia tulee poistaa yhteydet näihin kolmikoihin erikoisontologiasta, jotta ontologiassa ei ole viittauksia olemattomiin käsitteisiin. Nämä poistetut kolmikot listataan erikoisontologiakehittäjälle muutoslistausten yhteydessä.

Muutosten esittäminen

Muutosten listaamisen ja erikoisontologian konsistenttiuden varmistamisen jälkeen on edessä muutosten visualisointi. Muutostyyppilistojen muutokset käydään tyyppittäin ja prioriteeteittain läpi ja näiden muutokset ohjataan visualisointiluokille toteutettaviksi.

Tulostetiedostot

MUTU-ohjelma tuottaa ajossaan kolme tiedostoa:

- **päivitetty erikois-yleisontologia-tiedosto**
RDF/XML-muotoinen tiedosto, joka on vastaava kuin syöttötiedoston erikois- ja-yleisontologia-tiedosto. Tämä tiedosto eroaa syötetystä tiedostosta siten, että yhteydet yleisontologiasta poistettuihin käsitteisiin on poistettu.
- **yleisontologia merkityillä muutoksilla**
RDF/XML-muotoinen tiedosto, joka on vastaava kuin syötetty yleisontologia-tiedosto, mutta ontologian sisältää ylimääräisiä ryhmiä listaamaan ja ryhmittelemään muuttuneet käsitteet. Yksittäisen käsitteen tietää muuttuneen siitä, että sillä on `subClassOf`-suhde johonkin näistä ryhmistä.
- **muutoslistaus**
HTML-verkkosivu, joka listaa muutostyypeittäin ja prioriteeteittain muuttuneet käsitteet sekä näyttää, miten käsitteet ovat muuttuneet. Sisältää myös listan erikois- ja-yleisontologiasta poistetuista kolmikoista.

5.2 Muutostyypit ja muutosten tärkeys

MUTU pyrkii helpottamaan erikoisontologiakehittäjien päivitystyötä luokittelemalla yleisontologian muutokset sekä muutostavan että muutoksen tärkeyden mukaan. Muutostavan mukaan erotellessa on tavoitteena ryhmitellä yhteen samalla tavoin muuttuneet käsitteet, sillä ne mitä luultavimmin aiheuttavat samankaltaisia muutoksia erikoisontologiaan. Muutosten tärkeyden perusteella lajittelu mahdollistaa sen, että ontologiakehittäjät kohdistavat mahdollisesti rajalliset kehitysresurssit ensin sellaisiin muutoksiin, joilla on suurin todennäköisyys aiheuttaa muutoksia erikoisontologiaan.

Muutosten prioriteeteilla on kaksi arvoa: *tärkeä* tai *muu*. Tärkeät muutokset ovat sellaisia, jotka tulisi käydä läpi ennen muita muutoksia. Tärkeät muutokset on kuvattu alla muutostyyppien kohdalla ja muut muutokset ovat sellaisia, jotka eivät sovi tärkeiden muutosten määritelmään.

MUTUssa on kaksi päämuutostyyppiä: muutokset käsitejoukossa sekä yksittäisen käsitteen muutokset. Kuten alta muutostyyppilistasta voi huomata, jo päämuutostyypit eroavat toisistaan huomattavasti. Tämän takia ei ole mahdollista määrittää yhtä tiettyä testiä tai mittaa tärkeydelle, vaan se on määritettävä muutostyypeittäin.

Seuraavaksi esittelen kaikki MUTUssa olevat muutostyypit sekä listaan, millaiset muutokset ovat kyseisessä muutostyyppissä korkeammalla prioriteetillä. Ennen MUTU-ajoa on voitu määrittää, minkä ominaisuuksien muutokset eivät milloinkaan kiinnosta erikoisontologian kehittäjää, joten nämä muutokset eivät edes ilmesty muutoslistauksiin.

MUTUn muutostyypit ovat:

- käsitejoukon muutokset
 - lisätyt käsitteet
 - poistetut käsitteet
 - samannimiset käsitteet
- yksittäisen käsitteen muutokset
 - hierarkiamuutokset
 - nimikemuutokset
 - assosiatiivisuusmuutokset
 - osa-kokonaissuhdemuutokset
 - muut muutokset

Käsitejoukon muutokset

Käsitejoukon muutokset sisältävät kokonaisten käsitteiden lisäykset tai poistot sekä näihin liittyvät muut listaukset.

Lisätyt käsitteet

Muutos: Käsitteet, jotka on lisätty yleisontologiaan sen vanhan ja uuden version välillä. Käsite löytyy uudesta versiosta, muttei vanhasta.

Tärkeitä: Lisätyt käsitteet, joiden välillisellä tai välittömällä alakäsitteellä on yhteys erikoisontologian käsitteeseen. Erikoisontologian kehittäjän tulee tarkastella, onko koko hierarkiaketju linkitetystä erikoisontologiakäsitteestä ontologian juureen yhä oikein.

Poistetut käsitteet

Muutos: Käsitteet, jotka on poistettu yleisontologiasta vanhan ja uuden version välillä, eli käsite löytyy vanhasta versiosta, muttei uudesta.

Tärkeitä: Kaikki poistuneet käsitteet ovat tärkeitä, sillä käsitteet tulisi haudata tai arkistoida ontologiasta poistamisen sijaan. Täten kaikki poistetut käsitteet tulisi tarkistaa.

Samannimiset käsitteet

Muutos: Ei varsinainen muutos, vaan käsitteet, joilla on samoja nimikkeitä. Kaikkia nimikkeitä verrataan keskenään riippumatta nimikkeen ominaisuudesta. Täten jos toisessa käsitteessä nimike ”home” on `prefLabel`-ominaisuudessa ja toisessa `altLabel`-ominaisuudessa, ehdotetaan näitä kahta käsitettä samoiksi käsitteiksi. Lisäksi nimikkeet muunnetaan tarkastelua varten perusmuotoon ja nimikkeitä verrataan vain kielittäin.

Tärkeitä: Yleisontologiaan käsitteet, joilla on sama nimikkeen perusmuoto kuin joillakin erikoisontologian käsitteellä, mutta näiden välillä ei ole ekvivalenssiyhteyttä. Nämä käsitteet voivat olla sama käsite, jolloin näiden välille tulisi joko lisätä ekvivalenssiyhteys tai poistaa päällekkäinen käsite erikoisontologiasta.

Käsitteiden muutokset

Käsitteiden muutokset ovat ominaisuuksien arvojen lisäämisiä, poistoja tai korvauksia käsitteillä, jotka esiintyvät sekä yleisontologian vanhassa että uudessa versiossa.

Hierarkiamuutokset

Muutos: Käsitteen yläkäsite on lisätty, poistettu tai korvattu toisella käsitteellä.

Tärkeitä: Muutos käsitteessä, jonka välillisellä tai välittömällä alakäsitteellä on hierarkkinen yhteys erikoisontologian käsitteeseen. Tällaisessa tapauksessa on syytä tarkistaa onko hierarkiaketju vielä hyväksyttävä erikoisontologian puolella.

Nimikemuutokset

Muutos: Käsitteeseen on lisätty, poistettu tai korvattu oletusnimike (`prefLabel`) tai vaihtoehtoisia nimikkeitä (`altLabel`). Muutoksia vertaillaan kieltäin.

Tärkeitä: Oletusnimike (`prefLabel`) korvattu käsitteessä, jolla on ekvivalenssiyhteys johonkin erikoisontologian käsitteeseen. Tässä tapauksessa hyvä tarkistaa, tarvitseeko myös erikoisontologian käsitteen nimikettä päivittää.

Assosiatiivisuusmuutokset

Muutos: Käsitteeseen on lisätty, poistettu tai korvattu assosiatiivinen suhde toiseen käsitteeseen.

Tärkeitä: Muutos käsitteessä, jolla on ekvivalenssiyhteys johonkin erikoisontologian käsitteeseen. On hyvä tarkistaa tarvitseeko myös erikoisontologian käsitteen assosiatiivisia suhteita päivittää.

Osa-kokonaissuhdemuutokset

Vastaava kuin hierarkiamuutos, mutta `partOf`-ominaisuudelle.

Muut muutokset

Muut muutokset ovat muutoksia muissa kuin yllä kuvatuissa ominaisuuksissa. Näitä ei ole lajiteltu prioriteetin mukaan.

Yhteenvedona tärkeydestä voidaan tiivistää, että on tärkeää tarkistaa transitiivisten ominaisuusketjujen paikkansapitävyys lisätyillä käsitteillä, sekä hierarkia- ja osa-kokonaissuhteiden osalta. Muissa tapauksissa ekvivalenssikäsitteiden muutosten tarkastelu on riittävää.

5.3 Muutosten visualisointitavat

MUTU-työkalulla lajitellut muutokset visualisoidaan erikoisontologiakehittäjille kahdella eri tavalla: verkkosivulla esitettävänä muutoslistoina sekä yleisontologiaan

merkittyinä käsitteinä, jotka sisältävät `rdfs:subClassOf`-suhteen johonkin muutosryhmään.

Muutoslistaus

Muutoslistauksen perusajatuksena on listata aliluvussa 5.2 esiteltyt muutokset. Kuvakaappauksia muutoslistauksesta löytyy kuvista 5.3-5.6. Muutoslistojen tuottohetkellä tärkeitä muutoksia kutsuttiin vielä ”kiinnostaviksi muutoksiksi”.

Muutoslistauksen tavoite on esittää erikoisontologiakehittäjälle selkeästi ja ymmärrettävästi yleisontologian muutokset muutostyypeittäin ja prioriteeteittain. Muutokset esitetään verkkosivuna, josta voi avata ja sulkea muutoslistoja tarpeen mukaan (kuva 5.2). Listan nimen vieressä on suluissa numero, joka kuvaa kuinka monta käsitettä on muuttunut kyseisellä tavalla. Tämän otsikkorivin alla on vielä kuvaus siitä, millaisia kyseiset muutokset ovat ja tärkeissä muutoksissa vielä sanallinen kuvaus siitä, millaiset muutokset ovat luokiteltu tärkeiksi.

Käsitteistä esitetään aina vähintään oletusnimike ja sen perässä suluissa käsitteen URI-tunniste lyhennytyssä muodossa. URI-tunnisteet toimivat linkkeinä ja ne ohjaavat ontologiaselaimen, josta pääsee tarkastelemaan käsitteen tietoja.

Seuraavaksi esittelen sanallisesti, miten eri muutostyyppit on verkkosivulla esitetty sekä missä kuvakaappauksessa ominaisuus on esillä.

Lisätyt käsitteet	Esitellään itse lisätyn käsitteen lisäksi ne lisätyn käsitteeseen lähimmät välilliset tai välittömät alakäsitteet, joilla on yhteys erikoisontologiaan. (Kuva 5.3)
Poistettut käsitteet	Jokaisen poistetun käsitteen vierestä pystyy laajentamaan näkyville kolmikot, jotka on poistettu ontologian konsistenttiuden takaamiseksi. (Kuva 5.4)
Samannimiset käsitteet	Löydettyjen käsitteiden listassa on ensin listattu nimikkeen perusmuoto, jonka perusteella käsitteet on tunnistettu samoiksi. (Kuva 5.5)
Yksittäisen käsitteen muutokset	Käsitteen kohdalla on aina listattu käsitteen kaikki muutokset, mutta kyseisen muutostyyppin muutokset on korostettu. Tämä mahdollistaa yleiskuvan saamisen käsitteen muutoksista, sillä käsitteen eri ominaisuuksien muutokset vaikuttavat toisiinsa. (Kuva 5.6)
Nimikemuutokset	Muutokset on eroteltu kielen mukaan. Tämän lisäksi erikoisontologian kehittäjän työtä on pyritty helpottamaan siten, että kielet on järjestetty prioriteettijärjestykseen erikoisontologiassa käytettyjen kielten mukaan. Koska erikoisontologiakehittäjä korjaa yleensä koko käsitteen kerralla, olisi turhaa näyttää sama muutos moneen kertaan alempien prioriteettisten kielten listoissa. Täten listojen sisältävät muutokset ovat aina sellaisia, joita ei ole aikaisemmissa kielissä esitelty. (Kuva 5.6)
Assosiatiiiviset muutokset	Muista muutostyypeistä poiketen, assosiatiiiviset muutokset on jaoteltu edelleen lisäyksiin, poistoihin ja korvauksiin. Ontologiakehittäjien toive oli pitää assosiatiiiviset ominaisuudet erillään, sillä ne ovat ylimääräistä tietoa. (Muutokset on esitetty vastaavasti kuin kuvassa 5.5)

MUTU - Muutoslistaus

[Evaluatiotaulukot](#)
[Kirjoituslusta](#)

Tämä tiedosto listaa YSOon päivityksessä tehdyt muutokset.

Osasta muutostyyppistä on eroteltu muutokset, jotka mahdollisesti johtavat erityisontologian päivittämiseen.

Luotu: 2012/07/18 09:59:53

Lisätyt käsitteet (5276)

YSOon päivityksessä lisätyt käsitteet.

[-]

Kiinnostavat (31)

YSOon lisätyt käsitteet, joiden jokin välitön tai välillinen alakäsite on yhteydessä erikoisontologiaan. Nämä alakäsitteet on listattu YSOon lisätyn käsitteen alla.

[+]

Muut (5245)

[+]

Poistetut käsitteet (21)

YSOon käsitteet, joita ei ole enää päivitettyssä YSOon versiossa.

Käsitteen vieressä olevaa [+] -painiketta painamalla näkee mitkä poistettuun käsitteeseen viittaavat ominaisuudet on automaattisesti poistettu erikoisontologiasta.

[+]

Samannimiset käsitteet (633)

YSOon sekä erikoisontologian käsitteellä on jokin sama nimitys, mutta käsitteiden välille ei ole merkitty ekvivalenssia.

[+]

Hierarkiamuutokset (806)

Käsitteet, joiden yläkäsite on muuttunut.

Kuva 5.2: Kuvakaappaus MUTU-työkalun tuottamasta muutoslistauksesta. Lisätyt käsitteet -kategoria on avattu, jotta nähdään tärkeiden (kuvassa ”Kiinnostavat”) ja muiden muutosten määrät.

- 4. aiotumaiset ([yso:p20765](#))
 - eläimet ([yso:p2023](#))
 - sienet ([yso:p90](#))
 - kasvit ([yso:p1755](#))
 - alkueliöt ([yso:p21312](#))
 - limasienet ([yso:p7909](#))
- 5. B-vitamiinit ([yso:p25125](#))
 - foolihappo ([yso:p4751](#))
 - B6-vitamiini ([yso:p12869](#))
- 6. eläinten ruoat ([yso:p21278](#))
 - rehut ([yso:p3267](#))

Kuva 5.3: Kuvakaappaus tärkeiksi luokitelluista lisätyistä muutoksista. Ensimmäisellä rivillä on lisätty käsite ja sen alla on listattu sen lähimmät alakäsitteet, joilla on yhteys erikoisontologiaan.

3. arvopapereidenvälittäjät ([yso:p13694](#)) [-]
 - arvopapereidenvälittäjät ([yso:p13694](#)) , type , - ([yso-meta:Concept](#))
 - arvopapereidenvälittäjät ([yso:p13694](#)) , overlaps , 1.0
 - arvopapereidenvälittäjät ([yso:p13694](#)) , prefLabel , arvopapereidenvälittäjät
 - arvopapereidenvälittäjät ([yso:p13694](#)) , overlappedBy , 1.0
 - arvopapereidenvälittäjät ([yso:p13694](#)) , definedConcept , - ([yso:Y11282](#))
 - arvopapereidenvälittäjät ([yso:p13694](#)) , subclassOf , ammatit ([yso:p1179](#))
 - arvopapereidenvälittäjät ([yso:p13694](#)) , semanticTag , arvopapereidenvälittäjät
4. asumiseen liittyvä objekti ([yso:p2098](#)) [+]
5. hirtvahingot ([yso:yso_hirtvahingot](#)) [+]

Kuva 5.4: Kuvakaappaus yleisontologiasta poistuneista käsitteistä. ”arvopapereidenvälittäjät”-käsitteen alla on avattu käsitteen poiston takia yleis-jä-erikoisontologiatiedostosta poistetut kolmikot.

28. charolais
 - charolais ([afo:charolais](#))
 - charolais ([yso:p21879](#))
29. deforestation
 - metsien häviäminen ([afo:metsien_häviäminen](#))
 - metsäkato ([yso:p23052](#))
30. depreciation
 - poisto ([afo:poisto](#))
 - arvonalasku ([yso:p22007](#))
31. developed country
 - teollisuusmaat ([afo:teollisuusmaat](#))
 - kehittyneet maat ([yso:p21596](#))

Kuva 5.5: Kuvakaappaus samannimisiksi tunnistetuista käsitteistä. Ensimmäisellä rivillä on perusmuoto, jonka perusteella käsitteet on luokiteltu samaksi käsitteeksi. Tämän alla ovat käsitteiden oletusnimikkeet sekä URI-tunnisteet.

1. aarnimetsät ([yso:p3622](#))
 - **prefLabel @fi**
aarnimetsät => aarniometsät
 - oldLabel @
--- => aarnimetsät
 - subclassOf
metsät ([yso:p5454](#)) => metsät iän mukaan ([yso:p14251](#))
--- => metsät hoidon mukaan ([yso:p5453](#))
 - altLabel @
aarnimetsät => ---
 - prefLabel @sv
--- => urskogar
 - associativeRelation
--- => sademetsät ([yso:p3487](#))
2. alkoholi ([yso:p13929](#))
 - **prefLabel @fi**
alkoholi => alkoholi (nautintoaineet)

Kuva 5.6: Tärkeiksi luokiteltuja nimikemuutoksia esimerkkinä käsitteen muutoksien esittämisestä. Kyseisen muutosluokan ominaisuus on korostettu.

Ontologiaan merkityt muutokset

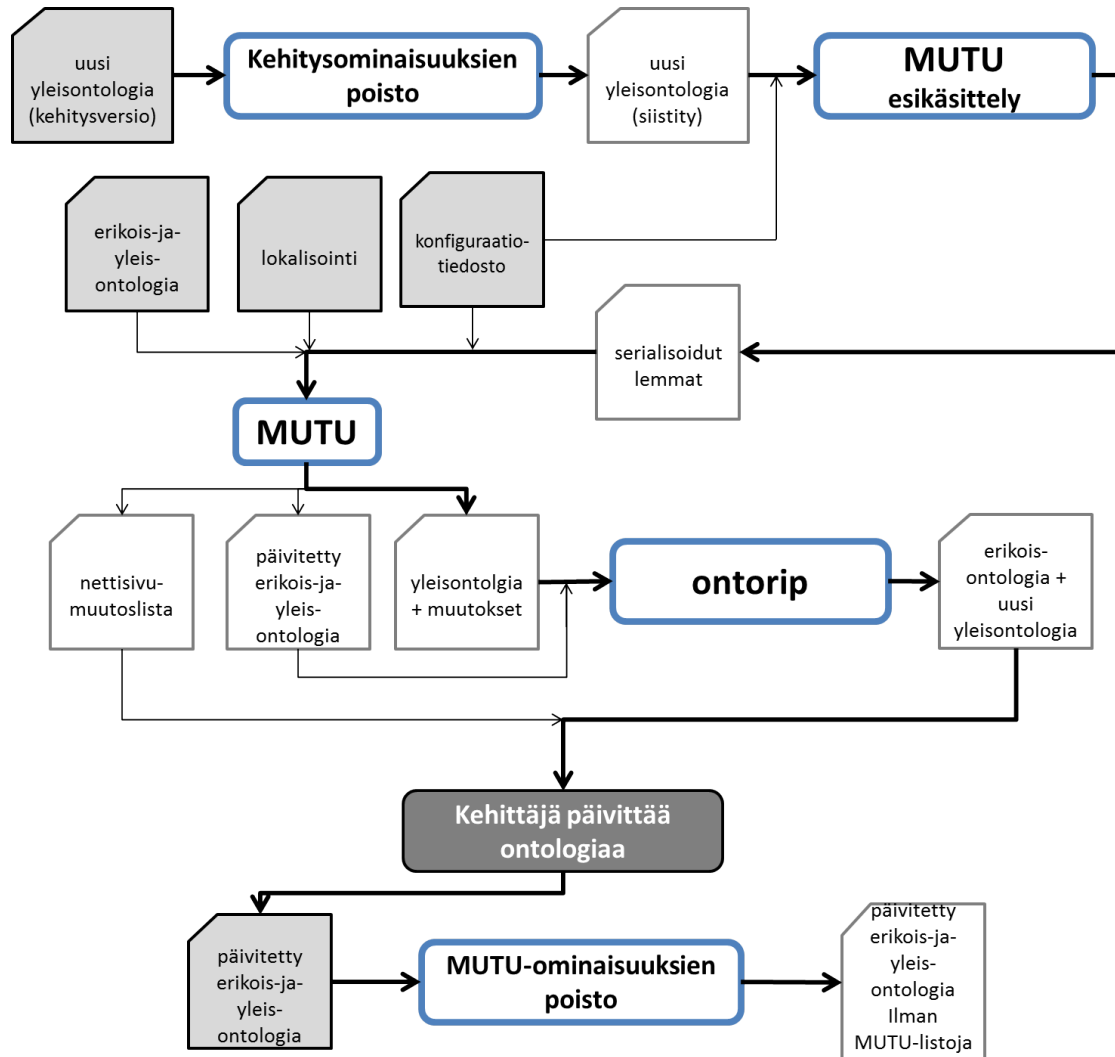
Erikoisontologian kehitystiedostoon lisättävät muutokset on tarkoitettu tukemaan muutoslistauksen käyttöä. Ontologiaan luotiin ylimääräisiä käsitteitä edustamaan eri muutostyyppisiä ja näiden tärkeyksiä. Yleisontologian muuttuneet käsitteet linkitettiin sitten näihin eri muutostyyppeihin `subClassOf`-ominaisuuden avulla. Ajatuksena oli, että erikoisontologiakehittäjä voisi ontologiaa muokatessaan nähdä myös ontologiaeditorissa muuttuneen käsitteen muutostyypin sekä listan muuttuneista käsitteistä muutostyypeittäin. Tätä muutostyypilistaa voisi hyödyntää päivityksen edistymisen seuraamiseen. Nämä muutostyypilistat voisi toteuttaa myös ontologiaeditorissa tehtävillä SPARQL-kyselyillä, mutta kaikkissa ontologiaeditoreissa ei ole tuke SPARQL-kyselykielelle. Tämän takia muutostyypit piti lisätä fyysisesti ontologiatiedostoon, eikä niitä voinut tuottaa ja päivittää automaattisesti ontologiaa muokatessa. On myös huomioitava, että ontologiaan lisättävän muutosmerkinnän avulla nähdään vain että käsite on muuttunut tietyllä tavalla, eikä miten itse arvo on muuttunut.

Muutostyypilistojen toteutus tehtiin noudatellen muutoslistausta, eli jokaiselle muutostyypille ja muutoksen prioriteetille luotiin oma käsite, esimerkiksi käsite `mutu:lisatyt_kiinnostavat`. Kyseisellä tavalla muuttuneet käsitteet yhdistettiin kyseiseen muutosluokkaan `rdfs:subClassOf`-ominaisuutta käyttäen. Samannimisissä käsitteissä luotiin ylimääräinen rakennekäsite yhdistämään samannimisten käsitteiden joukkoa. Tästä esimerkkinä on `mutu-temp:i237`-rakennekäsite, johon on linkitetty ominaisuudella `rdfs:subClassOf` yleisontologian ja erikoisontologian ”re-huherne”-käsitteet. `mutu-temp:i237`-rakennekäsite taas on alakäsite muutostyypikäsitteelle `mutu:samannimiset_kiinnostavat`.

5.4 MUTU päivityksen osana

Tässä aliluvussa esitellään koko prosessi, jolla tuotettiin luvun 6 MUTUn arvioinnissa käytetyt ontologiat (kuvassa 5.7). Kaavion harmaat tiedostot ovat MUTUn ja muiden ohjelmien tarvitsemat tiedostot. Valkoiset tiedostot kuvaavat ohjelmien tuottamia tulostetiedostoja. Paksuin nuoli edustaa datavirtaa ohjelmasta toiseen ja sitä seuraamalla näkee ohjelmien suoritusjärjestyksen sekä näiden tärkeimmän tulostetiedoston. Ohuemmat nuolet osoittavat ohjelmien tarvitsemat syöttötiedostot sekä muut tulostetiedostot.

Jos ontologian kehitysversio on sama kuin sen julkaisuversio, voidaan MUTU-työkaluun antaa suoraan ontologian kehitysversio. Tällainen tilanne ei ollut evaluatiossa käytetyissä yleis- ja erikoisontologioista, joiden kehitysversiot poikkesivat



Kuva 5.7: MUTU ja muut skriptit, joiden avulla tuotettiin tämän työn arvioinnissa käytetyt tiedostot. Harmaapohjaiset tiedostot ovat valmiina päivitykseen lähettäessä ja valkopohjaiset ovat skripteistä tuotettuja tiedostoja.

huomattavasti julkaisuversioista. Lisäksi evaluaation erikoisontologiatiedostot sisälsivät kopion vanhasta yleisontologiasta, mikä jouduttiin korvaamaan uusimalla yleisontologian versiolla MUTU-päivityksen aikana.

Esikäsittelyvaiheessa yleisontologian kehitysversiota muokattiin poistamalla siitä yleisontologian kehitykseen tarkoitettuja väliaikaisluokkia. Kehitysversiota käytettiin muutosten tunnistamisen pohjana, sillä se oli teknisesti lähimpänä erikoisontologioiden kehitysmuotoa. Yleisontologian sopivaan muotoon muokkaamisen jälkeen tehtiin MUTUn etukäteisajo ontologian nimikkeiden perusmuotoistamiseen. Koska verkkoyhteydellä perusmuotojen pyytäminen lemmatisointiserveriltä koko yleisontologialle kesti kymmeniä minutteja, perusmuodot kyseltiin ja tallennettiin ennen varsinaisen MUTU-ajoa. Erikoisontologiat olivat kooltaan huomattavasti yleisontologiaa pienempiä, joten näiden käsitteiden perusmuotoistaminen pystyttiin tekemään MUTU-ajon aikana. Ajatuksena oli myös, että käytettäessä samaa yleisontologian versiota useassa erikoisontologian päivitykseen, aikaavievä yleisontologian perusmuotoistaminen tarvitse tehdä vain kerran. Yleisontologian esikäsittelyn ja nimikkeiden perusmuotoistamisen jälkeen suoritettiin varsinainen MUTU-ajo, joka on esitelty tarkemmin luvussa 5.1.

MUTU-ajon jälkeen erikois-yleisontologia-tiedoston vanha yleisontologia-versio korvattiin MUTU-ajoon käytetyllä uudella yleisontologian versiolla. Tähän korvaamiseen käytettiin Osma Suomisen ohjelmoimaa *ontorip*-työkalua¹. Yleisontologian vaihtoprosessin jälkeen erikois-yleisontologiatiedosto oli valmis erikoisontologiakehittäjän tarkasteltavaksi.

Lopuksi, kun erikoisontologiakehittäjä oli saanut tehtyä tarvittavat muutokset erikoisontologiaan, erikois-yleisontologiatiedostosta poistettiin *mutu-* ja *mutu-temp-*nimiavaruuksissa olevat muutosmerkinnät sekä ylimääräiset luokat, jotka oltiin luotu muutoslistauksia varten.

¹<https://github.com/NatLibFi/Finto-data/tree/master/tools/ontorip>

Luku 6

MUTUn suoriutuminen

MUTU on ”proof of concept” -tyyppinen prototyyppi uudenaikaiseen kevytontologioiden muutosten esittämiseen. Tässä luvussa esitetään, miten MUTUn muutosten lajittelualgoritmit toimivat ja miten ontologiakehittäjät suhtautuvat prototyyppiin. Varsinainen käytettävyydestutkimus on ajankohtainen vasta myöhemmässä vaiheessa, kun työkalun peruseriaatteet ovat kypsyneet.

MUTUn arvioinnissa kolme erikoisontologiakehittäjää päivitti omia erikoisontologioitaan MUTUn muutoslistauksia käyttäen. Arviointi perustuu heidän tekemiin muutoksiinsa sekä heidän haastatteluihinsa päivittämisen jälkeen.

6.1 Testikäyttäjät ja erikoisontologiat

MUTUn muutoslistojen testaamiseen osallistui yhteensä neljä erikoisontologiakehittäjää, joista kolme kävi läpi testiin kuuluvat muutokset. Kehittäjien ontologiakehityskokemus vaihteli noin vuodesta viiteen vuoteen, ja kaikkien kehittäjien ontologiakehitys on ollut enemmän tai vähemmän katkonaista. Heidän ylläpitämiään erikoisontologioita kehitettiin käyttäen Protégé-ontologiaeditorin¹ versioita 3.3.1 tai 3.4.4.

Testikehittäjien erikoisontologiat olivat kaikki eri aloilta ja he olivat ontologioidensa ainoita kehittäjiä. Erikoisontologiat oli luotu olemassa olevien tesaurusten pohjalta 2-5 vuotta sitten ja niissä esiintyy termejä suomeksi, ja joissakin lisäksi ruotsiksi ja englanniksi. Ontologiat sisälsivät muutamia tuhansia käsitteitä. Kaikki erikoisontologiat oli ripustettu samaan yleisontologiaan, Yleiseen suomalaiseen ontologiaan (YSO) [54], joka sisältää lähes 30 000 käsitettä. Yleisontologia oli fyy-

¹<http://protege.stanford.edu/>

sisesti kopioitu erikoisontologiaan kehitystä helpottamaan [64].

Mitään näistä erikoisontologioista ei ollut aikaisemmin päivitetty vastaamaan yleisontologian muutoksiin. Jokaisella kehittäjällä oli edessään tuhansien muutosten läpikäyminen, sillä YSO itsessään laajenee suunnilleen tuhannella käsitteellä vuodessa, minkä lisäksi myös olemassa olevia käsitteitä muokataan.

Testikäyttäjät olivat taustaltaan kirjastoammattilaisia eikä heillä ollut teknistä koulutusta. Aloittaessaan erikoisontologian kehittämisen, heille oli annettu aikaisemmin koulutusta ontologiaeditorin käyttöön sekä ontologioiden rakentamisen periaatteisiin.

Testausasetelma

MUTUN luomia muutoslistauksia testattiin todellisessa erikoisontologioiden päivitystilanteessa. Erikoisontologiakehittäjät päivittivät omia erikoisontologioitaan yleisontologian muutosten perusteella hyödyntäen MUTU-työkalun tuottamia muutoslistoja.

Erikoisontologiakehittäjille valittiin läpikäytäväksi 25 muutosta jokaisesta muutostyyppistä. Muutokset valittiin aakkosjärjestyksessä ensimmäisistä muutoksista. Mikäli muutoksia ei ollut niin paljoa, otettiin kaikki muutokset, mitä muutostyyppissä oli.

Testausprosessi koostui kolmesta osasta: aloitustapaamisesta, kehittäjän itsenäisestä päivittämisestä sekä lopputapaamisesta.

Aloitustapaaminen

Aloitustapaamisessa erikoisontologiakehittäjälle annettiin merkityt muutokset sisältävä erikois- ja yleisontologiatiedosto, muutoslista sekä suunnilleen kymmensivuinen ohjelehtinen, joka sisälsi kuvaukset muutostyypeistä ja muutosten tärkeydestä. Lisäksi ohjeessa oli tietoa evaluointitaulukoiden ja kirjoituslustan käytöstä sekä ohjeet miten toimia, kun on saanut omat testausmuutoksensa tehtyä. Lopuksi kehittäjän kanssa käytiin vielä muutama muutos läpi, jotta listojen käyttö konkretisoitui. Aloitustapaaminen kesti noin tunnin ja se tehtiin kehittäjän omaa tietokonetta käyttäen.

Itsenäinen päivittäminen

Aloitustapaamisen jälkeen erikoisontologiakehittäjä alkoi päivittää ontologiaansa yleisontologian muutoslistojen avulla. Erikoisontologioita ei oltu vuosiin päivitetty vastaamaan yleisontologiaa, joten muutoslistaukset sisälsivät tuhansia muutoksia.

Testissä läpikäytävien muutosten määrä vaihteli 404 muutoksesta 523 muutokseen. Tarkat läpikäytyt muutosmäärät on listattu liitteessä C kehittäjittäin kpl-sarakkeessa. Läpikäytävien muutosten määrä vaihteli ontologioittain, sillä erikoisontologiat olivat kiinni YSON eri versioissa. Luonnollisesti myös tärkeiksi luokiteltujen muutosten määrät vaihtelivat, sillä erikoisontologiat olivat linkittyneet eri kohtiin yleisontologiaa.

Käydessään muutoksia läpi ja päivittäessään ontologiaansa erikoisontologiakehittäjät lisäksi merkitsivät jokaisen yleisontologian muutoksen kohdalla oliko kyseinen muutos kiinnostava vai ei. Mikäli kehittäjillä tuli päivityksen aikana kysyttävää, he pystyivät olemaan yhteydessä testin järjestäjään kirjoituslunastan kautta tai sähköpostitse.

Lopputapaaminen

Kun erikoisontologian kehittäjä oli saanut käytyä hänelle määrättyt muutokset läpi katsottua, päätettiin testaus kyselylomakkeella ja haastattelulla. Kyselylomake ja haastattelurunko löytyvät litteistä A ja B. Lopputapaamisen jälkeen kehittäjien ei tarvinnut enää päivittää ontologiaansa, mutta yksi kehittäjistä halusi jatkaa päivittämistä.

Prosessi aloitustapaamisesta lopputapaamiseen kesti kaikilla osallistujilla useita kuukausia. Kaksi kehittäjää aloitti kesä-heinäkuussa vuonna 2012 ja heidän lopputapaamisensa olivat syys-lokakuussa samana vuonna. Kolmas päivityksen läpi tehnyt kehittäjä aloitti lokakuussa 2012 ja hänen lopputapaamisensa oli tammikuussa 2013.

Arviointidata

MUTUN muutoslistojen käytön arvioinnista kerättiin jokaiselta kehittäjältä seuraavaa dataa: päivitetty erikois-yleisontologiatiedosto, kiinnostavien muutosten taulukko, arviointilomake sekä haastattelunauhoite.

Päivitetty erikois-ja-yleisontologiatiedosto

Päivitetty erikois-ja-yleisontologiatiedosto sisälsi ne muutokset, joita käyttäjä oli tehnyt käydessään muutoslistoja läpi. Vertaamalla tätä tiedoston tilaan ennen päivitystä, saatiin lista erikoisontologiaan tehdyistä muutoksista.

Kiinnostavien muutosten taulukko Erikoisontologiakehittäjät listasivat erilliseen Google Sheets -taulukkoon mitkä muutokset he haluaisivat vastaisuudessaakin nähdä muutoslistauksessa, eli mitkä muutokset olivat heille ”kiinnostavia”. Kehittäjät kopioivat taulukkoon käsitteen URI-tunnisteen joko

kiinnostavat- tai *muu*-kategoriaan. Lisäksi taulukko sisälsi tiedon siitä, minkä numeroiset muutokset kehittäjän tulisi käydä läpi muutoslistauksestaan. Eri muutostyypit oli erotettu taulukon eri välilehtiin.

Kiinnostavien muutosten merkitsemisen lisäksi kehittäjien tuli merkitä, mikä yleisontologian käsitteiden muutokset johtivat muutoksiin erikoisontologiassa. He merkitsivät muuttuneen käsitteen URI-tunnisteen sekä sanallisen kuvailun muutostaulukkoon. Selitteet eivät olleet valmiiksi määriteltyjä, vaan kehittäjät saivat kuvailla muutoksia parhaaksi katsomillaan lauseilla.

Arviointilomake

Lopputapaamisen alussa täytetty arviointilomake sisälsi erilaisia kysymyksiä MUTUN muutoslistausten käytöstä. Kysymysten suunnittelussa oltiin otettu vaikutteita muun muassa System Usability Scale -arviointimenetelmästä [9] ja se sisälsi sekä positiivisia että negatiivisia väittämiä. Tässä työssä käytetty arviointilomake löytyy liitteestä A.

Haastattelunauhoite

Loppuhaastattelun kysymykset ja keskustelun aiheet kartoittivat muun muassa muutoslistojen käyttöä, tukimateriaaleja, päivitysprosessia sekä kehittäjän suhtautumista muutoslistoihin päivityksen tukena. Haastattelut nauhoitettiin kehittäjien luvalla.

Lisäksi kehittäjät pystyivät kommunikoimaan testin järjestäjän kanssa myös kirjoitusalan kautta, mutta näitä chat-keskusteluja ei ole käytetty tässä työssä arviointidatana.

6.2 Tulokset

Seuraavaksi käydään läpi eri arviointidatan lähteet. Ensin käydään läpi millaiset yleisontologian muutokset erikoisontologiakehittäjät kokivat kiinnostaviksi ja muutoksia aiheuttaviksi. Tämän jälkeen tarkastellaan, millaisia muutoksia erikoisontologiakehittäjät tekivät eri muutostyypeissä. Lopuksi käydään läpi haastattelut sekä muut arviointijakson aikana tulleet huomiot.

Muutostaulukot

Erikoisontologiakehittäjät merkitsivät taulukkoon, mitkä muutokset he kokivat *kiinnostaviksi* eli haluaisivat nähdä kehittäessään, sekä mitkä yleisontologian muutokset aiheuttivat muutoksia erikoisontologiaan. Kiinnostavien muutosten tauluk-

ko löytyy liitteestä C ja luokitellut muutoksia aiheuttaneet yleisontologian muutokset taulukossa D.

Kiinnostavien ja muutosmäärien taulukoissa riveinä ovat muutostyypit lyhennetyin nimin. Muutostyyppin alla t-kirjain vastaa tärkeää muutosta ja m-kirjain muita muutoksia. Lisäksi nimikemuutoksilla on kielikoodit ja assosiatiivisilla muutoksilla lisäys-poisto-korvaus-lajittelu. Taulukon sarakkeissa on testiin osallistuneet ontologiakehittäjät K1-K3. ”kiin” tai ”muutoksia” ilmoittaa kiinnostavaksi merkittyjen tai muutoksia aiheuttavaksi merkittyjen yleisontologiamuutosten määrän. Suluisassa oleva numero esittää, kuinka monta prosenttia tämä on kyseisen muutostyyppin läpikäydystä kokonaismäärästä, joka on esitetty sarakkeessa ”kpl”. Lukua ei ole ilmoitettu tapauksissa, kun määrä on 100% tai 0%. Taulukon alla on vielä yhteenlaskettu muutosmäärä kustakin sarakkeesta.

Kehittäjiä kiinnostavat muutokset

Kiinnostavien muutosten taulukkoon on laskettu yhteen kehittäjien kiinnostaviksi merkityt muutokset sekä yleisontologian muutokset, jotka aiheuttivat muutoksia erikoisontologiassa. Tämä johtuu siitä, että luonnollisesti myös muutokseen johtaneet käsitteet ovat kehittäjien näkökulmasta kiinnostavia.

Kehittäjä K3 ilmoitti haastattelussa, että hän laittoi kiinnostavien muutosten listaan kaikki läpikäymänsä käsitteet, joten taulukkoa voidaan hyödyntää vain kehittäjien K1 ja K2 osalta. Kuitenkin kehittäjien K1 ja K2 merkinnöistään saa käsityksen, että kehittäjät toivoisivat näkevänsä lähes kaikki muutokset, minkä he mainitsivat myös haastatteluissa. Lisäksi K1 ei toivonut näkevänsä ruotsinkielisten nimikkeiden muutoksia, sillä hänen erikoisontologiassaan ei käytetä tätä kieltä. Kehittäjä K2 mainitsi haastattelussa, että hän oli unohtanut käydä poistuneet muutokset läpi, minkä vuoksi näistä ei ollut yhtään merkitty kiinnostaviksi.

Kuten kiinnostavien muutosten taulukosta voi huomata, joitakin muutoksia on merkitty kiinnostavaksi enemmän kuin kyseisessä muutostyypissä oli läpikäytäviä muutoksia. Testitaulukoihin ei merkitty selvästi läpikäytäviä käsitteitä, joten inhimillisiä virheitä tapahtui. Koska taulukoilla ja erikoisontologioiden muokkauksilla ei ollut suoraa yhteyttä, ei näitä ylimääräisiä merkkauksia pystytty erottamaan muokatuihin erikoisontologioihin, joten nämä poikkeukset esiintyvät datassa.

Yhteenvetona, erikoisontologian kehittäjät haluavat mahdollisuuden nähdä kaikki ontologian muutokset, eivätkä vain tärkeiksi luokiteltuja muutoksia. He ovat kiinnostuneet tapahtuneista muutoksista sekä, kuten K2 mainitsi loppuhaastattelussa, myös ei-tärkeiksi merkityt yleisontologian muutokset voivat aiheuttaa muutoksia erikoisontologiaan.

Muutosmäärät

Muutosmäärätaulukossa, liitteessä D, on listattu kuinka monta yleisontologian muutosta aiheutti muutoksen erikoisontologiaan kussakin muutostyyppissä. Yleiskuva taulukosta on, että muutoksia aiheuttivat lisätyt käsitteet, samannimiset käsitteet, hierarkiamuutokset sekä asiasanamuutokset. Taulukkoon 6.1 on laskettu muutosten tarkkuudet, saannit ja F-mitat. *tp*-sarake (engl. *true positive*) sisältää määrän tärkeiksi luokitelluista muutoksista, jotka aiheuttivat muutoksen erikoisontologiassa. *fn*-sarake (engl. *false negative*) sisältää erikoisontologian muutoksiin johtaneet muutokset, jotka oltiin luokiteltu ei-tärkeiksi. *fp*-sarake (engl. *false positive*) ilmoittaa kuinka monta tärkeäksi luokiteltua muutosta ei aiheuttanut muutoksia erikoisontologiaan.

	tp	fn	fp	tarkkuus	saanti	F-mitta
käsitejoukon muutokset	70	18	112	0.38	0.8	0.52
yksit. käsitteen muutokset	41	1	419	0.09	0.98	0.16
kaikki muutokset	111	19	531	0.17	0.85	0.29

Taulukko 6.1: Muutosmäärien tarkkuudet, saannit ja F-mitat

Arviointitaulukosta voidaan huomata, että varsinkin yksittäisten käsitteiden muutoksilla saanti on lähes täydellinen. Vain yksi erikoisontologian muutokseen johtaneista muutoksista luokiteltiin väärin. Tämä virhe johtui siitä, että kyseisen nimikkeiden kielimääre oli virheellinen. Saanti on korkeahko myös käsitejoukon muutoksilla.

Toisin kuin saanti, luokittelun tarkkuus on hyvin huono varsinkin yksittäisen käsitteen muutoksilla. Täten huonon tarkkuuden takia myös käsitteiden luokittelun F-mitat ovat alhaisia. Luonnollisesti tarkkuuden parantaminen vähentäisi ontologiakehittäjien turhaan läpikäymiä muutoksia, mutta tässä sovelluskohteessa on tärkeämpää, että väärin negatiivisten määrä on mahdollisimman alhainen väärin positiivisten kustannuksella. Tämä tarkoittaa, että mahdollisimman paljon erikoisontologiaan oikeasti vaikuttavista yleisontologian muutoksista tulisi luokitella tärkeiksi muutoksiksi.

Käsittemuutosten lisättyjen ja samannimisten käsitteiden huonoille lajittelutuloksille on olemassa intuitiivinen syy: lisätyillä ja samoilla käsitteillä ei ole kattavia yhteyksiä erikoisontologiaan, joten ontologioiden välisten linkkien hyödyntäminen ei toimi yhtä hyvin, kuin muuttuneilla käsitteillä. Samannimisten käsitteiden muuttokategoria, jossa listataan yhtäläisyyksiä erikoisontologian käsitteiden ja yleisontologiassa jo aikaisemminkin olleiden käsitteiden välillä, ei esitä varsinaista uutta muutostietoa, vaan puutteita erikois- ja yleisontologian välisissä linkeissä. Tällaisten ongelmien löytäminen kuuluisi varsinaisesti ontologian laadunvalvontaan

eikä muutostyökaluun. Lisäksi lisätyissä käsitteissä ja tärkeissä samannimisissä käsitteissä on paljon päällekkäisyyksiä. Tämän takia on hankala tietää mitkä varsinaiset muutosmäärät ovat näissä kahdessa kategoriassa, koska tärkeiden samannimisten käsitteiden tunnistukseen osallistuvat yleisontologiasta vain lisätyt käsitteet. Nämä olivat kuitenkin suurin syy erikoisontologian muuttumiseen, joten samojen käsitteiden tunnistusta tullaan jatkossakin tarvitsemaan päivityksissä.

Kehittäjien tekemät muutokset

Kehittäjien tekemät muutokset on listattu liitteeseen E. Riveinä ovat MUTU-muutoslistauksen muutostyypit. Sarakkeet kuvaavat, millaisia muutoksia erikoisontologiaan tehtiin. Taulukko on koostettu erikoisontologiakehittäjien tekemien muutosten sanallisista kuvauksista.

Kehittäjien tekemiä muutoksia tarkastellessa nähdään, että kaikki ontologiakehittäjät lisäsivät erikoisontologian ja yleisontologian käsitteiden välille ekvivalenssiyhteyksiä sekä tekivät käsitteiden termeihin muutoksia. Kehittäjät pääsääntöisesti joko linkittivät erikoisontologian käsitteet yleisontologiaan ilmestyneeseen käsitteeseen tai korjasivat käsitteen nimikkeen vastaamaan yleisontologian nimikettä.

Erikoisontologioiden kehityseriaatteet vaikuttivat siihen, millaisia muutoksia erikoisontologioihin tehtiin. Kehittäjät K2 ja K3 pyrkivät pitämään erikoisontologian kyseistä alaa edustavana käsittekokonaisuutena, joka voi sisältää myös alalle tärkeitä yleistermejä tai -käsitteitä. Nämä yleiskäsitteet on sitten linkitetty vastaaviin käsitteisiin yleisontologiassa. Kehittäjä K1 taas pyrki pitämään erikoisontologiansa tiukasti yleisontologian laajenuksena olettaen, että yleisontologia on aina käytössä yhdessä erikoisontologian kanssa. Tämän takia kehittäjä K1 poisti erikoisontologiastaan paljon liian yleiseksi kokemiaan käsitteitä, jotka esiintyivät jo yleisontologiassa.

Käsitejoukon muutokset näyttävät aiheuttavan erilaisia muutoksia kuin yksittäisen käsitteen muutokset. Käsitejoukon muutoksissa, viivan yläpuolella, muutoksia tosin aiheuttivat vain lisätyt ja samannimiset käsitteet, jotka muutenkin nivoutuivat tiiviisti yhteen. Nämä aiheuttivat pääasiassa ekvivalenssiyhteyksien lisäämistä. Käsitteen muutoksissa, viivan alapuolella, ei esiintynyt juurikaan ekvivalenssiyhteyksien lisäämistä, vaan itse käsitteen ominaisuuksien muutoksia.

Lisäksi kehittäjä K1 mainitsi, että yhdessä käsitteen poistoksi merkityssä muutoksessa hän ei tehnyt tätä varsinaisen yleisontologian käsitteen muutoksen takia, vaan selaillessaan erikoisontologiaa tämän muutoksen takia hän löysi liian yleiseksi kokemiaan käsitteitä. Päivityslistat voivat siis myös aiheuttaa muita ontologian ylläpitotoimenpiteitä, sillä ne saavat kehittäjän selailemaan ontologiaa tarkasti.

Myös ontologiakehittäjien tekemissä muutosten määrissä oli eroja. Kehittäjä K1 teki selvästi enemmän muutoksia kuin kehittäjät K2 ja K3. Tämä on tosin luonnollista, sillä K1-kehittäjän ontologia on ontologioista vanhin ja täten siihen tuli eniten muutoksia. Lisäksi vain K1-kehittäjän ontologiassa on käytössä osakokonaisuussuhteita ja K1 oli kehittäjistä ainoa, joka teki muutoksia assosiativisten tai poistettujen käsitteiden johdosta. Kehittäjä K3 ei tehnyt yhtään muutosta hierarkiamuutosten takia ja mainitsikin haastattelussa, että hänestä kyseinen muutoskategoria on turha.

Loppukysely

MUTU-työkalun arviointiin osallistui vain kolme kehittäjää, joten sen pohjalta ei voida tehdä juurikaan päätelmiä loppukyselydatasta. Lisäksi kehittäjien vastaukset asettuivat pääsääntöisesti likert-asteikon keskimaille, joten kehittäjien vastaukset olivat lähellä toisiaan. Esittelen seuraavaksi yhteenvedon kehittäjien vastauksista sekä kerron missä kysymyksissä käyttäjien vastaukset hajaantuivat eri ääripäihin.

Loppukyselyssä kehittäjät vastasivat, että muutoslistoja opi nopeasti hyödyntämään ja eniten aikaa päivityksessä meni ontologian selaamiseen Protégé-editorissa. Heidän mielestään muutoslistauksia ei voi lähteä käyttämään ilman ohjeistusta, mutta päivitystä edeltävä ohjeistus oli heidän mielestään riittävä. Kehittäjät K2 ja K3 pitivät muutostyyppisiä ja tärkeät-muut-jaottelua intuitiivisena, mutta K1-kehittäjällä oli ollut haasteita tärkeiden muutosten lajittelun ymmärtämisessä. K2-kehittäjän mielestä muut muutokset olivat välillä kiinnostavia, kun taas K1- ja K3-kehittäjät eivät kokeneet näitä kiinnostaviksi. Kehittäjät K2 ja K3 olivat sitä mieltä, että muutoslistaukset olivat helppokäyttöisiä, kun taas K1 oli tästä eri mieltä.

Loppuhaastattelu

Erikoisontologiakehittäjien antama yleiskuva muutoslistauksista oli positiivinen. Kehittäjät pitivät muutosten jaottelua muutostyyppisiin hyvänä; varsinkin samannimisten käsitteiden tunnistaminen oli tärkeäksi koettu ominaisuus. K3-kehittäjä oli sitä mieltä, että hän ei tarvitse muut-muutoksia ollenkaan ja tärkeäksi lajitelluista muutoksistakin vain asiasana-, samannimiset-muutokset sekä lisätyt käsitteet. Muut kehittäjät pitivät kyllä muut-muutoksia hyödyllisinä olla näkyvillä, mutta riippuen mahdollisesta päivitysjajasta sivuttaisivat muut muutokset. Heidän mielestään kuitenkin muut-osiossa voi olla myös erikoisontologiaan vaikuttavia yleisontologian muutoksia, joten ne kategoriat olisi hyvä myös käydä läpi, mikäli aikaa riittää. Kehittäjien mielestä ominaisuudet, joita ontologiassa ei ole käytössä

sekä kielet, joita ei ontologiassa aktiivisesti kehitetä voisi laittaa automaattisesti ”muut muutokset”-kategoriaan.

Kehittäjä K3:n mielestä muutoslistaukset olivat turhan monipuolisia ja K1 oli sitä mieltä, että tällaisessa massiivisessa tuhansien muutosten päivityksessä oli hyvä, että muutokset oli jaoteltu tarkemmin. Näin ei kuitenkaan tarvitsisi olla kun päivitys on tiheämpää, jolloin muutoksiakin on vähemmän.

Haastattelussa kävi ilmi, että kehittäjä K1 oli luullut, että kaikki muutoslistaukset perustuivat lisättyihin käsitteisiin. Hänelle oli myös jäänyt epäselväksi tärkeät-muut-muutosten listat. Kokonaisuutena kehittäjät olivat kuitenkin sitä mieltä, että tärkeät-muut-muutoslajittelu piti paikkansa. Jatkossa tulisikin kiinnittää enemmän huomiota muutosten lajitteluperiaatteiden kommunikoimiseen, sillä K2-kehittäjällä oli haasteita poistettujen käsitteiden kategorian hahmottamisessa. Lisäksi kehittäjät pitivät siitä, että he pystyivät näkemään käsitteen kaikki muutokset riippumatta siitä, mitä muutoskategoriaa he tarkastelivat. Myös poistettujen käsitteiden vaikutukset erikoisontologiaan jäivät useammalla kehittäjällä epäselväksi.

Muutoslistaus

MUTUn seuraavan version muutoslistaukseen tulisi haastattelujen perusteella tehdä muutamia selviä muutoksia. Pienetkin yksityiskohdat vaikuttavat listauksen käytettävyyteen oleellisesti.

Käyttäjä K1 kertoi, että hän ei käyttänyt muutoslistauksen käsitelinkkejä, sillä ne avautuivat ONKI-ontologiaselaimessa² eri näkymään kuin mitä hän yleensä käyttää. ONKIn korvanneessa sanasto- ja ontologiapalvelussa FINTOssa³ käsitteen katsomiseen ei ole useampia näkymiä, joten käsitteenäkymää ei tarvitse vaihtaa, kuten vanhassa ONKI:ssa olisi tarvinnut.

Muutoslistauksessa käsitteen tiedot avautuivat suoraan samaan välilehteen kuin missä itse listaus on. Käyttäjien aloitustapaamisen harjoitteluosiossa huomattiin, että kehittäjillä ei ollut rutiinia avata käsitelinkkejä uusiin välilehtiin. Useilla käyttäjillä oli ONKI samanaikaisesti auki käsitteiden selailua varten, mutta K3-kehittäjä mainitsi suoraan, että ei halua avata käsitteitä muutoslistauksesta, sillä silloin listaus jää useamman näkymän taakse. Tämä voitaisiin korjata sillä, että käsitetietosivut avautuisivat automaattisesti uuteen välilehteen.

Suurin osa käyttäjien ontologiaan tekemistä päivityksistä oli ekvivalenssimuutosten lisäyksiä. K3-kehittäjä ehdotti, että kiinnostaviksi luokiteltujen samannimis-

²<http://onki.fi/fi/browser/>

³<http://www.finto.fi/>

ten käsitteiden välille luotaisiin automaattisesti ekvivalenssiyhteys ja kehittäjä voisi käydä poistamassa ei-toivotut suhteet. Ontologiaeditoreissa ominaisuuksien arvojen poistaminen on yleensä nopeampaa kuin uuden arvon lisääminen, joten tämä säästäisi aikaa sekä vähentäisi mekaanista työtä.

Yleisontologiaan merkityt muutokset

Käyttäjille ei ollut selvää, mitä muutoslistauksiin liittyviä väliaikaiskäsitteitä ja ominaisuuksien arvoja he saisivat poistaa. Jo päivityksen aikana päädyttiin tulokseen, että käyttäjät voivat jättää muutoslistausominaisuudet ontologiaan ja ne poistettaisiin myöhemmin koneellisesti. Ontologiaan luodut muutostyyppikäsitteet eivät siis toimineet edistymisen seurantaan, kuten työkalua luodessa oli ajateltu.

Muutoslistojen läpikäyminen

Kehittäjät arvioivat, että muutoslistoja läpikäydessään he pystyivät tietämään 5-30 sekunnin sisällä muutoslistauksen käsitteen näkemisestä, onko käsite kiinnostava erikoisontologian kannalta. Mikäli muuttunut käsite oli kehittäjälle tuttu, hän pystyi tietämään kiinnostavuuden heti, mutta ennestään tuntemattomien käsitteiden kohdalla kiinnostavuuden tietäminen vaati käsitteen merkityksen selvittämistä internethauilla. Muutostarpeen kartoittamiseen kehittäjät arvioivat vievän 1-2 minuuttia. Tässä vaiheessa käyttäjät siirtyivät tarkastelemaan erikoisontologian käsitteiden tietoja itse ontologiaeditorissa.

Muutosten optimaaliseen läpikäymisjärjestykseen ei muutoslistojen arviointi tuonut lisää tietoa. Kaksi kehittäjistä kävi muutokset läpi verkkosivuna esitetyn muutoslistauksen mukaisessa järjestyksessä. Vain kehittäjä K3 kehittäjistä kävi muutostyyppit läpi omassa järjestyksessä. Hän pyrki tekemään mielestään helpoimmat muutostyyppit ensin:

1. samannimiset
2. nimikkeet
3. poistetut
4. lisätyt
5. hierarkia
6. assosiatiiivisuus

7. osasuhde
8. isättömät
9. muut

Kysyttäessä mahdollista muutosten läpikäyntijärjestystä, kehittäjät K1 ja K2 pitivät nimikkeiden ja hierarkian muutoksia tärkeimpinä, minkä jälkeen assosiatiivisia muutoksia. Heidän mielestään ensin tulisi käydä läpi kiinnostavat muutokset, minkä jälkeen ei-kiinnostavat, mikäli ylimääräistä aikaa on. K3-kehittäjän mielestä edes hierarkiamuutoksia ei tarvitsisi käydä läpi, vaan nekin tulisi käydä läpi, jos ylimääräistä aikaa on.

Koska yleisontologian läpikäytäviä muutoksia oli sen verran paljon, että muutoksia ei voinut tehdä kerralla, erikoisontologiakehittäjät pitivät kirjaa edistymisestään. Kehittäjät K1 ja K2 merkitsivät paperille, mitkä käsitteet ovat käyneet läpi. Kehittäjä K3 hyödynsi muutoslistauksen linkkien värejä selvittäessään, mitkä käsitteet on jo käsitelty.

Protégé-ontologiaeditorin lisäksi päivityksen aikana kehittäjillä oli auki ontologioiden sekä asiasanastojen selailuverkkosivuja, kuten ONKI-ontologiaselain, VESA⁴ tai erikoissanaston oma selailunäkymä. Syy omien erikoisontologioiden sekä YSON ONKI:ssa aukipitämiseen oli se, että sieltä he pystyvät hahmottamaan ontologian paremmin kuin Protégé-editoria käyttäen.

Kehittäjä K3 tiedusteli kehityksen aikana, päivittyvätkö muutetun käsitteen alakäsitteen automaattisesti, mikäli erikoisontologian käsitteen nimikettä muutetaan ja kyseinen nimike esiintyy käsitteen alakäsitteissä. Tämä voisi olla kiinnostava lisä ontologiaeditorien muutosten tekemiseen.

Muutoslistausten ohjeistus

Kaikki erikoisontologiakehittäjät olivat olleet loppukyselylomakkeessa sitä mieltä, että MUTU-muutoslistoja ei voi käyttää ilman ohjeistusta. Kehittäjät olivat tyytyväisiä päivityksen alussa annettuun ohjeistukseen, mutta väärinymmärryksiäkin syntyi.

Erikoisontologiakehittäjät saivat kehityksen tueksi kymmensivuisen ohjekirjan muutostyyppien ja kiinnostavien muutosten esittelystä. Kirjanen sisälsi vain teoreettisen tason tietoja ja esittelyjä muutostyypeistä ja tärkeistä muutoksista. Kehittäjät kuitenkin olisivat kaivanneet ohjeisiin konkreettisempia esimerkkejä ja muutosten

⁴<http://vesa.lib.helsinki.fi/>

vaikutusmahdollisuuksia erikoisontologian käsitteisiin. Lisäksi jälkikäteen huomattiin, että muutoslistauksen muutosten esitystapoja ei ollut ohjeistettu kirjasessa.

Päivitysten aikana kävi myös ilmi, että kehittäjät eivät olleet tietoisia ajantasaisen yleisontologian ominaisuuksista, kuten vanhentuneiden käsitteiden merkitsemisestä. Nämä aiheuttivat osalle kehittäjistä ihmettelyä, sillä ominaisuuksia ei ollut yleisontologian aikaisemmassa versiossa, eikä konkreettista tietoa uusista ominaisuuksista ollut helposti saatavilla.

Osa kehittäjistä ei ollut varmoja siitä, miten tulisi toimia, kun erikoisontologiassa jo ollut käsite on tullut uutena YSOon: poistetaanko oman ontologian käsite vai luodaanko tästä ekvivalenssiyhteys YSO:n käsitteeseen. Kehittäjät päätyivät lopulta erilaisiin ratkaisuihin: kehittäjät K2 ja K3 muodostivat päällekkäisissä tapauksissa ekvivalenssiyhteyksiä yleisontologiaan, kun taas K1 poisti liian yleiseksi kokemansa käsitteet. Jatkossa yhteisen kehityslinjan sopiminen olisi etu erikoisontologioiden yhtenäisyydelle. Lisäksi kehittäjät kommentoivat, että kun käsite on sekä YSOssa että erikoisontologiassa, niin työ on päällekkäistä ja kyseisen ajan voisi käyttää muuhunkin.

Luku 7

Yhteenvedo ja lopputulokset

Tässä työssä luotiin MUTU-niminen prototyypityökalu yleisontologian muutosten lajitteluun ja esittämiseen siihen linkittyneen erikoisontologian ylläpitäjälle. Työkalua testattiin kolmella erikoisontologiakehittäjällä aidossa päivitystilanteessa.

Tässä luvussa tarkastellaan vastauksia johdannossa esitettyihin tutkimuskysymyksiin. Lisäksi listataan kehityskohteita MUTU-työkalulle sekä mahdollisia jatkotutkimusaiheita.

7.1 Tutkimuskysymyksiin vastaaminen

Päättökysymys oli, miten erikoisontologiakehittäjien päivitystyötä voi helpottaa, kun yleisontologiassa on tapahtunut paljon muutoksia. Kysymyksen vastauksena tässä työssä esiteltiin uudenlainen työkalu, MUTU, yleisontologian muutosten välittämiseen ja esittämiseen erikoisontologiakehittäjälle. Työkalun toiminnan lisäksi myös sen muutosten esitystapoja kuvailtiin. Tämä toimii avauksena kentälle, jotta tulevassa tutkimuksessa esitettäisiin ja perusteltaisiin myös käyttöliittymäratkaisuja muutosten havainnollistamiseen. Myös ontologiakehittäjien tarpeita ja näkemyksiä on tuotu esiin enemmän kuin tutkimuksissa yleensä.

Ensimmäinen alatutkimuskysymys oli onko mahdollista erotella yleisontologian muutoksista ne, jotka aiheuttavat muutoksia erikoisontologiaan. MUTU-työkaluun ohjelmoitujen heuristiikkojen avulla muutokset lajiteltiin tyypeittäin ja näiden muutostyyppien sisällä vielä ennustettiin onko todennäköistä, että kyseinen yleisontologian muutos johtaisi muutokseen erikoisontologiassa, jolloin muutos katsotaan tärkeäksi. Tuloksena oli, että yksittäisen käsitteen muutoksilla tärkeät-kategoria sisälsi lähes kaikki yleisontologian muutokset, jotka johtivat erikoisontologian muutoksiin. Varjopuolena lajittelussa oli tosin se, että tärkeät-kategoriaan

tuli myös paljon muutoksia, jotka eivät vaikuttaneet erikoisontologiaan.

Toisessa alatuokimuskysymyksessä haluttiin selvittää, onko mahdollista käydä muutuskategoriat läpi tietyssä järjestyksessä. Tähän kysymykseen ei saatu vastausta, sillä suurin osa muutoslistojen arviointiin osallistuneista kehittäjistä kävi muutokset läpi muutoslistoissa annetussa järjestyksessä. Muutosten läpikäyminen tietyssä järjestyksessä mahdollistaisi, että sama käsite ei esiintyisi monessa muutuskategoriassa, jolloin päivittäminen tehostuisi, koska käsite käydään läpi vain kerran.

7.2 MUTU-työkalun kehityskohteita

Tutkimusprototyypiksi kehitetyssä MUTU-muutoslistaustyökalussa on paljon mahdollisia kehityskohteita.

Luvussa 6 ehdotin, että lisättyjen ja samannimisten käsitteiden muutoksia voisi yhdistää. Tämä voisi selkiyttää päivityslistojen läpikäymistä. Lisäksi ISO 25964-2 -standardi ilmestyi vasta MUTUN ohjelmoimisen jälkeen, joten sen linkitettyjen tesaurusten muutoksiin liittyvät maininnat tulisi huomioida [26, s. 43-45].

MUTUun olisi hyvä lisätä enemmän tukea ontologian versiointiin, esimerkiksi Mädchen ja kumppanien [40] ehdottamaa versiometatietoa, kuten versionumero, evoluutiolokin fyysinen URI sekä fyysinen URI muutoksia listattavista ontologioista. Myös yhteentoimivuus skos-historyn¹ kanssa ja tuki Dataset Versioning Ontology -ontologialle² olisi harkittavissa. Lisäksi muutoksista voisi luoda automaattisia tilastoja LODStats-tyylisesti [1].

MUTUN koodista tulisi tehdä modulaarisempi ja geneerisempi. Olisi hyvä erottaa muutosten tunnistaminen, lajittelu ja tärkeyden tunnistaminen selvästi erillisiksi, jotta MUTUa voisi käyttää myös yksittäisen ontologian muutoksen tunnistamiseen ilman toisen ontologian ”näkökulmaa”. Myös olisi kiinnostavaa testata voisiko muutostyypit ja niiden tärkeydet esittää SPARQL-kyselyinä, jolloin työkalu olisi helposti muokattavissa.

MUTUN nykyversiossa samojen käsitteiden tunnistus on tehty hyvin yksinkertaisella ja naiivilla tavalla. Olisikin hyvä jatkossa hyödyntää olemassaolevia ontologioiden linkitystyökaluja, kuten AgreementMakerLight [13] tai LogMap [30], joiden kehittyneemmät algoritmit voisivat auttaa samojen käsitteiden löytämisessä yleis- ja erikoisontologian välillä.

Ontologioiden laadunvalvonnan tarve nousi selvästi esiin muutoslistoja käydessä, jolloin käsitteiden tai ontologioiden välisten linkkien virheet tulevat helposti ke-

¹<https://github.com/jneubert/skos-history>

²<https://github.com/JohanDS/Dataset-versioning-for-KOS-data-sets>

hittäjän nähtäväksi. Olisi kuitenkin parempi tarkentaa MUTUN fokusta siten, että ontologioiden laadunvalvonta pidetään erillään muutosten tunnistamisesta.

Ekvivalenssisuhteiden lisäksi sekä käsitteiden nimikkeiden muutoksia tehtiin erikoisontologioissa eniten. Yksi kehittäjistä ehdotti, että ehdotetut ekvivalenssisuhteet luotaisiin automaattisesti ja näistä sitten vain poistettaisiin ne, jotka eivät ole oikein. Tämä olisi yksi mahdollisuus nopeuttaa muutoslistojen läpikäyntiä, sillä arvojen lisääminen tai korvaaminen kestää kauemmin kuin olemassaolevan arvon poistaminen.

Tällä hetkellä erikoisontologiakehittäjät eivät ole voineet nähdä yleisontologian muutoksia erikoisontologiassa. Muutokset on mahdollista saada näkyviin lisäämällä ylimääräisiä ominaisuuksia. Toisaalta tässä kohtaa tulisi miettiä, että jos muutokset saa näkyviin ontologiassa, onko erillinen nettisivumuutoslista tarpeellinen kehittäjälle. Lisäksi tämän hetkisissä muutoslistoissa ei ole minkäänlaista mahdollisuutta merkitä, mitkä muutokset kehittäjä on käynyt läpi. Olisi hyvä tarkastella teknologisia mahdollisuuksia tähän.

Tämän työn tulosten ja kehityskohteiden pohjalta lähdetään rakentamaan MUTU-työkalusta seuraavaa versiota. Kehitys aloitetaan vuonna 2015 Kansalliskirjastossa FINTO-projektissa³.

7.3 Jatkotutkimusaiheita

Alla listaan avoimia kysymyksiä sekä lisätutkimusta vaativia asioita, jotka nousivat esille tämän työn aikana.

Työssä tuotettu data

MUTU-muutoslistojen arvioinnista tuotettiin myös muutosdataa itse erikoisontologioista. Tämän työn puitteissa tätä dataa ei ehditty analysoida, sillä erikoisontologiakehittäjien tekemät muutoskuvailut eivät ole linkitetty erikoisontologian muutoksiin. Voisi olla kuitenkin mielenkiintoista tutkia tätä dataa ja saada konkreettisempi kuva kehittäjien tekemistä muutoksista.

Muutosten jaottelu

Tässä työssä luotiin tärkeyden käsite ontologioiden muutoksiin, jonka avulla pystytään priorisoimaan kehitystyötä. Jatkossa olisikin hyvä saada lisätutkimusta erilaisten muutosten tärkeydestä sekä onko tärkeys mahdollisesti yleistää enemmän kuin tässä työssä on tehty. Varsinkin käsitejoukon muutosten tärkeyden tutkiminen olisi hyödyllistä.

³<https://wiki.helsinki.fi/display/ONKI/Finto-projekti>

Päivitysprosessin sujuvoittamiseksi tulisi myös selvittää, missä järjestyksessä muutoskategoriat tulisi käydä läpi. Myös millaisia muutoksia yleisontologian muutokset vaikuttavat yleisontologiaan linkittyneissä erikoisontologioissa sekä miten ontologioiden liitoskohdat vaikuttavat muutosten tärkeyteen olisi syytä tutkia tarkemmin. Lisäksi työkaluarviointien tekeminen suuremmalla muutosmäärillä ja ontologioiden oikeilla kehittäjillä voisi osoittautua hedelmälliseksi.

Muutosten visualisointi

Muutosten visualisointitavoista ei ole juurikaan ollut keskustelua tiedeyhteisössä, vaan alla oleva tekniikka on ollut tutkimuksen kohteena. Menetelmät alkavat kuitenkin olla sen verran kypsiä, että myös käyttöliittymiin ja esitystapoihin tulisi alkaa paneutumaan. Intuitiivisempiin esitystapoihin, muutoskokonaisuuksiin ja käyttöliittymiin tulisi panostaa.

Ontologian vaikutus muihin tahoihin

Ontologioita, ja varsinkaan linkitettyjä ontologioita, ei luoda olemaan eristyksissä muusta maailmasta. Linkitettyjen ontologioiden muutokset vaikuttavat toisiinsa. Tästä on jo mainittu lyhyesti ISO 25964-2 -standardissa [26]. Lisää työkaluja kuitenkin tarvitaan linkitettyjen ontologioiden kehittämiseen. Lisäksi olisi hyvä huomata, että tesaurusmaisten ontologioiden muutokset eivät vaikuta pelkästään muihin ontologiakehittäjiin, vaan myös näitä käytäviin annotoijiin. Olisi tärkeää, että selvitetäisiin miten annotoijille saataisiin tehokkaimmin välitettyä tietoa ontologioiden muutoksista.

Kirjallisuutta

- [1] AUER, S., DEMTER, J., MARTIN, M., AND LEHMANN, J. LODStats – An Extensible Framework for High-performance Dataset Analytics. In *Knowledge Engineering and Knowledge Management (EKAW 2012)* (2012), vol. 7603, pp. 353–362.
- [2] BAI, F. *Collaboration Support for the Distributed Development of Ontologies*. PhD thesis, University Duisburg-Essen, 2013.
- [3] BECKET, D., AND BERNERS-LEE, T. Turtle – Terse RDF Triple Language. Team submission, 28.3.2011, W3C, 2011. <http://www.w3.org/TeamSubmission/turtle/> (Luettu 4.1.2015).
- [4] BERNERS-LEE, T., FIELDING, R., AND MASINTER, L. Uniform Resource Identifier (URI): Generic Syntax. Tech. Rep. RFC 3986, STD 66, RFC Editor, 2005. <http://www.rfc-editor.org/rfc/rfc3968.txt> (Luettu 18.09.2013).
- [5] BERNERS-LEE, T., HENDLER, J., AND LASSILA, O. The Semantic Web. *Scientific American* 284, 5 (2001), 28–37.
- [6] BHIDE, M., DEOLASEE, P., KATKAR, A., AND PANCHBUDHE, A. Adaptive push-pull: Disseminating dynamic web data. *IEEE Transactions on Computers* 51, 6 (2002), 652–668.
- [7] BIZER, C., HEATH, T., AND BERNERS-LEE, T. Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems* 5, 3 (2009), 1–22.
- [8] BRAUN, S., SCHMIDT, A., WALTER, A., AND ZACHARIAS, V. The Ontology Maturing Approach to Collaborative and Work Integrated Ontology Development: Evaluation Results and Future Directions. In *Proceedings of the International Workshop on Emergent Semantics and Ontology Evolution at (ISWC/ASWC 2007)* (2007), pp. 5–18.

- [9] BROOKE, J. SUS-A Quick and Dirty Usability Scale. *Usability Evaluation in Industry 189* (1996), 189–194.
- [10] COPELAND, M., GONÇALVES, R. S., PARSIA, B., SATTLER, U., AND STEVENS, R. Finding fault: Detecting Vssues in a Versioned Ontology. In *Proceedings of the 2nd International Workshop on Debugging Ontologies and Ontology Mappings (WoDOOM 2013)* (2013), pp. 9–20.
- [11] EHRIG, M. *Ontology Alignment: Bridging the Semantic Gap*. Springer, 2006.
- [12] FALCONER, S., TUDORACHE, T., AND NOY, N. F. An Analysis of Collaborative Patterns in Large-Scale Ontology Development Projects. In *Proceedings of the 6th International Conference on Knowledge Capture (K-CAP 2011)* (2011), pp. 25–32.
- [13] FARIA, D., PESQUITA, C., SANTOS, E., CRUZ, I. F., AND COUTO, F. M. AgreementMakerLight: A Scalable Automated Ontology Matching System. In *Proceedings of the 10th International Conference on Data Integration in the Life Sciences (DILS 2014)* (2014), pp. 29–32.
- [14] FLOURIS, G., PLEXOUSAKIS, D., AND ANTONIOU, G. A Classification of Ontology Change. In *Proceedings of the 3rd Italian Semantic Web Workshop, Semantic Web Applications and Perspectives* (2006).
- [15] FLUIT, C., SABOU, M., AND VAN HARMELEN, F. Supporting User Tasks Through Visualisation of Light-Weight Ontologies. In *Handbook on Ontologies in Information Systems*. Springer, 2004, pp. 415–432.
- [16] FROSTERUS, M., TUOMINEN, J., PESSALA, S., AND HYVÖNEN, E. Linked Open Ontology Cloud – Managing a System of Interlinked Cross-domain Light-weight Ontologies, 2015. Submitted.
- [17] GARSHOL, L. M. Metadata? Thesauri? Taxonomies? Topic maps! Making sense of it all. *Journal of Information Science* 30, 4 (2004), 378–391.
- [18] GIARETTA, P., AND GUARINO, N. Ontologies and Knowledge Bases: Towards a Terminological Clarification. *Towards Very Large Knowledge Bases: Knowledge Building & Knowledge Sharing* (1995), 25–32.
- [19] GIUNCHIGLIA, F., AND ZAIHRAYE, I. Lightweight Ontologies. *The Encyclopedia of Database Systems* (2007).
- [20] GOMEZ-PEREZ, A., FERNÁNDEZ-LÓPEZ, M., AND CORCHO-GARCIA, O. *Ontological Engineering*. Springer, 2004.

- [21] GRUBER, T. R. A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition* 5, 2 (1993), 199–220.
- [22] HAHN, U., AND SCHULZ, S. Building a Very Large Ontology from Medical Thesauri. In *Handbook on Ontologies in Information Systems*. Springer, 2004, pp. 133–150.
- [23] HAMEED, A., PREECE, A., AND SLEEMAN, D. Ontology Reconciliation. In *Handbook on Ontologies in Information Systems*. Springer, 2004, pp. 231–250.
- [24] HENDLER, J. Agents and the Semantic Web. *IEEE Intelligent Systems* 16, 2 (2001), 30–37.
- [25] ISO. Information and documentation—Thesauri and interoperability with other vocabularies—Part 1: Thesauri for information retrieval. ISO 25964-1:2011, International Organization for Standardization, 2011.
- [26] ISO. Information and documentation—Thesauri and interoperability with other vocabularies—Part 2: Interoperability with other vocabularies. ISO 25964-2:2013, International Organization for Standardization, 2013.
- [27] JASPER, R., USCHOLD, M., ET AL. A Framework for Understanding and Classifying Ontology Applications. In *Proceedings of the 12th International Workshop on Knowledge Acquisition, Modelling, and Management (KAW 1999)* (1999), pp. 16–21.
- [28] JIMÉNEZ RUIZ, E., CUENCA GRAU, B., HORROCKS, I., AND BERLANGA, R. Supporting Concurrent Ontology Development: Framework, Algorithms and Tool. *Data & Knowledge Engineering* 70, 1 (2011), 146–164.
- [29] JIMÉNEZ-RUIZ, E., CUENCA GRAU, B., HORROCKS, I., AND LLAVORI, R. B. Building Ontologies Collaboratively Using ContentCVS. *Description Logics* 477 (2009).
- [30] JIMÉNEZ-RUIZ, E., AND CUENCA GRAU, B. LogMap: Logic-Based and Scalable Ontology Matching. In *The Semantic Web – ISWC 2011*, vol. 7031 of *Lecture Notes in Computer Science*. 2011, pp. 273–288.
- [31] KHATTAK, A. M., LATIF, K., KHAN, S., AND AHMED, N. Ontology Recovery and Visualization. In *Proceedings of the 4th International Conference on Next Generation Web Services Practices* (2008), pp. 90–96.
- [32] KLEIN, M. *Change Management for Distributed Ontologies*. PhD thesis, Vrije Universiteit Amsterdam, 2004.

- [33] KLEIN, M., FENSEL, D., KIRYAKOV, A., AND OGNANOV, D. OntoView: Comparing and Versioning Ontologies. In *Collected Posters of the 1st International Semantic Web Conference (ISWC 2002)* (2002).
- [34] KLEIN, M., AND NOY, N. F. A Component-Based Framework for Ontology Evolution. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence Workshop on Ontologies and Distributed Systems* (2003), vol. 71.
- [35] KLEIN, M. C., AND FENSEL, D. Ontology Versioning on the Semantic Web. In *Proceedings of the 1st International Semantic Web Workshop* (2001), pp. 75–91.
- [36] LANTHALER, M., SPORNY, M., AND KELLOGG, G. JSON-LD 1.0. W3C Recommendation, W3C, 2014. <http://www.w3.org/TR/2014/REC-json-ld-20140116/> (Luettu 4.1.2015).
- [37] LASSILA, O., AND MCGUINNESS, D. The Role of Frame-Based Representation on the Semantic Web. Tech. Rep. Knowledge Systems Laboratory Report KSL-01-02, Stanford University, 2001.
- [38] LASSILA, O., AND SWICK, R. R. Resource Description Framework (RDF) Model and Syntax Specification. Tech. rep., World Wide Web Consortium, Tammikuu 1999. <http://www.w3.org/TR/1999/PR-rdf-syntax-19990105/> (Luettu 8.10.2013).
- [39] LEHMANN, J., ISELE, R., JAKOB, M., JENTZSCH, A., KONTOKOSTAS, D., MENDES, P. N., HELLMANN, S., MORSEY, M., VAN KLEEF, P., AUER, S., ET AL. DBpedia—A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia. *Semantic Web* (2014).
- [40] MÄDCHÉ, A., MOTIK, B., STOJANOVIC, L., STUDER, R., AND VOLZ, R. An Infrastructure for Searching, Reusing and Evolving Distributed Ontologies. In *Proceedings of the 12th International Conference on World Wide Web* (2003), pp. 439–448.
- [41] MÄDCHÉ, A., MOTIK, B., STOJANOVIC, L., STUDER, R., AND VOLZ, R. Ontologies for Enterprise Knowledge Management. *IEEE Intelligent Systems* 18, 2 (2003), 26–33.
- [42] MCBRIDE, B. RDF Vocabulary Description Language 1.0: RDF Schema. Tech. rep., W3C, 2004. <http://www.w3.org/TR/2004/REC-rdf-schema-20040210/> (Luettu 17.10.2013).

- [43] MILES, A., AND BECHHOFFER, S. SKOS simple knowledge organization system reference. W3C Recommendation, W3C, 2009. <http://www.w3.org/2009/08/skos-reference/skos.html> (Luettu 4.1.2015).
- [44] MILLER, E., SEPPA, C., KITTUR, A., SABB, F., AND POLDRACK, R. A. The Cognitive Atlas: Employing Interaction Design Processes to Facilitate Collaborative Ontology Creation. *Frontiers in Neuroinformatics* 5 (2010).
- [45] MIZOGUCHI, R., VANWELKENHUYSEN, J., AND IKEDA, M. Task Ontology for Reuse of Problem Solving Knowledge. *Towards Very Large Knowledge Bases* (1995), 46–59.
- [46] NILES, I., AND PEASE, A. Towards a Standard Upper Ontology. In *Proceedings of the International conference on Formal Ontology in Information Systems (FOIS 2001)* (2001), pp. 2–9.
- [47] NOY, N. F., CHUGH, A., LIU, W., AND MUSEN, M. A. A Framework for Ontology Evolution in Collaborative Environments. In *Proceedings of the 5th International Semantic Web Conference (ISWC2006)* (2006), pp. 544–558.
- [48] NOY, N. F., AND KLEIN, M. Ontology Evolution: Not the Same as Schema Evolution. *Knowledge and information systems* 6, 4 (2004), 428–440.
- [49] NOY, N. F., KUNNATUR, S., KLEIN, M., AND MUSEN, M. A. Tracking Changes During Ontology Evolution. In *Proceedings of the 3rd International Semantic Web Conference (ISWC-04)* (2004), pp. 259–273.
- [50] NOY, N. F., MCGUINNESS, D. L., ET AL. Ontology Development 101: A Guide to Creating your First Ontology. Tech. Rep. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, Stanford University, 2001.
- [51] PESSALA, S., SEPPÄLÄ, K., SUOMINEN, O., FROSTERUS, M., TUOMINEN, J., AND HYVÖNEN, E. MUTU: An Analysis Tool for Maintaining a System of Hierarchically Linked Ontologies. In *Proceedings of the Workshop on Ontologies come of Age Workshop in the 10th International Semantic Web Workshop (ISWC 2011)* (2011).
- [52] PINTO, H. S., TEMPICH, C., AND STAAB, S. Ontology Engineering and Evolution in a Distributed World using DILIGENT. In *Handbook on Ontologies*. 2009, pp. 153–176.
- [53] RÜGER, S. *Multimedia Information Retrieval*. Morgan & Claypool Publishers, 2009.

- [54] SEPPÄLÄ, K., AND HYVÖNEN, E. Asiasanaston muuttaminen ontologiaksi. Yleinen suomalainen ontologia esimerkkinä FinnONTO-hankkeen mallista. Tech. rep., 2014. <https://www.doria.fi/handle/10024/96825> (Luettu 6.1.2015).
- [55] SERAFINI, L., AND TAMILIN, A. Composing Modular Ontologies with Distributed Description Logics. In *Modular Ontologies*. Springer, 2009, pp. 321–347.
- [56] SHAH, N., CHAO, K.-M., ZLAMANIEC, T., AND MATEI, A. Ontology for Home Energy Management Domain. In *Proceedings of the Digital Information and Communication Technology and its Applications* (2011), pp. 337–347.
- [57] SMITH, B., AND WELTY, C. Ontology: Towards a New Synthesis. In *Formal Ontology in Information Systems* (2001), pp. 3–9.
- [58] SOLDATOVA, L., AND KING, R. An Ontology of Scientific Experiments. *Journal of the Royal Society Interface* 3, 11 (2006), 795–803.
- [59] STOJANOVIC, L., MÄDCHE, A., MOTIK, B., AND STOJANOVIC, N. User-driven ontology evolution management. In *Proceedings of the International Conference on Knowledge Engineering and Knowledge Management (EKAW 2002)* (2002), pp. 133–140.
- [60] STUCKENSCHMIDT, H., AND KLEIN, M. Integrity and Change in Modular Ontologies. In *Proceedings of the International Joint Conference on Artificial Intelligence* (2003), pp. 900–908.
- [61] STUDER, R., BENJAMINS, V. R., AND FENSEL, D. Knowledge Engineering: Principles and Methods. *Data & Knowledge Engineering* 25, 1 (1998), 161–197.
- [62] SUNAGAWA, E., KOZAKI, K., KITAMURA, Y., AND MIZOGUCHI, R. An Environment for Distributed Ontology Development Based on Dependency Management. In *Proceedings of the International Workshop on Emergent Semantics and Ontology Evolution in the 2nd International Semantic Web Workshop (ISWC 2003)* (2003), pp. 453–468.
- [63] TUOMINEN, J. Helppökytkentäiset ontologiapalvelut semanttisessa webissä. Master’s thesis, University of Helsinki, Department of Computer Science, May 2010.
- [64] VALO, A., KOMULAINEN, V., AND HYVÖNEN, E. A Tool for Collaborative Ontology Development for the Semantic Web. In *Proceedings of International Conference on Dublin Core and Metadata Applications* (2005).

- [65] VAN HEIJST, G., SCHREIBER, A. T., AND WIELINGA, B. J. Using Explicit Ontologies in KBS Development. *International Journal of Human-Computer Studies* 46, 2 (1997), 183–292.
- [66] W3C OWL WORKING GROUP. OWL 2 Web Ontology Language - Document Overview (Second Edition). W3C Recommendation, W3C, 2012. <http://www.w3.org/TR/2012/REC-owl2-overview-20121211/> (Luettu 17.10.2013).
- [67] WONG, W., LIU, W., AND BENNAMOUN, M. Ontology Learning from Text: A Look back and into the Future. *ACM Computing Surveys (CSUR)* 44, 4 (2012), 20.
- [68] WORLD WIDE WEB CONSORTIUM. RDF/XML Syntax Specification (Revised). W3C Recommendation, W3C, 2004. <http://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/> (Luettu 7.10.2013).
- [69] ZAMAZAL, O., BÜHMANN, L., AND SVÁTEK, V. Checking and Repairing Ontological Naming Patterns Using ORE and PatOMat. In *Proceedings of the 2nd International Workshop on Debugging Ontologies and Ontology Mappings (WoDOOM 2013)* (2013).

Liite A

Loppukysely

Arvioinnin lopuksi erikoisontologiakehittäjien täyttämä kyselylomake. Käyttäjät ovat vastanneet kyselyyn Google Form -lomakkeena. Arvioinnin aikana tärkeitä muutoksia kutsuttiin vielä *kiinnostaviksi muutoksiksi*, muutoslistauksia *nettisivumuutoslistauksiksi* ja ontologiaan merkittyjä muutoksia *ontologiamuutoslistauksiksi*.

Yleistä

Kuinka kokenut ontologiakehittäjä olet

1 2 3 4 5
aloittelija ○ ○ ○ ○ ○ konkari

Erikoisontologian päivittäminen

Eniten aikaa erikoisontologiaa päivittäessä vei:

- nettisivulla olevan muutoslistausraportin selaaminen
- Protégésta käsitteiden etsiminen ja tarkastelu
- ONKIn selaaminen
- Other: _____

Erikoisontologian päivittäminen sujui hyvin muutoslistauksia hyödyntäen.

1 2 3 4 5
täysin eri mieltä täysin samaa mieltä

Oppiminen

Alkuopastuksessa keskityttiin tilanteisiin ja ongelmiin, joita päivittäessä kohtasi.

1 2 3 4 5
täysin eri mieltä täysin samaa mieltä

Muutoslistauksia oppi nopeasti hyödyntämään.

1 2 3 4 5
täysin eri mieltä täysin samaa mieltä

Muutoslistaukset nettisivulla ja ontologiassa

Muutoslistaukset olivat turhan monimutkaisia.

1 2 3 4 5
täysin eri mieltä täysin samaa mieltä

Muutoslistaukset tekivät kaiken, mitä tarvitsen.

1 2 3 4 5
täysin eri mieltä täysin samaa mieltä

Nettisivu- ja ontologiamuutoslistauksien yhteyttä oli hankala ymmärtää.

1 2 3 4 5
täysin eri mieltä täysin samaa mieltä

Muutoslistauksia yhdessä käyttämällä oli helppo ymmärtää, mitä YSOssa oli muuttunut.

1 2 3 4 5
täysin eri mieltä täysin samaa mieltä

Nettisivun muutosraportti

Muutosraportti ei ollut hyödyllinen.

1 2 3 4 5
täysin eri mieltä täysin samaa mieltä

Muutosraportti oli helppokäyttöinen.

1 2 3 4 5
täysin eri mieltä täysin samaa mieltä

Muutosraportti-nettisivu oli tarpeeksi nopea.

1 2 3 4 5
täysin eri mieltä täysin samaa mieltä

Ontologiamuutoslista

Yhdistelmäontologiaan lisätyt muutoskategoriat eivät olleet hyödyllisiä.

1 2 3 4 5
täysin eri mieltä täysin samaa mieltä

Yhdistelmäontologian muutoslista oli helppokäyttöinen.

1 2 3 4 5
täysin eri mieltä täysin samaa mieltä

Informaatio

Muutosraportti-nettisivun selitetekstit olivat selkeät.

1 2 3 4 5
täysin eri mieltä täysin samaa mieltä

Oli helppoa löytää tarvitsemani tieto, jotta pystyin määrittämään tarvitseeko erikoisontologiaa muuttaa.

1 2 3 4 5
täysin eri mieltä täysin samaa mieltä

Muutoslistauksia voisi käyttää ilman erillisiä ohjeita.

1 2 3 4 5
täysin eri mieltä täysin samaa mieltä

Kiinnostavat-muut-jaottelu

Kiinnostavat-muutokset olivat kiinnostavia.

1 2 3 4 5
täysin eri mieltä täysin samaa mieltä

Muut-muutokset eivät olleet kiinnostavia.

1 2 3 4 5
täysin eri mieltä täysin samaa mieltä

Haluan tietää millä perusteella kiinnostavat- ja muut-muutokset on jaoiteltu.

1 2 3 4 5
täysin eri mieltä täysin samaa mieltä

Kiinnostavat- ja muut-muutokset jaottelu jäi minulle epäselväksi.

1 2 3 4 5
täysin eri mieltä täysin samaa mieltä

Muutostyypit

Minulle oli selvää mitä muutoksia eri muutostyypit tarkoittivat.

1 2 3 4 5
täysin eri mieltä täysin samaa mieltä

Jaottelu muutostyypeittäin oli intuitiivinen.

1 2 3 4 5
täysin eri mieltä täysin samaa mieltä

Asiasanamuutosten kielijaottelu oli hyödyllinen.

 1 2 3 4 5
täysin eri mieltä täysin samaa mieltä

Assosiatiivisuusmuutosten lisätyt/korvatut/poistetut-jaottelu oli hyödyllinen.

 1 2 3 4 5
täysin eri mieltä täysin samaa mieltä

Liite B

Loppuhaastattelun runko

- Miten päivittäminen meni ylipäätänsä
- Mitä mieltä olet MUTU-listauksista ylipäätänsä hyvää/huonoa/parannettavaa
- Perehdytystapaminen ja -materiaali
 - tapaaminen
 - * hyvää/huonoa/puuttuiko
 - ohjekirja
 - * hyvää/huonoa/puuttuiko
 - * ohje eteen
- Mitä muita ohjelmia/nettisivuja käytät päivittäessäsi
- Aika
 - miten kauan suunnilleen menee, jotta tietää, onko jokin erikoisontologian kannalta kiinnostava
 - kuinka kauan menee, että tietää tarvitseeko erkoisontologiaan tehdä muutoksia
- Mitä mieltä olit kiinnostavat-muut -jaottelusta?
- Mitä mieltä olit muutostyypeistä?
 - Lisättävää/poistettavaa/hiottavaa
 - Lista eteen

- Miten pysyit selvillä siitä, mitkä käsitteet olet jo käynyt läpi?
- Oliko helppoa tietää kumman ontologian käsitettä katsoo/muokkaa
- Asioita, joita ei ole osattu ottaa huomioon?
- Käytön epäloogisuuksia?
- Jotain muuta, mitä haluaisit sanoa

Liite C

Kiinnostavat muutokset -taulukko

			K1		K2		K3	
			kiin (%)	kpl	kiin (%)	kpl	kiin (%)	kpl
Lisätyt	k		27 (1.08)	25	27 (1.08)	25	4 (0.66)	6
	m		25	25	32 (1.28)	25	24 (0.96)	25
Poistetut			19 (0.90)	21	0 (0.00)	19	1	1
Saman	k		25	25	28 (1.12)	25	18 (0.72)	25
	m		24 (0.96)	25	25	25	24 (0.96)	25
Hierarkia	k		26 (1.04)	25	25	25	25	25
	m		24 (0.96)	25	23 (0.92)	25	25	25
Nimike	k	fi	27 (1.08)	25	25	25	14 (0.88)	16
		sv	0 (0.00)	25	25	25	19 (3.8)	5
		en	25	25	25	25	5 (0.26)	19
	m	fi	25	25	25	25	25	25
		sv	0 (0.00)	25	25	25	26 (1.04)	25
		en	25	25	25	25	25	25
		muut	2	2	2	2	2	2
Assosia	k	lisäys	25	25	25	25	25	25
		poisto	25 (1.04)	24	24	24	11	11
		korvaus	16 (0.94)	17	15	15	4	4
	m	lisäys	25	25	25	25	25	25
		poisto	25	25	24 (0.96)	25	25	25
		korvaus	25	25	25	25	25	25
Osa-koko	k		11	11	9	9	5	5
	m		23	23	24 (0.96)	25	10	10
Muut			25	25	5 (0.20)	25	25	25
Yht			474 (0.90)	523	488 (0.94)	519	392 (0.97)	404

Liite D

Muutosmäärät-taulukko

			K1		K2		K3	
			muutoksia	kpl	muutoksia	kpl	muutoksia	kpl
Lisätyt	t		4 (0.16)	25	1 (0.04)	25	2 (0.33)	6
	m		0	25	4 (0.16)	25	0	25
Poistetut			1 (0.05)	21	0	19	0	1
Saman	t		19 (0.76)	25	25	25	18 (0.72)	25
	m		5 (0.20)	25	4 (0.16)	25	5 (0.20)	25
Hierarkia	t		8 (0.32)	25	2 (0.08)	25	0	25
	m		0	25	0	25	0	25
Nimike	t	fi	6 (0.24)	25	1 (0.04)	25	12 (0.75)	16
		sv	0	25	0	25	0	5
		en	4 (0.16)	25	0	25	0	19
	m	fi	0	25	0	25	0	25
		sv	0	25	0	25	0	25
		en	0	25	0	25	0	25
		muut	1 (0.50)	2	0	2	0	2
Assosia	t	lisäys	2 (0.08)	25	0	25	0	25
		poisto	4 (0.17)	24	0	24	0	11
		korvaus	0	17	0	15	0	4
	m	lisäys	0	25	0	25	0	25
		poisto	0	25	0	25	0	25
		korvaus	0	25	0	25	0	25
Osa-koko	t		2 (0.18)	11	0	9	0	5
	m		0	23	0	25	0	10
Muut			0	25	0	25	0	25
Yht			56 (0.11)	523	37 (0.07)	519	37 (0.09)	404

Liite E

Erikoisontologiakehittäjien tekemät muutokset -taulukko

muutos kehittäjä	equi			lisäys + equi			equi + hierar			hierarkia			nimike + hierar			poisto			asso			nimike			yht			yht
	K1	K2	K3	K1	K2	K3	K1	K2	K3	K1	K2	K3	K1	K2	K3	K1	K2	K3	K1	K2	K3	K1	K2	K3	K1	K2	K3	
lisätyt	1	4	1	-	-	-	-	-	-	-	1	-	-	-	-	3	-	-	-	-	-	-	-	1	4	5	2	11
poistetut	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	1	0	0	1
samannimiset	17	23	21	1	3	-	2	2	-	-	1	-	-	-	-	4	-	1	-	-	-	-	-	1	24	29	23	76
hierarkia	-	-	-	-	-	-	-	-	-	4	2	-	-	-	-	4	-	-	-	-	-	-	-	-	8	2	0	10
nimike	1	-	-	-	-	-	-	-	-	1	-	-	-	-	-	6	-	-	-	-	-	3	1	12	11	1	12	23
asso	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	4	-	-	1	-	-	6	0	0	6
osa-koko	-	-	-	-	-	-	-	-	-	1	-	-	1	-	-	-	-	-	-	-	-	-	-	-	2	0	0	2
muut	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	0	0
yhteensä	19	27	22	1	3	0	2	2	0	6	4	0	1	0	0	19	0	0	4	0	0	4	2	14	56	37	37	
		68			4			4			10			1			19			4			20			130		