

Ontologioiden soveltuvuus henkilökohtaisten tekstiaineistojen jäsentämiseen - esimerkkinä elämäkerta-aineistot

Katriina Karkimo

Tampereen yliopisto
Luonnontieteiden tiedekunta
Informaation ja järjestelmien tohtoriohjelma, tietojenkäsittelyoppi
Lisensiaatintyö
Marraskuu 2018

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -ohjelmalla
Tampereen yliopiston laatujärjestelmän mukaisesti

Tampereen yliopisto

Luonnontieteiden tiedekunta

Informaation ja järjestelmien tohtoriohjelma, tietojenkäsittelyoppi

Karkimo, Katriina: Ontologioiden soveltuvuus henkilökohtaisten tekstiaineistojen jäsentämiseen - esimerkkinä elämäkerta-aineistot

Lisensiaatintyö, 104 sivua, 14 liitesivua

Marraskuu 2018

Työssä selvitetään yleisten, julkisten ontologioiden soveltuvuutta käyttäjien itse tekemien avainsanajäsennysten tukemiseen. Kysymystä tutkitaan selvittämällä, sisältävätkö julkiset ontologiat käyttäjien itsensä valitsemia avainsanoja, joilla he kuvaavat kirjoittamiaan elämäkerta-aineistoja. Vertailun vuoksi tutkija on valinnut avainsanat samoista aineistoista, ja myös niiden esiintymistä ontologioissa selvitetään. Samassa yhteydessä kirjattiin ylös minkälaisia piirteitä tai mahdollisia ongelmia nousi esiin, kun nykyisillä ontologioilla haluttiin jäsentää elämäkertoja. Saatuja tuloksia verrattiin lähdekirjallisuuteen. Aineistoina ovat kansalaisopiston opintopiirissä laaditut elämäkerta-aineistot ja Finto-ontologiakirjaston ontologiat.

Tutkimusmenetelmien osalta työ on suunnittelutieteellistä arvioivaa tutkimusta. Arvioinnin kohteena olivat ontologiat ja niiden kattavuus. Tutkimus on myös teorioita testaavaa tutkimusta sekä case-tutkimusta. Saatuja tuloksia verrataan lähdekirjallisuuteen ja tapauksina on viisi elämäkerta-aineistoa.

Tehdyissä vertailuissa havaitaan, että elämäkerta-aineistojen avainsanoista karkeasti sanottuna noin puolet sisältyy nykyisellään Finto-ontologiakirjaston ontologioihin. Tarkevammassa analyysissä voidaan erottaa pieni painotus siihen suuntaan, että todennäköisemmin ontologiat sisältävät tutkijan kuin käyttäjien valitsemia avainsanoja. Tämä kvantitatiivinen tulos on työssä kuitenkin marginaalinen.

Työn keskeisin anti on laadullisissa seikoissa, jotka ilmenevät tehtyjen kokeilujen yhteydessä. Käy ilmi, kuten lähdekirjallisuudessa mainitaan, että avainsanojen merkitsemisen tavoitteet ja käytännöt vaihtelevat suuresti ja ne voivat muuttua myös yksittäisellä henkilöllä työskentelyn edetessä. Tehtyjen kokeilujen yhteydessä havaitaan, että tällä hetkellä Finto-ontologiakirjaston ontologioihin tarvittaisiin lisää erityisesti erisnimiä, samoin on tarvetta terävöittää ontologiakirjaston sisältämien eri ontologioiden välistä linkittämistä. Suomenkielisten termien kohdalla olisi tarvetta lisätä eri taivutusmuotoja sekä puhekielen mukaisia sanamuotoja ja ilmaisuja. Termihakujen yhteydessä nousi myös esiin tarve työkaluille, jotka tukisivat yhdyssanojen ja sanaliittojen muodostamien ilmaisujen semanttista tulkintaa ja analysointia.

Avainsanat: ontologiat, folksonomiat, semanttinen web, web 2.0, elämäkerrat, muistitieto

University of Tampere

Faculty of Natural Sciences

Doctoral programme in Information and Systems, Computer Science

Karkimo, Katriina: The applicability of ontologies for structuring personal texts – using biographies as an example

Licentiate's thesis, 104 pages, 14 appendix pages

November 2018

In this work the applicability of general, public ontologies is examined. The question is whether these ontologies are suitable for supporting the keyword definitions done by users. This is examined by finding out whether the public ontologies contain the terms that the users themselves choose to be the keywords of their texts. The texts are the users' own biographies that they have written themselves. In order to compare the results, the researcher, too, has chosen keywords for the same texts, and the presence of the researcher's keywords in public ontologies is also examined. The biographies are written for a course at an adult education center, and the ontologies are all contained in the Finnish ontology library Finto.

Methodologically this represents design science and the ontologies of Finto library are the object of evaluation. They are measured by their extensiveness. This is also theory-testing research and a case study that consists of five cases. Each biography forms one of the cases and the findings are compared with the source material.

As a result, it is found out that roughly about half of the keywords can currently be found in the ontologies of the ontology library Finto. In a more detailed analysis, one can see that the researcher's keywords are contained in the ontologies with a slightly higher probability than the keywords chosen by the authors themselves. However, this quantitative result is marginal in this work.

The main outcomes are the qualitative findings noted while doing the experiments. It turns out, as in the referred sources, that the objectives and practices of keyword marking vary widely, and that they can also change during the work process of a single person. It is also found out that, at this moment, more proper nouns are needed in the ontologies of Finto. Additionally, there is a need for refined mapping between the various ontologies contained in the ontology library. For Finnish terms, a larger variety of inflected forms is needed, as well as a greater number of vernacular expressions and colloquialisms. While searching for terms it was also discovered that there is a need for tools that would help the semantic analysis and interpretation of compound words and set phrases.

Keywords: ontologies, folksonomies, semantic web, web 2.0, biographies, oral history

Esipuhe

Tätä työtä on tehty pitkään ja hartaasti. Kun tavoite on nyt viimeinkin saatettu päämääräänsä, olen suuren kiitoksen velkaa useille henkilöille. Ensimmäkin monet kiitokset työn ohjaajalleni, professori emeritus Pertti Järviselle. Kontakti jatko-opintoihin olisi saattanut helposti katketa kokoaan ilman hänen jatkuvaa jatko-opiskelijoille lähetettyä sähköpostiuutiskirjettä. Sen avulla oli helppo sopivan hetken tullessa vastata viestiin ja kysyä mahdollisuudesta tulla taas mukaan seminaariin.

Erityisen suuren kiitoksen olen velkaa Mäntsälän kansalaisopiston rehtorille, Airi Pekala-Mäkitalolle. Pidimme hänen kanssaan muisteluryhmää, jonka jäsenten tuotokset muodostavat tämän työn aineiston. Airi oli rehtorina luonnollisesti myös esimieheni ja hänen erittäin joustava ja kannustava suhtautumisensa tämän työn tekemiseen teki tilanteen minulle helpoksi. Tunnen saaneeni hänen suunnaltaan parhaan mahdollisen tuen.

Lämpimät kiitokset menevät myös muisteluryhmän jäsenille, erityisesti heille, jotka antoivat oman tekstinäytteensä minulle aineistoksi. Ilman heitä en olisi voinut toteuttaa koko tutkimusta. Työskentely ryhmän kanssa oli antoisaa, ilmapiiri oli välitön ja on ollut mukava kuulla, että he jatkavat omaehtoisia tapaamisia edelleen.

Lopuksi haluan osoittaa monet lämpimät kiitokset kärsivällisyydestä myös omalle perheelleni.

Porvoossa marraskuussa 2018,

Katriina Karkimo

Sisälllys

1	Johdanto.....	1
1.1	Tutkimuksen aihe	1
1.2	Työn yhteys muihin vastaaviin tutkimuksiin	2
2	Työn tavoitteet ja työskentelyn kulku	4
2.1	Tutkimuksen tarkoitus	4
2.2	Tutkimuksen kulku.....	5
3	Tutkimusmenetelmät	7
3.1	Tutkimuskysymys	7
3.2	Työssä käytetyt tutkimusmenetelmät	8
3.2.1	Teorioita testaava case-tutkimus.....	8
3.2.2	Suunnittelutieteellinen arvioiva tutkimus	9
3.3	Työskentelyssä sovellettuja periaatteita	10
3.3.1	Ennaltaohjailemattomuus	10
3.3.2	Kvalitatiivinen lähestymistapa.....	11
4	Tutkimuksen kannalta keskeisten käsitteiden määrittely	12
4.1	Ontologiat.....	12
4.1.1	Mitä ontologioilla tarkoitetaan?.....	12
4.1.2	Ontologia suhteessa muihin käsite- ja termikokoelmiin.....	15
4.1.3	Ontologioihin liittyviä ongelmia.....	18
4.2	Folksonomiat	22
4.2.1	Avainsanat ja folksonomiat	22
4.2.2	Folksonomioiden heikkouksia	25
4.3	Eri näkemyksiä ontologioiden suhteesta folksonomioihin.....	29
4.3.1	Ontologiat ja folksonomiat ovat eri asioita.....	29
4.3.2	Folksonomiat ovat eräänlaisia ontologioita	31
4.4	Semanttinen web	33
4.4.1	Semanttisen webin määritelmä	33
4.4.2	Semanttisen webin hyötyjä	35
4.4.3	Web 2.0.....	36
4.5	Elämäkerta, elämänkaarikirjoitus.....	39
4.6	Muistitieto	40
4.6.1	Muistitiedon merkitys tutkimuksessa	40
4.6.2	Muistitiedon määritelmä	41
4.6.3	Muistitiedon ominaispiirteitä.....	42
5	Aikaisempi aihetta koskeva tutkimus	44
5.1	Eri tapoja ontologioiden ja folksonomioiden yhdistämiseksi	44
5.1.1	Lähdeaineiston kerääminen	44
5.1.2	Limpensin, Gandonin ja Buffan viitekehys	45
5.2	Ontologioiden käyttäminen luokittelussa.....	46
5.2.1	Weblogien jäsentäminen ontologioiden avulla.....	46
5.2.2	Muita lähestymistapoja ontologioiden hyödyntämiseen avainsanaluokitteluissa	47

5.3	Folksonomian muuntaminen ontologiaksi	49
5.3.1	Automatisoitu ontologioiden johtaminen	49
5.3.2	Käyttäjien täydennettäviksi tarkoitetut järjestelmät	52
5.4	Folksonomian rikastaminen semanttisella sisällöllä	54
5.4.1	Van Dammen, Heppin ja Siorpaesin menetelmä.....	54
5.4.2	Specian ja Mottan menetelmä.....	57
5.4.3	Angeletoun ym:iden menetelmä.....	58
5.4.4	Tesconin ym:iden TDA-algoritmi	58
5.4.5	Passantin ja Laubletin tapa yhdistää folksonomiat ja ontologiat.....	59
5.5	Tämän työn asema suhteessa aiempaan tutkimukseen.....	60
6	Tutkimuksen toteutusympäristö	61
6.1	Tikutus-hanke ja sen tavoitteet.....	61
6.1.1	Tikutus-hanke	61
6.1.2	Pilottiryhmät	62
6.2	Työskentelyssä käytettyjä menetelmiä ja toimintatapoja.....	63
6.2.1	Talletusmuodon ja -välineen valinta.....	63
6.2.2	Mielleyhtymäkartoista ontologioihin.....	64
7	Työskentelyn eteneminen.....	66
7.1	Työskentelyn organisointi ja välineet.....	66
7.1.1	Ryhmän perustaminen ja työskentelyn aloittaminen.....	66
7.1.2	Mielleyhtymäkartat ja graafinen web-editori	67
7.1.3	Multimedia-aineiston muokkaamiseen käytettyjä työkaluja	67
7.2	Aineiston kerääminen ja jäsentäminen.....	68
7.2.1	Muistelemisen tukeminen.....	68
7.2.2	Aineiston jäsentäminen.....	70
7.3	Pilotin päättäminen ja jatkotyöskentely	70
8	Tulokset.....	72
8.1	Työn tavoitteet ja tutkimuksen kulku.....	72
8.1.1	Tavoitteet ja niiden saavuttamiseksi tehty toiminta.....	72
8.1.2	Aineiston kokoamisessa noudatetut toimintamallit.....	73
8.2	Aineiston käsittelyssä tehtyjä havaintoja	74
8.2.1	Aineiston määrällisen käsittelyn tulokset	75
8.2.2	Käytettyjen lähteiden kanssa yhtenevät tulokset	78
8.2.3	Aineiston käsittelyn yhteydessä tehdyt huomiot	80
8.2.4	Tulosten pohjalta syntyneet uudet kysymykset	87
8.3	Tulosten rajoitukset	89
9	Lopuksi.....	91
9.1	Yhteenveto tutkimuksen etenemisestä	91
9.2	Saadut tulokset	91
9.3	Uudet tutkimusongelmat	93
	Lähteet	97

Liite 1: Tutkimusaineistot

1 Johdanto

1.1 Tutkimuksen aihe

Työssä tutkitaan yleisten ontologioiden eli käsittekoelmien sopivuutta yksittäisten henkilöitten itse laatimien elämäkertatekstien jäsentämiseen. Tutkimuksessa toteutetussa järjestelyssä elämäkertojen kirjoittajat jäsensivät ensin tekstiään itse määrittämillään avainsanoilla, ja sen jälkeen tutkittiin, kuinka suuri osa avainsanoista sisältyi ontologiakirjasto Finton ontologioihin. Prosessi toistettiin myös niin, että avainsanajäsennyksen teki kirjoittajan sijasta tutkija.

Toteutetulla järjestelyllä selvitettiin, kuinka yhteensopivia julkiset ontologiat ovat suhteessa käyttäjien laatimiin avainsanajäsennyksiin. Miten kirjoittajan itsensä ja toisaalta myös tutkijan tekemä jäsenitys tekstin avainkäsitteistä toimii yhteen yleisesti käytettyjen ontologioiden kanssa? Sisältävätkö ontologiat niitä samoja käsitteitä, joita kirjoittajat itse käyttävät? Minkälaisia ominaisuuksia ja mahdollisia ongelmia liittyy loppukäyttäjien itse tekemiin avainsanajäsennyksiin?

Ovatko julkiset ontologiat käyttökelpoisia henkilöiden omaa elämää kuvaavien tekstien jäsentämisessä? Julkisten ontologioiden käyttökelpoisuus näkyy siinä, sisältävätkö ontologiat niitä käsitteitä, joita kirjoittajat tarvitsevat jäsentämistyöhönsä. Voidaanko yleisiä ontologioita käyttää apuna, kun esim. halutaan varmistaa, että elämäkertatekstin mukana säilyisi tietoa sen käsitteiden merkityksestä?

Tämä on nousemassa tärkeäksi ja hedelmälliseksi tutkimuskysymykseksi sitä mukaa, kun käyttäjien itse webiin tuottamien aineistojen määrä kasvaa. Samanaikaisesti myös ontologioita kehitetään ja laajennetaan.¹ Koska kehittämistyö on jatkuvaa, kertoo nyt saatu vastaus vain hetkellisestä tilanteesta. Tämäntyyppisten tarkastelujen tekeminen on kuitenkin tärkeää, jotta saadaan selville, missä määrin olemassa oleviin tarpeisiin on tähän mennessä pystytty vastaamaan.

Elämäkerrat kirjoitetaan tuleville sukupolville, ja kirjoittaja luonnollisesti toivoo, että tulevaisuuden lukija ymmärtää, mitä teksti käsittelee. Sisällön merkityksen säilyminen ei kuitenkaan ole varmaa, sillä kieli muuttuu koko ajan ja arkielämän käytäntöjen muut-

¹ Ks. esim. Finto-projekti <https://www.kiwi.fi/display/Finto/Organisointi> (viitattu 23.6.2017)

tumisen myötä myös osasta käsitteitä tulee uusille sukupolville tuntemattomia. Samaan käsitteeseen voidaan mahdollisesti myöhemmin viitata eri termillä kuin ennen.

Jotta välttyttäisiin näiltä ongelmilta, tekstin merkityksen säilymistä voidaan tukea sitomalla sisältö yleisesti tunnettuihin käsitteisiin, jotka on sisällytetty julkisiin ontologioihin. Merkityksen säilyminen edellyttää tietysti myös sitä, että ontologioista ei poisteta vanhoja käsitteitä, vaan ne tarpeen vaatiessa määritellään uudelleen uudempien käsitteiden avulla.

Tämän työn tarkoituksena on siis tutkia, missä määrin julkisia ontologioita voidaan käyttää yksityishenkilön omien, itse laatimien ja henkilökohtaisten elämäkerta-aineistojen jäsentämiseen. Tähän pyritään saamaan vastaus vertaamalla elämäkerta-aineistojen avainkäsitteitä ja julkisen Finto-ontologiakirjaston ontologioiden sisältöjä. Samalla saadaan kokemusperäistä tietoa, millaisia piirteitä ja mahdollisia ongelmia sisältyy käyttäjien tekemiin avainsanajäsennyksiin ja niiden yhdistämiseen julkisten ontologioiden kanssa. Aineistona ovat kansalaisopiston elämäkertapiirin jäsenten tuottamat tekstit.

1.2 Työn yhteys muihin vastaaviin tutkimuksiin

Kirjallisuuskartoituksen perusteella näytti, että samantyyppistä asiaa oli tutkittu, mutta hieman erilaisissa ympäristöissä. Tähän mennessä on tehty tutkimusta mm. blogikirjoitusten ja erilaisten verkossa toimivien yhteisö sivujen jäsentämisestä ontologian avulla, ks. [Passa2007]. On myös tutkittu eri tapoja yhteensovittaa käyttäjien itse laatimia avainsanalistoja ja formaaleja ontologioita, ks. esim. [Limpe2009]. Käyttäjien tekemien luokitteluiden ja ontologioiden yhdistämistä on tutkittu myös monella muulla eri tavalla, niitä kuvataan luvussa 5.

Aikaisemmassa tutkimuksessa on siis käsitelty samansuuntaisia kysymyksiä, mutta hyvin erilaisissa ympäristöissä ja erilaisilla aineistoilla. Tässä työssä mukana olevat aiemmat tutkimukset on toteutettu työympäristössä, mikä tekee sekä käytettävän kielen että terminologian virallisemmaksi. Voisiko vastaavaa soveltaa yksityishenkilöiden digitaalisten, elämänkaarta kartoittavien tekstiaineistojen jäsentämiseen? Onko käyttäjien omilla jäsenyksillä yhtymäkohtia yleisesti tarjolla oleviin ontologioihin?

Ontologioiden ja käyttäjien tekemien jäsenysten yhdistämiseen liittyvää tutkimusta on aiemmin tehty muissa kieliympäristöissä (työssä viitataan pääasiassa englanninkielisiin lähteisiin). Miten tähän mennessä sovelletut mallit toimivat suomenkielisen aineiston kanssa työskenneltäessä? Suomen kieli asettaa omia lisähaasteitaan sen vuoksi, että taipumusmuotojen avulla voidaan säätää semanttista sisältöä ja taipumusmuodot asettavat

lisähaasteita hakutoiminnoille. Lisäksi osa julkisista ontologioista on maa- ja kulttuurisidonnaisia. Toimintamalleja on siis testattava täällä meillä erikseen ja verrattava, toimivatko muihin kieliympäristöihin tarjotut toimintatavat myös suomenkielisen tekstin jäsentämisessä.

Tarkoituksena on pitäytyä suomenkielisissä ontologioissa, koska elämäkerta-aineistot on kirjoitettu suomen kielellä. Maailmalla on toki tehty englanninkielisiä, kansainvälisessä käytössä olevia ontologioita, joissa hyvinkin saattaisi olla hyödynnettävissä olevia käsitteistöjä. Englanninkieliset ontologiat päätetään kuitenkin sulkea kokonaan työskentelyn ulkopuolelle, ellei niitä ole valmiiksi käännetty suomen kielelle.

Perustelu suomenkielisissä aineistoissa pitäytymiseen on kääntämisen vaatima asiantuntemus. Koska elämäkerta-aineistojen sisällöissä käytettävät käsitteet ovat suurelta osin arkielämän alueella, ei tietojenkäsittelytieteen koulutuksen antama asiantuntemus riitä pohjaksi käsitteistön kääntämiseen suomesta englanniksi ja päinvastoin. Siinä pitäisi olla joko käännöstieteellistä asiantuntemusta tai lähellä äidinkieltä oleva englannin kielen taito. Ontologioita on kuitenkin jo käännetty valmiiksi muista kielistä suomeen, sillä osa meilläkin julkisesti käytössä olevista ontologioista on alunperin laadittu jollain muulla kielellä.

Tämä työ poikkeaa aiemmin tehdyistä vastaavista tutkimuksista sekä kohteena olevien tekstien aiheen että toteutusympäristön suhteen: tässä tutkitaan suomenkielisten elämäkerta-aineistojen jäsentämistä. Työn tarkoitusta ja tavoitteita kuvaillaan yksityiskohtaisemmin luvussa 2, jossa selvitetään työn tavoitteet ja työskentelyn etenemisen karkeat suuntaviivat.

Sen jälkeen luvussa 3 esitetään työn tutkimuskysymys ja määritellään, millä menetelmällä siihen pyritään vastaamaan. Luvussa 4 esitellään työn kannalta keskeiset käsitteet: ontologia, folksonomia, semanttinen web, elämäkerta sekä muistitieto.

Luvussa 5 käydään läpi aikaisempaa aiheeseen liittyvää tutkimusta. Aikaisemmasta tutkimuksesta on poimittu erilaisia näkemyksiä ja menetelmiä folksonomioiden ja ontologioiden yhdistämiseksi. Sen jälkeen esitellään tutkimuksen toteutusympäristö (luku 6). Työn läpivienti sai alkunsa Tikutus-hankkeen pilottiryhmänä ja jatkui sen jälkeen vielä itsenäisenä ryhmänä muutaman lukuvuoden ajan.

Luvussa 7 on kuvattu työskentelyn vaiheet, minkä jälkeen esitellään työn tulokset (luku 8) sekä loppuyhteenvedo ja keskustelu luvussa 9.

2 Työn tavoitteet ja työskentelyn kulku

Tässä luvussa määritellään aluksi tutkimuksen tavoite kohdassa 2.1. Kohdassa 2.2 esitetään, miten tutkimuskysymykseen vastataan esittelemällä tutkimuksen kulku pääpiirteittäin.

2.1 Tutkimuksen tarkoitus

Tutkimuksen tarkoitus kuvataan määrittelemällä tutkimuskysymys: minkälaista tukea ontologioista on tällä hetkellä saatavissa elämäkertatekstien jäsentämiseen ja mitä ominaispiirteitä sekä mahdollisia ongelmia ontologioiden käyttöön nykyisellään liittyy.

Miksi ontologioiden käyttö ylipäätään olisi tarpeellista? Elämäkertatiedon kyseessä ollessa olisi tärkeää, että merkitykset säilyisivät mahdollisimman yksiselitteisinä tuleville lukijoille. Mikäli lukijat ovat kirjoittajan jälkeläisiä, kiinnostaa heitä todennäköisesti esivanhempien elämä kaikkine yksityiskohtineen. Yleisellä, koko yhteiskunnan tasolla tarkasteltuna elämäkerrat ovat yksi dokumenttimuoto, jonka välityksellä meidän kulttuuriperintömme sekä arjen mikrohistoria säilyvät tuleville sukupolville.

Elämänvaiheita kuvaava multimedia-aineisto voi sisältää esim. valokuvia, joiden sisältö tulee myöhemmin olemaan katsojille epäselvä (kuvissa näkyviä esineitä tai laitteita ei enää käytetä tai ne näyttävät nykyään erilaisilta). Myös tekstissä voidaan kuvata menetelmiä tai välineitä, jotka ovat myöhempien sukupolvien lukijoille tuntemattomia. Yleiset ontologiat voivat olla välineitä, joiden avulla jo aktiivikäytöstä poistuneiden käsitteiden merkitys on mahdollisesti edelleen tulkittavissa.

Jotta yleiset ontologiat auttaisivat merkityksen välittämisessä tulevaisuuteen, tulee niiden nykyhetkellä sisältää käsitteitä, jotka ovat tarpeellisia arkielämää kuvaavien kirjoitusten jäsentämisessä. Sen vuoksi pyritään selvittämään, sisältävätkö julkiset ontologiat samoja käsitteitä, joita a) elämäkertojen kirjoittajat tai b) tekstejä lukeva tutkija käyttävät elämäkertatekstin jäsentämisessä. Työssä mukana olevat aineistot on kaikki ja sen netetty kahden eri henkilön toimesta, jotta saadaan samalla laajempi näkökulma.

Työn tarkoituksena on siis selvittää, voidaanko olemassa olevia ontologiakirjastoja hyödyntää henkilöiden vapaasti luomien elämäkertojen jäsentäjinä. Tähän kysymykseen ei suoraan löydy vastausta aiemmin tehdystä tutkimuksesta, vaikkakin läheisesti samansuuntaisia tutkimuksia on tehty (ks. luku 5). Aikaisemmissa tutkimuksissa ontologioita

on sovellettu erilaisiin aineistoihin kuin tässä työssä: suurin ero on, että aiemmissa tutkimuksissa aineistot ovat olleet työtehtäviin liittyviä.

Koska nykyään suuri osa arkipäivän dokumenteista tallentuu digitaaliseen muotoon, on tärkeää tuottaa ja hyödyntää välineitä, joilla säilyneitten aineistojen merkityssisältöjä saadaan välitettyä tulevaisuuteen. Tutkimuksessa ei kuitenkaan ole voitu tarkastella käsitteiden muuttumista tekemällä vertailua kirjoittamishetken ja jonkin sitä seuranneen tarkasteluajankohdan välillä, sillä tähän ei ole ollut resursseja käytettävissä. Sen sijaan aikaperspektiiviin päästään jossain määrin käsiksi sitä kautta, että tekstit kertovat menneistä tapahtumista mahdollisesti vielä sen ajan käsitteistöllä. Lukeminen ja tulkinta puolestaan tapahtuvat tässä päivässä. Työssä asetettuja kysymyksiä tutkitaan käyttämällä aineistona kansalaisopiston ikäihmisen elämäkertapiirin tuottamia elektronisia muisteluaineistoja, jotka on kirjoitettu hypertekstimuotoon.

Ontologinen jäsentäminen mahdollisesti auttaa myös hypertekstin rakenteen ja sisällön hahmottamisessa, vaikka se ei kuuluakaan tutkimuksen kannalta relevantteihin kysymyksiin. On olemassa työkaluja, joiden avulla avainsanoista voi helposti tehdä hypertekstidokumentin yhteyteen hakemiston. Tällainen hakemisto helpottaa lukijan kannalta sisällön hahmottamista sekä määrättyjen sisältökohtien etsintää. Seikka puoltanee osaltaan avainsanojen käyttöä ryhmän työkaluna, vaikkei työskenneltäisikään erityisesti niitä koskevien tutkimuskysymysten parissa.

Tässä työssä tehdyn kokeilun kontekstina oli Tikutus-hanke, jossa elämäkerta-aineistoa hyödynnettiin myös muistisairaitten hoidossa sekä maahanmuuttajien viranomaiskohtaamisissa. Silloin on oleellista, että elämäkerta-aineiston sisältö on käsitteellisesti yksiselitteisesti yhdistettävissä eri palveluiden sisältämään, kansalaista koskevaan tietoihin.

2.2 Tutkimuksen kulku

Tämän tutkimuksen aineistona olevia elämäkertatekstejä jäsentävät ensin vapaasti tekstien kirjoittajat heidän omien näkemystensä mukaan. Myös tutkija tekee teksteihin omat avainsanajäsennyksensä ilman, että on perehtynyt kirjoittajien merkitsemiin avainsanoihin. Tarkoituksena on yleisestä mielenkiinnosta käyttää hyväksi tilaisuus nähdä, miten paljon avainsanajäsennykset poikkeavat, kun niiden tekijöinä on eri näkökulmasta tekstejä tarkastelevat henkilöt.

Tutkija edustaa tässä tapauksessa lukijaa, jolle tekstin aihepiiri ei ole tuttu ja joka syntymäaikansa perusteella on kirjoittajien lasten ikäluokkaa. Tutkija myös pyrkii omissa jäsenyksissään ottamaan huomioon termit, joiden merkitys on nuoremman sukupolven

lukijoille epäselvä. Avainsanojen merkitsemisen jälkeen kummastakin jäsenyyksestä syntyneitä sanalistoja tarkastellaan käyttäen hyväksi suomalaisen Finto-ontologiapalvelun ontologioita [Finto2014].

Finto-asiasanasto- ja ontologiapalvelu on Kansalliskirjaston, opetus- ja kulttuuriministeriön sekä valtiovarainministeriön yhteisen ONKI-projektin kehittämä kansallinen ontologiakirjasto, johon on koottu sekä suomalaisia että kansainvälisiä ontologioita. Palvelun sisältöä ei ole rajattu mihinkään määrättyihin aihealueisiin, ja se on tarkoitettu palvelemaan yleisesti erilaisia ontologioiden käyttötarkoituksia: ontologioiden kehitystyötä, aineistojen indeksointia ja hakujen tekemistä [Hyvön2007]. (ks. myös D’Aquinin ja Noyn [2012] vertailu eri ontologiakirjastoista).

Tämän työn puitteissa tehtävien avainsanajäsennysten avulla verrataan, kuinka paljon yhteisiä käsitteitä on 1) kirjoittajien jäsenyyksissä ja ontologioissa sekä 2) tutkijan jäsenyyksissä ja ontologioissa. Näiden vertailujen avulla pyritään selvittämään, soveltuvatko saatavilla olevat ontologiat tämäntyyppisen aineiston jäsentämiseen.

Vertailut antavat suuntaa, onko semanttisesta jäsentämisestä apua hypertextimuodossa olevan elämäkerta-aineiston sisällön määrittelyssä, jos avainsanat halutaan yhdistää julkisiin ontologioihin. Sisältyykö ontologioihin sellaisia käsitteitä, jotka ovat keskeisiä elämäkerta-aineistoissa? Voisivatko ontologiat täten edes teoriassa auttaa merkityksen säilymisessä tulevaisuuteen? Tätähän emme voi testata käytännössä, koska emme voi nähdä, miten jäsenyykset toimivat tulevaisuudessa.

Työn kautta käy myös selville, missä määrin nimenomaan Finto-ontologiapalvelun tarjoamat ontologiat ovat toimivia elämäkertojen jäsentämisessä. Finto-kirjasto on jatkuvasti laajeneva kokoelma, johon kirjaston käyttäjät voivat ehdottaa lisäyksiä. Tarpeen vaatiessa voitaisiin siis saada aikaiseksi joko ehdotuksia uusista ontologioista tai täydennystä jo olemassa oleviin ontologioihin.

3 Tutkimusmenetelmät

Tutkimusmenetelmät-luvussa kuvataan, millä menetelmillä tutkimuskysymykseen haetaan vastauksia. Kohdassa 3.1 käsitellään tutkimuksen tutkimuskysymystä: miten julki-set ontologiat soveltuvat elämäkertatekstien jäsentämiseen. Kohdassa 3.2 kuvataan, minkälaisilla menetelmillä siihen haetaan vastauksia. Sen jälkeen käydään läpi työskentelyn periaatteita sekä työn tulosten yleistettävyyttä suhteessa aineiston pienuuteen (kohta 3.3).

3.1 Tutkimuskysymys

Työn tutkimuksen kohteena on kysymys ontologioiden toimivuudesta elämäkertojen jäsentämisessä. Tämä kysymys jakautuu vielä kahteen osaan: a) miten käyttäjien omat avainsanajäsennykset eli folksonomiat¹ ja yleiset ontologiat kohtaavat ja b) minkälaisia ominaispiirteitä ja mahdollisia ongelmia liittyy ontologioiden käyttöön tämäntyyppisten, käyttäjien laatimien ja arkipäivää kuvaavien aineistojen jäsentämisessä.

Ensimmäinen osakysymys kertoo käyttäjien valitsemien käsitteiden ja asiantuntijoiden laatimien käsittekirjastojen kohtaavuudesta. Tutkimusintressi on alunperin lähtenyt liikkeelle epäilystä, että ”viralliset”, asiantuntijoiden laatimat ontologiat ovat kaukana tavallisten webin käyttäjien laatimista avainsanalistoista, joihin internetin sisältöjen löytäminen pitkälti myös perustuu - ainakin silloin, kun kyseessä ovat käyttäjien itsensä laatimat aineistot.

Kysymyksen ensimmäisessä puoliskossa verrataan käyttäjän itsensä muodostamaa folksonomiaa ja yleistä ontologiaa. Tätä varmistetaan tekemällä myös toinen vertailu, jossa kohteena ovat tutkijan tekemät avainsanajäsennykset ja ontologiat. Jakamalla tämä vertailu kahteen osaan jätetään varaus sellaiselle tulokselle, että ontologiat saattaisivat sisältää elämäkerta-aineistojen käsitteitä, mutta käyttäjät eivät nosta kyseisiä käsitteitä avainsanoiksi tai ne eivät jostain syystä löydy siinä muodossa kuin käyttäjät niitä hakevat.

Kysymyksen toiseen osaan saadaan vastauksia edellä kuvatun käsitevertailun suorittamisen yhteydessä. Tämä kysymys on sekundäärinen suhteessa ensimmäiseen kysymyk-

¹ Folksonomian tarkempi määrittely on kohdassa 4.2.

seen, mutta yhtä lailla sitä pidetään tärkeänä kysymyksenä ja myös sen osalta tehdään vertailuja lähdekirjallisuuteen.

3.2 Työssä käytetyt tutkimusmenetelmät

Työn metodologisessa pohjassa yhdistetään monia eri menetelmiä: se on teorioita testaavaa tutkimusta, jossa vertailtava aineisto kootaan tapaustutkimuksesta. Saatuja tuloksia verrataan aikaisempien tutkimusten tuloksiin. Tutkimus edustaa myös suunnitteluteollista arvioivaa tutkimusta, sillä se arvioi tehtyä artefaktia ja esittää siihen kehittämisehdotuksia.

3.2.1 Teorioita testaava case-tutkimus

Tutkimuksen pohjaksi on ensin kartoitettu aiempaa tutkimusta, joka koskee ontologioiden käyttöä sekä ontologioiden ja folksonomioiden (eli käyttäjien itsensä tekemien avainsanaluokitteluitten) yhdistämistä. Niistä on koottu tuloksia verrattavaksi tämän tutkimuksen tuottamaan aineistoon.

Työ testaa siis aikaisempaa tutkimusta. Omaa tutkimustani vastaavia kysymyksiä on tutkittu aiemmin erilaisilla aineistoilla (nyt elämäkerta-aineistoilla) ja lopuksi on tarkoituksena verrata omia tuloksia sekä näitä aiempia tutkimuksia. Vertailun avulla selvitetään, pätevätkö lähdeaineiston aiemmista tutkimuksista löydetyt seikat myös tämän tutkimuksen tapauksissa. Tutkimusten tulosten vertailun osalta kyseessä on teorioita testaava tutkimus [Järvi2011].

Aineiston kannalta tarkasteltuna tutkimus on luonteeltaan case-tutkimusta, joka koostuu useasta tapauksesta [Järvi2011]. Tarkastelun kohteena tulee olemaan viisi näytettä, jotka on kaikki otettu eri muisteluaineistoista. Jokainen tutkimukseen osallistuva elämäkerta-aineisto on oma tapauksensa, sillä aineistot ovat keskenään hyvin erilaisia sekä lähestymistavaltaan että rakenteensa ja sisältönsä puolesta.

Vaikka kirjoituksilla on yhteinen teema – oman elämän tai suvun vaiheiden muistelu – on tuloksena hyvin erilaisia tekstejä. Kirjoittajien lähtökohdat vaihtelevat myös, sillä elämäkertapiirin osallistujien taustat ovat hyvin erilaiset, samoin heidän tietotekniset valmiutensa¹ sekä kokeneisuutensa kirjallisen tekstin tuottajana.

¹ Tietoteknistien valmiuksien ajatellaan tässä tapauksessa vaikuttavan lähinnä siihen, miten he hahmottavat avainsanojen merkityksen webin sisältöjen jäsentämisessä ja kuinka tuttu kyseinen käytäntö heille on.

Materiaali on kuitenkin siinä mielessä homogeenista, että kaikki osallistujat ovat saaneet saman verran ohjausta ja ohjaajina ovat olleet samat henkilöt. Kaikki aineistot myös sisältävät elämänvaiheita kuvailevaa tekstiä.

3.2.2 Suunnittelutieteellinen arvioiva tutkimus

Tutkimuksen käytännöllisessä osuudessa elämänvaiheita kuvaavia dokumentteja jäsenetään ensin valitsemalla aineistoille avainsanoja. Näin syntyneitä avainsanalistoja tutkitaan sen jälkeen suhteessa yleisen ontologiakirjaston, Finton, sisältöön.

Elämäkerta-aineistojen avulla haetaan vastauksia edellä esitettyihin, ontologioiden soveltuvuutta selvittäviin tutkimuskysymyksiin, joten työni on tältä osin luonteeltaan suunnittelutieteellistä arvioivaa tutkimusta. Suunnittelutieteellinen arvioiva tutkimus määrittelee tutkittavan artefaktin tai innovaation arvioinnissa tarvittavan mittarin/mittarit ja evaluoi tätä kohdetta kyseisten mittareiden avulla [Järvi2011].

Jotta tutkimuksen kysymykseen saadaan vastaus, tulee ontologioita arvioida niiden kattavuuden kannalta. Arvioinnin kohteena oleva artefakti on tässä tapauksessa käsitteikirjasto (eli ontologiat) ja sitä arvioidaan nimenomaan kattavuuden osalta: sisältääkö yleinen ontologiakirjasto sellaisia käsitteitä, joita tarvitaan elämäkertatekstien jäsentämiseen. Vastaus saadaan vertailemalla käyttäjien ja tutkijan tekemiä avainsanajäsennyksiä ontologiakirjaston sisältöön.

Suunnittelutieteellinen tutkimus käsittää yleensä jonkin ongelman ratkaisuksi tarkoitettua artefaktin suunnittelun, toteutuksen sekä arvioinnin. Peffersin ym:iden mukaan suunnittelutieteellinen tutkimus voi myös lähteä liikkeelle olemassa olevan artefaktin tai innovaation arvioinnista esimerkiksi niin, että sitä sovelletaan uudella tavalla. Tällöin tutkimuksen tuloksena on esim. arvio artefaktin sopivuudesta kyseiseen tarkoitukseen ja suunnitelma mahdollisista kehittämistoimenpiteistä, jotka myös toteutetaan. [Peffe2007]

Tässä työssä tutkija osallistui hypertekstin suunnitteluun ja aineistojen kokoamiseen yhteistyössä aineistojen laatijoiden kanssa. Tutkija myös tuotti oman näkemyksensä perusteella ehdotuksen elämäkertojen jäsennyksessä käytettävistä käsitteistä. Tutkimusmenetelmä sisältää näin suunnittelututkimuksen ja toimintatutkimuksen piirteitä, sillä tutkija osallistuu aktiivisesti artefaktin evaluointiin (ks. menetelmien tunnusmerkkejä [Järvi2012]).

Oma työni sisältää artefaktin arvioinnin ja ehdotuksia kehittämistoimenpiteiksi. Koska arvioinnin kohteena on käytössä oleva, yleinen palvelu (Finto-ontologiakirjasto), ei opinnäytetyön resursseilla ole mahdollista lähteä toteuttamaan siihen kehittämistoimia. Prosessi päättyy siis siihen, että tulokset julkaistaan avoimeen keskusteluun ja näin ol-

len myös Finto-palvelun kehittämisestä vastaavien tahojen saataville. Tässä mielessä työni ei edusta suunnittelutieteellistä tutkimusta siten kuin se tyypillisimmillään tunnetaan.

3.3 Työskentelyssä sovellettuja periaatteita

Käytettävistä tutkimusmenetelmistä käsitellään erityisesti ennaltaohjailemattomuuden periaatetta ja sen vaikutuksia sekä tutkimuksen laadullista luonnetta, joka mahdollistaa tietyissä määrin myös yleistämisen pienestä aineistosta.

3.3.1 Ennaltaohjailemattomuus

Liiallista havaintojen tai mielipiteitten etukäteen jäsentämistä kannattaa välttää, jotta saisi kohteesta aidomman kuvan, ei sellaista, joka on omien ennakkokäsitysten läpi suodatettu. Havainnoinnin heikkous on se, että omat asenteet vaikuttavat siihen, mitä tutkija havaitsee. Tästä puhuvat esimerkiksi Golden-Biddle ja Locke etnografisten tutkimustekstien ominaisuuksia käsittelevässä artikkelissaan [Golde1993]. Samaan on kiinnittänyt huomiota Eisenhardt [1989], joka jäsentää case-tutkimuksiin perustuvan teorianmuodostuksen vaiheita. Hän painottaa paljon teoreettisen joustavuuden tärkeyttä.

Pyrkimys ohjailemattomuuteen koskee myös elämäkerta-aineistojen jäsentämistä: autenttisemman tuloksen saavuttamiseksi aineiston laatijat saavat ensin itse määrittellä avainsanansa ja vasta sen jälkeen tutkija tekee aineistoon omat jäsenyskierroksensa. On tärkeää, että ensimmäistä avainsanojen määrittelyä ei ohjailta, sillä normaalisti käyttäjien itsensä webiin tuottama sisältö on aina jäsenetty laatijoiden itsensä toimesta ilman ulkopuolista ohjausta tai opastusta.

Ensimmäinen avainsanajäsennys tehdään siis ilman ontologioita ja vasta sitten toteutetaan ontologioiden avulla vertailu niistä avainsanoista, joita aineistoista on nostettu esiin. Näin pidetään huoli siitä, että tieto ontologioiden sisällöistä ei ohjaa sitä, mitä käyttäjät valitsevat avainsanoiksi.¹

Ennaltaohjailemattomuuden pyrkimys muistuttaa paljon niitä piirteitä, joita kuuluu grounded theory -tutkimuksiin, ks. [Järvi2011]. Aineiston käsittelyssä on muitakin haas-

¹ Huomautuksena lisättäköön, että tavallisesti ontologioiden "normaalikäytössä" toimitaan päinvastoin: tavoitteena on valita aineiston kuvailuun nimenomaan sellaisia avainsanoja, jotka sisältyvät ontologiaan/-oihin. Tarvittaessa valintaa voidaan tukea hakemalla ontologioista mahdollisia avainsanaehdokkaita tarjolle kuvailutietojen syöttämisen yhteydessä.

teita, jotka ovat samankaltaisia kuin grounded theory -lähestymistavassa: tutkijan on oltava riittävän kokenut ja tarkka päätelmissään, jotta hän pystyy erottamaan todelliset ilmiöt siitä, mitä ovat hänen omat oletuksensa tai aikaisempien teorioitten oletukset.

Tässä tutkimuksessa ei kuitenkaan ole tavoitteena luoda uutta teoreettista mallia kuten grounded theory -tutkimuksissa, joten perimmäisenä tutkimusmenetelmänä voimme pitää suunnittelutieteellistä tutkimusta. Siinä on tarkoituksena arvioida määrätyn toimintamallin toimivuutta [Järvi2011].

3.3.2 Kvalitatiivinen lähestymistapa

Tutkimus on luonteeltaan enemmän kvalitatiivinen kuin kvantitatiivinen. Esimerkkiaineistojen määrä on pieni, mutta oletuksena on, että tutkimuksen tulokset ovat suuntaantavia ja ne toistavat pienessä mittakaavassa samoja suuntaviivoja kuin suuremmilla-kin aineistoilla kokeiltaessa.

Tilastollista näyttöä tällaisella aineistolla ei saavuteta, mutta esim. käytettävyytsteissä pidetään usein riittävänä, että testit toteutetaan 4-5 henkilöllä kustakin käyttäjäryhmästä. Suurempien testiaineistojen on havaittu vain toistavan ensin saatuja tuloksia [Dumas1999]. Toki samalla täytyy muistaa, että käytettävyytstestit eivät ole tieteellisiä tutkimuksia.

Flyvberg [2011] nimeää jopa yhdeksi tieteelliseksi väärinkäsitykseksi sen, että tapaus-tutkimuksen perusteella ei voisi tehdä yleistyksiä. Yleistämisestä ovat kirjoittaneet myös Lee ja Baskerville [2012]. He ovat erottaneet erilaisia yleistämistyyppjejä, joista yksi on empiirinen yleistäminen otoksesta populaatioon. Tässä on kyse juuri tuosta yleistämistyyppistä: oletetaan, että pienen otoksen löydökset pätisivät myös suuremmassa populaatiossa.

Erityisesti tuloksella on merkitystä silloin, jos se on selkeän negatiivinen: folksonomiat ja ontologiat eivät kohtaa. Jos folksonomiat ja ontologiat osoittautuvat selkeästi eriäviksi käsittekokoelmiksi, on se huomionarvoinen tulos otoksen pienuudesta huolimatta.

Positiivinen tulos taas vahvistaa asian jatkokehittelyn tärkeyttä: kyseessä on toimintamalli, jonka käytöstä voi olla hyötyä ja jota tulisi sekä tutkia että kehittää edelleen. Vaikka tulokset eivät antaisikaan tukea alkuolettamuksille, on tutkimuksen läpiviemisestä hyötyä: Davisin mukaan tieto sekä kehittämistyön onnistumisesta että epäonnistumisesta on hyödyllistä [Davis2000].

4 Tutkimuksen kannalta keskeisten käsitteiden määrittely

Seuraavaksi määritellään käsitteet, jotka ovat tässä tutkimuksessa keskeisiä. Kyseiset käsitteet määrittävät tutkimuksen kohteen ja osa niistä on toiminut hakuterminä, kun on haettu aiempaa aihetta koskevaa tutkimusta. Vaikka kyse on tietojenkäsittelytieteen alalla tunnetuista termeistä, on niiden huolellinen määrittely ja pohtiminen tärkeää, sillä käsitteet ovat kohtalaisen uusia, ne määritellään osassa lähteitä toisistaan poikkeavasti eivätkä ne ole lukijoille tuttuja normaalin yleiskielen perusteella.

Keskeisinä käsitteinä määritellään ensin ontologiat (kohta 4.1) ja folksonomiat (kohta 4.2). Ontologioiden ja folksonomioiden välisestä suhteesta on olemassa hyvin erilaisia käsityksiä, niitä on koottu kohtaan 4.3. Tämän jälkeen määritellään semanttinen web sekä web 2.0 kohdassa 4.4. Lopuksi tehdään katsaus työn sovellusalueen kannalta keskeisiin käsitteisiin ”elämäkerta” (kohta 4.5) sekä ”muistitieto” (kohta 4.6). Nämä siksi, että olisi selkeästi määritelty, mitä tämän tutkimuksen esimerkkiaineistot sisälsivät ja mikä on niiden rooli ja merkitys tietona ylipäätään: miksi niitä kannattaa tutkia ja miksi niiden sisältöjen ymmärtäminen on tärkeää.

4.1 Ontologiat

Ontologia on perinteisesti tunnettu filosofian yhtenä osa-alueena, mutta viime vuosina ontologia-termi on esiintynyt yhä enemmän myös tietojenkäsittelytieteiden puolella. Kyse on eri asioista, joten on syytä esittää määritelmä, mitä ontologialla tarkoitetaan. Lopuksi käsitellään kirjallisuuden pohjalta myös erilaisia ontologioihin liittyviä ongelmia.

4.1.1 Mitä ontologioilla tarkoitetaan?

Ontologia filosofiassa

Filosofiassa ontologia on oppi olemassaolosta. Ontologian kiinnostuksen kohteisiin kuuluvat mm. se, mitä on olemassa, mitä olevainen on ja eri tavat olla olemassa. Ontologia tutkii siis kysymyksiä siitä, millaisista elementeistä olemassa oleva koostuu, mistä

aineesta maailma on tehty ja millä tavalla jokin voi olla olemassa (esim. täytyykö olemassaolon osoituksena olevaiselle olla konkreettinen vastine). [Palom1991]

Näin ollen ontologiassa voidaan erottaa useita eri lähestymistapoja: joidenkin määrittelyiden mukaan ontologian tehtävänä on määrittellä, miten olevainen jakautuu eri substansseihin tai asioihin. Esim. olevainen voidaan määrittellä tapahtumina ja prosesseina. Ontologia voidaan myös hahmottaa kaikkien entiteettien luokittelun sisältävänä taksonomiana¹. Eräiden lähestymistapojen mukaan kyse on olevaisen palauttamisesta niihin pienimpiin osasiin, joista se muodostuu. [Smith2004]

1900-luvulta lähtien on ajateltu, että ontologia ei ole metamalli olevaisesta, vaan ontologiassa mallintamisen kohteena ovat teoriat, kieli ja uskomusjärjestelmät, jotka puolestaan kuvaavat käsitystämme olevaisesta [Smith2004]. Enää ei edes oleteta, että pystyisimme tyhjentävästi kuvaamaan tai määrittämään jotain kohdetta, vaan ontologinen kuvaus perustuu nimenomaan kulloistakin kohdetta määrittelevään teoriaan.

Esimerkkinä edellisestä voidaan viitata paljon siteerattuun Karl Popperin jäsennykseen, jossa on erotettu kolme maailmaa: 1. fyysisten olevaisten maailma, 2. ihmisten tietoisuuden maailma ja 3. tiedon maailma. Maailma 3 (tiedon maailma) sisältää kaikki maailma 1:tä ja maailma 2:ta koskevat teoriat [Poppe1979]. Tämän jäsennyksen mukaan ontologia mallintaa maailmaan 3 kuuluvia käsityksiämme. Popperin ajatusmalliin kuuluu myös oletus siitä, että maailmoja 1 ja 2 koskeva tieto ja käsitykset muuttuvat ajan kuluessa.

Ontologia on siis kohdealuetta määrittelevien ontologisten esitysten sisältöä ja rakentamista koskeva teoria. Ontologisten kysymysten pohtiminen on laajentunut filosofiasta muihin tieteisiin: myös kielitieteilijät, psykologit ja antropologit ovat pyrkineet määrittelemään eri kulttuureiden ja ryhmien ontologisia sitoumuksia. [Smith2004]

Ontologia tietojenkäsittelytieteessä

Tietojenkäsittelytieteessä ontologia-käsitettä käytetään suppeamassa merkityksessä kuin filosofiassa: ontologia käsittelee eksistenssiin liittyviä kysymyksiä korkeintaan välillisesti, sen sijaan painopiste on käsitteiden määrittelyssä. Ontologioiden rakentaminen katsotaan kuuluvaksi käsitteellisen mallintamisen alle.

Gruberin mukaan ontologia on käsitteellisen määrittämisen tuloksena syntynyt kuvaus [Grube1995]. Käsitteet puolestaan ovat yksinkertaistettuja ja abstrakteja välineitä, joita me käytämme reaali maailman kuvaamiseen. Ontologia ei siis kuvaa reaali maailmaa, vaan meidän käsityksiämme siitä.

¹ Taksonomia = luokittelu, systematiikka ks. <http://www.kielitoimistonsanakirja.fi/>

Studerin, Benjaminsin ja Fenselin määritelmän [1998] mukaan ontologia on yhteisesti jaettu, määrämuotoinen ja eksplisiittinen, käsitteellistämisen tuloksena syntynyt määrittely. Heidän mukaansa kaikki em. määreet ovat oleellisia: määrämuotoisuus tukee mahdollisuutta lukea ontologioita koneellisesti, eksplisiittisyys tarkoittaa sitä, että käsitteet ja kaikki niiden ominaisuudet, rajoitukset, niitä koskevat väittämät, funktiot yms. on määritelty eksplisiittisesti. Käsitteellistäminen puolestaan tarkoittaa abstraktin ja yksinkertaistetun mallin rakentamista jostain reaali maailman ilmiöstä. Ontologian yhteinen jakaminen mahdollistaa yhteisymmärryksen perustuvan tietämyksen.

Ontologioiden yhteistä jakamista korostavat muutkin - tietojenkäsittelyssä ontologioiden tarve on lähtöisin eri järjestelmien tai niiden sisältämien tietojen yhteensovittamisongelmista. Kun eri järjestelmissä on suunnitteluvaiheessa käytetty samoja termejä eri merkityksissä, on järjestelmien sisältöjen yhteensovittaminen myöhemmässä vaiheessa ollut mahdotonta.

Myös web-palveluita kehitetään olettaen samalla itsestäänselvytenä, että kaikki käyttävät samoja sanastoja ja tulkitsevat ne semanttisesti samalla tavalla. Täyteen yhteensopivuuteen ei kuitenkaan päästä, ellei tunnusteta ja oteta tarkastelun alle semanttista heterogeenisyyttä. [Uscho1996]

Ontologioita on tarjottu ikään kuin vastauksena tähän. Ontologia on yhteisesti tunnustettujen sääntöjen mukaan koostettu termistö, joka sisältää myös termien määritelmät. Näin se muodostaa yhteisesti hyväksytyyn pohjan tietojen esittämiseksi ja tulkinnalle. Se myös mahdollistaa eri järjestelmien sisältämien tietojen yhdistämisen. [Smith2004] Ontologia toimii myös toiseen suuntaan: se auttaa paikallistamaan asiat tai ominaisuudet, joiden osalta järjestelmät eroavat toisistaan [Grube2007].

Ontologioita voidaan jopa käyttää osana ohjelmistojen vaatimusmäärittelyjä. Ontologioiden avulla saadaan yhteensopivuus myös sellaisten järjestelmien välille, joita ei suunnitteluvaiheessa ole ollut tarkoitus yhdistää. Vastaavasti olemassa olevaan ympäristöön ja järjestelmäkokonaisuuteen on helppo liittää uusia järjestelmiä. [Uscho1996] ja [Jaspe1999]

Ontologiakirjastot

Ontologiakirjastot ovat eräs ratkaisu ontologioiden jakamiseksi. Emme tule toimeen yhdellä ontologialla, ja kaiken kattavan ontologian luominen olisi loputon työ. Sen sijaan pienempien, rajattujen ontologioiden rakentaminen on yleistä ja niitä voidaan koota ontologiakirjastoiksi.

Ontologiakirjastot toimivat alustoina, joille voidaan koota useita eri alojen ontologioita, jolloin käyttäjä voi tarkistaa käsitteen samalla useasta eri ontologiasta. Jo edellisessä luvussa mainittu Finto-ontologiakirjasto on suomalainen ontologiapalvelu, johon on tarkoitettu koota kaikki käytettävissä olevat ontologiat [Finto2014].

D'Aquin ja Noy ovat käyneet läpi ontologiakirjastojen ominaisuuksia ja määritelleet, mitä seikkoja tulisi ottaa huomioon ontologian valinnan tai oman ontologian julkistamisen yhteydessä [D'Aqu2012]. Kirjastot voivat tarjota käsitteitä sekä manuaalista että automatisoitua hakua varten. D'Aquin ja Noy korostavat yhteisten ontologiakirjastojen käytön ja hyvän saavutettavuuden tärkeyttä: ilman niitä eri järjestelmien sisältäminen tietojen yhdistäminen ei onnistuisi.

Ontologia yhteisön välineenä

Gibsonin, Wolstencroftin ja Stevensin [2007] mukaan ontologiat on nimenomaan määriteltävä yhteisöjen välineiksi, joiden tarkoitus on mahdollistaa eri järjestelmien sisältämien tietojen yhdistäminen. Jotta ontologioiden olemassaolosta saataisiin irti kaikki potentiaaliset hyödyt, täytyy niiden olla massojen tavoitettavissa.

Ontologiakirjastot voivat tarjota yhteisölle mahdollisuuden ontologioiden arviointiin ja kommentointiin, mikä auttaa ontologioiden validoinnissa [D'Aqu2012]. Gibson, Wolstencroft ja Stevens ehdottavat, että ontologioiden yhteyteen pitäisi aina luoda tila käsitteiden määrittelyä koskevalle loogiselle ja kielitieteelliselle keskustelulle. Ylipäätään ontologioissa pitäisi erottaa loogiset ja kielitieteelliset aspektit, jotta kohdealueen asiantuntijoiden olisi helpompi keskustella määrittelyjen sisällöstä ja validiteetista. He muistuttavat myös, että ontologioiden arvioinnin kannalta on tärkeää, että ne ovat sekä ihmisen että koneen luettavassa muodossa [Gibso2007].

Edellä mainittujen lisäksi lisäksi ontologian yhteisyyttä tuodaan esiin mm. lähteissä [Fense2001] ja [Uscho1996]. Erityisesti ontologioiden sosiaaliseen ulottuvuuteen on kiinnittänyt huomiota Mika [2011].

4.1.2 Ontologia suhteessa muihin käsite- ja termikokoelmiin

Ontologian suhde käsitteelliseen malliin

Tietojenkäsittelytieteessä puhutaan usein käsitteellisistä malleista. Mikä on niiden suhde ontologioihin? Tätä rajanvetoa on käsitelty mm. Gruber [1993 ja 1995], hänen kirjoi-

tuksiaan käytetään usein lähteenä ontologioista puhuttaessa. Hänen mukaansa ontologia sisältää kohdealuetta koskevan tietämyksen esittämiseen tarvittavan terminologian. Ontologia sisältää kohdealueen ”entiteettien” nimet ja määritelmät sekä sääntöjä, jotka koskevat termien tulkintaa ja käyttöä. Kohdealueella (universe of discourse) tarkoitetaan Gruberin mukaan niitä objekteja, joita ontologinen määrittely edustaa [Grube1995].

Gruberin mukaan ontologia eroaa käsitteellisestä mallista siten, että käsitteellinen malli määrittää datan merkitystä ja käyttöä sääntöjä, kun taas ontologia on tätä laajempi: se koskee tietämyksen (knowledge) esittämistä. Data on hänen mukaansa yksittäisiä faktoja, mutta tietämys voi sisältää myös päättelysääntöjä. [Grube1993]

Myös Smith [2004] tuo esille, että kunnianhimoisesti määriteltynä ontologia ei käsitä vain termejä ja niiden määritelmiä. Ontologia sisältää myös joukon aksioomia, jotka muodostavat käsitteille niitä tukevan viitekehyksen.

Käytännössä Gruberin määritelmä ei kuitenkaan näyttäisi pitävän paikkaansa, sillä ontologiaksi ainakin kutsutaan usein sanastoja, jotka sisältävät hyvin vähän termien käyttöä sääntöjä (tämän Gruber toteaa myös itse). Toki todennäköisesti kaikkia termejä käytetään jossain virheellisesti, joten tästä ei pidä tehdä vielä pitkälle meneviä johtopäätöksiä.

Ontologiat vs. tesaurokset

Guarino ja Giaretta määrittelevät formaalin ontologian käsitteen [Guari1995]. He erottavat ontologian määrittelyssä kaksi puolta. Ensinnäkin he asettavat painon sanalle formaali: kyse on määrämuotoisesta esitystavasta. Toisaalta kyseessä on teoria eroavaisuuksista: eroavaisuudella voidaan tarkoittaa jakoa kohdealueen olioihin tai eroavaisuus voi määrittää metatason kategorioita, joita voidaan käyttää kohdealueen mallintamisessa. Ontologian avulla voidaan siis erotella, mihin luokkaan kohdealueen eri ilmentymät kuuluvat.

Eriyisesti informaatiotutkimuksen puolella käytetään paljon nk. kielellisiä ontologioita, esim. tesauroksia. Niitä hyödynnetään mm. kirjastoissa ja erilaisten tekstiarkistojen ylläpidossa. Nämä, nk. kielelliset ontologiat eroavat tietojenkäsittelyn puolella hyödynnettävistä formaaleista ontologioista sisällön yksityiskohtaisuuden suhteen. Sitä kautta myös niiden soveltamiskohde on eri. Formaali ontologiat sisältävät enemmän yksityiskohtia ja soveltuvat tietämyksen esittämiseen ja keräämiseen, sillä yleensä ne on määriteltä riittävän eksaktisti. [Guari1995]

Guarino ja Giaretta analysoivat myös termiä ”ontologia” ja käyvät läpi useita eri määrittäyksiä ontologialle. Lopulta he päätyvät suositteluun määritelmää, jonka mukaan ontologia-termiä voidaan käyttää synonyyminä ontologiselle teorialle. Tällaisen ontologi-

sen teorian tulee päteä kaikissa mahdollisissa maailmoissa. Toiseksi, ontologia voi tarkoittaa samaa kuin käsitteellistäminen. Tällöin kyse on tietämuskannan käsitteiden semantiikan määrittelystä, ja sitä varten tarvitaan käsitteiden intensionaalisen rakenteen analysointia.¹ [Guari1995]

Jasper ja Uschold [1999] myöntävät, että ontologioita on monenlaisia, samoin ontologiaa koskevia määritelmiä. Yksinkertaisimmassa muodossaan ontologioiksi on kutsuttu termeistä koostuvia luokittelurakenteita. Aina ei ole edes määritelty sitä, mikä on luokitteluperusteena. Itse he katsovat, että ontologiassa on olennaista termien muodostama sanasto, mutta sen tulee sisältää myös termien merkityksen määrittelyt sekä kuvaukset termien välisistä suhteista (miten ne liittyvät toisiinsa). Käytännössä myös määrittelyiden taso vaihtelee: vähimmillään merkitystä hahmotetaan aliluokka-yliluokkasijoittelun² kautta. Toisessa ääripäässä ovat ontologiat, jossa käsitteille on aina esitetty erityinen sanallinen määritelmä ja lisäksi käyttöön liittyvät säännöt ja rajoitukset.

Hyvönen ym. ovat esitelleet tesaurusten puutteita, joiden vuoksi niitä tulisi laajentaa ontologioiksi [2007]. Heidän mukaansa tesauksissa termien väliset suhteet esitetään yleensä niin löyhästi, että niiden semanttisen merkityksen ymmärtäminen edellyttää aina käsitteet tuntevaa ihmistä, pelkkä ohjelman suorittama tietojenkäsittely ei riitä.

Puutteita esitellessään he käyttävät esimerkkinä tesauksissa esiintyvää relaatiota ”laajempi termi” (esim. kissa **laajempi termi** lemmikkieläimet). Ongelmana on se, että termien välistä hierarkiaa ei ole määritelty eksplisiittisesti, joten voi olla epäselvää, tarkoittaako ”laajempi termi” osa_jostakin- vai aliluokka-suhdetta [Hyvön2007]. Esim. näppäin on näppäimistön osa, mutta langallinen näppäimistö ja langaton näppäimistö ovat näppäimistön aliluokkia. Kaikkia näitä suhteita voitaisiin kuvata samalla relaatiolla: ’näppäin **laajempi termi** näppäimistö’, ’langallinen näppäimistö **laajempi termi** näppäimistö’. Pelkkä ”laajempi termi” -relaatio ei siis kuvaa, minkä tyyppisestä suhteesta on kysymys.

Termien väliset relaatiot eivät myöskään ole transitiivisia (jos kultakala kuuluu kalojen ryhmään ja kalat kuuluvat ruoka-aineisiin, niin kultakala ei silti ole ruoka-aine). Lisäksi osa relaatioita kuvaavista termeistä on monimerkityksisiä. [Hyvön2007]

¹ Guarino on täsmentänyt tätä myöhemmässä julkaisussaan tähdentämällä, että ontologialla ja käsitteellisellä on hienoinen ero: ontologia sisältää epäsuorasti lauseista pääteltävissä olevan käsitteellistämisen, sillä ontologia on riippuvainen kielestä, mutta käsitteellistäminen on kielestä riippumaton. [Guari1998]

² Esim. Omenapuulla yläkasite hedelmäpuut, assosiatiiiviset käsitteet omena ja ruusukasvit (lähde finto.fi)

4.1.3 Ontologioihin liittyviä ongelmia

Ontologioihin liittyviä ongelmia on kerätty monesta eri lähteestä ja ne on järjestetty uudelleen sen mukaan, mitkä kulloinkin ovat mainitun ongelman aiheuttavat ontologioiden ominaisuudet.

Ontologioiden sisältö

Edellä on esitetty, että ontologioita tarvitaan, jotta erillisten järjestelmien sisältämien tietojen merkitys voidaan sovittaa yhteen ja järjestelmien sisältämät tiedot on mahdollista yhdistää. Smith on kuitenkin epäileväinen sen suhteen, kuinka yhteneväiseksi eri tahojen näkemykset voidaan saattaa. Tarvittaisiin yhteinen ontologia, joka olisi niin neutraali, että kaikki voisivat sen hyväksyä. Käytännössä tämä tarkoittaisi hyvin yleistä ontologiaa, joka sisältäisi sellaisia käsitteitä kuten aika, tila, identiteetti, määrä jne. [Smith2004]

Myös Uschold ja Gruninger ovat epäileväisiä sen suhteen, että koskaan voitaisiin saavuttaa yhteisymmärrystä yhdestä yhteisestä ontologiasta. Heidän mukaansa yhteiset yleisen tason ontologiat kuitenkin tukevat osaltaan järjestelmien yhdistettävyyttä. Se on helpompi toteuttaa, kun edes jonkinlainen perusta on yhteinen. [Uscho1996] Tässä tarkoituksessa Uschold tutkimusryhmineen on kehittänyt myöhemmin liiketoimintaa varten hyvin yleisen tason ontologian “The enterprise ontology” [Uscho1998].

Toisena sisältöihin liittyvänä ongelmana voidaan ottaa esille standardien sisällyttäminen ontologioihin: jotta ontologia palvelisi sovelluksia täysipainoisesti, täytyisi siihen sisällyttää aihealueeseen liittyvät standardit. Monien standardien käyttöehdoissa on kuitenkin kielletty niiden uudelleenjulkaisu edes johdannaistuotteiden muodossa, joten ontologioiden ylläpitäjä joutuu joko hankaliin sopimusneuvotteluihin tai ontologia jää standardin osalta vaillinaiseksi. [HeppM2007]

Kehitystyön hitaus

Sekä Hepp [2007] että Braun ym. [2007] ovat käsitelleet osittain samoja ontologioiden käyttämiseen liittyviä ongelmia. Molemmissa artikkeleissa todetaan, että valitettavasti ontologioiden kehitystyö on hidasta.

Braun ym:iden mukaan tämä johtuu siitä, että kehittäminen on annettu asiantuntijoiden, ei käyttäjien tehtäväksi. Tähän vaikuttavat osaltaan ontologian rakenteen monimutkaisuus ja vaatimukset tarkasta formalismista, sillä ne nostavat käyttäjien kynnystä osallistua kehittämistyöhön [Braun2007].

Aika käytännössä todetusta määrittelytarpeesta toteutukseen on näin ollen liian pitkä, joten yleensä ontologiasta puuttuvat sen valmistuessa kaikkein uusimmat käsitteet. Ne puolestaan olisivat usein semanttisen hyödyntämisen kannalta mielenkiintoisimpia. [Braun2007] [Angel2007] [Genda2007]

Usein myös oletetaan virheellisesti, että käsitteet pysyvät samoina. Kuitenkin, mitä monimutkaisempi käsiterakenne on, sitä todennäköisemmin siihen kohdistuu muutostarpeita. [HeppM2007], [Braun2007] Voidaan myös olettaa, että täysin uusien käsitteiden määritelmät tarkentuvat vähitellen ja niitä joudutaan tarkistamaan vielä sen jälkeen, kun ne on ensi kerran määritelty.

Kehittämistyön taloudellinen kannattavuus

Voidaan myös epäillä ontologioihin käytettyjen resurssien kannattavuutta: onko niiden käytöstä vielä niin suurta hyötyä, että panostus kannattaa kehitystyön rahoittajien kannalta katsottuna [HeppM2007]? Kun mietitään työn kannattavuutta, törmätään luonnollisesti nk. laajuusongelmaan: mitä kustannuksia ja mitä hyötyjä laskelmiin otetaan mukaan [Virkk1951].

Ontologioita käytettäessä osa hyödyistä on sellaisia, että niitä on vaikea mitata ja niiden arvoa on vaikea määrittää. Miten esim. mitataan hakutulosten hyvyys ja miten arvotetaan parantuneet hakutulokset verkkohauissa?

D'Aquin ja Noyn mukaan on myös kyseenalaista, kuinka hyvin ontologiakirjastot ylläpitäen tavoittavat käyttäjänsä - onko julkisin varoin kustannettu ontologiakirjastojen ylläpito hyödyllistä toimintaa yhteiskunnan kannalta, jos käyttäjät eivät löydä niitä? [D'Aqu2012]

Ontologioiden muodostamisesta aiheutuvia kustannuksia on pyritty pienentämään luomalla (osittain) automatisoituja prosedureja, joiden avulla ontologioita voitaisiin ohjelmallisesti muodostaa käytettävissä olevasta aineistosta (teksteistä, tietokannoista, web-sivuista yms.) Tähän (puoli)automatisoituun kehittämistyöhön on viitattu termillä ”ontology learning” (suomeksi ontologioiden johtaminen¹) vastineena termille ”ontology building/engineering” (ontologioiden rakentaminen/kehittäminen).

Esimerkiksi Maedche ja Staab ovat määritelleet viisivaiheisen menetelmän, jolla ontologioita voi johtaa ja edelleen kehittää käyttämällä hyväksi webissä jo olevia ontologioita ja muita dokumentteja. He arvelevat, että ontologioiden automatisoitu johtaminen parantaa ontologioiden hyödyntämismahdollisuuksia huomattavasti, koska näin niiden kehitystyö nopeutuu ja halpenee huomattavasti.[Maedc2001]

¹ oma suomennokseni

Vastaavia menetelmiä ontologioiden johtamiseksi on kehitetty lukuisia, mutta emme paneudu niihin tässä työssä tätä enempää.

Käyttäjien ja kehittäjien kohtaaminen

Ontologioiden hyödyntämisessä on ongelmallista myös se, että ontologioitten käyttäjät ja kehittäjät ovat eri tahoja. Kuinka hyvin suunnitelluista ajattelemalla semanttinen sisältö mahtaa välittyä käyttäjien suuntaan - erityisesti kun otetaan huomioon, että ontologiat on usein esitetty sellaisella formaalilla notaatiolla, joka ei ole kaikkien käyttäjien ymmärrettävissä [HeppM2007]?

Myös Weber tuo esille sen, että ontologioiden kehitystyössä ja ylläpidossa on erikseen ontologian ylläpitäjä ja se yhteisö, joka käyttää ontologiaa. Onko ylläpitäjä onnistunut tallettamaan ontologiaan yhteisön käsitteellisen näkemyksen vai oman näkemyksensä? [Weber2008]

D'Aquin ja Noy ovat [2012] tuoneet esille ontologioiden käyttäjien ja kehittäjien välisen kommunikaation puutteen: ontologioiden käyttäjien on usein mahdollista antaa palautetta ontologioista, mutta tuo viestintä on vain yhdensuuntaista - käyttäjiltä kehittäjille. Miten ontologioiden käyttäjät tavoittavat toisensa ja miten kehittäjät voivat olla yhteydessä käyttäjiin [D'Aqu2012]?

Gibson, Wolstencroft ja Stevens [2007] ehdottavat ontologioiden yhteyteen erityistä metadata-työtilaa, jossa olisi mahdollista tutkia olemassa olevien ontologioiden profiileja. Profiileissa olisi kuvattuna ontologian ala ja tarkoitus sekä tietoa ontologian kehitystyön tilanteesta. Työtilassa olisi myös kommentointimahdollisuus, joten käyttäjät voisivat lisätä omia kommenttejaan liittyen kokemaansa ontologian hyödyllisyyteen. Keskustelualueella voisivat ontologioiden ja kohdealueen asiantuntijat vaihtaa mielipiteitä liittyen termien määrittelyyn jne. Tätä varten ontologioiden määrittelyssä pitäisi myös erottaa selkeästi ontologis-loogiset määrittelyt kielellisistä määrittelyistä, jotta kohdealueen asiantuntijoiden olisi helpompi arvioida määrittelyksiä ja keskustella niistä.

Yhteistyömahdollisuuksien puutetta ontologioiden kehitystyössä ovat kritisoineet myös Braun ym. [2007]. He ovat suunnitelleet mallin, jossa ontologiaa kehitetään yhteistyön kautta ja kehittämistyö on sulautettu työprosesseihin.

Braunin ym.:iden mallissa uuden käsitteen lisäämisen tarve nousee käytännön toiminnasta: työntekijät tarvitsevat uusia avainsanoja merkitäkseen sisältöjä. Nämä kaikki avainsanat kerätään yhteiselle alustalle. Vähitellen uusien ja aikaisemmin lisättyjen avainsanojen välille tunnustetaan yhteyksiä, avainsanojen väliset relaatiot kypsyvät hie-

rarkioiksi ja lopulta niihin lisätään myös semanttiset määritelmät. Ajatuksena on, että näin ontologia kypsyy vähitellen luonnollisena osana työtä. [Braun2007]

Käyttäjät kehittäjinä

On myös muita tapoja ratkaista ontologioiden kehittäjien ja käyttäjien kohtaamisongelma: jos käyttäjät ovat itse kehittäjiä, ei em. ongelmaa ole [Genda2007] ja [Kotis2006]. Tällöin tarvitaan vain välineitä tai ympäristöjä, jotka mahdollistavat yhteisön osallistumisen kehittämistyöhön.

Gendarmin ym:iden mallissa käyttäjillä on apunaan avainsanoihin liittyvää metadataa ja semanttista sisältöä. Kun käyttäjä merkitsee jonkin aineiston avainsanalla, tulee hänen myös valita avainsanalle haluamansa paikka käsittehierarkiassa. Mikäli avainsana on uusi, täytyy se määritellä. Näin tulee mahdolliseksi se, että jollain avainsanatermillä on useita eri merkityksiä, eli se voi sijaita eri käyttäjillä eri ylituokkien alla. Määrittelytyökalun avulla käyttäjät saavat rakennettua omista avainsanoistaan itselleen henkilökohtaisen, hierarkiamuotoisen luokittelurakenteen. Tämän lisäksi on tarjolla myös tapa, jolla käyttäjät voivat jakaa keskenään omia luokittelurakenteitaan. [Genda2007]

Kotis ja Vouros ovat kehittäneet samantapaisen mallin, jossa käyttäjät laativat henkilökohtaiset ontologiansa ja jakavat ne sen jälkeen muille [2006]. Yhteisössä on jäseniä, jotka aktiivisesti pyrkivät sovittamaan ontologioita yhteen vertaamalla eri henkilöiden laatimia ontologioita toisiinsa ja evaluoimalla niitä oman työnsä ja kokemuksensa pohjalta.

Tässä työssä koetellaan erityisesti edellä mainittua käyttäjien ja kehittäjien kohtaamisongelmaa: tarkoituksena on tutkia, kohtaavatko yleiset ontologiat ja käyttäjien itsensä tekemät jäsenyykset. Ovatko yleiset ontologiat toimivia silloin, kun sisältö kertoo arkielämästä ja avainsanajäsennys on tekstin laatijan itsensä tekemä?

Ontologioiden ja käyttäjien tekemien jäsenyyksien kohtaamista tutkitaan yksinkertaisesti selvittämällä, sisältyvätkö jäsenyyksissä käytetyt käsitteet yleiseen ontologiakirjastoon. Kehitystyöstä ovat vastuussa tällöin Finto-palvelun tai siihen sisällytettyjen ontologioiden laatijat ja ylläpitäjät. Finton ontologiat on tehty vapaasti käytettäväksi internetissä ilman tunnuksia ja kirjautumisia, joten sen potentiaalisia käyttäjiä ovat kaikki internetin käyttäjät.

4.2 Folksonomiat

Folksonomia koostuu avainsanoista, joten ensin määritellään, mitä avainsanalla tarkoitetaan. Sen jälkeen tutkitaan erilaisia folksonomian määritelmiä ja käydään läpi kirjallisuudesta kerättyjä folksonomioiden heikkouksia.

4.2.1 Avainsanat ja folksonomiat

Avainsana

Ennen kuin pääsemme puhumaan folksonomioista, on syytä määritellä käsite "avainsana". Avainsana on verbaalinen määrite, joka voidaan liittää erityyppisiin kohteisiin (sekä konkreettisiin että abstrakteihin) ja joka omalla tavallaan edustaa sekä kohdetta että merkitsijänsä käsitystä siitä [Kieli2017].

Merkitsemisen tarkoituksena on yleensä mahdollistaa kyseisen kohteen löytäminen merkityn ominaisuuden perusteella. Esim. teksti merkitään avainsanoilla "reseptit", "kasvisruoka", "pääruoka". Kielitoimiston sanakirjan mukaan avainsana on "sana, jonka avulla etsitään tietoa julkaisuista, tietopankeista tms."¹ [Kieli2017].

Suomen kielessä avainsanalla viitataan yleensä sekä julkisiin, asiantuntijan tekemiin että käyttäjien itsensä tekemiin avainsanamäärityksiin. Asiantuntijoiden avainsanamäärittelyt ovat usein esim. kirjaston henkilökunnan tekemiä. Ne voivat pohjautua julkisiin sanastoihin, jolloin niitä yleensä kutsutaan asiasanoiksi. Avainsanojen katsotaan olevan osa kohdettaan kuvailevaa metatietoa.

Asiantuntijoiden lisäksi siis myös käyttäjät luovat itse erilaisia avainsanamäärityksiä. Käyttäjä voi merkitä esim. erilaisiin sähköisiin aineistoihin itse valitsemiaan määreitä, jotka jollain tapaa hänen mielestään edustavat tai luonnehtivat kyseistä kohdetta. Erityisesti käyttäjien tekemät omat avainsanamerkinnot ovat käytössä verkossa olevissa aineistoissa, jotka usein ovat käyttäjien itsensä tuottamia. On kuitenkin myös sivustoja, kuten esim. del.icio.us, joille käyttäjät voivat linkittää ja luokitella muiden tekemiä aineistoja [Hotho2006].

Englannin kielessä on avainsanan vastineeksi kaksi eri termiä, "keyword" ja "tag", joista jälkimmäistä on käytetty nimenomaan käyttäjälähtöisessä kontekstissa. Myös esim. lähteessä [Passa2007] "tag" mainitaan nimenomaan käyttäjän määrittelemäksi

¹ Ensijainen merkitys avainsanalle on "sana jonka avulla jokin ongelma, salakirjoitus tms. on ratkaistavissa." [Kieli2017]

avainsanaksi. Suomen kielessä ei tälle termille ole täsmällistä vastinetta. Puh kielessä toki puhutaan “tägeistä” tai “täggäämisestä” (=avainsanojen merkitseminen), mutta kirjoitettuun tekstiin nuo termit eivät oikein istu.

Lähes samaa tarkoittaa “aihetunniste”, mutta se liitetään lähinnä #-merkin käytön yhteyteen, ja se vastaa ensisijaisesti englannin kielen termiä “hashtag”. Näissä yhteyksissä esiintyy myös suomenkielinen termi “tunniste”, joka usein myös esitetään suomennoksena englanninkieliselle termille “tag”. Sanastokeskus TSK:n Tietotekniikan termitalkoot -sanaston [2018] mukaan “tunnisteena käytettävää sanaa kutsutaan avainsanaksi”¹ ja englannin kielen termi “tag” voidaan kääntää suomeksi sekä tunnisteeksi että avainsanaksi. Työn aloittaessani päädyin käyttämään suomen kielen termiä “avainsana” vastaamaan sekä sanaa “keyword” että “tag”, mutta voi olla, että jos tekisin valinnan nyt uudestaan, valitsisin avainsanan sijasta termin “tunniste”.

Folksonomian määritelmä

Folksonomiat ovat nousseet esille sosiaalisen median käytön myötä - kyse on avainsanoista ja luokitteluista, joilla käyttäjät itse kuvailevat tuottamiensa verkkoaineistojen sisältöjä [Saart2009], [Vande2007]. Folksonomioita syntyy esim. silloin, kun merkitään avainsanoja web-sivuille, verkossa oleviin valokuviiin tai vaikkapa tieteellisten artikkeleiden tiivistelmiin [VanDa2007].

Folksonomioiden käytön tavoitteena on helpottaa tiedonhakuja suurista aineistomassoista [Saart2009]. Gruberkin [2007] muistuttaa, että folksonomiat ovat yllättävän voimakas apuväline hakujen tekemisessä. Käyttäjien itsensä tuottamien aineistojen haku perustuu monissa yhteisöpalveluissa pitkälti juuri avainsanoihin, toki myös sisältöön kohdistuvaa analyysiä ja pelkkiä tekstihakuja on mahdollista hyödyntää.

Folksonomioiden etuna pidetään määrittelyjen tekemisen helppoutta [Limpe2009]. Myös Wu, Zhang ja Yu ovat korostaneet, että folksonomioilla on useita etuja huolimatta siitä, että ne ovat suurpiirteisiä, epämuodollisia ja täsmällisyydessä on toivomisen varaa. Ne ovat kuitenkin helpommin käyttäjän ulottuvilla ja kuvaavat ontologioihin verrattuna paremmin sitä, mikä merkitys aineistoilla on käyttäjien näkökulmasta katsottuna. [Wu-Zha2006]

Lähteissä on jonkin verran eroja sen suhteen, ajatellaanko folksonomian sisältävän vain yhden henkilön tekemiä avainsanaluokitteluita vai onko ennemminkin kyseessä kokonaisuus, joka sisältää useiden eri käyttäjien tekemiä luokitteluita. Useissa lähteissä

¹ http://www.tsk.fi/tsk/termitalkoot/haku-266.html?page=get_id&id=ID35&vocabulary_code=TSKTT, viitattu 10.7.2018

folksonomia määritellään selkeästi *kokoelmaksi*, joka muodostuu jonkin järjestelmän käyttäjien tekemistä avainsanaluokitteluista.

Yksi esimerkki julkaisuista, joissa folksonomia määritellään edellä kuvatulla tavalla, on Jäschke ym.:iden kirjoittama artikkeli [2008]. Heidän määritelmänsä mukaan yksittäisen käyttäjän tekemät avainsanaluokittelut muodostavat “personomian” (englanniksi *personomy*) ja monen käyttäjän *personomioista* muodostuu folksonomia. Samaa käsitteparia käyttävät myös Weller ja Peters [2008] sekä Hotho ym. [2006].

Folksonomian määrittely useiden henkilöiden tekemäksi kokoelmaksi perustuu ainakin osittain ajatukseen termin alkuperästä. Jäschke ym. nojaavat Vander Walin määritelmään [Vande2007], jossa folksonomian taustaksi esitetään sanat “folk” ja “taxonomy”. Heidän mukaansa tällöin kyse olisi selkeästi ihmisjoukon tekemistä luokitteluista. [Jäsch2008]

Tähän tuovat oman, mielenkiintoisen näkökulmansa Tanasescu ja Streibel [2007]: He eivät pidä folksonomia-termistä sen vuoksi, että se on moniselitteinen ja sisältää assosiaation taksonomiaan eli (systemaattiseen) luokitteluun. Heidän mielestään folksonomia juuri eroaa ontologioista mm. sen suhteen, että avainsanat eivät noudata mitään systematiikkaa. Toki esim. Gruber [2007] on tulkinnut niin, että Vander Wal on nimenomaan korostanut, että folksonomia ei ole sananmukaisesti mikään taksonomia.

Spyns ym. [2006] ovat puolestaan tehneet hienojakoisemman jaottelun, jossa otetaan huomioon se, onko näkökulma asiantuntijalähtöinen (semanttinen web) vai käyttäjälähtöinen (web 2.0)¹. Semanttisesta webistä puhuttaessa he tekevät eron ontologioiden ja taksonomioiden välille. Taksonomia eroaa heidän mukaansa ontologiasta niin, että taksonomiassa käsitteiden välillä on ainoastaan kahdenlaisia yhteyksiä: “osa-jotakin” ja “alatyyppe/-laji”. Ontologiassa on monenlaisia muitakin käsitteidenvälisiä yhteystyyppejä.

Web 2.0:n terminologiassa heillä on folksonomian lisäksi käytössä myös termi “folksologia”. He ovat sitä mieltä, että web 2.0:n käsitteissä folksonomia vastaa taksonomiaa ja sen lisäksi on käsite “folksologia”, joka vastaa ontologiaa [Spyns2006].

Tätä he täsmentävät myöhemmässä kohdassa kritisoimalla, että folksonomia on semanttisesti määritelty liian laveasti [Spyns2006]. Oikeasti folksonomiat eivät ole kunnollisia taksonomioita, lähinnä vain sanastoja. Muihin lähteisiin verrattuna Spyns ym. määrittelevät folksologian suunnilleen samalla tavalla kuin muut kirjoittajat kuvaavat folksonomiaa.

¹ Semanttisen webin ja web 2.0:n määritelmiä on käsitelty kohdassa 4.4

Vander Walin mukaan folksonomiat voidaan jakaa laajoihin (broad) ja kapeisiin (narrow) folksonomioihin sen perusteella, voivatko useat käyttäjät luokitella avainsanoilla samaa objektia vai ei. Kapeassa folksonomiassa luokittelu on vain yhden henkilön tuottama, laajassa folksonomiassa useat henkilöt voivat liittää omia avainsanojaan esim. samaan artikkeliin. [Vande2005]

Useissa lähteissä folksonomia tunnutaan määrittelevän laajemmaksi kuin vain yhden käyttäjän avainsanojen kokonaisuudeksi. Vaikka olisikin voimassa sääntö, että kukin voi määritellä vain itse lisäämiään aineistoja (esim. valokuvia tai weblinkkejä), ymmärretään folksonomia kuitenkin siksi *kokonaisuudeksi*, joka muodostuu verkkoyhteisön *useiden* käyttäjien itse tekemistä avainsanaluokitteluista, ks. [Echar2008], [Halpi2007], [Hotho2006], [Limpe2009], [Knerr2006], [Speci2007], [Spite2005] ja [Veres2006].

4.2.2 Folksonomioiden heikkouksia

Merkitsemistä ei ohjalla säännöillä

Folksonomioita ei pidetä erityisen hyvinä: väljät soveltamissäännöt johtavat siihen, että luokittelut eivät ole yhteismitallisia. Sen vuoksi niitä ei voida hyödyntää ohjelmallisesti semanttisen webin teknologioilla niin hyvin kuin olisi toivottavaa [Knerr2006].

Myös Spyns ym. kaipaavat folksologioihin (heidän vastaava terminsä folksonomialle) enemmän semanttista selvyyttä: on tehtävä ero luonnollisen kielen termien ja käsitteiden nimikkeinä toimivien, kieliriippumattomien ilmaisujen välille [2006]. Käyttäjien on siis tärkeää tietää, kummasta milloinkin on kyse.

Sekä avainsanojen merkitsemisen perusteet että merkintöjen abstraktiotasot vaihtelevat. Eri tarkkuudella merkityt avainsanat palvelevat eri käyttäjiä tai eri käyttötilanteita. Joissain tapauksissa riittää yleisemmän tason avainsana (esim. hedelmä), toisinaan taas tarvitaan hyvinkin tarkkaa luokittelua (esim. Raike, Suislepp, Syysviiru¹). [Knerr2006], [Spite2005]

Myös avainsanojen esittämisen syntaksi on vapaa, järjestelmät eivät yleensä sisällä ohjeita, miten ne merkitään: esim. kirjataanko määritelmät tyyllillä “gluteenittomat leivonnaiset” vai “leivonnaiset, gluteeniton”. Käyttäjien tekemien merkintöjen joukossa

¹ Omenalajikkeita

saattaa olla myös virheellisiä merkintöjä (esim. ihmiset usein sekoittavat vähälaktoosisen ja laktoosittoman.) [Spite2005], [Knerr2006]

Vaikka järjestelmissä olisi toimintoja, jotka pyrkisivät yhtenäistämään ja siistimään avainsanamerkintöjä (esim. välilyöntien poisto, isot alkukirjaimet, synonyymien käsittely jne.), voivat nämä käytännöt kuitenkin olla erilaisia eri sivustoilla. Sivustoilla voi olla muitakin eroavaisuuksia, esim. sen suhteen, miten avainsanamerkinnot yhdistetään käyttäjän tietoihin. [Grube2008]

Vapailla avainsanaluokitteluilla on monia muitakin puutteita: mm. se, että ihmiset ilmaisevat luonnostaan samoja asioita eri termeillä ja käyttävät samasta termistä eri muotoja (esim. yksikkö/monikko). Folksonomioiden sisältämien termien välisiä yhteyksiä ei ole kirjattu mihinkään ylös, ja kun avainsanoja ei valita valmiista listoista, jo pelkät kirjoitusvirheet voivat jättää tärkeitä sisältöjä hakujen ulkopuolelle. [Golde2006], [Passa2007]

Termien välisiä yhteyksiä ei ole määritelty

Yksittäinen avainsana saattaa olla monimerkityksinen (on huomioitava homonyymit ja polysemia)¹, jolloin tarvittaisiin lisätarkennusta käyttötarkoituksen varmistamiseen. Joskus sisällön aihepiiri voi tarjota tämän tarkennuksen, mutta se ei ole aina käytettävissä tai voi olla tähän tarkoitukseen liian niukka.

Mathes on tutkinut käyttäjien itse tekemiä avainsanaluokitteluita ja on tässä yhteydessä huomannut, että samaa termiä voidaan käyttää hyvin moneen eri tarkoitukseen. Esim. tekemällä hakuja sanalla ”suodattaminen” (alkup. sana filtering) voidaan saada tulokseksi sekä tilastotieteellisiä että vodkan valmistamista käsitteleviä aineistoja. [Mathe2004]

Avainsanojen avulla tehtävä luokittelu ei muutenkaan ole tarkkarajaista, joten on mahdollista, että osa luokittelun kohteista ei osu mihinkään kategoriaan tai voi kuulua yhtälailla useaan eri kategoriaan samanaikaisesti (esim. kasvissosekeitto voi kuulua sekä keittojen, kasvisruokien että alkuruokien kategoriaan). [Golde2006]

Spiteri huomauttaa, ettei folksonomioissa myöskään yleensä merkitä synonyymejä, eivätkä folksonomioitten ylläpitoon tarkoitettut järjestelmät tarjoa välineitä synonyymien määrittelyyn. Silloin eri käyttäjät voivat merkitä samaa asiaa eri avainsanalla eikä mi-

¹ Homonyymi tarkoittaa sanaa, jolla on useita eri merkityksiä, esim. "voi". Polysemiolla tarkoitetaan, että yksittäinen sana voi viitata useaan eri kohteeseen, mutta nämä kohteet ovat jollain tapaa samaa kantaa, esim. "jalka" voi viitata ihmisen jalkaan, pöydän jalkaan tai pyörän jalkaan. [Infim2015]

hinkään saa kirjattua ylös, että ne ovat keskenään synonyymejä (esim. vasta=vihta, flunssa=nuhakuume jne.) [Spite2005]

Folksonomioiden käytön rajoitukset tulevat erityisesti esiin juuri hakujen yhteydessä. Jos haetaan tietoja eläimistä, eivät avainsanalla ”kissa” merkityt tekstit tule mukaan, koska folksonomia ei pidä sisällään termien välisien suhteitten määrittelyä. Hakujen onnistuminen jää tällöin sen varaan, että avainsanojen merkitys on otanut huomioon eri abstraktiotasot ja merkinnyt tekstiin avainsanat ”kissa” ja ”eläin”.

Golderin ja Hubermanin [2006] mukaan avainsanojen abstraktiotaso riippuu merkintöjen tekijän asiantuntemuksesta: mitä enemmän tietoa on käytettävissä, sitä yksityiskohtaisempia avainsanoja käytetään. Heidän mukaansa suurin osa merkitsee luokittelun jostain hierarkian keskivaiheilta: esim. avainsanaksi merkitään kissat, ei eläimet eikä similaiset kissat. Hakujen tekemisen tueksi siis pitäisi olla käytettävissä käsittehierarkioita, jotta tulokseksi saataisiin kaikki käytettävissä olevat aineistot.

Tähän kohtaan haluaisin lisätä huomion, että tekstin abstraktiotaso todennäköisesti vaikuttaa myös siihen abstraktiotasoon, jolla kirjoittaja tai lukija määrittää avainsanoja. Tämä sama asia täytyy ottaa huomioon myös silloin, kun etsitään lähteitä tekstihaulla niin, että haku kohdistuu tekstiin eikä erikseen merkittyihin avainsanoihin. Silloinkin ollaan sen varassa, että haun ja lähteen abstraktiotaso on sama ellei haun tukena ole käytettävissä semanttisia hakukoneita, jotka automaattisesti etsivät ontologian perusteella myös ylempään sekä alemman abstraktiotason termeillä.

Jos avainsanamäärittelyt on tehty yksityiskohtaisella tasolla, ne eivät palvele sellaista hakujen tekijää, jonka oma asiantuntemus aihealueesta ei ole yhtä syvästä. Sen vuoksi yksityiskohtaiset määrittelyt eivät toimi ilman määrittelyä luokasta, johon ne puolestaan kuuluvat (esim. ’Nissan’ yksinään ei kerro vielä paljoa, tarvitaan laajempi luokka ’auto’). [Golde2006]

Avainsanaluokitteluilla erilaisia tarkoituksia

Kun avainsanoja tutkitaan tarkemmin, havaitaan, että ne on voitu määrittellä hyvin erilaisista näkökulmista. Golder ja Huberman [2006] ovat huomauttaneet, että avainsanoja käytetään monessa eri merkityksessä: ne voivat kertoa, mitä aineiston sisältö käsittelee tai kenestä se kertoo. Avainsanoilla voidaan esittää, kuka omistaa tai on laatinut aineiston. Ne voivat myös määrittellä, minkätyyppinen tai -muotoinen aineisto on kyseessä (blogi, valokuva tms.). Tähän voisin omana lisäyksenä liittää määreet, jotka kuvaavat aineiston luonnetta (huumori, fiktio, dokumentti).

Avainsanaluokittelut voivat olla myös adjektiiveja, jotka kuvaavat luokittelun kohdetta (iloinen, kesäinen, tieteellinen), ne voivat kuvata kirjoittajan suhdetta kuvattuihin asioihin (minun lemmikkini, minun työpaikkani) tai ne voivat olla määrittelyitä siitä, mitä pitäisi tehdä (kirja, joka pitäisi lukea). [Golde2006]

Avainsanat ovat käytössä monissa internetin yhteisöpalveluissa, joissa käyttäjät jakavat ottamiaan valokuvia. Näissä palveluissa on yleensä tapana merkitä avainsanaksi myös valokuvaan ja sen ottamiseen liittyvää metatietoa, esim. laajakulma, maisema, mustavalkoinen jne. [Mathe2004]

Mitä kauemmaksi mennään asiategististä ja faktasisällöistä, sitä enemmän korostuu kohteen tulkinnan subjektiivisuus. Mm. Saarti mainitsee artikkelissaan [2009], että kaunokirjallisuuden sisällön kuvailu on hankalaa, koska jokainen lukija voi tulkita sen hieman eri tavoilla. Tästä luonnollisesti seuraa se, että myös avainsanaluokittelut määräytyvät subjektiivisesti.

Avainsanat eivät myöskään aina perustu pysyviin luokitteluihin. Veres [2006] viittaa kognitiopsykologi Lawrence Barsaloun [1983] määrittelemään ad hoc -kategorian käsitteeseen. Ad hoc -kategorioita voisivat olla esim. ”tavarat, joita voisi myydä pihakirpputorilla” tai ”eräretkelle mukaan otettavat tavarat”. Ne eivät kuvaa asioiden pysyvää luokittelua ja ominaisuuksia, vaan ovat enemmänkin päämäärän ohjaamia kategorisointeja. Veres otaksuu, että internetin käyttäjien omissa avainsanoissa on suuri määrä tämän-tyyppisiä koosteisia yhdistelmiä.

Veres [2006] käsittelee myös kielitieteilijä Anna Wierzbickan [1984] luokitteluita, jotka eivät perustu siihen, mikä jokin asia on, vaan siihen, mihin jotain asiaa voidaan käyttää. Esimerkkejä tällaisista kategorioista hän antaa neljä:

1. Toiminnallisuuteen perustuvat kategoriat, kuten ”lelu”, ”ase” jne. Esim. käpyä voidaan käyttää leluna ja kaulinta voidaan käyttää aseena.
2. Toiminnalliset kokoelmat, kuten ”kalustus”, ”kattaus”. Sohva ei vielä kaupassa kuulu kalustukseen. Se muuttuu kalustuksen osaksi vasta kun se on tuotu taloon osaksi käyttöympäristöään. Samoin kaapista vain osa astioista otetaan osaksi kattausta.
3. Yhteiseen alkuperään perustuvat kokoelmat, esim. ”ostokset”, ”tiskattavat astiat”. Ne ovat päätyneet samaan paikkaan jostain yhteisestä syystä (esim. saman henkilön toimesta), mutta niillä ei ole muita yhdistäviä tekijöitä.
4. Tarkoitukseen sekä alkuperään perustuvat kokoelmat, esim. ”lääkitys”. Yksittäiselle henkilölle määrätty lääkkeen muodostavat lääkityksen, joka auttaa hänen tilanteesaan.

Monimuotoisuuden käytännön implikaatiot

Veresin mukaan edellä mainitun kaltaisten luokitteluperusteiden sisällyttäminen ontologioihin johtaa semantiikan hämärtymiseen [Veres2006]. Näitä monimuotoisia luokitteluperusteita kuitenkin hyvin todennäköisesti on folksonomioissa, joten tämä osoittaa osaltaan, että folksonomiat ja ontologiat ovat väistämättä jossain määrin eri asioita.

Näistä variaatioista tulee hyvin esille se kirjo, miten monella eri tavalla luokittelujen tekijät voivat tulkita avainsanojen funktion. Kun tiedostetaan, että sama variaatio sisältyy myös avainsanojen hyödyntäjien intentioihin, on helppo päätellä, että folksonomioiden käytöllä voidaan tavoittaa vain osa niistä hyödyistä, joita avainsanojen käyttö parhaimmillaan voisi tarjota.

Folksonomioiden käyttöä voidaan edistää muokkaamalla niiden sisältöä, poistamalla luokitteluista kirjoitusvirheitä, yhdenmukaistamalla kirjoitusasua, lisäämällä homonyymien omistaviin avainsanoihin tarkentavia määreitä jne. Näitä erilaisia toimenpiteitä on käsitelty Wellerin ja Petersin artikkelissa [2008] ja niistä ovat maininneet lyhyesti myös Tesconi ym [2008].

Wellerin ja Petersin tapaan myös Gruber ja hänen jälkeensä Knerr ovat ottaneet huomioon eri kirjoitusasuun liittyvän ongelman sekä folksonomiakäsitteiden välisten suhteiden esittämisen tärkeyden [Grube2007], [Knerr2006]. Weller ja Peters [2008] ovat tuoneet myös esille mahdollisuuden rikastaa folksonomioita yhdistämällä niitä ontologioihin ja tesauruksiin. Tällaisten toimintamallien käytännön sovelluksia on esitelty tarkemmin luvussa 5.

4.3 Eri näkemyksiä ontologioiden suhteesta folksonomioihin

Käsitykset ontologioiden ja folksonomioiden suhteesta vaihtelevat eri lähteissä: ensin tarkastellaan näkemyksiä, joiden mukaan kyseessä on kaksi eri asiaa. Sen jälkeen tuodaan esille vastakkainen kanta: folksonomiat ovat ontologioiden alalaji.

4.3.1 Ontologiat ja folksonomiat ovat eri asioita

Ontologioiden yhteydessä joudutaan tekemään rajanvetoa suhteessa folksonomioihin tai ainakin selventämään nämä kaksi käsitettä ja niiden väliset erot. Gruberin mukaan kyse on kahdesta täysin eri asiasta. Ontologiat ovat välineitä, joiden avulla ihmiset voivat esittää, mitä tarkoitetaan sellaisilla termeillä, joita he käyttävät tuottaessaan, käyttäessään tai jakaessaan tietoa. Folksonomia puolestaan muodostuu siitä datasta, joka

kertoo, miten ihmiset yhdistävät termit sisältöihin, joita he tuottavat, jakavat tai käyttävät. [Grube2007]

Gruberin mukaan ontologioita ei myöskään käytetä luokitteluvälineinä, joiden avulla löydetään sisältöjä - ne löydetään hakukoneiden ja hakemistojen avulla. Sen sijaan ontologioita tarvitaan erillään tuotettujen tietokantojen ja muiden tietovarantojen yhdistämiseen. Erona on myös se, että folksonomioissa termien väliset hierarkkiset suhteet jäävät epäselviksi eikä termien merkityksestä ole yleisesti hyväksyttyä yhteisymmärrystä. [Grube2007]

Myös Limpens ym. ovat päätyneet siihen, että folksonomioissa ja ontologioissa on kyse eri asioista: folksonomiat voivat sisältää semioottisen esityksen yhteisön tietämyksestä (folksonomioista näkee, mitä termejä yhteisössä käytetään), mutta folksonomioista puuttuu semanttinen jäsenys ja rakenne (käsitteiden välisiä suhteita ei määritellä) [Limpe2009]. Heidän mukaansa folksonomioita ei myöskään käsitellä ja arvioida, kun halutaan tehdä yhteistyötä. Sen takia folksonomiat eivät voi toimia eri tietovarantojen yhdistämisen perusteena.

Halpin ym. ovat ottaneet asiaan mielenkiintoisen näkökulman [2007]. Heidän mielestään ontologioiden ja folksonomioiden erossa on kyse kahden eri asian erosta: luokittelun (classification) ja kategorisoinnin (categorization)¹. Luokittelu on yksiselitteistä, ei riipu kontekstista eikä luokittelun tekijästä. Jokainen kohde kuuluu yhteen luokkaan, ja luokat ovat toisensa poissulkevia eivätkä missään tapauksessa päällekkäisiä².

Kategorisointi taas tarkoittaa asioiden jakamista eri ryhmiin jonkin tietyn näkökulman perusteella. Nämä ryhmittelyt voivat muuttua riippuen siitä, kuka tekee ryhmittelyä tai mikä näkökulma on ryhmittelyn perustana. Eri käyttäjillä voi olla eri näkemyksiä siitä, mihin kategoriaan jokin kohde kuuluu. Kategorisointi voi myös olla kontekstista riippuvaa ja perustua havaittavissa oleviin ominaisuuksiin. [Halpi2007]

Edellä kuvattuun perustuen ontologiat sisältävät luokitteluita ja folksonomiat kategorisointeja. Mikäli käy ilmi, että avainsanamääritystistä saavutetaan laajan käyttäjäjoukon kesken konsensus, joka säilyy ajan kuluessa, voidaan määritystistä jalostaa lisäys ontologiaan. Näin kategorisoinnista syntyy luokittelu. [Halpi2007]

Tätä kevyemmin asiaan suhtaudutaan Kimin ynnä muiden artikkelissa [2007]. Siinä avainsanojen merkitsemistä pidetään ainoastaan käsitteiden välisten assosiaatioiden luomisena, mikä ei heidän mukaansa edellytä edes kategorisointia. Assosiaation

¹ Mukailleen lähde Jacob, Elin K. "Classification and categorization: a difference that makes a difference." *Library trends* 52.3, 2004, ss. 515-540.

² Luokittelua on määritellyt samaan tapaan myös Bunge [1998]. Hänen mukaansa hyvä luokittelu on kattava ja luokitteluperuste on pysyvä, luokat ovat erillisiä ja niiden ekstensiot ovat keskenään yhteensopivat.

luomiseen riittää jo se, että sama henkilö on kiinnostunut kahdesta eri asiasta. (Toki tätäkin voidaan pitää kategorisointina: asiat kuuluvat “kiinnostavien asioiden kategoriaan”, oma huom.)

Veres puolestaan näkee, että käyttäjät eivät merkitse avainsanoja samojen periaatteiden pohjalta kuin “virallisten” luokitteluitten kehittäjät. Käyttäjien laatimissa folksonomioissa on paljon toiminnasta lähteviä kategorisointeja, kuten on tuotu esille kohdassa 4.2.2. [Veres2006]

Veresin mielestä folksonomiat ovat kuitenkin tarpeellisia. Hänen mukaansa folksonomia muodostaa usein ainakin osittaisen käsitteenmäärittelyn [Veres2006]. Esimerkiksi liittämällä kohteeseen “Helsingin Sanomat” avainsanat “suomenkielinen” ja “valtakunnallinen” saadaan jo määriteltyä kyseisestä kohteesta kaksi olennaista piirrettä.

Myös Hotho ym. kirjoittavat, että folksonomiat ovat osista kokonaisuuksiin rakentuvia täydentäjiä “virallisille” semanttisen webin tekniikoille [2006]. Ajatuksena on, että käyttäjiä ei vaivata määrämuotoisuuden vaatimuksilla ym., mutta älykkäillä tekniikoilla kerätään implisiittisesti tietoa siitä, miten käyttäjät tulkitsevat määrätyn sanaston. Ylipäätään haasteena on, miten saavuttaa semanttisen webin hyödyt ilman että käyttäjiä rasitetaan virallisten ja formaalien teknologioiden kankeudella.

Gruberin [2007] mukaan ontologioiden vastustajat vastustavat itse asiassa useimmiten top-down-luokittelua ja informaation organisointia. Folksonomioiden ylistäminen puolestaan pohjautuu siihen, että on huomattu käyttäjien osallistumisen olevan kokonaan uusi resurssi, jota voidaan hyödyntää tiedon hakemisessa ja organisoinnissa. Gruberin mukaan avainsanatieoja koskevan ontologian tuella voisi olla mahdollista kehittää teknologioita, jotka etsivät, kokoavat ja yhdistävät ihmiset sekä sisällöt, joita tuotetaan eri puolilla webiä.

Veresin [2006] mukaan folksonomioiden ja ontologioiden selvin ero on kuitenkin siinä, että folksonomiat sisältävät adjektiiveja, verbejä ja erisnimiä, joita taas ontologioihin ei sisälly. Tähän voisi mainita, että ainakin suomalaisessa Finto-ontologiakirjastossa on listattuna paljon erisnimiä. Muilta osin Veresin esitykseen ei ole huomauttamista.

4.3.2 Folksonomiat ovat eräänlaisia ontologioita

On myös edellisten kanssa vastakkaisia näkemyksiä: Mika [2011] on sitä mieltä, että folksonomiat ovat ontologioita. Hän on kehittänyt menetelmän, jolla folksonomioista saadaan johdettua ontologia. Kyseisen menetelmän soveltamisesta saatujen kokemusten perusteella hän väittää, että folksonomioissa on enemmän semantiikkaa kuin päälle päin

näyttää. Niissä on paljon vähemmän vaihtelevuutta kuin otaksutaan ja erityisesti folksonomioiden pysyviä osia voidaan kutsua kevytontologioiksi.

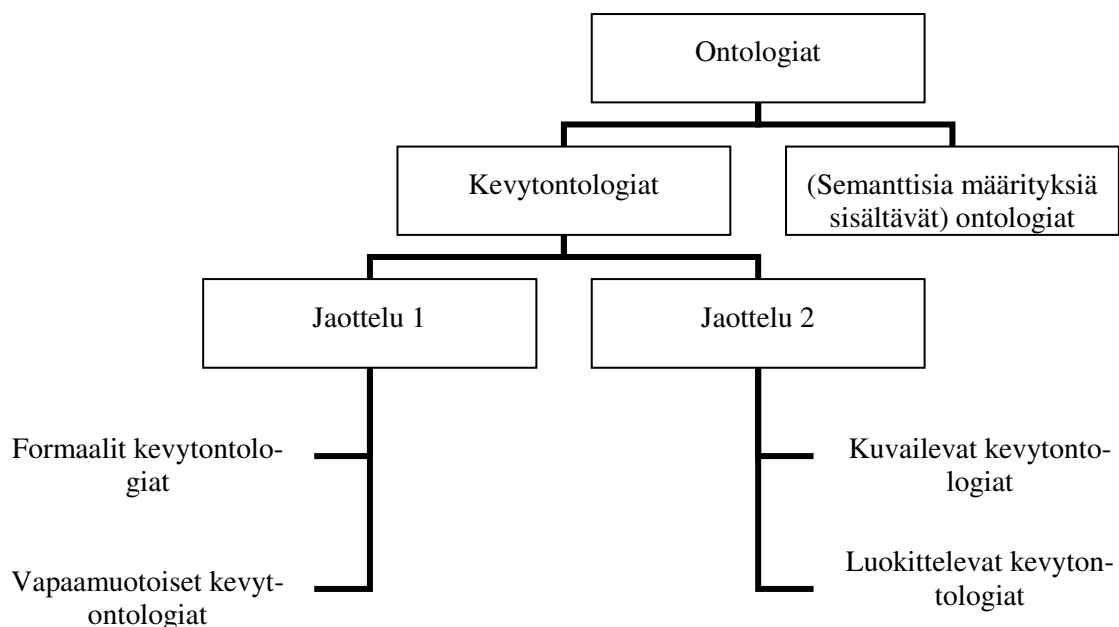
Kevytontologia-käsitettä ovat käyttäneet myös Giunchiglia ja Zaihrayeu, jotka ovat erottaneet niissä vielä erikseen formaalit ja vapaamuotoiset kevytontologiat [2009]. Näihin vapaamuotoisiin kevytontologioihin kuuluvat mm. tesaurukset, erilaiset luokittelut, yrityshakemistot ja käyttäjien itse tekemät avainsanaluokittelut.

Heidän mukaansa käyttäjien tekemistä luokitteluista on mahdollista tehdä automatisoidusti formaali *kevytontologia*, mutta täysiverisen, semanttisia määrittelyitä sisältävän ontologian muodostaminen on jopa kokeneelle asiantuntijalle suuri ja virhetulkinnoille herkkä tehtävä. Automatisoitua muunnosta haittaa esim. se, että sanat voidaan tulkita väärin niiden monimerkityksisyyden vuoksi. [Giunc2009]

Giunchiglia ja Zaihrayeu [2009] myös jakavat kevytontologiat kuvaileviin (descriptive) ja luokitteleviin (classification) ontologioihin. Kuvailevat ontologiat määrittelevät termien merkitystä ja kohdealueen luonnetta, kun taas luokittelevia ontologioita käytetään dokumenttikokoelmien tai tietovarastojen sisällön kuvailuun, luokitteluun ja hakemiseen.

Luokittelevat ontologiat ovat monimutkaisempia ja ne voivat toimia myös kuvailevina kevytontologioina, mutta päinvastoin ei voi vaihtaa: kuvaileva kevytontologia ei voi toimia luokittelevana kevytontologiana. [Giunc2009]

Tein yhteenvedon edellä esitellyistä Giunchiglia ja Zaihrayeu:n artikkelin [2009] mukaisista ontologioiden luokitteluista (kuva 1).



Kuva 1: Ontologioiden luokittelu Giunchiglian ja Zaihrayeu:n [2009] mukaan.

Giunchiglia ja Zaihrayeu määrittelevät niin, että luokittelevat ontologiat (joihin heidän mukaansa kuuluvat myös käyttäjien itsensä tekemät luokitukset) kuvaavat *dokumentteja*, eli niiden ekstensiona ovat nimenomaan erilaiset dokumentit [2009]. Tämä pitänee paikkansa monessa tapauksessa. Esimerkiksi elämäkertojen kohdalla se tarkoittaisi, että avainsanat määrittävät, mistä asioista elämäkertateksti kertoo.

Heräsin kuitenkin miettimään tätä viittaussuhdetta oman työni tapauksessa. Siinä avainsanan merkitys on mielestäni myös syvällisempi: avainsanat määrittävät, mitkä asiat ovat olleet erityisiä kirjoittajan elämässä. Ei niin, että avainsanojen kautta saataisiin tyhjentävä luettelo merkityksellisistä asioista, mutta niin, että avainsanojen esittämät asiat ovat olleet jossain elämän vaiheessa jollain tapaa merkityksellisiä kirjoittajan elämässä.

4.4 Semanttinen web

Semanttisen webin teknologioiden tarkoituksena on tehostaa webissä olevien sisältöjen hakuja ja yhdistämismahdollisuuksia. Yhdistämisen hyödyt ilmenevät erityisesti, kun ennestään suunnittelemattomat yhteistyömahdollisuudet avautuvat eri järjestelmien välille. Semanttinen web tukeutuu yhteisiin käytänteisiin, joita asiantuntijat ovat luoneet mahdollistamaan eri sisältöaineistojen yhdistäminen. Tälle täydennykseksi määritellään myös käyttäjälähtöinen näkökulma eli web 2.0.

4.4.1 Semanttisen webin määritelmä

Semanttisen webin idean isäksi mainitaan yleensä Tim Berners-Lee. Semanttinen web ei ole erillinen kokonaisuus, vaan se tarkoittaa olemassa olevaan webiin sisällytettävää määrittelykerrosta, joka avaa uusia hyödyntämismahdollisuuksia. Semanttisen webin teknologioihin kuuluvien käytäntöjen tarkoituksena on yhdistää web-sisältöihin tarpeellinen määrä koneellisesti luettavaa metadataa ja sisällön merkitystä kuvaavaa aineistoa, jotta materiaalin automaattista käsittelyä saadaan tehostettua. Tavoitteena on tuottaa sisältöaineiston hyödyntäjille parempia palveluita, jotka vastaavat käyttäjien tiedonhaku-tarpeisiin. [Berne2001]

Semanttisen webin toteutus koostuu useista eritasoisista määrittelytyökaluista: varsinaisen sisällön päälle voidaan lisätä sitä koskevaa metadataa eri toteutustavoilla. Yksi näistä toteutustavoista on perinteisesti ollut XML-kieli (eXtensible Markup Language), jonka avulla voidaan määrittää esim. web-sivujen tai niiden osien sisältöä luonnehtivia

avainsanoja ja muita ominaisuuksia. XML:n sijaan viime vuosina metadatan esittämisessä on yleistynyt JSON-formaatti (JavaScript Object Notation). JSONin etu on se, että se on koneen lisäksi helppolukuista myös ihmisille. Ohjelmat voivat hyödyntää näitä avainsanamerkintöjä, mutta ongelma on, että metadatatamerkintöjen merkitys tulee saada tietoon jostain muualta. [Berne2001] [Hyvön2018]

Aineistojen sisältöjä kuvaava metadata voidaan esittää RDF-määrittelyinä (Resource Description Framework). RDF-mallin mukaiset määrittelyt koostuvat tyypillisesti kolmikoista (triple): tällainen voisi olla esim. kolmikko {dokumentin_tunniste, ”otsikko on”, otsikon_arvo}. W3C-konsortio on luonut näille määrittelyille yhteisen rakenteen. [RDF2004]

Jotta eri sisältöjä voidaan yhdistää, tulee talletetun metadatan noudattaa yhteistä rakennetta. Sen lisäksi on oleellista, että metadata perustuu samoihin semanttisiin sanastomäärittelyihin. Yksi tällainen sanasto on luotu Dublin Core –yhteisön toimesta [Dubli2013]. Heidän sanastomäärittelynsä käsittää joukon erilaisia web-aineistojen ominaisuuksia.

Näihin sanastomäärittelyihin on merkitty, mitä ominaisuuksia aineistoista kirjataan ylös ja kuinka kyseiset ominaisuudet merkitään. Tätä kautta halutaan luoda mahdollisuus yhdistää semanttisten web-hakujen avulla sellaisia sisältöjä, joiden välille ei eksplisiittisesti ole luotu linkkiä. Tarkoituksena on, että yhtenäisten merkitsemiskäytäntöjen avulla nämä sisällöt ovat kuitenkin löydettävissä ja yhdistettävissä. Dublin Core –sanaston vaihtoehdoksi on myöhemmin noussut verkkopalvelutuottajien yhteistyönä syntynyt schema.org.¹ [Schem2017]

Yksi tärkeä komponentti semanttisen webin rakentamisessa ovat ontologiat. Metadataa abstraktimmat määrittelyt muodostavat ontologiakerroksen, jonka tarkoituksena on kuvata tiedon merkityssisältöjä. Jotta edellä kuvatuissa kolmikoissa käytettyjen termien merkitys olisi selvillä, täytyy termit yhdistää niitä vastaaviin käsitelmäärittelyksiin. Nämä määrittelyt puolestaan on kuvattu ontologioissa. Tietojen semanttiseen määrittelyyn W3C-konsortio suosittelee OWL-kieltä [OWL2012].

Yhdistämisessä käytetään apuna web-osoitteita, joista käytetään myös lyhennettä URI (Uniform Resource Identifier). Jokaisella käsitelmäärittelyllä ja jokaisella sisältöaineistolla on oma yksiselitteinen ja pysyvä osoite. URI-osoitteita voidaan käyttää siis sekä itse sisällön että siihen liittyvän käsitelmäärittelyn paikantamiseen. On siis tärkeää, että termien määrittelyt ovat kaikkien luettavissa ja ne on mahdollista löytää URIen avulla.

¹ Sanastoprojekteja on historian kuluessa ollut useita. Eräs tällainen on FOAF (Friend of a Friend). Se yhtäläillää käsittää sanaston ja kieliopin, joiden avulla voidaan kuvata henkilöiden tekemisiä ja heidän välisiä yhteyksiään siten, että eri lähteiden sisältämien tietojen yhdistäminen on mahdollista, ks. <http://www.foaf-project.org/>. Vielä yhtenä esimerkkinä voi mainita SIOC-sanastokokoelman, joka on mainittu jäljempänä kohdassa 5.2.1

Käsitteisiin liitetyt URI:it mahdollistavat myös sen, että uusia käsitteitä voidaan vaivattomasti tuoda webiin muiden hyödynnettäväksi. [Berne2001]

Edellä kuvattuja semanttisen webin määrittelytasoja on kuvannut havainnollisesti mm. Steve Bratt [2007]. Brattin mukaan on tärkeää, että kaikki määrittelyt ja standardit ovat avoimia ja ilmaisia. Vain näin taataan niiden laajamittainen käyttö, mikä puolestaan edistää semanttisen webin teknologioiden hyödyntämisestä saatavia hyötyjä.

4.4.2 Semanttisen webin hyötyjä

Semanttisen webin tarkoituksena on mahdollistaa internetissä olevien sisältöjen käyttö siten, että sisällön merkitys voidaan ottaa huomioon myös koneellisessa käsittelyssä – ei vain ihmisen suorittamassa. Semanttisen webin hyödyntäminen edellyttää yhteisten käsitteiden käyttöä, mihin tarkoitukseen voidaan soveltaa yleisesti saatavilla olevia ontologioita [Hyvön2007]. Nämä yleiset ontologiakäsitteet takaavat sen, että hakuja tekevät tahot puhuvat samaa kieltä kuin ne käyttäjät, jotka tekevät sisältöjä internetin web-sivuille.

Kun yhdistetään sisältöjä eri lähteistä, on toiminta paljon tehokkaampaa silloin, kun sisällöstä on käytettävissä rakenteista metadataa verrattuna siihen, että jouduttaisiin aina tekemään tekstianalyysiä sisällön selville ottamiseksi [Grube2008]. Hakujen määrittämisen pitäisi olla helpompaa ja tuloksellisempaa, kun aineisto on jäsennetty yleisesti tunnetuilla ontologiakäsitteillä.

Semanttisen webin teknologioiden tavoitteena on tarjota kieli, jolla pystytään esittämään sekä data että sen merkitykseen, käsittelyyn ja eheyteen liittyvät säännöt. Näiden sääntöjen tuella voidaan tehdä dataan perustuvia päätelmiä ja siirtää sääntöjä järjestelmien välillä. Sovellukset siirtävät myös tietoja, ja silloin tarvitaan todisteita sekä tietojen oikeellisuudesta että niiden alkuperästä (digitaaliset allekirjoitukset ja muut varmenneet). [Berne2001]

Kun dataan on liitetty sen merkityksen avaavaa semanttista aineistoa, yhteistyötä voivat tehdä myös sellaiset ohjelmat, joita ei ole suunniteltu toimimaan yhdessä ja ontologiat tarjoavat tähän yhteistyöhön tarvittavan sanavaraston [Berne2001]. Näitä uusia yhteistyömahdollisuuksia voidaan hyödyntää eri toimijoiden ylläpitämien tietovarantojen välillä koko internetin laajuudessa. Niitä toki voidaan hyödyntää myös yhden organisaation sisällä yhdistämään eri sovellusten tuottamia tietoja [Bratt2007].

On mahdollista tarkastella yleisellä tasolla sitä internetin web-palvelimilla olevaa aineistoa, joka on potentiaalisesti yhdistettävissä semanttisen webin teknologioiden avul-

la. Yhtenäisten määrittelytyökalujen ja -mallien avulla webiin muodostuu abstrakti kokonaisuus, johon viitataan termillä ”linked data” [Linke2013].

Tavoitteena on, että sovellukset voisivat hyödyntää tätä webissä sijaitsevaa, toisiinsa yhdistettävissä olevaa dataa, mikäli metadata noudattaa yhtenäistä tietomallia ja on koodattu yleisesti tunnetuilla määrittelykielillä. [Dubli2013] Datan yhteiskäyttöä tukemaan on kehitetty myös kyselykieli SPARQL¹, jonka on tarkoitus tukea webissä olevaan ”linked data” -kokonaisuuteen kohdistuvia kyselyitä [W3C2017].

Semanttisen webin todellinen voima ilmenee vasta, kun on laadittu paljon ohjelmia, jotka keräävät sisältöä eri lähteistä ja vaihtavat sisältöjä muiden sovellusten kanssa [Berne2001]. Jo tällä hetkellä on tehty valtavasti yhteistyötä sen eteen, että saadaan yhdistettyä verkossa olevat tietovarannot tavalla, joka mahdollistaa niiden monipuolisen käytön eri sovelluksissa. Tätä globaalia, verkossa olevaa tiedon kokonaisuutta kutsutaan myös nimellä Giant Global Graph (GGG). Sen keskeinen osa on avoimista data-aineistoista muodostettu pilvi, Linked Open Data Cloud², jossa on jo haettu semanttiset vastaavuudet eri tietosisältöjen välille. [Hyvön2018]

Gruber [2008] huomauttaa kuitenkin, että vaikka semanttisen webin teknologiat mahdollistavat tietojen yhdistämisen eri järjestelmistä, täytyy semanttisen yhteentoimivuuden lisäksi ottaa huomioon, että yhdistettävä data on voitu näissä järjestelmissä tallettaa eri tietomallien ja talletusformaattien pohjalta.

W3-konsortion sisälle on syntynyt joitakin alaryhmiä, joiden tavoitteena on edistää erityisesti oman alansa semanttisen webin resurssien ja teknologioiden kehitystyötä. Yksi tällainen on esimerkiksi terveydenhuoltoalan ja biotieteiden ryhmä ”Semantic Web in Health Care and Life Sciences Community Group”.³ [W3C2017]

4.4.3 Web 2.0

Käyttäjät tiedon tuottajina

Kun puhutaan semanttisesta webistä, kyseessä on enemmän kehittäjien, ylläpitäjien ja asiantuntijoiden näkökulma, ei niinkään käyttäjien. Semanttisen webin tavoitteena on mahdollistaa internetin sisältöjen yhdistäminen automatisoitujen järjestelmien avulla, ja nämä järjestelmät on yleensä laadittu asiantuntijoiden toimesta.

¹ https://www.w3.org/standards/techs/sparql#w3c_all

² ks. <https://lod-cloud.net/>

³ <https://www.w3.org/community/hclscg/>, viitattu 20.6.2017

Web 2.0 lähestyy samaa asiaa käyttäjien näkökulmasta. Web 2.0:lle tyypillisiä sovelluksia ovat sellaiset, joissa julkaiseminen ja tiedonhallinta on käyttäjäkeskeistä, kuten on laita esim. wikien, blogien ja muiden yhteisöpalveluiden kohdalla [Hotho2006]. Web ei kuitenkaan ole pelkkä julkaisemisen foorumi, vaan osallistuminen ja käyttäjien toiminta ovat tärkeitä tekijöitä web 2.0:n kokonaisuudessa [O'Rei2007].

Semanttisen webin teknologiat on suunniteltu automatisoituja järjestelmiä varten, web 2.0 puolestaan tukee ihmisen toimintaa ja sosiaalista vuorovaikutusta. Web 2.0:n tavoitteena on yhdistää internetissä olevat sisällöt epämuodollisimmilla keinoilla, kuten folksonomioiden avulla ja niin, että käyttäjät ovat tärkeä lenkki yhdistämisessä. Folksonomiat mahdollistavat hajautetun semantiikan kehittymisen, jossa siis semanttisia määrittäjiä tuotetaan usean tahon yhteistyönä. [Spyns2006]

Osallistumisen lisäksi myös tiedonhallinta on tärkeää: web 2.0 -tyyppiset sovellukset hyödyntävät avointa ja jaettua dataa ja sitä dataa, joka kertyy käyttäjien käyttäytymisestä. Tiedonhallinta on toteutettu niin, että se tukee yhdistämistä ja osallistumisen hyödyntämistä. Näin saadaan myös hyödynnettyä kollektiivisen tietämyksen, kollektiivisen älykkyyden¹ voimavarat. [O'Rei2007]

Seuraavassa on yksi esimerkki web 2.0:lle tyypillisestä tavasta hyödyntää toiminnan kautta syntyvää palautetta, joka koskee järjestelmään kertyneen tietämyksen laatua: Hakutuloksia tarjotessaan järjestelmät käyttävät hyväkseen muiden käyttäjien merkitsemiä avainsanoja. Kun käyttäjä on tehnyt haun, järjestelmä pyytää häntä tämän jälkeen liittämään klikkaamaansa hakutulokseen jonkin avainsanan. Näin järjestelmä pyytää tavallaan välillisesti palautetta avainsanoista, joita muut käyttäjät ovat kyseiseen materiaaliin liittäneet [Halpi2007].

Mitä enemmän käyttäjät vahvistavat yhteyttä joidenkin asioiden välillä, sitä vahvempaa yhteys näkyy web 2.0:n sisällöissä [O'Rei2007]. Järjestelmä oppii koko ajan käyttäjien toiminnasta. Toimintohistoriaa hyödynnetään kahdella tapaa: käytetään hyväksi tietoja yksittäisen käyttäjän omasta aikaisemmasta käyttäytymisestä ja yhdistetään siihen myös tietoja kaikkien käyttäjien toiminnasta.

Ohjelmistot palveluina

O'Reillyn mukaan World Wide Web voidaan nähdä eräänlaisena sovellusalustana, ja Web 2.0:n sovellukset ovat luonteeltaan palveluita (käsite ”ohjelmisto palveluna” eli ”software as a service”, SaaS). Web 2.0 -sovelluksissa ei siis perinteiseen tapaan ole varsinaisesti eri versioita ja julkaisuajankohtia. Nämä palvelut ovat koko ajan kehitys-

¹ Käsite "collective intelligence"

työn alla, sillä ne keräävät tietoa käyttäjän toiminnasta ja täydentyvät sitä mukaa, mitä enemmän niitä käytetään. Palvelut uusiutuvat viikoittain, jopa päivittäin. [O'Rei2007]

Esimerkkinä käyttäjien käyttäytymisen mukaan päivittyvästä palvelusta voisi olla roskapostin suodatusjärjestelmä. Postin vastaanottaja voi merkitä saamansa viestin roskapostiksi. Mikäli riittävän monet käyttäjät merkitsevät saman viestin roskapostiksi, saa se yleisen ”roskapostin statuksen”, jolloin sitä ehdotetaan roskapostiksi myös kaikille muille käyttäjille.

Palvelut sisältävät eri käyttäjien tuottamaa sisältöä, joka on jo valmiiksi talletettu sellaiseen muotoon, että se on mahdollista yhdistää muiden käyttäjien sisältöjen kanssa. Myös palveluiden toteutus näkyy avoimesti¹. Näin monet pienet tekijät voivat yhdessä muodostaa todella merkittäviä kokonaisuuksia². Yhdistämisessä hyödynnetään sitä tietoa, joka on talletettu käyttäjien käyttäytymisestä [O'Rei2007].

Esimerkiksi karttasovellus näyttää kartalla yrityksiä, nähtävyyksiä jne. sen perusteella, mitä muualla internetissä on merkitty näiden kohteiden osoitteiksi. Kun käyttäjät usein klikkaavat määrättyä nettisivua näiden yritysten nimeä haettuaan, liitetään sekin kartassa oleviin yrityksen tietoihin. Näin käyttäjien toiminta vahvistaa talteen jäävää informaatiota, jonka mukaan kyseinen yritys ja hakutuloksissa näytetty verkkosivusto liittyvät toisiinsa.

Tällainen ”osallistumisen arkkitehtuuri” tarkoittaa käytännössä sitä, että palvelu paranee koko ajan sitä mukaa, mitä enemmän sitä käytetään. Jotta käyttökertoja kertyisi paljon, on web 2.0 -sovelluksilla tavoitteena, että ne mukautuvat kaikenlaisiin käyttötilanteisiin ja laiteympäristöihin. Nimenomaan on tärkeää, että sovellus ei ole pelkkä pc-sovellus, vaan sitä voidaan käyttää myös kaikilla mobiililaitteilla missä tahansa. [O'Rei2007]

Gruberin mukaan perinteinen tapa käyttää webin sisältöä on tehdä hakuja ja käyttää sitä tietolähteenä. Tulevaisuudessa web ymmärretään yhä enemmän inhimillisen toiminnan ja automaation yhdistelmänä, jossa opitaan ja tiedetään asioita yhteisöllisesti. [Grube2008]

Web 2.0 on jo vakiintunut käsite, mutta sen rinnalle jo kymmenisen vuotta sitten lanseerattu web 3.0 hakee vielä vakiintunutta määritelmäänsä. Joidenkin mukaan semanttinen web on sama kuin web 3.0, ks. esim. [Hyvön2018]. Toiset lähteet näkevät web 3.0:n yhdistelmänä, jossa semanttisen webin teknologiat ja web 2.0:n käyttäjien osallistuminen kohtaavat.

Esim. Gruberin mukaan web 3.0:lle on ominaista osallistumisen ja yhteistyön korostaminen [2008]. Siinä yhdistetään semanttisen webin tietovarantoja kollektiiviseen älyk-

¹ Esim. web-sivun lähdekoodin voi nähdä komennolla "view source".

² Esim. ebay.com ja amazon.com ovat suuria kauppapaikkoja, jotka yhdistävät useita pieniä toimijoita

kyyteen. Hän arvioi, että osallistuminen ja web 2.0:n työskentelytavat sopivat myös teellisen tiedon tuottamiseen.

Gruber jopa määrittelee erityisen sovellustyyppin: kollektiivisen tietämyksen järjestelmät [2008]. Niissä avaintekijöinä ovat käyttäjien sosiaalisessa kanssakäymisessä tuottamat sisällöt ja koneellisen käsittelyn tehokkuus, joka mahdollistaa ennennäkemättömän kohdealueen kattavuuden, näkökulmien monimuotoisuuden ja informaation määrän.

Tämän työn aihepiiriin suhteutettuna folksonomiat edustavat web 2.0 -henkistä näkökulmaa ja ontologiat puolestaan ovat semanttisen webin asiantuntijavetoinen lähestymistapa. Edellä mainituilla käsitteillä ilmaistuna tämä työ vertaa semanttisen webin ja web 2.0:n sisältöjen yhteensopivuutta, kun aineistona ovat elämänvaiheita kuvaavat tekstit.

4.5 Elämäkerta, elämänkaarikirjoitus

Minkälaiset ominaisuudet ovat tyypillisiä elämäkertoille ja elämänkaarikirjoitukselle? Mikä on elämäkerta-aineistojen merkitys tutkimukselliselta näkökannalta?

Elämäkerrat eivät ole merkityksellisiä pelkästään yksilön näkökulmasta: tutkimuksen kannalta elämäkertoja on pidetty kiinnostavina aineistoina 1970-80-luvuilta lähtien. Sosiaalitieteissä siirryttiin Suomessa 1980-luvulla elämäntavan tutkimuksesta elämäkertojen tutkimukseen, samalla vuosikymmenellä niistä kiinnostuttiin myös etnologian ja folkloristiikan piirissä. [Pelto2006]

Folkloristisessa tutkimuksessa muisteluaineistot ja elämäkerrat voivat olla hyvin monenlaisten tutkimuskysymysten aineistoina yhdessä erilaisten tekstien, haastatteluiden, päiväkirjojen ja mediatekstien kanssa. Aineistolajina ne ovat narratiivisia, kerronnallisia ja niistä voidaan hakea merkityssuhteita, jotka yhdistävät tapahtumat toisiinsa. Ne sisältävät faktatietoa, mutta tärkeintä muisteluteksteissä on tulkinnan ja merkityksen antaminen menneisyydelle. [Ukkon2006]

Elämäkertoja voidaan kirjoittaa myös omille jälkeläisille ilman minkäänlaista tutkimusintressiä. Minkälaiset piirteet ovat tyypillisiä elämänkaarikirjoitukselle ja elämänvaiheita dokumentoivalle aineistolle? Kenelle aineisto on kirjoitettu? Montonen muistuttaa elämänkaarikirjoitusta käsittelevässä oppaassaan kirjoittajaa siitä, että todellinen lukijakunta saattaa olla aivan eri kuin mitä kirjoittaja mielessään ajattelee [Monto2008].

Montosen huomautus tuo esiin seikan, joka puoltaa semanttisen sisällön tukemista johonkin yleiseen viitekehykseen, kuten ontologiaan. Tämän työn aineistona ovat elämäkertakirjoitukset, joita jäsennetään yleisillä ontologioilla.

Montosen mukaan elämäntarinalla voi olla erilaisia rakenteita: kronologisen järjestyksen sijasta aineisto voidaan järjestää teemoittain (lapsuus, työ, perhe, harrastukset jne.). Kerronta voi edetä myös psykologisesti, esim. kipein asia ensin. Aineiston jaottelu voidaan tehdä henkilöiden tai paikkojen perusteella, tai tarina voi edetä sirpalemaisesti vaipaitten assosiaatioiden mukaan, ilman mitään erityistä ”punaista lankaa.” Kirjoitusjärjestys voi myös olla eri kuin lopullinen esittämisjärjestys. [Monto2008] Nämä seikat huomioon ottaen hyperteksti tuntuu luontevalta organisointimuodolta elämänvaiheita kuvaavalle materiaalille.

Kirjoittamista voidaan ruokkia sitomalla aineistoa sen aikakauden yleisiin tapoihin ja ilmiöihin. Yhdessä muiden kanssa voi kerätä muistikuvia siitä, miten tietynä aikana pukeuduttiin, minkälaisia kodit tai autot olivat, mitä musiikkia kuunneltiin, minkälaista ohjelmaa tuli televisiosta jne. Ottamalla ajan ilmiöitä mukaan kuvaukseen yksinkertaisesta arjenkuvauksesta voi saada kiinnostavan ja monipuolisen [Monto2008].

Tässä tulee esille tämän työn yksi mahdollinen jatkotutkimustavoite: muisteluaineiston laajentaminen ja rikastaminen semanttisen webin hakujen avulla. Jäsentämällä aihealuiden käsitteitä voi käyttää hyväksi internetin tietovarastoja ja hakea muistamisen tueksi tietoa esim. siitä, mitkä asiat olivat ajankohtaisia tietynä menneisyyden ajanhetkenä.

4.6 Muistitieto

Muistitietoa käytetään useilla eri tieteenaloilla, erityisesti historian ja kulttuurien tutkimuksessa. Sen perusteella esitetään muistitiedon määritelmä historian tutkimuksen näkökulmasta ja käydään läpi muistitiedon ominaispiirteitä.

4.6.1 Muistitiedon merkitys tutkimuksessa

Historiantutkimuksen, etnologian ja folkloristiikan tutkimuksessa yksi käytettävistä tiedon lajeista on muistitieto. Näistä tieteenaloista ainakin etnologian piirissä muistitieto on ollut jo viime vuosisadan alkupuolelta lähtien yksi tutkimuksen kohde. Alkuun kiinnostus muistitietoon keskittyi kaikille yhteisen muistitiedon, eräänlaisen kollektiivisen muistin tutkimukseen, ja muistitietoa käytettiin mm. perinteiden tutkimisessa. Myöhemmin alettiin kiinnittää huomiota yksilön muistitiedon ja kollektiivisen muistitiedon eroihin. [Korkia2006]

Näille em. tieteenaloille on viime vuosikymmenien aikana kehittynyt suuntauksia, jotka ottavat historiankirjoitukseen mukaan yksilötason mikrohistorian. Niissä ei tavoitella

yhteistä ja yhtenäistä historiankirjoitusta, vaan ennemminkin moniäänistä ja näkökulmiltaan rikasta tarinaa. Tästä johtuen mikrohistoriallisessa tutkimustavassa annetaan arvoa myös muistitiedolle [Finge2006-1].

Muistitietoa on pidetty tutkimuksen kannalta merkittävänä sen vuoksi, että historian ja kulttuurien tutkimuksessa on 1980-luvulla alettu arvostaa tavallisen ihmisen näkökulmaa: on esitelty käsite ”historia alhaalta” (Jorma Kalelan suomennos), englanniksi ”history from below”. Muistitietoa on kerätty myös muista kuin tutkimuksellisista lähtökohdista: 1970-luvulta alkaen Yhdysvalloissa ja Britanniassa on kouluissa kerätty kotiseutuhistoriaa haastatteleamalla vanhoja ihmisiä. On myös toteutettu kirjakahviloiden yhteydessä hankkeita, joissa on tuotettu erilaisten ihmisten omaelämäkertoja. [Kalel2006]

Fingerroosin ja Peltosen mukaan muistitietotutkimus on monitieteistä ja –metodista ja tyypillisesti muistitietotutkijat ovat joltain muulta kuin historian tutkimuksen alalta. [Finge2006-2] Koska oma tutkimukseni käyttää aineistonaan juuri muistitietoa, luodaan seuraavassa lyhyt katsaus siihen, mitä muistitieto on ja mitä siitä sanotaan historian tutkimuksessa.

4.6.2 Muistitiedon määritelmä

Muistitieto (oral history) ei englanninkielisestä nimestään ”suullinen historia” huolimatta rajoitu pelkkään suulliseen perinteeseen, vaan se määritellään ”tiedoksi, joka ei pohjaudu kirjallisiin lähteisiin, vaan tiedonantajan muistiin” [Pelto2013]. Suullinenkin tieto tallennetaan usein tekstinä, joten muistitietolähteiden piiriin kuuluvat sekä puhe että teksti [Finge2006-2].

Fingerroosin ja Haanpään mukaan lähes kaikissa heidän lukemissaan muistitietotutkimuksissa muistitietoa pidetään ”erilaisissa muistelutilanteissa tuotettuna konstruktiona” ja tiedon tuottajina toimivat muistelijat. Motivaatio muisteluun voi olla esim. tietyn yhteisön halu kirjoittaa ylös omaa menneisyyttään. [Finge2006-1]

Muistitietoa voidaan siis kerätä sekä haastatteleamalla että kirjallisessa muodossa, jolloin kirjoittajina toimivat muistelijat itse. Suomessa on aika ajoin kerätty johonkin asiaan liittyvää muisteluaineistoa järjestämällä kirjoituskilpailu, esim. vantaalaisten omaelämäkertoja kerättiin tällä menetelmällä vuonna 1978¹.

¹ Keruun järjestivät Yleisradio ja Suomen Akatemia. Tietoja tästä sekä muista perinteen ja nykykulttuurin keruuaineistoista on Suomalaisen kirjallisuuden seuran verkkopalvelussa www.finlit.fi

Kirjoittamistilanne on haastatteluun verrattuna yksityinen ja ohjailematon, eikä tutkijalla ole mahdollisuutta tarkentaviin kysymyksiin. Aineiston analysoinnissa on näin ollen käytettävä apuna muita lähteitä. Tutkijan kannalta tällaisen muistelijoiden itsensä kirjoittamista teksteistä koostuvan aineistokokonaisuuden suurin ongelma on se, että kirjoitukset ovat hyvin heterogeenisiä sekä tyyliltään että sisällöltään. Ne eivät kommunikoi keskenään, eivätkä tekstit anna tukea toisten tekstien analysointiin. [Pöysä2006]

4.6.3 Muistitiedon ominaispiirteitä

Voiko muistinvaraiseen tietoon luottaa? Tutkimuksen kannalta muistitietoa pidetään luotettavana, kun henkilö kertoo itse omista kokemuksistaan ja lisäksi muisteltavasta asiasta on käytettävissä useita yhtäpitäviä todistajalausuntoja [Pelto2013]. Kiinnostavia eivät kuitenkaan ole pelkät faktatiedot tapahtuneesta, vaan muisteluitten ylöskirjaamisessa on kaksi ulottuvuutta: toisaalta se asia, jota muistellaan ja toisaalta se, mitä ihmiset muistavat ja uskovat. Myös tämä jälkimmäinen on historiallinen fakta.

Se, mitä ihmiset nostavat esiin ja miten he sen tulkitsevat, on kiinnostavaa muistitiedon kannalta. Muistitieto ”kertoo enemmän tapahtumien merkityksestä kuin itse tapahtumista” [Porte1998]. Tähän on kiinnittänyt huomiota myös Korkiakangas [2006]. Hänen mukaansa muistitieto ei kerro todellisuudesta, vaan se välittää koetun todellisuuden tulkintoja.

Muistitieto kertoo usein myös yhteisöllisestä kokemuksesta, vaikka kertojana on yksityishenkilö ja hän kertoo omasta elämästään. Yleensä kuitenkin historialliset tapahtumat kerrotaan yksityisissä muisteluissa siten kuin yhteisö on ne aikanaan kokenut ja tulkinut. [Ukkon2000]

Vaikka elämäkerrallinen aineisto kertoo asioista siten kuin yksittäinen henkilö on ne kokenut, muodostuu erillisten muisteluaineistojen kautta myös kokonaiskuva siitä, miten suuremmat ryhmät, yhteisöt tai koko kansakunta ovat asiat kokeneet tai arvottaneet. Kertoja yleensä sitoo oman tarinansa yleisiin merkityksiin. [Tuoma2006] Käsitteiden tasolla folksonomiat voivat edustaa kertojan omaa kokemushorisonttia ja yleiset ontologiat taas kulttuurin yhteistä tulkintaa ja merkityksenantoa eri käsitteille.

Tietoteoreettisesti muistitietotutkimuksessa voidaan Ukkosen [2006] mukaan erottaa realistinen ja tulkinnallinen suuntaus. Realistinen suuntaus olettaa, että maailma on olemassa tietynlaisena havaintojen tekijästä riippumatta. Muistitiedon avulla haetaan faktoja sekä tietoa siitä, mitä on tapahtunut ja miten asiat ovat olleet. Tulkinnallinen muistitietotutkimus hakee vastauksia siihen, mitkä asiat ovat muistelijan mielestä muis-

telemisen arvoisia ja mikä on muistelijan suhde yleisesti hyväksytyihin menneisyyden tulkintoihin. [Ukkon2006]

Mikäli muisteluaineistoja halutaan käyttää faktatietojen lähteinä, on aineistojen sisältöihin liittyvien verkkolähteiden hakemisella tärkeä funktio faktojen todentamisessa. Tällöin olisi tärkeää, että elämäkertojen ja verkossa olevien tietoaaineistojen käyttämät käsitteet ovat yhteneväisiä.

Tämän tutkimuksen elämäkerta-aineistot ovat pääosiltaan muistitietoa, sillä niistä suuri osa on talletettu kirjoittajan muistin pohjalta. Lisäksi osa kirjatuista tapahtumista on talletettu kollektiivisen muistelutuokion pohjalta: esim. sisarukset ovat yhdessä kokoontuneet muistelemaan jotain määrättyä asiaa, henkilöä tai tapahtumaa. Myös kansalaisopiston elämäkertapiirin parissa on järjestetty muistelutuokioita yhteisistä aiheista. Mukana on kuitenkin myös muunlaista tietoa-aineistoa: valokuvia, karttoja, asiakirjoja, väestörekisteritietoja ym.

5 Aikaisempi aihetta koskeva tutkimus

Seuraavaksi esitellään aiheeseen liittyvää aiempaa tutkimusta. Alussa kerrotaan lähdeaineiston kartuttamisesta sekä eri tavoista yhdistää ontologiat ja folksonomiat (kohta 5.1). Sen jälkeen tarkastellaan erityisesti ontologioiden käyttämistä luokittelussa (kohta 5.2). Kohdassa 5.3 tutustutaan menetelmiin, joilla niiden kehittäjien mukaan folksonomia voidaan muuntaa yleiseksi ontologiaksi.

Tämän jälkeen kuvataan tapoja, joilla folksonomiaa saadaan terävöitettyä lisäämällä siihen semanttista sisältöä (kohta 5.4). Lopuksi suhteutetaan toisiinsa tämä työ ja aiemmin esitetyt tutkimukset (kohta 5.5).

5.1 Eri tapoja ontologioiden ja folksonomioiden yhdistämiseksi

Tässä kuvataan, miten aikaisempaa tutkimusta on kartoitettu ja esitellään viitekehys, jonka avulla voidaan hahmottaa ontologioiden ja folksonomioiden suhdetta.

5.1.1 Lähdeaineiston kerääminen

Kun tätä työtä koskeva tutkimusidea oli syntynyt, täytyi aluksi selvittää, onko asiaa tutkittu jo aikaisemmin. Tein hakuja tieteellisistä artikkeleista siten, että käytin hakusanoina termejä ”folksonomy” ja ”ontology”. Hakutulosten tarkastelu on rajattu koskemaan vain sellaisia töitä, joissa puhutaan molemmista: sekä folksonomioista että ontologioista.

Näiden lisäksi on tehty tutkimuksia, joissa on käsitelty folksonomioita, mutta tavoitteenä ei ole ollut niiden lähentäminen ontologioiden suuntaan. Sellaiset tutkimukset on kuitenkin rajattu pois tästä tarkastelusta. Jatkoin hakujen tekemistä niin kauan, että senhetkinen tilanne aiheen tutkimuksessa vaikutti olevan katettu.

Lähdeaineistoon tutustuttaessa voi käyttää viitekehystenä Limpensin, Gandonin ja Buffan [2009] tekemää yhteenvetoa eri tavoista yhdistää ontologiat ja folksonomiat. Perehdymme seuraavassa siihen.

5.1.2 Limpensin, Gandonin ja Buffan viitekehys

Limpens, Gandon ja Buffa [2009] ovat tutkineet yleisellä tasolla folksonomioiden ja ontologioiden välistä yhteyttä ja erottaneet asiaa käsittelevässä lähdeaineistossa erilaisia lähestymistapoja. Ensinnäkin, on tehty kokeiluja, joissa ontologioita johdetaan folksonomioista. Toisissa lähestymistavoissa folksonomioita käytetään ontologioiden levittämiseen tai uusien ontologioiden eräänlaiseen testaamiseen. Folksonomioihin voidaan myös lisätä semanttista aineistoa, joka rikastaa niitä ontologioiden suuntaan.

Seuraavassa tarkastellaan hieman tarkemmin neljää eri tapaa, joilla Limpensin ym.:iden mukaan folksonomiat ja ontologiat voidaan yhdistää toisiinsa [Limpe2009].

1) Ontologiat johdetaan folksonomioista (ks. myös tämän työn kohta 5.3): lähestymistapa johtaa usein suppeisiin ja pinnallisiin ontologioihin, joiden semanttinen sisältö jää vähäiseksi.

2) Ontologioiden ja folksonomioiden kehitys ja käyttö synkronoidaan keskenään: ontologioiden käyttöä tuetaan ja markkinoidaan tarjoamalla sen sisältämiä käsitteitä folksonomioiden laatijoille. Ontologioita puolestaan laajennetaan folksonomioiden edellyttämällä tavalla.

3) Lisätään folksonomioihin semanttista sisältöä (ks. kohta 5.4).

4) Yhdistetään folksonomiat ja ontologiat semanttisen webin tekniikoilla (ks. myös esim. [GoodB2007]).

Koska semanttisen informaation sisällyttäminen on tärkeää webin yhteisten tietovarantojen hyödyntämiseksi, ehdotetaan Limpensin ym.:iden tutkimuksessa mahdollisina suositeltavina toimintamalleina, että folksonomioista muodostetaan eräänlainen ontologian kevytversio tai folksonomioiden muotoutumista on jo etukäteen ohjailtu valmiiden sanastojen avulla [Limpe2009].

Yllä kohdassa 2 esitetty ontologioiden ja folksonomioiden yhdistäminen voidaan toteuttaa myös puolittain: aineistoa viedään vain folksonomioista ontologioiden suuntaan, jolloin ontologioita laajennetaan folksonomioiden perusteella. Esim. Torniai ym. ovat [2008] yhdistäneet verkko-oppimistyökalussa opiskelijoiden tekemiä folksonomioita asiantuntijoiden ylläpitämiin ontologioihin. Tavoitteena on ollut tuottaa apuvälineitä ontologioiden kehitystyöhön.

Torniain ym.:iden tutkimuksessa opiskelijoille annettiin verkko-oppimistyökalun puitteissa mahdollisuus merkitä kurssisisältöjä avainsanoilla. Automatisoitu työkalu keräsi tietoa opiskelijoiden käyttämien avainsanojen suosiosta ja yhdisti niitä ontologiarakenteeseen. Yhdistämisessä käytettiin hyväksi Torniain ym.:iden laatimaa

metodia, joka määrittä avainsana-ontologia -käsitteille kertoimen. Tämä kerroin kuvasi sitä, miten läheisesti avainsana ja ontologiakäsite liittyivät toisiinsa. [Torni2008]

Kun havaittiin, että ontologia ei sisältänyt jotain sellaista avainsanaa, joka oli tullut suosituksi ja jolle olisi tarvetta, työkalun avulla tuotettiin ontologian ylläpitäjälle automaattisesti ehdotus, mikä avainsana olisi hyvä lisätä ontologiarakenteeseen. Näin ontologioiden ylläpitämisestä vastaavat opettajat voivat päivittää ontologian vastaamaan senhetkisiä tarpeita. [Torni2008]

Myös seuraavana esiteltävä Passantin tutkimus [2007] sekä tämä työ suosivat tätä samaa toisena esitettyä vaihtoehtoa: folksonomioitten käyttöä terävöitetään ontologioiden hyödyntämisellä (folksonomioita verrataan ontologioihin). Tarvittaessa ontologioita voitaisiin myös täydentää folksonomioitten esiin tuomalla tavalla. Lisäksi jatkotutkimuksena voitaisiin soveltaa yllä esitetyistä Limpensin yhdistämismahdollisuuksista tapaa 4, eli voitaisiin tutkia aineiston laajentamista semanttisten web-hakujen avulla.

5.2 Ontologioiden käyttäminen luokittelussa

On tehty useita eri tutkimuksia, joissa ontologioita on hyödynnetty käyttäjien tekemien avainsanajäsennysten tukena. Näistä esitellään perusteellisimmin Passantin [2007] tutkimus, sen lisäksi tehdään katsaus muihin samankaltaisiin kokeiluihin.

5.2.1 Weblogien jäsentäminen ontologioiden avulla

Passant on tutkinut weblogien jäsentämistä ontologioilla ja folksonomioilla, ks. [Passa2007]. Jäsentämisessä on käytetty sekä kohdealueen omia ontologioita että SIOC-sanastoa (SIOC= Semantically interlinked online communities¹), joka on tarkoitettu käytettäväksi, kun halutaan jäsentää verkossa toimivien yhteisöjen tuottamia sisältöjä.

Passantin [2007] tutkimuksessa suuren yhtiön työntekijöiden annettiin vapaasti jäsentää ammatillisiin asioihin liittyviä blogitekstejään omilla avainsanoilla. Sen jälkeen jokainen avainsana linkitettiin joko kyseisen kohdealueen oman ontologian sisältämään käsitteeseen tai suoraan johonkin esiintymään (esim. toiseen yritykseen, web-palveluun jne.) Täten saatiin terävöitettyä sisällön semanttista luokittelua, mikä johti parempiin mahdollisuuksiin saattaa yhteen samansisältöisiä blogitekstejä.

¹ <http://sioc-project.org/>

Passantin suunnittelema systeemi tarjosi myös mahdollisuuden tehdä myöhemmin blogiteksteihin hakuja avainsanojen avulla. Tällöin huomattiin, että samalla avainsanalla on voitu viitata useaan eri ontologiakäsitteeseen. Termien monimerkityksisyydestä johtuen hakujen yhteyteen täytyi myös luoda toiminto, jolla hakujen tekijä pystyi valitsemaan useasta vaihtoehdosta, missä merkityksessä hän tarkoitti käyttää valitsemaansa hakusanaa. [Passa2007]

Käyttäjää palveltiin hakujen yhteydessä tuomalla tarjolle tuloksia, jotka saatiin laajentamalla hakuja ohjelmallisesti. Jos käyttäjä haki avainsanalla A, hänelle tarjottiin myös tuloksia, joissa esiintyy hakusana B, mikäli A ja B oli merkitty saman blogitekstin avainsanoiksi. Käyttäjälle voidaan tuoda myös tekstejä, jotka on luokiteltu samalla ontologiakäsitteellä (esim. ohjelmointikieli), vaikka tekstit edustavat kyseisen käsitteen eri esiintymää kuin millä haku on tehty.

Esimerkiksi haku on tehty hakusanaalla ”PHP”, jolloin käyttäjälle tuodaan myös avainsanalla ”Perl” merkittyjä sisältöjä, koska molemmissa on luokitteluna ”ohjelmointikieli”. Käyttäjälle saatetaan myös tarjota hänen haussa käyttämänsä käsitteen aliluokan esiintymiä (esim. hakemalla maa=Ranska tuodaan tarjolle myös tuloksia, joissa kaupunki=Pariisi). [Passa2007]

Edellä mainittujen hakulaajennusten lisäksi Passantin [2007] systeemi tarjosi välineitä, joilla ontologioita voidaan muokata ja täydentää käytön aikana tulleitten muutostarpeitten mukaisesti.

Tämän tutkimuksen asettelu on hyvin paljon samankaltainen kuin Passantilla lukuun ottamatta hakujen toteuttamiseen liittyviä toiminnallisia laajennuksia sekä ontologioiden muokkaustyökaluja. Tekstien luokittelu tehdään kuitenkin samaan tyyliin kuin Passantilla: käyttäjät merkitsevät ensin avainsanat vapaasti, oman mielensä mukaan, ja vasta sen jälkeen avainsanakäsitteitä verrataan ontologioiden sisältämiin käsitteisiin. Näin olen mielenkiintoista nähdä, tuleeko vastaan samanlaisia kokemuksia liittyen avainsanojen valintaan sekä synonyymien ja homonyymien käyttöön.

5.2.2 Muita lähestymistapoja ontologioiden hyödyntämiseen avainsanaluokitteluissa

Good ym. [2007] ovat luoneet hieman samankaltaisen systeemin kuin Passant. He kehittivät yhteisöllisiin avainsanaluokitteluihin tarkoitetun järjestelmän, jossa biolääketieteen asiantuntijat saivat luokitella avainsanoilla oman alansa lähteitä. Luokittelut yhdistettiin alan yleiseen ontologiaan ja tehdyt luokittelut talletettiin verkkoon avoimesti kaikkien saatavilla olevaan tietokantaan.

Yhdistämällä yhteisön tuottamat avainsanaluokittelut ja asiantuntijoiden laatima ontologia päästään heidän mukaansa lähemmäksi eheää ja laadukasta luokittelua, vaikka yhteisöllisesti tuotettu luokittelu ei koskaan ole täysin verrattavissa ammattimaisesti tehtyyn vastaavaan. Syyksi he mainitsevat yksinkertaisesti sen, että käyttäjien taustat ovat kirjavia eikä heillä ole käytössään samantasoisia välineitä kuin ammattilaisilla. Tämän seurauksena käyttäjien tekemät avainsanaluokittelut ovat ”virhealttiita, omalaatuisia ja epäjohdonmukaisia.” [GoodB2007, s.17]

Tähän samaan ideaan perustuvat myös jo aiemmin tässä työssä (ks. kohta 4.1.3.) mainitut, Braunin ym:iden kehittämät kaksi järjestelmäkokeilua [2007]. Niissä käyttäjät merkitsevät löytämiään sisältöjä avainsanoilla, jotka voivat olla uusia tai peräisin yhteisesti käytössä olevasta taksonomiasta. Tämä taksonomia käsittää sekä yleisen ontologian että käyttäjien siihen tekemät lisäykset. Ontologia siis täydentyy ja muokkautuu koko ajan sitä mukaa, kun uusia avainsanoja kerätään yhteiselle alustalle.

Spyns ym. ovat kehittäneet edellisten kaltaisen järjestelmän [2006], jossa selaimen liitettävällä lisäosalla mahdollistetaan ontologiakerroksen yhdistäminen yhteisen verkkosivuston sisältämiin avainsanamäärittelyksiin. Lisäosan avulla käyttäjä voi merkitä avainsanoiksi ontologiakäsitteitä, joiden määrittelyt on koottu yhteiseen tietokantaan.

Ontologialla on erikseen määritelty ylläpitäjätaho, ja käyttäjät voivat tehdä ehdotuksia uusista käsitteistä. Käyttäjät eivät voi suoraan lisätä käsitteitä, jotta tietokanta ei paisuisi hallitsemattomasti. Kirjoittajat ovat maininneet, että kyseessä on kevytontologia: se sisältää käsitteiden määrittelyt, mutta ei niiden välisiä päättelysääntöjä. [Spyns2006]

Kim ym. ovat [2007] puolestaan luoneet systeemin, jossa web-sivuja voidaan luokitella yhteiseen ontologiaan perustuvilla avainsanoilla. Luokittelut jaetaan saman palvelun käyttäjille, ja käyttäjät voivat ryhmitellä toisten tekemiä luokitteluita esim. liittämällä yhteen samoihin aihepiireihin liittyviä aineistoja.

Kun vertaillaan eri tutkimuksia, joissa ontologioita on hyödynnetty, tulee kiinnittää huomiota ontologian määrittelyyn. Koska eri lähteet saattavat määrittellä ontologian eri tavoilla, voi toteutuksissa olla eroja sen suhteen, kuinka paljon semanttista sisältöä ontologian ajatellaan sisältävän. Osassa töitä ontologia nähdään vain käsittekirjastona, jolla on tietty formaali esitysmuoto, mutta johon ei välttämättä kuulu käsitteiden määrittelyä.

Esimerkiksi tässä Kimin ym:iden työssä ontologia sisältää RDF-muodossa olevaa metadataa, mutta ainakaan tekstissä ei missään vaiheessa mainita semanttista sisältöä, joka määrittäisi tarkemmin käsitteiden merkitystä. Metadata sisältää ainoastaan asioita, kuten avainsanan merkitsemispäivämäärä, merkitsijä, suosikki-merkinnät jne. [KimYa2007]

5.3 Folksonomian muuntaminen ontologiaksi

Tarkastelemme aluksi malleja, joissa ontologioiden johtaminen on tarkoitus tehdä automatisoidusti folksonomian pohjalta. Sen jälkeen esitellään menetelmiä, joissa järjestelmä ylläpitää avainsanakokoelmaa, mutta kokonaisuuteen tarvitaan myös käyttäjän osuutta. Käyttäjä joko täydentää kokoelmaa lisäämällä sinne avainsanojen välisiä suhteita tai tarkastaa relaatiot, joita ohjelma on merkinnyt eri avainsanojen välille.

5.3.1 Automatisoitu ontologioiden johtaminen

Jos eri folksonomiat on kirjattu ja muodostettu hyvin erilaisilta pohjilta, on niiden vertailu ja yhdistäminen hankalaa. Gruber on esittänyt [2007], että folksonomia tulisi esittää jäsennellysti, yhtenäisessä muodossa, jotta eri folksonomioitten yhdistäminen olisi mahdollista. Tämänäsuuntaisia jäsennyksiä on tehty useita, ja niistä muutamia esimerkkejä tarkastellaan seuraavassa. Ensin esitellään Echarten ym:iden ja sen jälkeen Knerrin menetelmä. Lopuksi esitellään lyhyesti muutamia muita vastaavia toimintamalleja.

Ontologiarakenne folksonomioiden esittämiseksi

Echarte ym. ovat kehittäneet menetelmän, jolla folksonomia saadaan muokattua formaalimpaan muotoon, ontologiaksi [Echar2008]. Heillä on algoritmi, jolla folksonomiasta saadaan muodostettua yleisen tason ontologia. Tähän muodostettavaan ontologiaan voidaan sisällyttää folksonomiaan liittyvää tietoa, josta on apua, kun folksonomiaa käytetään hauissa. Algoritmi tukee myös ontologian myöhempää muokkaamista sitä mukaa, kun tarpeita ilmenee.

Työssään Echarte ym. ovat kehittäneet yleisen ontologiarakenteen, jonka on tarkoitus sopia kaikkien folksonomioitten esittämiseen. Ontologiarakenne on määritelty OWL-ontologiakielellä [OWL2012], ja sen on tarkoitus korjata monia sellaisia puutteita, joita on kirjattu folksonomioitten osalle, mm. synonyymien ja eri kirjoitusmuotojen aiheuttamat sekaannukset (ks. tarkemmin tämän työn kohta 4.2 Folksonomiat). Kirjoittajien mukaan heidän rakenteensa avulla voidaan esittää täydellisesti mikä tahansa folksonomia [Echar2008].

Echarten ym:iden ideaan kuuluu, että ontologia muodostetaan folksonomian pohjalta automaattisesti, ohjelman avulla. Sitä myös täydennetään automaattisesti sitä mukaa, kun uusia avainsanaluokitteluita lisätään aineistoon. Kyse on kuitenkin pelkästään teknisestä muunnosprosessista: jokainen avainsana esitetään OWL-kielen mukaisella

notaatiolla ja siitä kirjataan ylös seikkoja, kuten avainsanaluokittelun tekijä, pvm, polaarisuus (onko merkintä positiivinen vai negatiivinen) sekä avainsanan sijainti (kuinka monentena sana esiintyy luokiteltavan kohteen kohdalla) [Echar2008].

Näistä ei kuitenkaan käy vielä ilmi kaikkia sellaisia seikkoja, jotka ovat ontologioille tunnusomaisia ja jotka erottavat ontologiat folksonomioista. Näitä ovat mm. termien takana olevien käsitteiden määrittelyt ja käsitteiden väliset suhteet. Erityisesti käsitteiden/termien välisten suhteiden puuttuminen aikaansaa sen, että ei voida puhua oikeista ontologioista (ks. kohta 4.3 Eri näkemyksiä ontologioiden suhteesta folksonomioihin). Näin ollen Echarten ym:iden tutkimus ei tarjoa tämän työn kannalta relevanttia sisältöä: kyse on pinnallisesta ja teknisestä muunnosprosessista.

Folksonomioiden yhdistäminen samanlaisen rakenteen avulla

Echarten ym:iden lisäksi aiemmin hyvin samankaltaisen systeemin on kehittänyt Knerr [2006]. Hän on myös laatinut avainsanamäärityksien kuvauksen OWL-kielellä ja tarkoittanut sen sovellettavaksi kaikkiin avainsanoihin. Tavoitteena on, että samanlaisen rakenteen avulla eri folksonomioista saataisiin yhteismitallisia ja niitä olisi mahdollista yhdistää verkkopalveluiden kesken, ei ainoastaan palveluiden sisällä.

Knerrin jäsenyksessä avainsana määritellään hieman erilaisella rakenteella (siitä on kerrottu eri ominaisuuksia), muuten asetelma on sama kuin Echarten ym:iden [2008] työssä. Echarten ym:iden työstä poiketen Knerrin ontologiassa on kuvattuna myös avainsanojen välisiä suhteita (laajempi termi - suppeampi termi). Gruberia [2007] mukailten Knerr on sisällyttänyt ontologiaan myös rakenteen, joka mahdollistaa samalle avainsanalle useita muotoja tai kirjoitusasuja, joista yksi on aina valittu ensisijaiseksi.

Knerrin mallissa avainsanamäärityksen ontologia tukee avainsanojen ja sen kuvaamien sisältöjen yhdistämistä, samoin on mahdollista yhdistää avainsanaviittauksen tekijä avainsanoihin ja avainsanan määrittämiin sisältöihin. Mallissa on kuitenkin tämän työn näkökulmasta sama ongelma kuin Echartella ym:illa: avainsanoihin ei yhdistetä semanttista sisältöä, joka määrittäisi avainsanan merkitystä. Näin ollen näillä malleilla ei ole sellaista annettavaa, joka olisi hyödynnettävissä tämän työn kannalta.

Avainsanamerkintöjen yhdistäminen henkilöiden tai kohteiden kautta

Jäschke ym. ovat kehittäneet algoritmin [2008], jolla voidaan tutkia avainsanajäsenyskolmikoita (kohde, käyttäjä, avainsana) siten, että tarkasteluun suodatetaan vain sellaiset avainsanamerkinnot, jotka ovat yhteisiä useille käyttäjille. Heidän esittämässään kokeilussa tätä ensimmäistä suodatusta jatkettiin valitsemalla vain

määrätyt avainsanat ja ottamalla mukaan ne kohteet, joihin määrätyn raja-arvon ylittävä määrä käyttäjiä (esim. vähintään 5 käyttäjää) oli merkinnyt saman avainsanan. Tämän lisäksi asetettiin ehto, että jos johonkin kohteeseen oli liitetty avainsanoja, vähintään 50 %:n käyttäjistä on täytynyt käyttää tarkastelussa mukana olevaa avainsanaa, muuten kyseistä avainsanaa ei oteta huomioon. Tavoitteena oli siis hakea sellaisia merkintöjä, joista vallitsee jonkinasteinen yksimielisyys usean käyttäjän kesken.

Kuvaamalla nämä poimitut kolmikot graafisesti hilarakenteena saatiin esitys, jossa samoihin kohteisiin on liitetty useampia avainsanoja. Avainsanat ovat termejä, jotka liittyvät rakenteen solmukohtina oleviin käsitteisiin ja käsitteiden ekstension muodostavat ne kohteet, joita avainsanoilla on määritelty. Tarkastelemalla kuvaukseen suodattuneita avainsanoja ja niiden välisiä yhteyksiä, voidaan huomata, että hilarakenne muodostaa kuvauksen käsitteiden välisistä yhteyksistä ja sitä voidaan käyttää hyväksi, kun laaditaan käsittehierarkioita. Heidän mukaansa tätä voidaan edelleen jalostaa niin, että saadaan muodostettua formaaleja ontologiarakenteita. [Jäsch2008]

Myös Mika [2011] on kehittänyt automatisoidun menetelmän ontologioiden johtamiseksi folksonomioista, mutta korostaa, että siihen ennakoedellytyksenä kuuluva ihmisen osuus on tärkeä. Menetelmän ideana on kerätä verkkosivuilla olevista avainsanamerkinnoistä tietoa siten, että noteerataan henkilö, käsite ja käsitteeseen liittyvä kohde.

Kun halutaan muodostaa ontologia jostain tietystä käsitteestä, seulotaan kyseisestä kokoelmasta ensin ne henkilöt, joilla on paljon kyseiseen käsitteeseen liittyviä merkintöjä. Sen jälkeen katsotaan näiden henkilöiden tekemistä avainsanaluokituksista, miten heidän merkitsemiensä aineistojen avainsanat liittyvät toisiinsa. Mikan havaintojen mukaan edellä kuvatulla menetelmällä saadaan eräänlainen kevytontologia, jossa avainsanaluokitusten perusteella toisiinsa liittyvät käsitteet ovat oikeasti käsitteellisesti lähellä toisiaan. [MikaP2011]

Vertailun vuoksi tutkimuksessa muodostettiin kaksi ontologiaa, joista toisessa yhdistävänä tekijänä oli *avainsanamerkintöjen laatija*, toisessa yhdistettiin *samaan kohteeseen* liittyvät avainsanat, jotka ovat siis eri henkilöiden asettamia. Tutkimukseen sisältyi syntyneiden ontologioiden arviointi: semattisen webin asiantuntijoita (kyseisen aihepiirin tutkijoita) pyydettiin kertomaan, kumpi syntyneistä ontologiarakenteista on heidän mielestään parempi. Tällöin asiantuntijat valitsivat mieluummin sen, jossa yhdistäminen oli tehty merkintöjen tekijän perusteella.

Kritiikkinä kyseiseen arviointiin voisi tuoda esille sen, että asiantuntijoilta kysyttiin ainoastaan, kumpi kahdesta vaihtoehdosta on parempi. Ei kysytty, onko se parempana pidetty hyvä, jolloin jää myös se mahdollisuus, että arvioijat valitsivat “kahdesta pahasta pienemmän.”

Eräänlaisena heikkoutena voidaan nähdä myös se, että aineistojen koostamisessa oletettiin avainsanaluokitusten tekijöiden olevan tunnistettavissa. Internetissä kuitenkin tehdään paljon asioita nimimerkin suojissa ja silloin ei voida olla varmoja, että eri verkkopalveluissa esiintyvä sama nimimerkki on oikeasti sama ihminen tai että yksittäinen ihminen ylipäättään rekisteröityy samalla nimimerkillä eri verkkopalveluihin. Mika tuo itsekin esille, että folksonomiat ovat ikäänkuin saarekkeita, joita on vaikea yhdistää yli verkkosivustojen rajojen [MikaP2011].

5.3.2 Käyttäjien täydennettäviksi tarkoitetut järjestelmät

Seuraavassa esitellään ensin menetelmiä, joissa käyttäjällä on mahdollisuus lisätä yhteyksiä avainsanojen välille. Toinen alakohta sisältää lähestymistapoja, joissa aineistosta haetaan avainsanat koneellisesti, mutta käyttäjä loppujen lopuksi evaluoi järjestelmän tuottamat ehdotukset ja päättää niistä.

Käsitteiden välisten suhteiden määrittäminen

Avainsanojen keräämiseen rajoittuneisiin menetelmiin verrattuna hieman kehittyneempiä ovat sellaiset järjestelmät, joissa käyttäjille annetaan mahdollisuus täydentää avainsanamäärityksiä lisäämällä aineistoon avainsanojen keskinäisiä relaatioita (esim. omena - laajempi termi - hedelmä tai porsas=possu). Tällainen on esim. Tanasescun ja Streibelin kehittämä systeemi [2007].

Heidän kehittämässään ETS-järjestelmässä (Extreme Tagging System) voidaan avainsanoille merkitä avainsanoja, samoin voidaan merkitä avainsanojen välisiä yhteyksiä [Tanas2007]. Avainsanojen välisten suhteiden merkitsemisen ansiosta termien monimerkityksisyys ei muodostu ongelmaksi.

Avainsanojen yhteyksien merkitseminen johtaa nk. semanttisten assosiaatioiden määrittelyyn. Voidaan sanoa, että kahden kohteen välillä vallitsee semanttinen assosiaatio, jos niiden välillä on käyttäjien merkitsemistä avainsanojen välisistä yhteyksistä koostuva polku. Sosiaalinen yhteisö huolehtii tehtyjen merkintöjen evaluoinnista. [Tanas2007]

Toisesta näkökulmasta katsottuna yhteydet esittävät käyttäjien tietämystä, ja kirjoittajat kutsuvatkin niitä tietämyspoluiksi (knowledge path) [Tanas2007]. Omasta mielestäni tämä näkemys on erityisen mielenkiintoinen, sillä voidaanhan olettaa, että käsitteiden välisten yhteyksien tunnistaminen ja merkityksen määrittely riippuvat näkökulmasta ja tulkitsijasta.

Myös Huynh-Kim Bang ym. ovat laatineet systeemin, johon käyttäjät voivat itse lisätä avainsanojen välisiä relaatioita [2008]. Tämä eroaa muista järjestelmistä siinä, että tekijät eivät halua yhteisen ontologian rajoittavan avainsanojen valintaa, vaan haluavat säilyttää käyttäjien vapauden määrittää avainsanat omalla tavallaan.

He ovat itse määritelleet, että monissa muissa lähestymistavoissa ontologioiden käyttöön lisätään web 2.0 -tyylistä käyttäjien toimintaa, mutta he sen sijaan haluavat omassa menetelmässään säilyttää käyttäjät ensisijaisina toimijoina. Käyttäjät ovat edelleen niitä, jotka tekevät avainsanamerkinnot, mutta avainsanamerkintöihin lisätään ontologioiden piirteitä (mahdollisuus määrittää avainsanojen välisiä yhteyksiä). [Huynh2008]

He myös sallivat eri käyttäjillä keskenään ristiriitaiset käsitykset [Huynh2008]. Tällainen järjestelmä ei kuitenkaan tue parhaalla mahdollisella tavalla hakujen tekemistä tai aineistojen yhdistämistä: jotta jaetuista avainsanaluokitteluista olisi hyötyä niiden yhteydessä, täytyy luokitteluiden perustua samanlaisiin käsityksiin siitä, mikä on kunkin termin merkitys.

Mikäli käyttäjät määrittelevät jonkin termin eri tavoilla, kyseessä on jonkinasteinen homonymia ja järjestelmän tulee erottaa nämä merkitykseltään eroavat termit jollain tavalla. Kielenkäytön kannalta homonyymien esiintyminen on hyvinkin yleinen ja normaali tilanne. Niiden kohdalla täytyy kuitenkin aina selvittää myös se, mistä tulkinnasta tässä on kyse: missä merkityksessä kyseistä termiä juuri tässä tapauksessa käytetään.

Järjestelmän löytämien avainsanayhteyksien evaluointi

Van Damme, Coenen ja Vandijck [2008] ovat kehittäneet menetelmän, jolla folksonomioista saadaan eräänlainen ontologian kevytversio. Aineistona kyseisessä tutkimuksessa on ollut erään yrityksen viestitietokanta, jossa viestin aihekenttä on tallettamisen yhteydessä korvattu sisältöä kuvaavilla avainsanoilla. Tutkijat ovat kehittäneet kuusivaiheisen menetelmän, jonka avulla näin syntyneitä folksonomiaa pelkistetään ja tiivistetään.

Kyseisen menetelmän alkuvaiheissa saman termin eri kirjoitusasut yhtenäistetään, mutta arvioinnin viimeinen päätös on jätetty ihmisvoimin tehtäväksi, sillä muuten uudet termit tai uudet merkintätavat eivät koskaan pääsisi pintaan. Menetelmän viimeisenä vaiheena avainsanakokoelmasta tuotetaan visuaalinen kuvaus, jossa näkyvät eri käsitteet ja niiden väliset suhteet. [VanDa2008]

Van Dammen, Coenenin ja Vandijekin menetelmässä on positiivista se, että folksonomian käsittelyä ei ole jätetty pelkästään koneellisen yhdenmukaistamisen varaan. Lopputuloksena saatava ontologia sisältää kuitenkin käsitteiden välillä lähinnä vain yläkäsite-alakäsite-relaatioita, joten se ei vielä täysin vastaa ajatusta siitä, mikä on ontologialle ominaista. Ontologian on tärkeää sisältää monia erityyppisiä relaatioita, niiden merkitystä on käsitelty edellä kohdassa 4.1.2.

Schmitz [2006] käyttää avainsanojen yhteyksien muodostamisessa hyväkseen tilastollista käsittelyä: hän hakee avainsanoista pareja, joiden välillä voidaan ajatella olevan subsumptiorelaatio¹ silloin, kun niiden yhdessä esiintyminen mahtuu määrättyjen raja-arvojen välille. Yhdessä esiintyminen voi paljastaa termien välillä hieman erilaisia yhteyksiä: jotkut ovat toisiinsa liittyviä termejä (esim. yleiskäsite ja sen alakäsite), mutta tilastollinen käsittely yhdistää myös toisiinsa käsitteellisesti liittymättömät termit, joten ihmisen toimesta tehtävä tarkistus on välttämätön.

Tarkoituksena on kerätä toisiinsa liittyvistä termeistä verkko tai mahdollisesti hierarkkinen puurakennelma, jota voi käyttää hyväkseen hakuja tehdessään [Schmi2006]. Tämän menetelmän tuoma lisäapu vaikuttaa kuitenkin pieneltä suhteessa siihen työmäärään, joka aiheutuu ohjelman tuottamien sanaparien tarkistamisesta.

5.4 Folksonomian rikastaminen semanttisella sisällöllä

Edellä folksonomioihin ei varsinaisesti ole yhdistetty mitään ulkopuolista sisältöä, sen sijaan niitä on käsitelty tilastollisesti ja niihin on merkitty käsitteiden välisiä yhteyksiä. Tässä esitetään ratkaisuja, joissa folksonomioita rikastetaan mm. internetissä olevien sanastojen ja ontologioiden avulla.

5.4.1 Van Dammen, Heppin ja Siorpaesin menetelmä

Van Dammen, Heppin ja Siorpaesin muodostama ryhmä on kehittänyt lähestymistavan, jolla folksonomiasta saadaan johdettua ontologia yhdistämällä eri lähteitä ja tekniikoita. [VanDa2007]

¹ Käsitteiden välillä subsumptiorelaatio tarkoittaa käsiteluoikkien välistä sisällymistä: käsite on toisen käsitteen alakäsite [Nardi2007]

Folksonomioista kohti ontologioita

Heidän kehittämänsä lähestymistavan tarkoituksena on lähentää folksonomioita ontologioiden suuntaan lisäämällä folksonomioihin semanttista sisältöä. He hyödyntävät ontologian muodostamisessa folksonomian käyttöä koskevaa tilastointia, webissä olevia sanastoja ja ontologioita sekä ontologioiden valintaa, määrittelyä ja yhdistämistä helpottamaan tehtyjä menetelmiä. Folksonomiat tarjoavat heidän mukaansa hyvän alkuaineiston ja lähtökohdan ontologioiden kehitystyölle, sillä uudet käsitteet esiintyvät tyypillisesti ensin folksonomioissa ja folksonomiat kertovat, mitä käsitteitä luokitteluisissa halutaan käyttää. [VanDa2007]

Van Dammen ym.:iden menetelmä kuvataan lyhyesti seuraavissa kohdissa.

Avainsanojen tunnistaminen web-lähteiden avulla

Kun käyttäjä merkitsee avainsanoja johonkin aineistoon, tallettuu siihen tietoa myös käyttäjästä, hänen tietämyksensä tasosta ja kiinnostuksen kohteistaan. Avainsanoja täydennetään sen jälkeen webissä olevien lähteiden perusteella. Ensin kuitenkin varmistetaan termien olemassaolo ja oikeinkirjoitus käyttämällä erilaisia online-sanakirjoja, Googlea ja Wikipediaa. [VanDa2007]

Van Damme ym. [2007] nojaavat paljon Wikipediaan ja sen määrittelyyn. He pitävät Wikipediaa luotettavana lähteenä, koska artikkeleilla on yksiselitteinen web-tunniste (URI 1. Universal Resource Identifier), käsitteiden merkitys on kuvattu luonnollisella kielellä, kuvausta voidaan täydentää multimedia-aineistoilla ja kuvaus on laajan yhteisön hyväksymä.

He myös tuovat esille, että Wikipedia-aineistoissa on tarkentavia linkkejä, joiden avulla voidaan homonymien kohdalla ratkaista kysymykset siitä, missä merkityksessä termiä käytetään tässä kohdassa. Myös käännökset eri kielillä oleviin vastaaviin sivuihin auttavat tässä tarkoituksessa. Van Damme ym. katsovat, että Wikipedia on luotettava juuri mm. sen vuoksi, että se sisältää niin paljon käsitteiden välisiä linkkejä eri kielillä [2007].

Omasta mielestäni taas Wikipediaan nojautuminen tuo mukanaan epävarmuustekijän käsitteiden tunnistamiseen ja määrittelyyn: ei esim. ole ollenkaan varma, että Wikipedia-artikkelin laatijat ovat huomanneet ottaa mukaan *kaikki* mahdolliset homonyymit. Myöskään mikään ei takaa aukottomasti sitä, että erikieliset Wikipedia-sivut on linkitetty toisiinsa oikein.

Van Dammen ym.:iden mukaan myös Google on käyttökelpoinen termien oikeinkirjoituksen tarkistamiseen [2007, 12]. Itse olen myös tämän suhteen kriittinen:

jos jokin väärä muoto on vakiintunut suuren yleisön keskuudessa, tarjoaa Google sitä pelkästään yleisyyden perusteella. Esim. Suomessa käytetään paljon termiä “kovalevy” puhuttaessa kiintolevystä eikä Googlea käyttämällä huomaa ilmaisun heikkouksia, ellei satu törmäämään aineistoon, jossa asiasta on mainittu.

Käsitteen semanttisen sisällön täsmentäminen

Tunnistamisen jälkeen Van Damme ym. tutkivat termiä tarkemmin tekemällä semanttisia web-hakuja Swooglella ja käyttämällä Wordnet-palvelua. Swoogle on semanttinen hakukone, joka käy läpi webissä olevia RDF-muodossa olevia aineistoja [Swoog2013].

Wordnet puolestaan on avoin englannin sanakirja, joka sisältää tiedon termien synonyymeistä. Wordnetissä samaa tarkoittavien sanojen synonyymiryhmistä on yleensä myös lyhyt määritelmä. Lisäksi termien välille on rakennettu erilaisia relaatioita, mm. hierarkioita, jotka kuvaavat termiä vastaavien käsitteiden välisiä abstraktiotaseroja (esim. pöytä kuuluu huonekalujen luokkaan jne.) [WordN2013]. Näiden em. palveluiden avulla saadaan haettua termien ympärille niiden semantiikkaa koskevaa aineistoa.

Wordnetin käyttäminen eittämättä auttaa folksonomian täsmentämisessä. Valitettavasti kyseisenkaltaista palvelua ei ole käytettävissä kaikilla kielillä. Wordnetistä on käytettävissä suomenkielinen versio¹, mutta se on käännetty englanninkielisestä Wordnetistä, joten kaikki meidän kieli- ja kulttuuriympäristölle omintakeinen terminologia puuttuu. Tämän vuoksi Van Dammen ym:iden kehittämää menetelmää ei suoraan voida soveltaa kaikissa ympäristöissä.

Seuraavaksi heidän menetelmässään siirrytään ontologioiden pariin: web-lähteiden avulla tapahtuneen täsmentämisen jälkeen käsitteet voidaan yhdistää olemassa oleviin ontologioihin. Tässä yhdistämistyössä Van Damme ym. [2007] hyödyntävät eri lähteissä esiteltyjä menetelmiä.

Yhdistäminen perustuu yksinkertaisimmillaan käsitettä vastaavien termien yhdenmukaisuuteen. Yhdistämisessä voidaan myös ottaa huomioon laajempi rakenne, josta ilmenee, mihin asiaan avainsanalla viitataan ja mitkä ovat sen yhteydet muihin avainsanoihin tai ontologiakäsitteisiin. Yhdistämistyö tehdään puoliautomasoidusti: menetelmä soveltaa sekä automatisoituja tekniikoita että tulosten kollektiivista arviointia. [VanDa2007]

¹ <http://www.ling.helsinki.fi/en/lt/research/finnwordnet/index.shtml>, viitattu 27.6.2017

Mutta mitä tehdään, jos käsite on niin uusi, että sitä ei vielä ole käytettävissä olevissa web-lähteissä? Tilanteeseen ei kuitenkaan ole esitetty ratkaisuksi mitään muuta toimintatapaa kuin se, että kyseisten tapausten arviointi tulee tehdä jonkin asiantuntijan toimesta. Uuden käsitteen prosessointia ei siis voida jättää automatisoidun algoritmin huolehdittavaksi.

5.4.2 Specian ja Mottan menetelmä

Van Dammen ym:iden lisäksi on muitakin ryhmiä, jotka ovat tutkineet folksonomioiden rikastuttamista semanttisella sisällöllä. Esimerkiksi Specia ja Motta [2007] ovat kehittäneet mallin, jossa avainsanaluokitteluihin yhdistetään käsitteellistä tietoa. Heidän yhtenä tavoitteenaan oli osoittaa, että ontologioilla voidaan jäsentää folksonomioita ja folksonomioista saatavalla tietämyksellä voidaan laajentaa ontologioita.

Specian ja Mottan menetelmässä avainsanat esikäsitellään ennen kuin ne niputetaan yhteen ja niiden välisiä yhteyksiä analysoidaan. Esikäsitelyyn kuuluu mm. sellainen vaihe, jossa lähellä toisiaan olevat, samanlaisilta vaikuttavat avainsanat yhdistetään yhdeksi avainsanaksi. Artikkelissa oli esimerkkinä “cat” ja “cats”, jotka yhdistettiin samaksi avainsanaksi. [Speci2007].

Tässä on pieni vaara, että samalla menetetään oikea semanttinen merkitys - ainakin jos yhdistäminen tehdään koneellisesti. Esim. lava (esiintymislava), lavat (roskalavat) ja lavat (hevosen lavat) saattavat kaikki vaikuttaa samalta käsitteeltä ennen semanttista tarkastelua¹. Tämä avainsanojen yhdistäminen tehdään siis Specian ja Mottan menetelmässä ennen kuin termin merkitystä analysoidaan ontologioiden ja verkkolähteiden avulla.

Edellä kuvatun “siivousoperaation” jälkeen avainsanat kerätään yhteen nipuiksi sen perusteella, että ne esiintyvät toistensa yhteydessä. Seuraavaksi nippujen avainsanoja tutkitaan pareittain. Avainsanoja etsitään saatavilla olevista ontologioista semanttisen hakukoneen, Swooglen [Swoog2013] avulla. [Speci2007]

Jos termejä ei ole ontologioissa, pyritään niille löytämään synonyymi tai niiden monimerkityksisyys ratkaistaan Wikipedian [Wikip2015], Wordnetin [WordN2013], Googlen tai erilaisten online-sanastojen avulla. Mikäli tästä ei ole apua, avainsana poistetaan tutkittavasta nipusta. Lopuksi avainsanoihin yhdistetään saatavilla olevista ontologioista semanttista sisältöä, joka määrittää avainsanan merkitystä. [Speci2007]

¹ Finto-ontologiakirjastossa lavat-termin haku antaa vastaukseksi seuraavanlaiset tarkenteet: ampumalavat, laululavat, maitolavat, makuulavat, soittolavat, tanssilavat (haettu 14.4.2017). Joukosta puuttuvat siis oikeastaan kaikki ne merkitykset, jotka yllä olevassa tekstissä oli esitetty: esiintymislava (esim. teatterissa), roskalava ja hevosen ruumiinosa.

5.4.3 Angetoun ym:iden menetelmä

Edellä kuvattua toimintamallia on viety eteenpäin osittain samojen kirjoittajien artikkelissa [Angel2007]. Heidän tavoitteenaan on tutkia, voidaanko folksonomioita rikastaa automaattisella algoritmilla, joka käyttää hyväkseen semanttisen webin teknologioita. Artikkelissa kuvataan tehtyjä kokeiluja, joissa kyseinen algoritmi saa aina syötteenä nipun avainsanoja.

Algoritmin syötteeksi annettava avainsananippu oli saatu verkkopalvelu Flickr.com:sta käyttämällä lähtökohtana jotain yksittäistä avainsanaa (esimerkkeinä olivat mm. sienet ja hedelmät) [Angel2007]. Flickr-palvelu antaa käyttäjälle tämän niin halutessa nipun uusia avainsanoja, jotka liittyvät ensin haettuun avainsanaan. Näin artikkelin kokeiluissa saatiin tutkittavaksi joukko avainsanoja, joiden välillä saattaa olla implisiittisiä relaatioita.

Angetoun ym:iden algoritmi tutkii avainsanoja verkossa saatavilla olevien ontologioiden avulla ja muuttaa saamiensa tietojen perusteella implisiittiset relaatiot eksplisiittiseksi tietämysrakenteeksi. Tämän lisäksi algoritmi palauttaa tiedon siitä, minkä avainsanojen välillä ei tunnistettu olevan yhteyksiä. Algoritmi listaa myös erikseen ne avainsanat, joista ei löytynyt lisätietoja semanttisen webin hauilla. [Angel2007]

5.4.4 Tesconin ym:iden TDA-algoritmi

Edellä kuvattujen kanssa hyvin samanlaiseen ratkaisuun ovat päätyneet Tesconi ym. [2008]. He ovat kehittäneet TDA-algoritmin (Tag Disambiguation Algorithm), jolla voidaan selvittää avainsanojen monimerkityksisyys, sillä kyseinen algoritmi hakee avainsanalle välittömästi sitä vastaavan käsitteen Wikipediasta¹. Koska käsitteeseen voidaan viitata monella eri termillä, ovat kirjoittajat kehittäneet Tagpedian. Se on Wikipedian laajennus, joka sisältää kaikki termit, joilla Wikipedia-käsitteisiin voidaan viitata.

TDA-algoritmissa käytetään hyväksi Wordnetiä, Wikipediaa sekä YAGOn luokkia. YAGO (Yet Another General Ontology) on yleinen ontologia, joka on muodostettu Wikipedian ja Wordnetin pohjalta. Tavoitteena on näiden lähteiden avulla päätellä automaattisesti TDA-algoritmin avulla, mikä on avainsanan merkitys. [Tesco2008]

¹ Jälleen kerran täytyy huomauttaa, että emme voi olla varmoja siitä, että Wikipedia sisältää kaikkien termien lukuisat merkitykset. Tesconi ym. pitävät sitä kuitenkin mitä suurimmassa määrin luotettavana lähteenä.

Tekemissään kokeiluissa he ovat onnistuneet selvittämään n. 91,5 %:sta avainsanoja niiden tarkoituksen. Selvittämättä jääneet 8,5 % avainsanoista olivat sellaisia, että niitä ei ollut Tagpediassa. [Tesco2008] Vaikka tämä näyttää hienolta prosenttiluvulta, on se mielestäni kuitenkin vielä liian kaukana ihanteellisesta tilanteesta. Jos asiasta lähes 10 % jää ymmärtämättä, voi se joskus olla vakavalla tavalla ratkaiseva osuus, joka johtaa kohtalokkaihin virhetulkintoihin.

Tesconin ym:iden ajatuksena on, että jokaisella henkilöllä on FOAF-profiilissaan oma URI ja jokaiselle avainsanalle saadaan Wikipedia-määritelmän URI TDA-algoritmilla. Näin saadaan RDF-kolmikoita, jotka yhdistävät kohteen, avainsanan ja avainsanamäärittelyn tekijän. Näiden kolmikoiden avulla käyttäjien tekemät avainsanamääritykset saadaan yhteensopiviksi semanttisen webin Linked data -sisältöjen kanssa. [Tesco2008]

5.4.5 Passantin ja Laubletin tapa yhdistää folksonomiat ja ontologiat

Passant ja Laublet [2008] ovat kehittäneet menetelmän, jolla saadaan laadittua folksonomian sisältämille käsitteille RDF-kieliset määrittelyt. Määrittely kuvaa termin merkitystä ja se voi sisältää linkkejä jo olemassa oleviin ontologioihin.

Menetelmän tavoitteena on hyödyntää semanttisen webin teknologioita ja tuottaa folksonomiakäsitteiden määrittelyjä niin, että kuitenkin säilytetään web 2.0:n ihanteet. Näihin ihanteisiin kuuluu, että käyttäjät tuottavat itse sisältöjä ja päättävät myös sisältöjen organisoinnista. Tätä tarkoitusta varten Passant ja Laublet [2008] esittelevät kehittämänsä MOAT-ympäristön (MOAT= Meaning Of A Tag), joka tallettaa jokaisesta avainsanasta myös sen merkityksen sekä tiedon siitä, ketkä käyttäjät ovat liittäneet kyseiseen avainsanaan tämän merkityksen.

Käytännössä MOAT-järjestelmä otetaan käyttöön asentamalla sitä tukeva lisäosa omaan blogityökaluun. Kun tekstejä merkitään avainsanoilla, MOAT-lisäosa hakee välittömästi nähtävillä kaikki kyseiseen avainsanaan liittyvät määrittelyt. Käyttäjä voi valita niistä sopivimman tai luoda kokonaan uuden määrittelyn. Liittämällä sisältöihin metadataa ja käyttämällä pysyviä URI-osoitteita saadaan folksonomioihin luotua semanttisen webin käytäntöjen mukainen määrittelykerros¹, joka tukee hakujen tekemistä ja sisältöjen yhdistämistä. [Passa2008]

Näissä kaikissa edellä kuvatuissa tutkimuksissa on hyvin paljon samankaltaisuutta oman työni kanssa. Yhteistä on se, että niissä kaikissa tutkitaan rinnakkain sekä folksonomioiden että ontologioiden sisältöä. Erona on puolestaan se, mitä tarkoitusta varten vertai-

¹ MOAT-järjestelmä käyttää metadatan määrittelyssä SIOC-sanaston rakennetta (SIOC on mainittu myös aiemmin kohdassa 5.2.1). <http://sioc-project.org/>

lua tehdään ja minkälaisia aineistoja avainsanoilla on jäsennetty. Samoin erona on järjestelmäympäristö, jossa kokeiluja suoritetaan. Monissa edellä esitellyissä tutkimuksissa on ollut käytettävissä tutkimusta varten räätälöityjä aputyökaluja, jotka esim. hakevat automaattisesti saataville tarvittavia käsitelmääriytyksiä Wikipediasta, Wordnetistä yms. palveluista. Omassa työssäni tällaisia automatisoituja aputoimintoja ei ole käytetty.

5.5 Tämän työn asema suhteessa aiempaan tutkimukseen

Edellä on esitelty useita menetelmiä, jotka pyrkivät jalostamaan avainsanaluokittelujen tuloksena syntyneitä folksonomioita. Niitä kokoavana analyysityökaluna toimii hyvin Limpensin ym:iden esittämä jäsennys [2009]. Se esiteltiin kohdassa 5.1, joten ei ole tarvetta ryhtyä toisintamaan sen esittämiä vaihtoehtoja folksonomioiden ja ontologioiden yhdistämiseksi. Kokonaisuudesta voidaan kuitenkin todeta, että kaikki tässä läpi käyty tutkimukset lähtevät siitä oletuksesta, että folksonomiat tarvitsevat jatkojalostamista ennen kuin ne ovat kunnolla hyödynnettävissä esim. hauissa tai aineistojen yhdistämisessä.

Kirjallisuuskatsaus osoittaa, että tämän työn aihetta lähellä olevaa tutkimusta on tehty englannin- ja ranskankielisessä ympäristössä, mutta ei Suomessa. Erityisesti Passantin tutkimus blogikirjoitusten folksonomioiden vahvistamisesta ontologioiden avulla [2007] on lähellä tämän työn näkökulmaa.

Muualla tehdyn tutkimuksen tulosten ei voi sellaisenaan olettaa sopivan myös tšekäläiseen ympäristöön. Suomessa on omat kansalliset ontologiansa ja suomen kieli on englannista tai ranskasta poikkeavan rakenteensa vuoksi niin erilainen toimintaympäristö, että muualla kehitettyjen käytäntöjen toimivuutta on testattava täällä erikseen. Samalla saamme tietoa myös omien ontologioidemme kattavuudesta.

Aiemmista tutkimuksista poiketen tässä työssä kohdealueen muodostavat *elämäkert*-aineistot, jotka kirjoitetaan *tulevaisuuden* lukijoille. Se tuo käsitteelliseen luokitteluun omat lisähaasteensa ja -motivaationsa. Koska aineistojen lukeminen tapahtuu ehkä pitkälläkin tulevaisuudessa, on sisällön merkitysten säilymiseen kiinnitettävä erityistä huomiota. Sen vuoksi yleisten ontologioiden käyttöä voidaan ylipäätään pitää perusteltuna ja toivottavana.

Aiemmin esitetyssä luvussa 3 tämän työn yhdeksi tutkimusmenetelmäksi on mainittu teorioita testaava tutkimus, ja juuri Passantin tutkimus on tällainen ”testattava teoria”. Tarkoituksena on selvittää, pitävätkö siinä tehdyt havainnot paikkansa myös tämän tutkimuksen aineistossa.

6 Tutkimuksen toteutusympäristö

Aluksi (kohta 6.1) kerrotaan Tikutus-hankkeesta, jonka pilottiryhmässä tämä työ on tehty. Sen jälkeen, kohdassa 6.2, kuvataan tutkimuskohteena olleen ryhmän työskentelyn pääperiaatteet.

6.1 Tikutus-hanke ja sen tavoitteet

Tutkimuksen aineisto on kerätty Tikutus-hankkeen pilottiryhmän toiminnan yhteydessä. Niinpä seuraavassa esitellään ensin koko Tikutus-hankkeen päätavoitteet ja sen jälkeen kerrotaan pilottiryhmien toiminnan tavoitteista ja periaatteista.

6.1.1 Tikutus-hanke

Työ liittyy Helsingin yliopiston Koulutus- ja kehittämiskeskus Palmenian ja Lappeenrannan teknillisen yliopiston Lahti School of Innovationin yhteiseen Tikutus-hankkeeseen. Hanke oli käynnissä neljän vuoden ajan vuosina 2010-2014 ja sen tarkoituksena oli ”kehittää palvelusisältöjä, jotka tukevat omaehtoista terveyden edistämistä ja hyvinvoinnista huolehtimista sekä järkevöittävät palvelukontakteja.” Käytännössä hankkeessa kehitettiin kansalaisen henkilökohtaisia sähköisiä arkistoja, joita tarvittaessa voidaan hyödyntää eri palveluntarjoajien kanssa asioitaessa. [Tikut2012]

Hankkeessa tuotettiin konsepti kahdelle henkilökohtaiselle tietovarannolle, jotka nimettiin terveystikuksi ja muistelutikuksi. Terveystikulle on tarkoitus tallentaa omia elämäntapoja (esim. ravitsemusta ja liikuntaa) tai esim. pitkäaikaissairauden kotihoitoa koskevia seurantatietoja. Muistelutikkua voidaan käyttää muistisairaiden hoidon tukena tai ylipäätään välineenä hahmottamaan ja jäsentämään laatijan oman elämänsä kulkua. [Tikut2012]

Muistisairailta elämänvaiheitten dokumentointi voidaan aloittaa siinä vaiheessa, kun kehittymässä oleva muistisairaus on todettu. Silloin henkilö voi itse huolehtia aineiston kokoamisesta. Muistelutikku voidaan tarvittaessa myös laatia ulkopuolisen toimesta sairauden myöhäisemmässä vaiheessa, kun henkilö on jo laitoshoidon tarpeessa. [Tikut2012]

Muistelutikkua voidaan käyttää muutenkin kuin muistisairauden yhteydessä. Elämänvaiheitten sähköisestä kirjaamisesta on erityisesti hyötyä esim. maahanmuuttajille, jolloin he voivat hyödyntää yhdessä paikassa sijaitsevia dokumentteja asioidessaan eri palveluntarjoajien kanssa. [Tikut2012]

Hankkeen Tikutus-nimi on mukaelma muistitikusta, mutta tutkijat ovat korostaneet, että tulosten kannalta ei ole oleellista, mitä talletusvälinettä käytetään. Kansalaisten tietovarannot voidaan yhtä hyvin tallentaa pilvipalveluun tai johonkin henkilökohtaiseen laitteeseen (tietokoneelle, puhelimeen) [Kuosm2014]. Ensisijaisesti on kyse tietovarannon sisällön määrittelemisestä ja sisältöä hyödyntävien käytäntöjen kehittamisestä.

6.1.2 Pilottiryhmät

Tikutus-hankkeen puitteissa toteutettiin erilaisia kokeiluja, joiden avulla haettiin sopivia tapoja tuottaa muisteluaineistoa. Eri tarkoituksiin tehdyille muistelutikuille kehitettiin myös omanlaisiansa sisältörakenteita. Tähän työhön liittyvässä pilottiryhmässä tarkoituksena oli kokeilla ikääntyneiden muisteluprosessia siten, että he tuottavat itse oman muisteluaineistonsa suoraan digitaaliseen muotoon. Myös aineiston rakenne oli laatijan itsensä määrittämä.

Piloteissa tutkittiin myös syntyneitä sisältöjä pyrkimällä arvioimaan aineiston merkitystä ja käytettävyyttä eri elämänvaiheissa. [Tikut2012] Tässä työssä sovellettavalla elämäkertaa-aineiston avainsanojen käytöllä ja ontologisella jäsentämisellä voisi olla funktionensa nimenomaan niissä tilanteissa, joissa aineistoa käytetään kontaktissa ulkoisen palveluntarjoajan kanssa. Ulkopuolisen lukijan on helpompi hahmottaa sisältöä jäsentävien termien avulla, mitä aineisto sisältää ja miten löytää mahdollisesti laajasta materiaalikokoelmasta tarvitsemansa tiedot.

Laitoskäyttöön tarkoitettu muistelutikku voisi puolestaan olla ontologioiden avulla integroitavissa muuhun asiakasta koskevaan tietoaaineistoon. Tämän tyyppinen avainsanojen hyödyntäminen olisi kuitenkin kokonaan oma tutkimusaiheensa, eikä siihen ole voinut paneutua tämän työn puitteissa.

Tämä työ on siis käyttänyt omana aineistonaan Tikutus-hankkeen pilottia, joka toteutettiin Mäntsälän kansalaisopiston opintopiirinä. Työskentely kesti kahden lukuvuoden ajan alkaen syksyllä 2012 ja päättyen kevätlukukauden 2014 lopussa. Syksyllä 2014 alkoi uusi, kolmevuotinen muisteluryhmä, joka pohjautui pääosin samanlaisiin työtapoihin kuin aiempi ryhmä. Myös aikaisemman ryhmän osallistujien oli mahdollista osallistua sen työskentelyyn ja jatkaa oman aineistonsa kartuttamista.

Työpajamuotoisen opintopiirin tavoitteena oli, että osallistujat laativat itselleen digitaalisessa muodossa olevan multimedia-aineiston, joka kuvaa heidän elämänvaiheitaan ja elämänkaarelle merkityksellisiä asioita. Tikutus-hankkeen pilottina toimineesta opintopiiristä ilmoitettiin Mäntsälän kansalaisopiston ohjelmassa lukuvuosille 2012-13 ja 2013-14. Ryhmän osallistujat otettiin ilmoittautumisjärjestyksessä eikä toisena lukuvuonna otettu enää uusia jäseniä. Minimivaatimukseksi mukaan tulijoille oli asetettu tietokoneen peruskäytön hallitseminen.

Vuosina 2014-17 toimineen jatkoryhmän organisointi noudatti hyvin pitkälle Tikutus-hankkeen pilotin toteutusta. Ryhmä oli mukana Mäntsälän kansalaisopiston kurssiohjelmassa nimellä ”Muisteluryhmä” ja siihen otettiin halukkaita osallistujia aina lukuvuoden vaihtuessa. Osallistumisen edellytykset pidettiin samoina kuin aikaisemmallakin ryhmällä. Seuraavassa kuvauksessa keskitytään kuitenkin hankkeen pilottiryhmän toiminnan kuvaukseen, koska se on dokumentoitu jo tutkimuksellisista syistä tarkemmin kuin jatkoryhmän työskentely.

6.2 Työskentelyssä käytettyjä menetelmiä ja toimintatapoja

Vaikka tutkimuksen aineistona olevia elämäkertoja ei ollut tarkoitus sijoittaa internetiin, olivat sekä työskentelyssä käytetyt välineet että menetelmät samankaltaisia kuin web-aineistojen tuottamisessa käytettävät. Seuraavassa kuvataan ensin työskentelyvälineitä ja sen jälkeen menetelmiä.

6.2.1 Talletusmuodon ja -välineen valinta

Pilottiryhmän työskentelyssä osallistujien tuottamat elämäkerta-aineistot talletettiin hypertekstimuotoon html-sivustoksi siten, että jokainen laati oman, henkilökohtaisen sivustokokonaisuuden. Sivusto sijoitettiin usb-muistitikulle, eikä sitä ollut missään vaiheessa tarkoitus laittaa internetiin.

Jälkipolville tarkoitettun materiaalin toivotaan luonnollisesti olevan mahdollisimman pitkään lukukelpoista, joten siihen tarkoitukseen pääasiassa html-kielellä toteutettu sivusto vaikuttaa sopivalta vaihtoehdolta. Koska koko yhteiskunnan tasolla on tällä hetkellä talletettu suuri määrä materiaalia web-sivujen muotoon, on niiden luettavuus todennäköisesti turvattu pitkälle tulevaisuuteen.

Sivuston laatimiseen käytettiin KompoZer-ohjelmaa, joka on avoimen lähdekoodin graafinen web-editori [Kompo2013]. Pilottiryhmän jäsenille suositeltiin, että sivujen

ulkoasu talletetaan erikseen ulkoiseen css-tiedostoon¹, jolloin varsinaisen sisällön käsittelevien web-sivujen koodi on mahdollisimman yksinkertaista ja säilyttää todennäköisemmin luettavuutensa myös tulevaisuudessa.

Aineiston talletusmuotoa suurempi ongelma on sopivan muistivälineen valinta, sillä tietokoneen oheislaitteet ja liitännät muuttuvat koko ajan eikä nykyisten tietovälineiden lukukelpoisuutta voida taata tulevaisuudessa. Siihen liittyviä riskejä voidaan pienentää tallettamalla aineisto muistitikun lisäksi myös muille muistivälineille (optinen levy, verkkotallennustila) ja tulostamalla materiaali siltä osin kuin se soveltuu tulostettavaksi.

6.2.2 Mielleyhtymäkartoista ontologioihin

Mind map -kartat

Muistelemisen ja elämäkerta-aineiston tuottamisen alkuvaiheissa yhtenä työskentelymenetelmänä käytettiin Mind map –menetelmää eli mielleyhtymäkarttoja². Materiaalin laatiminen lähti liikkeelle siitä, että jokainen ryhmän osallistuja teki itselleen mielleyhtymäkartan, joka kuvasi tulevan aineiston sisältöaiheita.

Ajatuksena oli, että mielleyhtymäkarta toimisi myös hypertekstin rakenteen suunnittelutyökaluna. Montosen mukaan mielleyhtymäkarta toimii ”ideoinnin apuna, ideoiden jäsentämisen välineenä ja asioiden välisen suhteen ilmentäjänä.” [Monto2008, 55] Kartta sopii erityisen hyvin tilanteisiin, joissa suunnittelun alussa ei tiedetä, millaiseksi kokonaisuus muodostuu.

Mindmappingia on suositeltu apuvälineeksi myös esim. hakujen tekemiseen ja aineiston jäsentämiseen [Beel2009], mitkä ovat hyvin samankaltaisia pyrkimyksiä kuin tässäkin työssä. Mind map –jäsenitys voi muodostaa rungon myös myöhemmin tehtävälle avainkäsitteiden merkitsemiselle ja käsiteontologian soveltamiselle.

Mielleyhtymäkarttoihin palattiin vielä työskentelyn loppuun siinä vaiheessa, kun aineistolle tehtiin aloitussivu ja sisällysluettelo. Silloin kartta auttoi hahmottamaan syntyneitä kokonaisuutta ja sen rakennetta.

¹ css-tiedosto sisältää css-kielellä tehtyjä ulkoasumäärittelyitä, jotka sopivat yhteen html-kielellä tuotetun dokumentin kanssa.

² Menetelmän esittely: ks. esim. Tony Buzan, Käytä järkeäsi - tehokkaammin, 1982.

Käsitteiden määrittely ja tutkiminen

Kun muisteluryhmän jäsen oli saanut valmiiksi haluamansa määrän aineistoa, tarkoituksena oli, että hän voi itse esittää, mitkä hänen mielestään olivat keskeisiä sisältöä kuvaavia käsitteitä (syntyi aineistoa kuvaava folksonomia). Tämän jälkeen myös tutkija kävi läpi kyseiset aineistonäytteet, muodosti omat ehdotuksensa avainsanoiksi ja teki näin oman ehdotuksensa folksonomiaksi.

Nämä kaksi folksonomiaa laitettiin rinnakkain ja niistä kirjattiin ylös käsitteiden määrä suhteessa toisiinsa. Kyseisten folksonomioiden pohjalta tutkija teki hakuja Finto-ontologiapalveluun ja selvitti, mitkä avainsanat sisältyvät kirjastossa oleviin ontologioihin – tällöin sovitettiin yhteen folksonomioita ja ontologioita.

Ontologisessa jäsentämisessä tutkija oli ensisijainen toimija - hän selvitti, mitkä folksonomian avainsanoista olivat mukana jossain Finto-palvelun ontologiassa. Tutkija muodosti tarvittaessa myös omia ehdotuksiaan vaihtoehtoisista avainsanoista, jotka sisältyivät ontologiakirjaston ontologioihin. Jäsentämistyön aikana tutkija ja aineiston laajittaja olivat tarvittaessa vuorovaikutuksessa, jotta oikea tulkinta olisi tullut varmistetuksi.

Kun kunkin aineiston kahta folksonomiaa oli verrattu ontologiakirjasto Finton sisältöön, tehtiin siitä määrällinen yhteenveto. Samalla kirjattiin ylös myös kaikki muut mahdolliset huomiot, joita aineiston käsittelyn yhteydessä oli syntynyt. Näistä tehtiin lopuksi yhteenveto, jossa tehtyjä huomioita verrattiin myös lähdekirjallisuuden sisältöön.

7 Työskentelyn eteneminen

Seuraavassa käydään läpi koko työskentely kronologista kulkua mukailevassa järjestyksessä. Aloitetaan kertomalla, miten työskentely organisoitiin ja mitä välineitä käytettiin aineiston tallettamiseen ja muokkaamiseen (kohta 7.1). Sen jälkeen kuvataan menetelmiä, joita käytettiin muistelemistyön tukemiseen (kohta 7.2). Lopuksi (kohdassa 7.3) kerrotaan pilotin loppuvaiheista sekä sitä seuranneesta jatkotyöskentelystä.

7.1 Työskentelyn organisointi ja välineet

Pilottiryhmän työskentely on tapahtunut osana Mäntsälän kansalaisopiston toimintaa. Seuraavassa kuvataan ryhmän työskentelyn alkuvaiheet ja kerrotaan tehdyistä välinevälinnoista.

7.1.1 Ryhmän perustaminen ja työskentelyn aloittaminen

Työskentely aloitettiin syksyllä 2012 perustamalla Muistelutikkutyöpaja-niminen opintopiiri Mäntsälän kansalaisopistoon. Ryhmä oli suunnattu ikäihmisille, mikä käytännössä tarkoittaa 63 vuotta täyttäneitä tai eläkeläisiä. Osallistujien pohjatiedoiksi edellytettiin tietokoneen käytön perusteiden osaamista.

Ryhmään ilmoittautui 9 osallistujaa, joista 8 oli naisia ja yksi mies. Muutaman viikon jälkeen miespuolinen osanottaja jäi pois ryhmästä, mutta syytä ei erikseen tiedusteltu häneltä. Joulun 2012 aikoihin yhden osallistujan puoliso kuoli, joten muuttuneen elämäntilanteen vuoksi leskeksi jäänyt jätti ryhmän.

Näiden muutosten jälkeen osanottajajoukko pysyi samana eikä uusia jäseniä otettu, jotta ryhmässä pysyisi luottamuksellinen ilmapiiri. Muistelussa tuli esille paljon jokaisen henkilökohtaista historiaa ja osittain vaikeitakin asioita. Niitä oli helpompi kertoa ihmisille, jotka ovat jo jossain määrin tuttuja ja jotka vastaavasti ovat avautuneet omista muistoistaan.

Työskentely aloitettiin tutustumalla muihin ryhmäläisiin ja avaamalla muistoja pienen kotitehtävän avulla: toiseksi tapaamiskerraksi jokaisen piti tuoda mukanaan jokin esine tai kuva, josta halusi kertoa muille ryhmäläisille. Tämä osoittautui hyväksi menetelmäk-

si päästä nopeasti asian ytimeen, ja tehtävä sai osallistujat avautumaan hyvin henkilökohtaisistakin muistoista.

7.1.2 Mielleyhtymäkartat ja graafinen web-editori

Heti ensimmäisillä kerroilla aloitettiin oman aineiston sisällön suunnittelu mielleyhtymäkarttojen (mind map) avulla. Karttojen avulla hahmoteltiin sisällön pääaiheita ja painopistealueita ja sitä käytettiin aineiston jaottelussa alakohtiin. Hyvin pian alettiin myös käydä läpi tietotekniikkaan liittyviä asioita: muistitikun käyttö ja kansiorakenteen luominen olivat näistä aivan ensimmäisiä.

Tietotekniikkaan tutustumista jatkettiin puhumalla varmuuskopioinnista ja tulostamisesta. Varmuuskopiointi osoittautui myöhemmin erittäin tarpeelliseksi, sillä yhdellä osallistujista oli muistitikku katkennut siten, että sitä ei voinut enää laittaa tietokoneen usb-liitäntään.

Aineisto kirjoitettiin alusta alkaen KompoZer-ohjelmalla [Kompo2013], joka on avoimen lähdekoodin graafinen web-editori. Materiaalia ryhdyttiin siis heti työstämään hypertekstimuotoon. Aineiston aloitussivu tehtiin yhdessä, ja sivun sisältö määräytyi kullakin oman mind map -kartan perusteella. Mind map -kartta toimi hyvin sisällön hahmottamisen apuvälineenä. Sen avulla voitiin keskittyä sisältöön ilman ajatusta teknisestä toteutuksesta.

Myöhemmin myös kokeilimme yhdessä, miten sivuille liitetään kuvia tekstin sekaan ja miten sivuja linkitetään toisiinsa. Sivujen linkittämisessä käytettiin apuna mind map-karttaa. Pitkille sisältösivuille, joissa on väliotsikoita, teimme sivun alkuun sisällysluettelon.

Ulkoasu laadittiin vasta työskentelyn loppuksi. Kirjoittamisen aikana sivuille merkittiin ainoastaan periaatteelliset korostuskeinot (lihavointi ja kursivointi) sekä otsikkotyylit. Internetistä haettiin ilmaisia ulkoasumäärittelyitä sisältäviä tyylitiedostoja, ja ne asennettiin aineiston kanssa samalle muistitikulle. Kun muisteluaineisto oli valmis, kyseiset tyylimäärittelyt yhdistettiin sisältösivuihin.

7.1.3 Multimedia-aineiston muokkaamiseen käytettyjä työkaluja

Koska tarkoituksena oli tehdä multimediaelämäkerta, otettiin kuvamateriaalin hankkimiseksi paperivalokuvien skannaus mukaan ohjelmaan jo aikaisessa vaiheessa. Samoin käsiteltiin digikameran käyttöä, jotta osallistujat voivat tarvitessaan ottaa valokuvia ai-

neistonsa täydennykseksi. Valokuvaamalla voi esim. tallentaa vanhoja esineitä tai merkityksellisiä paikkoja sen näköisinä kuin ne nyt ovat, kaiken kuvamateriaalin ei tarvitse olla vanhaa.

Jotta valokuvien hyödyntäminen olisi ongelmaton, opeteltiin myös kuvatiedostojen perusmuokkaus. Kuvankäsittelyssä käytettiin ilmaista Picasa-ohjelmaa [Picas2015]. Kuvien muokkaamisessa keskityttiin sellaisiin toimintoihin, jotka ovat tarpeellisia skannattuja kuvia käsiteltäessä: kuvien suoristaminen, rajaaminen, punasilmäisyyden poisto, valaistuksen säätäminen ja retusointi.

Myöhemmin myös opeteltiin kirjoittamaan tekstiä kuvan päälle. Siitä oli hyötyä esim. kun haluttiin merkitä, keitä henkilöitä on kuvassa, tai kun haluttiin lisätä selittäviä tekstejä osaksi kuvaa. KompoZer-työkalulla opeteltiin liittämään kuvat osaksi muisteluaineistoa.

Ensimmäisen puolen vuoden työskentelyn jälkeen otettiin mukaan myös äänitiedostot. Opettelimme yhdessä äänittämään tietokoneella puhetta sekä myös muuntamaan vanhojen c-kasettien sisältöä digitaaliseen muotoon. Äänen tallettamiseen ja muokkaamiseen käytettiin ilmaista Audacity-ohjelmaa. [Audac2015]

7.2 Aineiston kerääminen ja jäsentäminen

Seuraavassa kerrotaan työskentelytavoista, joilla ryhmässä tuettiin osallistujien muistelemista. Kokoonumisissa tarjottiin runsaasti erilaisia virikkeitä hyödynnettäväksi työskentelyssä. Kuvataan myös se tapa, jolla osallistujat evästettiin semanttisen webin periaatteisiin ja avainsanojen merkitsemiseen.

7.2.1 Muistelemisen tukeminen

Muistin avaamiseksi kerättiin yhdessä erilaisia muisteluaineistoja, jotka puolestaan ruokkivat kunkin henkilökohtaista muistelutyötä. Ryhmässä esim. koottiin yhteistyönä lista erilaisista sukupolvikokemuksista, jotka olivat jääneet osallistujien mieleen merkityksellisinä tapahtumina ja jotka ovat sellaisia, että ne ovat vaikuttaneet kaikkien samaan ikäluokkaan kuuluvien elämään. Tällaisia kokemuksia olivat esim. sotavuodet, pula-aika, teknologian kehitys, päivähoito, kuulennot jne.

Toinen tapa herättää muistoja oli keskustella jostain yhteisestä teemasta ensin pareittain ja sitten koko ryhmä yhdessä. Tällaisia teemoja olivat esim. ruoka ja siihen liittyvät asi-

at, koulu ja opiskelu, entiset asuinpaikat, korujen käyttö, ravintolassa käyminen. Esille tulleista muistoista koottiin lyhyt yhteenveto ryhmän yhteiseen blogiin käytettäväksi oman aineiston kirjoittamistyössä.

Muistelutyön tueksi myös luettiin yhdessä tarinoita ja keskusteltiin sekä tarinan sisällöstä että sen herättämistä ajatuksista. Lisäksi ryhmän osanottajilta pyydettiin toivomuksia yhteisten muisteluhetkien teemoiksi. Ehdotuksina kiinnostaviksi aiheiksi saatiin mm. matkat, pukeutuminen ja kodin juhlat.

Ensimmäisen toimintavuoden lopuksi ryhmä teki yhdessä retken Helsinkiin, jossa käytiin tutustumassa Kalevala-koron tehtaaseen. Sen jälkeen vierailtiin Hotelli- ja ravintolamuseossa sekä Kansallismuseossa, jossa oli parhaillaan meneillään 1900-luvusta kertova näyttely. Lopuksi tehtiin lyhyt ajelukierros parin muisteluryhmäläisen entisille asuinpaikoille.

Toisen toimintavuoden loppupuolella ryhmä sai vieraakseen kirjailija ja historioitsija Kirsti Mannisen, joka evästi ryhmän jäseniä kirjoitustyössä. Hän korosti, että suurta tarinaa tärkeämpi on värikäs tilkkutäkki ja kaikki autenttiset aineistot kannattaa liittää mukaan sellaisenaan, ei tarinoiden sisällä. Tätä tukee hyvin myös valitsemamme työskentelytapa: se, että aineistot ovat multimedia-aineistoja eivätkä perustu pelkkään tekstimuotoiseen dokumentointiin.

Internetin hyödyntäminen otettiin mukaan ensimmäisen lukukauden aikana. Kävimme yhdessä läpi erilaisia arkistolähteitä ja karttoja, lehtiarkistoja sekä yleisiä historiasivustoja. Tämän jälkeen internet-lähteiden hyödyntämistä jatkettiin koko työskentelyn ajan.

Internetin lisäksi ryhmäläiset käyttivät monipuolisesti myös muuta materiaalia: kirjoja, lehtiartikkeleita, historia-aiheisia näyttelyitä, radio-ohjelmia yms. Kirjoista sekä tietoa kaunokirjallisuus koettiin hyödylliseksi. Kaunokirjallisuudessa on myös talletettuna ajankuvaa, ja se voi toimia hyvin muistin avaajana omaa elämää muistellessa.

Ryhmän jäsenet jakoivat auliisti keskenään tietoja, joita olivat löytäneet, ja näin saatiin levitettyä hyviä vinkkejä muistelutyön kannalta hyödyllisistä aineistoista. Osallistujat hyödynsivät myös muuta kurssitarjontaa, joka tuki muisteluaineistojen kirjoittamista: ryhmän jäsenten ohjelmassa oli mm. opistolla järjestettäviä vierailuluentoja ja sukututkimuksen sekä luovan kirjoittamisen kursseja.

Yhdessä käytiin myös läpi kirjoittaessa vastaan tulleita kysymyksiä. Yksi esimerkki näistä oli kysymys siitä, miten tekstissä kannattaa viitata lähisukulaisiin. Kirjoittajalle itselleen olisi luontevaa kutsua äitiä äidiksi, siskoa siskoksi jne. Lukijan kannalta tätä pidettiin kuitenkin huonona, sillä lukijan suhde kyseessä oleviin henkilöihin on aivan jokin muu. Näin ollen päädyttiin siihen, että pääasiassa henkilöistä kannattaa puhua heidän nimillään.

7.2.2 Aineiston jäsentäminen

Ryhmäläisille oli kerrottu pienissä osissa matkan varrella erilaisista semanttisen webin periaatteista ja hyödyntämismahdollisuuksista. Aivan aluksi hypertekstirakenteen esittelyn yhteydessä kuvattiin, miten rakenne mahdollistaa myös toisiinsa linkittämättömien sisältöjen yhdistämisen. Kun meillä oli vieraita Tikutus-hankkeesta, Lahdesta, keskusteltiin yhdessä avainsanoista ja sanapilvistä.

Kun työskentelyssä oli päästy toisen lukuvuoden loppupuolelle, osallistujilta pyydettiin vapaaehtoisuuteen perustuen näytteitä omasta muisteluaineistostaan käytettäväksi tämän työn tutkimusaineistona. Kirjoittajat saivat vapaasti valita näytteen pituuden itse. Käytännössä ne vaihtelivat n. 2,5 sivusta 9 sivuun.¹ Kyseiseen aineistonäytteeseen kirjoittajia myös pyydettiin valitsemaan ne avainsanat, jotka heidän mielestään kuvaavat tekstin sisältöä.

Avainsanojen valintaan ei annettu mitenkään yksityiskohtaisia ohjeita: esimerkkejä katsottiin joistain blogeista, kuvakirjastoista ja tunnetuista verkkosivustoista. Ajatuksena oli, että kirjoittajat merkitsisivät avainsanat mahdollisimman itsenäisesti ja ilman laajoja pohjatietoja, kuten internetin käyttäjät muutenkin toimivat.²

Avainsanojen on tarkoitus toimia myös toisin päin: niiden avulla mahdollistetaan se, että muut lukijat löytäisivät nyt laaditun, uuden aineiston, mikäli se sijoitettaisiin internetiin. Ontologioiden käytön hyödyllisyys näiden muiden toimijoiden kannalta jää testaamatta. Siihen ei tässä tapauksessa ole mahdollisuutta, sillä aineistoille ei ole toistaiseksi suunniteltu muuta jatkokäyttöä eikä niitä ole tarkoitus sijoittaa avoimesti suuren yleisön luettavaksi. Tikutus-hanke on saattanut poikia mahdollisia jatkoideoita ja silloin voidaan eri tutkimuksissa testata ontologioiden käytön hyödyllisyyttä myös muiden toimijoiden (esim. palveluntarjoajien) näkökulmasta.

Muisteluryhmän osallistujilta saadut aineistonäytteet käytiin läpi tavalla, joka kuvataan seuraavassa luvussa 8.

7.3 Pilotin päättäminen ja jatkotyöskentely

Vaikka muisteluryhmän toiminta on jatkunut edelleen useiden lukuvuosien ajan³, loppui tutkimushankkeen pilottina toimiminen kevätlukukauden 2014 päättyessä. Toisen toi-

¹ Näytteiden pituudet on mainittu liitteessä 1 aina kunkin näytteen avainsanalistojen yhteydessä.

² Näin ajattelevat myös esim. Angeletou ym. [2007] Heidän mukaan folksonomiat ovat kontrolloimattomia ja niiden käyttäjiltä ei edellytetä ennakkotietoja.

³ Ryhmän toimintaa jatkettiin kevääseen 2017 saakka.

mintavuoden lopuksi järjestimme huhtikuussa 2014 tiedotustilaisuuden, jossa esitimme työn tuloksia. Paikalla olivat sekä Tikutus-hankkeen edustaja Lahdesta että Mäntsälän Uutisten toimittaja.

Osallistujat kertoivat työskentelystä ja omista aineistoistaan, ja tilaisuuden pohjalta julkaistiin artikkeli Mäntsälän Uutisissa [Niemi2014]. Ryhmäläisten aineistoja oli mukana myös Tikutus-hankkeen loppuraportissa [Kuosm2014].

Toimintavuoden päätteeksi järjestettiin myös historia-aiheinen retki: ryhmä kävi yhdessä Riihimäellä. Tutustumiskohteina olivat Suomen lasimuseo, lasinpuhaltajien työhuoneet, Riihimäen rautatieasema, Rautatiepuiston puutaloalue sekä Riihimäen taidemuseo.

Koska kuluneen kahden vuoden aikana oli monella ryhmän jäsenistä omien muisteluitten työstäminen päässyt vasta hyvään alkuun, perustettiin ”yleisön toivomuksesta” jatkoryhmä lukuvuodelle 2014-15. Se ei ollut enää tekemisissä Tikutus-hankkeen kanssa ja siihen otettiin uusiakin osallistujia. Suurin osa ensimmäisen muisteluryhmän osallistujista jatkoi työskentelyä vielä jatkoryhmänkin kanssa.

Jatkoryhmän työskentely noudatteli pääpiirteissään samoja linjoja kuin aiempikin ryhmä, mutta sitä ei kuvata enää tässä työssä, koska se ei kuulunut alkuperäiseen pilottiin eikä se tuonut varsinaista uutta muistelutikkukonseptiin. Lähinnä toiminta vain toisti alkuperäisiä, hyviksi koettuja toimintatapoja, joita ei koettu tarpeelliseksi muuttaa.

Myöhemmin muisteluryhmän toimintaa jatkettiin aina kevätlukukauden 2017 loppuun saakka.

8 Tulokset

Kohdassa 8.1 kerrataan lyhyesti työn tavoitteet ja kuvataan tutkimuksen kulku. Sen jälkeen (kohta 8.2) käydään läpi työskentelyn aikana tehdyt havainnot jaoteltuina sen mukaan, olivatko ne yhteneviä lähdemateriaalin kanssa vai ei. Tämän jälkeen kuvataan myös työn tuottamat uudet huomiot. Lopuksi kohdassa 8.3 keskustellaan työn rajoituksista.

8.1 Työn tavoitteet ja tutkimuksen kulku

Aluksi kertaamme tutkimuskysymyksen ja kuvaamme sitä toimintaa, joilla kysymykseen on pyritty löytämään vastauksia. Käymme lyhyesti läpi sekä toimenpiteet että työskentelyn periaatteet.

8.1.1 Tavoitteet ja niiden saavuttamiseksi tehty toiminta

Työn aineistona ovat henkilöiden itsensä laatimat, digitaaliset multimedia-elämäkerrat ja muisteluaineistot. Tavoitteena oli saada vastaus kysymykseen, sisältävätkö julkiset, yhteiseen käyttöön tarkoitetut ontologiat elämäkertatekstien jäsentämiseen tarvittavia käsitteitä. Selvittämisen tarkoitus oli tutkia, missä määrin asiantuntijavetoisesti kehitetyt ontologiat ja käyttäjien itsensä laatimien aineistojen avainsanat ovat yhteneviä.

Kirjallisuuskartoituksen perusteella samansuuntaisia tutkimuksia on tehty aiemmin, tosin hyvin eri kontekstissa ja eri kielialueilla (lähteissä oli englanninkielisissä ja ranskan-kielisissä ympäristöissä tehtyjä kokeiluja). Tarkoituksena oli selvittää, tuottavatko suomenkielisiin elämäkertateksteihin tehdyt avainsanajäsennykset ja niiden vertailu yleisiin ontologioihin (käytännössä Finto-ontologiakirjaston ontologioihin) samanlaisia tuloksia kuin aikaisemmat tutkimukset. Lisäksi tavoitteeksi otettiin kirjata ylös myös vertailussa tehdyt omat uudet huomiot.

Sitä varten suoritettiin kokeilu, jossa viisi henkilöä tuotti omasta elämästään hypertekstimuotoista muisteluaineistoa, valitsi siitä tutkimuskäyttöön haluamansa katkelman ja jäseni sen sisältöä omilla avainsanoilla.

Tutkija luki saamansa aineistokatkkelmat kahteen kertaan ja valitsi myös itse jokaisesta tekstistä joukon avainsanoja. Kirjoittajan ja tutkijan keräämät avainsanalistat koottiin yhteen, niitä verrattiin toisiinsa ja Finto-ontologiakirjaston sisältöön. Saadut tulokset koottiin yhteen ja niitä verrattiin lähdeaineistosta saatuihin vastaaviin tuloksiin. Näitä tuloksia sekä edellä kuvattujen käsittelyvaiheiden varrella tehtyjä huomioita kuvataan edempänä kohdissa 8.2 ja 8.3.

8.1.2 Aineiston kokoamisessa noudatetut toimintamallit

Kuten edellä (alakohdat 3.3.1 ja 7.2.2) oli kuvattu, elämäkerta-aineistojen kirjoittajille ei annettu kovin syvällistä evästystä avainsanojen merkitsemistä varten. Perusteluna oli, että yleisesti webin käyttäjät jäsentävät omat verkkomateriaalinsa ilman sitä koskevaa perehdytystä.

Tutkimuksen toteutussuunnitelmaan kuului myös se, että tutkija teki itse omat avainsanalistansa. Tarkoituksena oli verrata, miten ”tavallisen käyttäjän” ja tietojärjestelmäammattilaisen tekemän avainsanajäsennykset poikkeavat toisistaan ja vaikuttaako valitsijan tausta siihen, kuinka suuri osa valituista avainsanoista sisältyy yleisen Finto-ontologiakirjaston ontologioihin.

Ylipäätään kaksi henkilöä saattaa merkitä aineiston avainsanat varsin eri tavoilla. Angeloun ym:iden [2007] mukaan asiantuntija ja ei-asiantuntija¹ saattavat merkitä saman kohteen hyvin erilaisilla avainsanoilla. Tässä tapauksessa ei ole tarkoitus nimetä tutkijaa asiantuntijaksi, kummallakin toimijalla on asiantuntemusta eri asioista. Elämäkertojen kirjoittajat ovat sisältöjen eli omien elämänvaiheittensa asiantuntijoita eli he vastaavat tässä substanssialueen asiantuntijaa. Tutkijalla on asiantuntemusta semanttisen webin ja tiedonhaun periaatteista, joten näkökulmat aineistoon ja avainsanoihin voivat olla erilaiset. Lähinnä tarkoituksena on hakea kaksi eri näkökulmaa siihen, mitä avainsanat voisivat olla.

Kahden eri henkilön tekemän avainsanajäsennyksen tarkoituksena oli myös vahvistaa ontologiassa mahdollisesti olevien avainsanojen löytämisen todennäköisyyttä. Onhan mahdollista, että aineistoa kuvaava avainsana olisi kyllä ontologiassa, mutta tekstin laatija ei ole nostanut kyseistä käsitettä avainsanojen joukkoon. Kahden eri henkilön tekemä avainsanajäsennys lisää näkökulmien määrää ja tuottaa näin ollen laajemman avainsanalistan.

Avainsanojen merkitsemisessä tuli välittömästi vastaan kysymys siitä, mikä merkitsemisen tavoitteena on. Avainsanojen merkitsemistä voidaan verrata myös tietokirjalli-

¹ Tässä tarkoitettiin substanssialueen asiantuntijaa.

suudessa usein käytettäviin asia-, henkilö- tai paikkahakemistoihin. Asiahakemistoon suositellaan laitettavaksi sellaisia käsitteitä, joita on käsitelty tekstissä enemmän kuin pelkän maininnan verran. [Merta2007]

Tutkijan roolissa tämä vaikutti sopivalta ohjenuoralta myös avainsanojen merkitsemiseen¹, mutta joissain kohdin siitä tuli tehtyä poikkeamia. Lähinnä poikkeamisen tarve tuli vastaan tapauksissa, joissa jokin tekstin keskeinen käsite oli jo nykyisellään tuntematon tai sen saatettiin olettaa muuttuvan tulevaisuudessa tuntemattomaksi. Tällöin se tuli merkittävä avainsanaksi, sillä tässä tapauksessa oli tiedossa, että avainsanojen kautta voitaisiin mahdollisesti luoda yhteys ontologiakirjaston sisältämiin määrittelyihin.

Mikäli ei olisi mahdollista hyödyntää ontologian tarjoamaa semanttista tukea tekstin sisällön tulkitsemiseen, olisi avainsanojen merkitsemisellä pelkästään hakemistotarkoitusta palveleva rooli. Koska tiedetään ontologioiden kautta saatavan tukea myös sisällön merkityksen avaamiseen, tutkija merkitsi myös sellaisia avainsanoja, joiden roolina oli enemmänkin tuottaa selittävää, merkitystä avaavaa sisältöä. (Esim. aineiston 1 tekstissä raskausmyrkytyksestä ei kerrottu kovin paljoa, mutta tekstin sisältö avautuu eri tavalla, kun tietää, mitä raskausmyrkytys tarkoittaa.)

8.2 Aineiston käsittelyssä tehtyjä havaintoja

Vaikka tutkimusaineisto oli pieni, nousi sen käsittelystä esiin asioita, jotka auttavat ymmärtämään käyttäjien tekemien omien avainsanajäsennysten ominaispiirteitä ja eroja suhteessa ontologioihin. Kohdassa 8.2.1 tutkimuksen määrälliset tulokset on vedetty yhteen ja esitelty kvantitatiivisen analyysin yhteydessä tehtyjä huomioita. Sen jälkeen, kohdassa 8.2.2, on kuvattu asioita ja huomioita, jotka ovat nousseet esiin aineiston laadullisen käsittelyn yhteydessä.

Aineiston perusteella tehdyistä havainnoista on ensin esitetty sellaiset asiat, jotka olivat vain tälle tutkimukselle ominaisia. Sen jälkeen on kuvattu ne ominaisuudet, jotka olivat yhtenevät lähdeaineiston kanssa. Lopuksi on kuvattu työn yhteydessä tehdyt täysin uudet huomiot.

¹ Elämäkertaa-aineistojen kirjoittajille kyseistä suositusta ei esitetty.

8.2.1 Aineiston määrällisen käsittelyn tulokset

Aluksi esitellään taulukkona määrällisen käsittelyn tulos. Sen jälkeen kuvataan lyhyesti, mitä vastaavia tuloksia oli lähdemateriaaleissa. Lopuksi keskustellaan siitä, mikä on saatujen tulosten ja lähdeaineistojen suhde.

Aineistojen vertailu suhteessa Finton sisältämiin käsitteisiin

Kaikkien aineistojen avainsanaluettelot on kerätty liitteeseen 1. Seuraavassa on määrällinen yhteenveto, joka kertoo avainsanojen esiintyvyydestä Finto-ontologiakirjaston ontologioissa.

Aineiston numero	Kirjoittajan valitsemat avainsanat (kpl)	Tutkijan valitsemat avainsanat (kpl)	Kirjoittajan valitsemista Finton ontologioissa (kpl)	Kirjoittajan valitsemista Finton ontologioissa (%)	Tutkijan valitsemista Finton ontologioissa (kpl)	Tutkijan valitsemista Finton ontologioissa (%)
1	25	61	11	44 %	37	61 %
2	59	44	25	42 %	28	64 %
3	9	58	6	67 %	33	57 %
4	19	25	8	42 %	14	56 %
5	53	90	27	51 %	68	76 %
yht.	165	278	77	47 %	180	65 %

Tuloksissa on oletettu, että taivutetussa muodossa olevat avainsanat on muunnettu perusmuotoon ennen hakujen tekemistä.

Määrällisessä analyysissä kiinnittää huomiota, että osa käyttäjistä (aineisto 3) merkitsi hyvin vähän avainsanoja, osa taas tähän verrattuna paljon (aineisto 2). Golder ja Huberman [2006] ovat tutkineet tätä asiaa del.icio.us-palvelun käyttäjistä ja heidän mukaansa osa käyttäjistä on sellaisia, että heillä on tyylinä merkitä vähän avainsanoja, jotkut toiset taas liittävätkin niitä kohteeseen runsaammin. He eivät kuitenkaan havainneet trendiä, että tämä taipumus muuttuisi käyttäjillä ajan myötä sitä mukaa, mitä enemmän he ovat liittäneet aineistojaan yhteiseen palveluun.

Yleiskuva, jonka määrällinen analyysi antaa on seuraava: karkeasti ottaen yksittäisen avainsanan esiintymistodennäköisyys Fintossa on 50 %:n luokkaa, mikä ei tavallaan vahvasta esiintyvyyden todennäköisyyttä mihinkään suuntaan. Se on normaali odotusarvo todennäköisyydelle silloin, kun ei ole mitään tekijöitä, jotka painottavat tilannetta johonkin suuntaan. Tuloksena se tarkoittaa, että käyttäjän aineistosta poimittu avainsana voi yhtä hyvin löytyä Finton ontologioista kuin olla löytymättä.

Käyttäjän ja tutkijan tekemissä avainsanamäärityksissä on pieni ero suhteessa Fintossa esiintymiseen. Jos avainsanamääritykset olivat tutkijan tekemiä, useimmiten niiden esiintyvyys Fintossa kohosi hieman korkeammaksi kuin käyttäjällä, ja tulos oli aina selvästi yli 50 %.

Vaikuttaa siis siltä, että tietojärjestelmäammattilaisen ja ontologioihin perehtyneen jatko-opiskelijan näkökulma vaikuttaa jossain määrin tehtyihin avainsanamäärityksiin. Omalta osaltaan tämä vahvistaa epäilyä siitä, että ontologiat ovat kehittäjien ja it-asiantuntijoiden näkökulmasta laadittuja, ei käyttäjien.

Muista tutkimuksista saatuja tuloksia

Tesconi ym. [2008] antavat muutamia esimerkkejä tutkimuksista, jotka hyödyntävät webissä olevia ontologioita folksonomioiden terävöittämisessä. Heidän mukaansa niiden ongelmana on ontologioiden huono kattavuus joitakin erityisaloja lukuunottamatta. He eivät kuitenkaan maininneet mitään prosenttilukuja, joten on mahdotonta tietää, kuinka huonosta kattavuudesta oli kyse.

Angeletoun ym:iden tutkimuksessa [2007] tutkittiin avainsanakimppuja, jotka oli saatu verkkopalvelusta seuraavanlaisen haun tuloksena: annettiin haluttu avainsana -> valittiin sen kohdalta lisähaku, joka tuottaa nipun jollain tapaa yhteen liittyviä avainsanoja (tag cluster). Näitä avainsananippuja pyrittiin rikastamaan semanttisesti käyttämällä hyväksi verkosta löytyviä ontologioita.

Tutkimusta tehdessä havaittiin, että avainsanoista vain osa sisältyi ontologioihin. Mahdollisina syinä annettiin avainsanojen mahdollinen uutuus, avainsanojen sisältämä yksityisluonteinen aineisto (yksityishenkilöiden nimet ja päivämäärät), yhdistelmäkäsitteet (esim. kouluruokakokeilu vs. kouluruoka ja kokeilu) ja monikielisten ontologioiden puute (folksonomiat osoittautuivat monikieliseksi). Lisäksi aineistojen laatijat käyttivät avainsanoina paljon tuottamistekniikkaan (esim. valokuvausalaan) liittyviä termejä, jotka eivät ole ontologioissa keskeisiä. [Angel2007]

Angeletoun ym:iden kokeilussa ei kuitenkaan voida ajatella, että ensisijaisesti avainsanojen uutuus aiheuttaisi sen, että niitä tai niihin liittyviä termejä ei ole ontologioissa. Käytetyt avainsanat olivat nimittäin sieni, hedelmät, virvoke ja nisäkäs, joten ne ovat termeinä hyvin vakiintuneita samoin kuin suurin osa niihin liittyvien avainsananippujen sisällöistä. Näistä kaikista avainsanoista kuitenkin vain keskimäärin 23 % oli ontologioissa. Tutkijat kuitenkin ajattelivat, että tilanne paranee koko ajan sitä mukaa, kun enemmän ontologioita tulee verkkoon saataville. Kokonaisuutena he kokivat kuitenkin ontologiat hyödyllisinä folksonomioitten semanttisen rikastamisen apuvälineinä. [Angel2007]

Specian ja Mottan tutkimuksessa valittiin kahdesta verkkopalvelusta (del.icio.us ja Flickr.com) käyttäjien merkitsemiä avainsanoja ja tutkittiin näiden esiintymistä ontologioissa. Tavoitteena oli jalostaa näin syntynyttä folksonomiaa ontologioiden suuntaan määrittelemällä folksonomioihin avainsanojen välisiä suhteita. Kun löydettyjä avainsanoja haettiin ontologioista, vain alle puolet oli löydettävissä sieltä: 42 % del.icio.us-palvelun avainsanoista ja 23 % Flickr.com-palvelun avainsanoista sisältyi johonkin ontologiaan. Suurin osa avainsanapareista oli löydettävissä ainoastaan WordNetin tietokannasta. [Speci2007]

Vertailu suhteessa muihin tutkimuksiin

Omassa tutkimuksessani sain siis määrällisten mittausten osalta hyvin samansuuntaisia tuloksia kuin lähdemateriaaleissa. Toki prosenttilukemat ovat jo parantuneet lähdeaineistojen kirjoittamisajankohdasta, mutta vieläkään ei olla käyttäjien tekemien jäsennysten kohdalla päästy selvästi yli 50 %:n.

Tutkin itse omilla aineistoillani termien esiintyvyyttä Finto-palvelussa pitkän aikavälin kuluessa (vuosina 2016-18). Finto-kirjasto on koko ajan täydentyvä kokoelma, joten on mahdollista, että viimeisenä tarkistetun, viidennen aineiston paremmat lukumäärät kertovat siitä, että kirjaston kattavuus on parantunut. Toisaalta kyse on vain yksittäisestä aineistosta, joten sitä ei missään tapauksessa voi käyttää todistusaineistona paremmasta kattavuudesta.

Täytyy myös huomauttaa, että Fintossa on koottuna eri alojen ontologioita. Mikäli teksti sattuu käsittelemään sellaista erikoisalaa, jonka ontologia on Finto-palvelussa, on termien esiintyminen huomattavasti todennäköisempää, ainakin jos termille halutaan sanallista määritelmää¹. Siinä mielessä myös tekstin aihe vaikuttaa termien esiintyvyyteen.

Ylipäätään määrällinen tarkastelu on omassa tutkimuksessani kuitenkin niin vähäistä, että siitä ei voida vetää mitään vahvoja tutkimuksellisia johtopäätöksiä. Tutkimusasetelmaa helpottaa se, että lähdekirjallisuuden perusteella on käytettävissä tähän rinnastettavia tutkimustuloksia. Tällöin kyseessä on Seddonin ja Shepersin [2012] mukaan nk. analyttinen yleistäminen. Siinä jotain oletusta tutkitaan käyttämällä hyväksi lähdekirjallisuutta ja huolellisesti analysoitua empiiristä otosta. Empiirinen otos voi olla tällöin niin pieni, että siltä ei voida edellyttää tilastollista selitysvoimaa.

¹ Fintoon sisältyvät myös Yleinen suomalainen asiasanasto (YSA) ja Yleinen suomalainen ontologia (YSO), jotka ovat varsin laajoja yleissanastoja, mutta niissä ei yleensä ole termien määritelmiä. Niissä semantiikka välittyy ainoastaan tutkimalla termin sijaintia suhteessa muihin termeihin (yläkäsitteet, alakäsitteet, assosiatiiviset käsitteet jne.)

Lähinnä kuitenkin on kyse mielenkiintoisesta selvityksestä, jota ei malttanut olla tekevä, mutta tilastollista todistusvoimaa näin pienellä aineistolla ei ole. Toki tästä voisi saada validin tuloksen myös tilastollisilla menetelmillä ottamalla useita eri aineistoja ja toistamalla samaa toimintamallia tarpeeksi paljon. Tämä esitetäänkin yhtenä jatkotutkimusaiheena.

8.2.2 Käytettyjen lähteiden kanssa yhtenevät tulokset

Osa tehdyistä huomioista oli hyvin samanlaisia kuin lähdekirjallisuudessa. Niitä käsitellään seuraavassa.

Yksikkö- ja monikkomuodot vaihtelevat

Samoin kuten lähteinä olevissa tutkimuksissa, myös tämän työn esimerkeissä tuli esille vaihtelevia käytäntöjä sen suhteen, kirjoitetaanko avainsana yksikkö- vai monikkomuotoon, ks. esim. [Golde2006] ja [Mathe2004]. Käytäntö ei riipu pelkästään kirjoittajasta, vaan sama kirjoittaja voi tilanteesta riippuen käyttää kumpaa tahansa muotoa.

Huomasin itsekin, että määrätyt sanat tuntuvat luonnollisemmilta monikossa kuin yksikkömuodossa, toiset taas päinvastoin. Silti oma logiikkani ei mennyt aivan yksiin Finton ontologioiden kanssa: esimerkiksi ”keittiö” tai ”saari” ovat termejä, joita mielestäni yleensä käytetään luontevasti puhuttaessa yhdestä keittiöstä tai yhdestä saaresta. Sen vuoksi on luonnollisen kielen näkökulmasta erikoista, että kyseiset käsitteet ovat lähes kaikissa ontologioissa nimenomaan monikkomuodossa.

Kaiken kaikkiaan Finto-ontologiakirjastossa oli vaihtelevuutta tämän suhteen, mikä johtuu Seppälän ja Hyvösen mukaan siitä, että Finto on kooste eri ontologioista ja eri kehittäjäryhmillä on erilaisia käsityksiä siitä, kumpi muoto on perusmuotona - yksikkö vai monikko. Perinteisesti suomalaisissa asiansastoissa on suosittu monikkomuotoa, kun taas semanttisen webin ontologioissa käytetään yksikkömuotoa, mikä mukailee kansainvälistä käytäntöä. [Seppä2014]

Yleislinja Fintossa vaikutti olevan, että ainesanoja (ja useita abstrakteja käsitteitä) lukuunottamatta substantiivit olivat ontologioissa ensisijaisesti monikon nominatiivimuodossa. Tähän oli kuitenkin joitain poikkeuksia, esim. ”nuotio” oli ontologiassa yksikkömuodossa. Suurimmasta osasta termejä haku onnistui myös yksikön nominatiivin avulla. Tässäkin oli poikkeuksia, esim. ”riippusilta” ei tuottanut tulokseksi ”riippusillat”-termiä, vaikka se oli ontologiassa. Samoin kävi myös jäljempänä mainitun termin ”uittohaka” kohdalla.

Finto on ontologiakirjasto, joka sisältää useita eri ontologioita. Nimenomaan yksikkö-/monikko-käytännön suhteen nousi esiin yksi ontologia - ikonografian ontologia, joka on tarkoitettu käytettäväksi taidetta ja kuvia käsittelevien tekstien yhteydessä. Useasti jokin yksittäinen termi tuli esiin vain ikonografian ontologiasta yksikkömuodossa, muissa ontologioissa se sama oli kirjattu monikkomuodossa. Käyttäjän onneksi haku toi yleensä automaattisesti nähtäväksi sekä yksikkö- että monikkomuodot.

Finto-ontologiakirjastossa oli myös tapauksia, joissa yksikkö ja monikko viittasivat eri käsitteisiin. Esimerkiksi radio ja televisio yksikkömuodossa viittasivat joukkotiedotusvälineisiin ja niihin liittyvään toimintaan. Monikkomuodossa vastaavat termit puolestaan viittasivat vastaanotinlaitteisiin. Jos siis automatisoitu tekstinkäsittely muuntaa yksikkömuotoiset avainsanat monikkoon ennen ontologiahakua, tulee tulokseksi eri käsitteitä kuin kenties oli tarkoituksena.

Kirjoittajan motiivit ja kiinnostus vaikuttavat valintaan

Kuten myös Mathes [2004] on tuonut esille, käyttäjien omat avainsanaluokittelut osoittavat, mitä termejä käyttäjät itse haluavat käyttää. Omissa aineistoissani tuli esille esimerkiksi murre sanojen käyttö. Joissakin tapauksissa sanasta saattaa olla olemassa vaihtoehtoinen kirjoitusasu (pätee erityisesti murre sanojen kohdalla), ja käyttäjien omista luokitteluista näkyy, mikä kirjoitusasu on heille luontevin.

Huomasin myös selkeästi saman asian kuin mitä oli tuotu esille Kimin ym:iden artikkeleissa [KimYa2007]: avainsanat kertovat, mistä me olemme kiinnostuneita, ne eivät niinkään kuvaa kohdetta objektiivisesti tai neutraalisti. Ylipäätään se, että merkitsemme jonkin asian avainsanaksi, voi olla merkki siitä, että asia on meille kiinnostava.

Samoilla linjoilla ovat Specia ja Motta, joiden mukaan avainsanojen merkitsemisen motiivina ovat omat tarpeet [2007]. Myös Golder ja Huberman [2006] ovat empiiristen tutkimustensa perusteella sitä mieltä, että huomattava osa avainsanamerkinnoista tehdään itselle, ei muille webin käyttäjille.

Jäljempänä (kohdassa 8.2.3) puhun lisää avainsanamerkintöjen intentiosta: se vaikuttaa huomattavasti siihen, mitä asioita aineistosta nostetaan esille. Onko avainsanojen tarkoituksena kuvata tekstin sisällöstä tällä hetkellä tärkeimmät käsitteet vai halutaanko ottaa huomioon myös tulevaisuuden lukutilanteet? Elämäkertojen kohdalla motiivi saattoi olla käsitteen merkityssisällön epäselvyys nykypäivän tai tulevaisuuden lukijalle. Esimerkiksi kasetti ei vielä nykypäivänä ole vieras termi/käsite, mutta tulevaisuudessa on hyvin mahdollista, että sen merkityksen tietää enää harva.

Itse huomasi, että joitakin käsitteitä tuli merkittyä avainsanaksi, vaikka ne eivät olleet tekstissä keskeisiä, esimerkkinä mm. ”jääsaha”. Jos kuitenkin lukeminen tapahtuu vaikkapa kahdenkymmenen vuoden päästä, niin silloin niille saatetaan kaivata selitystä, jotta tekstikokonaisuuden merkitys avautuisi. Jos tällainen käsite on tekstissä merkittävässä asemassa siten, että kokonaisuuden idea jää vaillinaiseksi ilman sen ymmärtämistä, kannattaa kyseinen asia merkitä avainsanaksi, jotta ontologiakirjaston kautta saadaan kyseiselle termille määritelmä ja voidaan hakea muita sitä koskevia aineistoja.

Ontologioiden automatisoitu johtaminen on hankalaa

Tässä tutkimuksessa tuli esille useita seikkoja, jotka puhuivat inhimillisen tulkinnan tärkeyden puolesta (eri kirjoitusasut, yhdyssanavirheet, murre-erot, homonyymit, synonyymit, eri tavat käyttää kieltä jne.). Myös Weber [2008] kirjoittaa tästä: hänen mukaansa luonnollinen kieli on monimutkaista, mistä johtuen automatisoidulla ontologioiden johtamisella on mahdollista saada irti ainoastaan osa aineiston sisältämästä tietämyksestä.

Ylipäätään aineistojen puutteellinen laatu, virheet ja kielen monimerkityksisyys aiheuttavat paljon epätarkkuutta silloin, kun luonnollisesta kielestä halutaan johtaa rakenteellisia faktoja ja määrittämiä. Mikäli web 2.0 -lähestymistavan mukaisissa palveluissa halutaan hyödyntää aineistojen sisältämää tietämystä, tarvitaan työkaluja, joilla on mahdollista yhdistää rakenteista, formaalia tietämystä (esim. ontologiat) luonnollisen kielen avulla esitettävään tietämykseen. [Weber2008]

Tässä astuvat kuvaan web 3.0 -teknologiat, joissa yhdistetään semanttisen webin työkaluja ja web 2.0 -mallin mukainen käyttäjälähtöisyys. On kehitetty paljon erilaisia tapoja johtaa tietämystä tekstistä, mutta kun tekstit ovat käyttäjien tuottamia, tulee kuvaan mukaan runsaasti lisähaasteita. Käyttäjien tuottamissa sisällöissä kirjoitusasu ei aina ole täsmällinen, teksti muistuttaa usein puhekieltä, sisältää murteita jne. Myös tämän tutkimuksen esimerkit osoittavat, että haasteet ovat moninaiset huolimatta siitä, että koneellinen tekstin tulkinta tuottaa koko ajan lupaavampia tuloksia.

8.2.3 Aineiston käsittelyn yhteydessä tehdyt huomiot

Seuraavat asiat nousivat esiin tehtäessä suunnitelman mukaista avainsanojen merkitsemistä sekä avainsanojen vertailua ontologioiden sisältöön.

Intention vaikutus avainsanojen valintaan

Aineistojen käsittelyn yhteydessä tuli moneen otteeseen esille se, miten paljon avainsanojen valintaan vaikuttaa se tarkoitus, jossa avainsanoja merkitään. Ensinnäkin aineistojen kirjoittajan näkökulmasta aivan eri asiat voivat olla tärkeitä kuin lukijan näkökulmasta. Kirjoittaja näkee tekstin oman elämänsä tarinana, ja silloin avainsanojen valintaan voi vaikuttaa se, mitkä asiat ovat olleet hänelle kokemusten yhteydessä tärkeitä.

Ulkopuolinen henkilö voi taas nähdä tekstissä ajankuvaa tai kiinnittää huomiota oman kokemuspohjansa perusteella erikoisena näyttäytyviin asioihin. On myös eri asia merkitä avainsanoja, jotka määrittävät tekstin keskeistä sisältöä kuin sanoja, jotka voivat olla tulevaisuuden lukijalle vieraita.

Merkitsemisen käytännöt kehittyvät työskentelyn myötä

Avainsanojen valinta ei ylipäätään ollut mikään itsestään selvä toimenpide, jossa lopputulos on helppo määritellä. Varmuuden vuoksi tein jokaisen aineiston kohdalla avainsanojen valinnan kaksi kertaa, ja tuloksena ollut avainsanajoukko muuttui kaikkien aineistojen kohdalla. Mitä enemmän sanoja valitsin, sitä enemmän aloin irtautua tekstin alkuperäisistä sanamuodoista. Esimerkiksi itselleni tuli käytännöksi kirjoittaa verbit aina samassa määrättyssä muodossa: valitsin avainsanaksi infinitiivimuodon ”ravustaminen”, vaikka tekstissä oli ”ravustus”.

Verbien kohdalla oma käytäntöni osoittautui olevan täysin päinvastainen kuin Finton ontologioiden laatijoilla. Ontologioissa verbit ovat yleensä substantivoidussa muodossa, esim. kalastus. Usein käsite löytyy Finto-palvelusta myös hakemalla verbin infinitiivimuodolla ”kalastaminen”, mutta aina tämä ei onnistu.

Avainsanat alkoivat myös luonnostaan muotoutua niin, että ne olivat keskenään johdonmukaisesti merkittyjä ja usein hieman tekstiä yleisemmällä tasolla (sen aihepiiriä kokoavia). Esim. kun teksti kertoi erilaisista kalastusvälineistä, en laittanut avainsanaksi niitä, vaan esim. termin ”kalastaminen” (jos oli kyse toiminnasta) tai ”kalastus” (jos viitattiin enemmän välineisiin yms.).

Edellä kuvatun perusteella on helppo päätellä, että verrattaessa eri käyttäjien tekemiä avainsanajäsennyksiä heillä on ollut hyvin erilaisia käytäntöjä merkitsemisessä. Samoin myös yksittäisen henkilön kohdalla eri aikoina tehdyt jäsenykset voivat noudattaa eri-laista logiikkaa.

Erisnimien rooli

Jos avainsanojen tarkoituksena on toimia eräänlaisena hakemistona aineiston sisältöön, on hyvin tärkeää, että erisnimiä merkitään avainsanoiksi. Erisnimillä voi olla myös tärkeä rooli yhdistettäessä eri lähteitä, esim. paikkakunnan nimen perusteella voidaan saada sitä koskevaa, tarkempaa tietoa muista web-aineistoista. Samoin aineistoissa esiintyvistä julkisuuden henkilöistä voi olla paljonkin tietoa muualla internetissä.

Yhdistettäessä web-aineistoja erisnimien perusteella voidaan hyödyntää aineistosivuilla mahdollisesti olevaa metadataa, joka on jäsennetty yhteisen, tunnetun rakenteen perusteella. Nimissä on kuitenkin paljon homonyymejä, joten ontologian URI-osoitteen kautta tapahtuva linkittäminen tuottaisi varmuuden siitä, että haun tuloksena saatu henkilö, paikka, tapahtuma tms. on juuri se, jota tarkoitetaan (toki laajalti tunnettujen kohteiden kohdalla voidaan linkitys tehdä myös esim. Wikipedian kautta).

Jätin ontologiatarjottimien yhteydessä pois sellaiset erisnimet, joita ei voisi olettaa löytyvän ontologioista, mutta silti yleisesti ottaen ontologioissa oli selkeästi puutteita erisnimien suhteen. Paikkakuntien nimiä oli suhteellisen hyvin, tosin joitakin yleisiä saattoi puuttua, esim. Västeråsia eikä Irkutskia ei ollut Finton ontologioissa¹. Sen sijaan tutkimuksen tekoaikaan henkilöitä, organisaatioita ja yleisesti tunnettuja tapahtumia oli ontologioissa hyvin vähän.

Eri taivutusmuotojen vaikutus

Kirjoittajat käyttivät avainsanojen merkitsemisessä erilaisia käytäntöjä. Toiset kirjoittivat sanat näkyviin tekstin yhteyteen (esim. sivujen alareunaan tai sivumarginaaliin), toiset alleviivasivat tekstistä ne sanat, jotka he valitsivat avainsanoiksi.

Erityisesti silloin, kun avainsanat oli merkitty alleviivaamalla, olivat ne taivutetussa muodossa, ts. siinä luvussa ja sijamuodossa kuin ne tekstissä esiintyvät. Tällöin käsitteiden etsiminen Finton ontologioista edellyttää mielellään sanojen muuntamista perusmuotoon. Mikäli haku tehtiin täysin siinä samassa muodossa, jossa sana esiintyi tekstissä, ei tulosta läheskään aina löytynyt, vaikka perusmuodossa sana oli sisällytetty ontologioihin.

Esimerkkeinä edellisestä voisi mainita termit, jotka esiintyvät tekstissä muodoissa ”maanvyörymän”, ”riippusillan” ja ”saaveissa”. Nuo kaikki sisältyvät kyllä Finto-ontologiakirjastoon, mutta taivutetussa muodossa kyseisiä sanoja ei ollut mahdollista löytää. Yleisesti haku pitää mieluiten tehdä monikon nominatiivimuodossa.

¹ Tarkistettu viimeksi 16.4.2017

Tarkastelun yhteydessä tuli vastaan muitakin esimerkkejä, joissa erikseen kirjoitetut avainsanat toimivat paremmin kuin sellaisenaan tekstissä esiintyvät sanat, sanaparit tai useasta sanasta muodostuvat ilmaisut. Esimerkkinä voisi olla aineistossa 4 esiintynyt sanallinen kuvaus ”ei osannu lukea”. Jos ontologiaan tehtäisiin hakuja kyseisellä kolmen sanan yhdistelmällä tai kullakin sanalla erikseen, ei haku tuottaisi tuloksia. Sen sijaan termi ”lukutaidottomuus” sisältyy Finto-ontologiakirjastoon, joten sillä haettaessa sisälön merkitys aukeaa.

Jos aineiston laatija oli merkinnyt sanaparin alle yhtenäisen viivan, tulkittiin sanapari yhdessä avainsanaksi. Tällaisia sanapareja oli kuitenkin ontologioissa hyvin harvoin, joten näissä tapauksissa ontologiahaku täytyi tehdä erikseen molemmille sanoille.

Edellä kuvatun perusteella voidaan päätellä, että avainsanat tulisi merkitä perusmuodossa (mieluiten monikon nominatiivissa) ja useasta termistä muodostuvalle ilmaisulle olisi hyvä löytää jokin selkeä käsitevastine.

Eri kirjoitusasut

Suomi on maantieteellisesti laaja maa ja eri murteissa on omanlaisensa tyyli käyttää sanoja. Teksteissä oli useita hyvinkin yleisiä arkipäivän termejä, joiden kirjoitusasu vaihteli sen mukaan, mistä päin maata kirjoittaja on lähtöisin, esim. ”kamari” vs. ”kammari”. Myös vierasperäiset tai vierasperäisiä osia sisältävät sanat saatetaan kirjoittaa joko alkuperäistä kirjoitusasua tai ääntämisasua mukailevalla tavalla, esim. ”limusiini” tai ”limousiini” tai ”puucee” vs. ”puusee”. Nämä kaikki tulisi olla huomioituna ontologioissa, jotta käsitteen hakeminen onnistuisi millä variaatiomuodolla tahansa.

Sanojen esiintyminen taivutetussa muodossa vaikutti tuloksiin yllättävän paljon. Jos siis tarkoituksena olisi, että tekstissä olevia sanoja voisi klikata ja hakea termille merkityksen, täytyisi sanat muuttaa perusmuotoon, jotta haku onnistuisi. Tästä on mainittu esimerkkejä edellä. Joskus myös perusmuodossa oleva sana jäi löytymättä, mikäli hakuja tehtiin yksikkömuodossa olevalla sanalla. Tästä on esimerkki termi ”uittohaka”, joka löytyy kyllä Finto-ontologiakirjastosta hakemalla muodossa ”uittohaak”.¹

Myös puhekieli muuttaa kirjoitusasua yllättävän paljon. Varsinkin käyttäjien itsensä tuottamat aineistot ovat usein ilmaisultaan lähempänä puhekieltä kuin ”viralliset aineistot”. Tällöin esim. paikkojen nimet saatetaan usein korvata puhekielessä tavallisesti käytetyillä vastineilla, esim. Helsinki on Hesa tai Stadi. Tämän tutkimuksen aineistoista yksi oli kirjoitettu murteella ja siinä esim. Aronkylä muuntui paikallisten suussa tavalliseen muotoon ”Aroonkylä”.

¹ Tarkistettu viimeksi 7.7.2017.

Sanaliittojen vaikutus merkityksen tulkintaan

Suomen kielessä on sanapareja, jotka kirjoitetaan erikseen, mutta joiden oikea merkitys avautuu vain ottamalla huomioon molemmat puoliskot. Joskus sanaparin toinen osakas on merkitsevempi kuin toinen (esim. ”ensimmäinen ulkomaanmatka”). Toisinaan taas sanojen käsittely erikseen vie aivan eri jäljille kuin niiden käsittely yhdessä (esim. ”viimeinen ehtoollinen” tai ”kissan silittäminen”).

Jos folksonomioiden ja ontologioiden yhdistäminen tehdään automaattisesti (tekstin laatijan merkitsemä avainsana linkitetään ontologiassa vastaavaan käsitteeseen), jää välistä pois ihmisen suorittama valinta sen suhteen, kannattaako sanaparissa tehdä hakuja kummallakin sanalla erikseen vain pelkästään yhdessä.

Tämä johtaa edelleen kysymykseen siitä, miten yhdyssanavirheiden yleistyminen vaikuttaa: viime aikoina on valitettavasti yleistynyt yhdyssanojen kirjoittaminen erikseen¹. Voidaan siis olettaa, että myös verkkojulkaisujen avainsanoissa on usein kirjoitettu yhdyssanaksi kuuluva sanaliitto erikseen.

Suomen kielessä on olemassa paljon tapauksia, joissa merkitys muuttuu riippuen siitä, kirjoitetaanko sanat yhteen vai erikseen. Tällaisia esimerkkejä on sekä yleis- että erisnimissä. Yleisnimitystä esimerkkinä voisi olla tapaus, jossa ”ilta” ja ”puku” on kirjoitettu erikseen eikä yhteen ”iltapuku”. Erisnimityksinä ”Hämeen linna” ja ”Hämeenlinna” viittaavat eri asioihin. Sanapari ”talonmiehen rouva” on puolestaan hyvä esimerkki tilanteesta, jossa oikea merkitys selviää ainoastaan siten, että ”talonmies” on kirjoitettu yhteen ja sanaparia ”talonmiehen rouva” tutkitaan kokonaisuutena.

Teksteissä oli myös hyviä esimerkkejä verbeistä, jotka ovat vastaavanlaisia sanojen yhdistelmiä. Ne tulee aina tulkita kokonaisuutena, jotta oikea merkitys tulee esiin, esim. ”antoi ylen” tai ”voi pahoin”. Myös ilmaisu ”meni näkö” on hyvä esimerkki tapauksesta, joka on hankalasti tulkittavissa koneellisesti ja jonka merkitys avautuu vain tarkastelemalla koko yhdistelmää, ei molempia sanoja erikseen.

Inhimillisen tulkinnan tärkeys

Jo edellä on kuvattu useita tapauksia, joissa korostui inhimillisen tulkinnan tärkeys, mutta niitä on vielä muitakin. Esimerkiksi hyvin yleiset sanat, kuten ”odotusaika”, voivat merkitä eri yhteyksissä hyvin eri asioita. Mikäli tiivis teksti ei muuten kerro synnyttämisestä ja lapsen saamisesta, voi odotusaika koneellisessa käsittelyssä helposti assosioitua esim. palvelun asiakkaan odotusaikaan.

¹ Ks. esim. Sari Maamies, Ylioppilasaine ja ajan henki - sensorien ajatuksia aineista, Kielikello 2/2002, <http://www.kielikello.fi/index.php?mid=2&pid=11&aid=1338>, viitattu 29.6.2017

Tässä tapauksessa tekstissä oli muita sanoja, jotka tukevat odotusaika-termin tulkintaa, mutta samalla tuli esiin, kuinka helposti oikean tulkinnan tuki voi jäädä puuttumaan, etenkin, jos asia mainitaan lyhyesti monen muun asian joukossa. Vastaavia helposti väärin tulkittavia homonyymejä ovat ”vuori” ja ”koski”.

Toinen konkreettinen, teksteihin sisältynyt esimerkki oli ”saha”. Tekstissä siitä kerrottiin merkityksessä ”sahalaitos”, mutta ontologiahaku saha-termillä toi vastaukseksi vain työkaluun saha liittyviä vastauksia. Tarvitaan lisätulkintaa kertomaan, että hakuun saadut vastaukset eivät ole tässä tapauksessa oikeita.

Aineistojen käsittelyssä tuli eteen myös muita esimerkkejä inhimillisen tulkinnan tarpeellisuudesta: esimerkkinä termi ”työmaaruokailu”. Se oli valittu avainsanaksi ja muoto mukaili tekstin kirjoittajan sanavalintoja. Sitä ei suoraan ollut Finto-ontologiakirjastossa, mutta vastaavasti siellä oli kyllä samaa tarkoittava ”työpaikkaruokailu”. Jos ontologiakirjastoihin ei ole ehditty kerätä kaikkia variaatioita synonyymeistä, voi näiden termien vastaavuus jäädä koneellisessa käsittelyssä huomaamatta.

Ontologiahakujen kohdalla vastaavia esimerkkejä tuli esiin useita: Finto-ontologiakirjastossa oli (lähes) samaa tarkoittava termi, mutta se ei löytynyt, mikäli hakuja tehtiin tekstistä lähtöisin olevalla hakusanalla. Esimerkkejä tällaisista ovat mm. uimapaikka vs. uimaranta, koskimylly vs. vesimylly. Ihmisen tekemää päättelyä tarvittiin myös valitsemaan oikea vaihtoehto haun tuloksista, kun avainsanana oli ”mäenlasku” ja tulokset viittasivat sekä mäkihyppyyn että kelkalla laskemiseen.

Käsitteillä saattaa myös olla todella erilaisia määritelmiä eri ontologioissa, ks. seuraava esimerkki. Kun tein hakuja Finto-palveluun, kiinnitti huomiota, että eri ontologioiden antamat vastaukset samaan termiin saattoivat olla keskenään hyvinkin erilaisia. Esimerkiksi haku termillä ”kanat” antoi seuraavia vaihtoehtoisia vastauksia¹ (suluissa ontologian tunniste):

- yläkäsite: aitokanat, alakäsitteet: kana ja punaviidakkokana (lähde: AFO ja KO-KO)
- johtaa käsitteeseen "kana", jonka yläkäsite: kanat, alakäsitteet: kananpoikaset, kukot, leghorn (lähde: YSA)
- yläkäsitteet: Galliformes, siipikarja (lähde: MeSH/FinMeSH)
- yläkäsite: kanalinnut (lähde: MAO/TAO)
- yläkäsite: kanalinnut, alakäsite: kukot (lähde: TERO)

¹ Tarkistettu 1.7.2018

Jossain vaiheessa aineistojen käsittelyä on tarpeellista tehdä valinta eri ontologioiden välillä: mikä monista vaihtoehtoista edustaa missäkin tapauksessa parasta tulkintaa. Kun eri aineistoja halutaan yhdistää, tarkoituksena on, että ontologialinkityksen avulla voitaisiin varmistaa, että aineistot puhuvat samasta asiasta. Koska saman termin ylä- ja alakäsitteet eroavat ontologiasta riippuen, herää kysymys siitä, onko termiä vastaavien käsitteiden ala sama eri ontologioissa.

Kun sama käsite esiintyy monessa eri ontologiassa, eivät ontologiaviittaukset välttämättä auta yhdistämisessä, ellei ontologioiden välillä ole tehty linkitystä eli siltausta. Fintossa on jo toteutettu ontologioiden välistä linkittämistä, mutta ei vielä täysin kattavasti. Tämä nousi esiin uutena tarpeena ja esitetään työn tulosten joukossa.

Toinen asia, joka liittyy ontologioiden väliseen linkittämiseen, koskee linkityksen olemassaolon tarkistamista. Tällä hetkellä on hyvin työlästä ihmissilmin tarkistaa, mitkä jonkin hakutermien tuottamista ontologiavastineista on sillattu keskenään. Olisi hyvä, jos tämä olisi visualisoitu hakutuloksissa jollain tapaa, esim. väreillä. Visualisointia voisi käyttää helpottamaan myös sen hahmottamista, milloin samalla termillä on tarkistetusti viitattu eri ontologioissa hieman eri käsitteeseen (ts. osoitettaisiin visuaalisesti, että vastaavuus on tarkistettu, mutta tuloksena on, että saman termin merkitys on eri ontologioissa määritelty eri tavoin).

Vastaan tuli myös sellainen tilanne, että ontologiahaku tuotti vastauksen, mutta se ei liittynyt tehtyyn hakuun ollenkaan. Aineistossa oli puhuttu kalastamisesta käytettävistä verkoista ja kirjoittaja oli merkinnyt avainsanan siihen muotoon, jossa se oli tekstissä ("verkkoa"). Kun haku tehtiin tässä muodossa olevalla termillä, saatiin vastaukseksi asioita kuten "nettiadressit" ja "verkkoaineistot". Kumpikaan näistä ei liittynyt kalastamiseen, itse asiassa kalastamiseen liittyviä termejä ei tullut vastauksiin ollenkaan.

Syynä on arvatenkin se, että hakua käsittelevä toiminto ei ota huomioon mahdollisuutta, että hakutermi on partitiivimuodossa ja se pitäisi vaihtaa nominatiivimuotoon. Sen sijaan haku tarjoaa tuloksia, joissa on alkuosana "verkkoa". Fintossa on oletuksena, että hakutermi on joko yksikön tai monikon nominatiivissa. Kuitenkin voisi olla mielekäästä sisällyttää Fintoon termien eri taivutusmuodot¹ (ja siis hakutermien muunnos perusmuotoon).

Tällöin eri palveluissa voitaisiin helpommin tarjota ontologialinkkiä niihin sanoihin, jotka käyttäjä merkitsee avainsanaksi. Ei ehkä ole realistista ajatella, että kaikki palvelut, joihin käyttäjä voi lisätä omia aineistojaan, sisältäisivät suomen kielen sanojen muunnoksen perusmuotoon. Ainakin oman kokeiluni perusteella käyttäjillä on taipumus

¹ Ellei kaikkia taivutusmuotoja, niin ainakin joitain yleisempiä.

merkitä avainsanoja taivutetuissa muodoissa, erityisesti partitiivi vaikuttaa olevan suosittu.

Käsitteiden muuttuminen

Koska tässä työssä käytetyt esimerkit on kirjoitettu tulevaisuuden lukijoille, on erityisen tärkeää mahdollistaa se, että tekstin tulkinta säilyy sellaisena kuin se on kirjoittamishetkellä ollut. Työn tekemisen aikana nousi esiin, miten tärkeää ontologioihin on tehdä näkyväksi niiden muuttumisen historia. Jokaisen käsitteen määrittelyyn pitäisi lisätä myös tietoa siitä, miten termien merkitys on muuttunut aikojen saatossa. Tällöin ontologia auttaisi myös vanhojen aineistojen tulkinnassa.

Tämä asia on onneksi otettu huomioon Finton ontologioiden kehitystyössä. Periaatteeksi on otettu, että vanhentuneeksi käsitteeksi muuttuneita käsitteitä ei poisteta, vaan uusi, korvaava käsite lisätään ontologiaan edellisten rinnalle ja käsitteet merkitään joko synonyymeiksi tai siten, että uudempaan versioon lisätään merkintä ”suositeltava nimitys”. [Seppä2014]

Käsitteille on ontologiakirjastossa usein määritelmä ja rinnakkaiskäsitteitä. Elämäkerran laatija on saattanut käyttää sellaista termiä, joka ei myöhemmin enää ole yleisesti tunnettu, mutta ontologiakirjaston avulla lukija saattaa löytää termille synonyymien, jonka tunnettuus on kantanut pitemmälle tulevaisuuteen. Käsitteiden määritelmien avulla tekstin tai kuvan aihe tulee lukijalle tutuksi monipuolisemmin kuin ilman ontologian apua.

Jatkuva muutos aiheuttaa myös sellaisen ongelman, että vaikka käsitteen merkitys ei sinänsä muuttuisi, muuttuvat ne reaali maailman ilmentymät (ekstensiot), joihin käsitteellä viitataan. Esimerkiksi muistelutekstissä on kuvattu terveystarkastuksia 1950-luvulla ja avainsanaksi merkitään ”terveystarkastus”. Jos tämä aineisto olisi verkossa ja se tulisi hakutulokseksi haettaessa terveystarkastuksen kuvauksia, eivät sisältö ja käytännöt ole ollenkaan samanlaisia kuin tänä päivänä. Tästä herää kysymys, pitäisikö verkossa olevaan aineistoon aina pyrkiä lisäämään joitakin kontekstia määrittäviä ominaisuuksia, joilla hakua voi tarvittaessa tarkentaa.

8.2.4 Tulosten pohjalta syntyneet uudet kysymykset

Seuraavassa kuvataan lyhyesti joitakin kysymyksiä ja uusia tutkimustarpeita, jotka ovat syntyneet edellä kuvattujen huomioiden pohjalta.

Sanaliittojen tulkinta

Edellä, kohdassa 8.2.3, on kuvattu avainsanojen käsittelyn yhteydessä tehtyjä huomioita. Siinä on kirjoitettu merkityseroista, joita syntyy riippuen siitä, kirjoitetaanko sanaparin osat yhteen vai erikseen. Aineistojen käsittelyn yhteydessä pitäisi tulkita, onko kyseessä tapaus, jossa sanapari (tai useamman sanan muodostama ilmaisu) pitäisi kirjoittaa yhteen ja/tai tulkita yhdessä.

Mitä enemmän asiaa kuitenkin pohtii, sitä monisyisemmältä se alkaa näyttää. Miten tekstistä voidaan erottaa potentiaaliset sanaparit, jotka olisi pitänyt kirjoittaa yhteen, mutta jotka on jätetty erilleen? Entä miten tunnistaa sellaiset usean sanan yhdistelmät, joita ei kuulukaan kirjoittaa yhteen, mutta jotka yhdessä muodostavat kuitenkin käsitteen koko merkityksen? Esim. ”savupiipun puhdistus”. Asiaa voisi laajentaa ottamalla aiheeksi yleisesti implisiittisten sisältöjen ja sanomien tulkinnan. Esimerkkinä ilmaisu ”vetää yhtä köyttä”.

Ontologioiden kohdalla olisikin haasteena löytää hyvä rakenne tällaisten käsitteellisten kokonaisuuksien erottamiseen ja niiden merkitysvaihtoehtojen hahmottamiseen. Tavoitteena olisi siis sisällyttää ontologioihin hakutoimintoja, joiden avulla voi tutkia merkityseroja, jotka muuttuvat riippuen siitä, kirjoitetaanko määrätty termit yhteen vai erikseen.

Eksplisiittisen ilmaisun puuttuminen

Usein esiintyy tilanteita, joissa jotain asiaa vastaavaa käsitettä ei eksplisiittisesti mainita tekstissä, mutta se tulee esille toiminnan kuvauksen kautta. Esimerkiksi tekstissä kerrotaan, että pankkivirkailija painaa nappulaa ja kadulla soi hälytyskello. Hän on siis tehnyt hälytyksen, mutta sanaa ”hälytys” ei sellaisenaan ole tekstissä.

Vastaava esimerkki liittyy kohtaan, jossa on puhe vuoden 1963 rahanmuutoksesta. Elämäkertakirjoittajan tekstissä luki: ”Vuonna 1963 jätettiin rahasta kaksi nollaa pois, siinä oli varsinkin vanhemmilla ihmisillä taas paljon oppimista...” Termiä ”rahanmuutos” ei siis esiintynyt tekstissä, mutta kirjoittaja oli laittanut sen avainsanaksi. Toisessa tekstissä oli puolestaan kerrottu paljon erilaisista kalastusvälineistä, mutta ei kertaakaan mainittu sanaa ”kalastaminen” tai ”kalastus”.

Muitakin vastaavia esimerkkejä sisältyi analysoitavina oleviin aineistoihin, esim. luku-taidottomuus ilmaistiin kirjoittamalla ”ei osannu lukea”. Erityisesti silloin, kun teksti kertoi vaikeasta asiasta, esim. surusta, oli tavallista, että asia ilmaistiin rivien välissä. Esim. tekstissä ei mainittu sanaa kuolema tai kuolla, vaan kirjoitettiin: ”seuraavana keväänä häntä ei enää ollut.” Jos tavoitteena on tekstin koneellinen käsittely, ei voida luottaa siihen, että kaikki keskeiset käsitteet on mainittu tekstissä nimeltä.

Monissa palveluissa on avainsanojen vaihtoehtoksi tarjolla mahdollisuus muodostaa aineiston sisällöstä sanapilvi tms., joka esittää tekstissä esiintyvät termit sitä enemmän korostettuina, mitä keskeisempiä ne ovat. Klikkaamalla sanapilvessä olevaa termiä luki- ja löytää ne kohdat, joissa kyseinen termi esiintyy. Ajatuksena on, että sanapilvi esittää samalla myös tiivistelmän aineiston keskeisestä sisällöstä. Olisi mielenkiintoista perehtyä aihetta koskeviin tutkimuksiin ja selvittää, missä määrin sanapilvet oikeasti esittävät keskeistä sisältöä, ts. kuinka relevanttia sisältöä niissä todella on.

Erikoistermien käyttö tuo hankaluuksia

Käytettyjen avainsanojen abstraktiotaso vaihteli. Mikäli teksti oli erikoistunut jonkin asian selvittämiseen niin, että käytetty terminologia oli erikoisalan sanastoa, ei avainsanoja aina ollut ontologiakirjastossa (poikkeuksia toki on). Monta kertaa myös suhteellisen ”tavallisessa” tekstissä huomasi, että tarkkuutta vähentämällä saa enemmän tuloksia ontologioista. Esimerkkejä tällaisista sanoista voisivat olla rantalaidun (laidun on kyllä Finto-kirjastossa, rantalaidun ei), tukkisumat (sumat on ontologiassa, tukkisumat ei) ja joenranta (ranta on ontologiassa, joenranta ei).

Kaiken kaikkiaan ainakin toistaiseksi tuli sellainen vaikutelma, että joitain poikkeuksia lukuunottamatta Finto-ontologiakirjastossa ovat ainoastaan yleisimmät termit. Esim. ”veisaaminen” on esimerkki termistä, jota ei ainakaan tätä työtä tehdessä ollut ontologiakirjastossa. Nykytilanteessa täytyisi tekstissä veisaamisen sijaan ilmaisuksi vaihtaa ”virsien laulaminen”, mikäli tekstin jatkokäsittelylle halutaan taata ontologiayhteys. Jos siis tulevaisuudessa olisi tarkoituksena kytkeä tekstin sisältö ontologiamäärittelyihin (ehkä jopa automatisoidusti), täytyisi ontologioita laajentaa vielä runsaasti.

Käyttäjä voisi saada tähän tilanteeseen apua myös toiminnosta, joka tuloksettoman yhdyssanahaun jatkona ehdottaisi hakutuloksia, jotka on tehty yhdyssanojen puolikkailla tai osilla. Ontologioiden hakutoimintoihin olisi hyvä saada tällainen ”jatkohaku”. Käyttäjän täytyisi tällöin itse valita, onko jatkoahan tarjoamista vaihtoehtoista hyötyä hänen tapauksessaan.

8.3 Tulosten rajoitukset

Tehty tutkimus on pieni, tehty muun työn ohessa lähes kokonaan vapaa-ajalla ja se on ensisijaisesti opinnäytetyö, joten tieteellisenä tutkimuksena se on hyvin rajoittunut. Vastaavan tutkimuksen toteuttaminen laajemmassa mittakaavassa olisi edellyttänyt jo erilaisia resursseja. Siihen nähden pienikin aineisto oli yllättävän antoisa. Se toi esiin mo-

nia asioita, joita ei pelkän teoreettisen spekuloinnin avulla olisi tullut ajatelleeksi. Pie-nimuotoisuuden lisäksi tuloksiin liittyy myös muitakin rajoituksia, joita tarkastelen seuraavassa.

Vaikka tarkasteltavia aineistoja on useita, ovat ne kaikki syntyneet saman ryhmän jäsen-ten toimesta ja saman vetäjän ohjauksessa. Jos kyseinen työskentelyjärjestely siirrettäisiin toiseen ympäristöön, pysyisivätkö tulokset samoina? Kuinka yleistettäviä tulokset ovat?

Tutkimus on tehty määrätyn aikavälin sisällä, ja Finto-ontologiakirjaston kattavuuden osalta tulokset ovat aikaan sidottuja, sillä kirjaston kokoelmat kehittyvät kaiken aikaa. Joka kuukausi sisältöön tulee uusia käsitteitä, joten toistettuna tämän tutkimuksen tulokset eivät olisi määrällisten analyysien osalta täysin samoja. Mielestäni tärkeintä on kuitenkin tulkita tuloksia vain suurien suuntaviivojen osalta, eksaktit lukumäärät eivät ole ratkaisevia. Aineiston määrällisen käsittelyn osalta on myös voimakkaasti tuotu esiin näin pienen aineiston riittämätön todistusvoima.

Yleistettävyyden oikeutetusta kyseenalaistamisesta huolimatta täytyy muistuttaa, että työn suurin anti ovat ne laadulliset seikat, jotka tulivat konkreettisesti esiin tehtyjen ko-keilujen yhteydessä. Jo niiden tiedostaminen vie ontologioiden ja avainsanojen käyttöä ainakin pienen askeleen eteenpäin.

9 Lopuksi

Seuraavassa esitetään hyvin tiivistetty yhteenvedo tehdystä tutkimuksesta (kohta 9.1) ja saaduista tuloksista (kohta 9.2). Lopuksi kohdassa 9.3 esitetään näiden pohjalta nousseita uusia tutkimusongelmia.

9.1 Yhteenvedo tutkimuksen etenemisestä

Tässä tutkimuksessa on Finto-ontologiakirjaston ja elämäkerta-aineistojen avulla haettu vastauksia kysymykseen, miten yleiset ontologiat ja käyttäjien itse muodostamat folksonomiat suhteutuvat toisiinsa. Onko Finto-ontologiakirjaston ontologioilla mahdollista jäsentää subjektiivisia, elämäkerta kartoittavia aineistoja? Kun käyttäjät jäsentävät omia elämäkertatekstejään, sisältyvätkö heidän avainsanansa Finton ontologioihin? Onko tutkijan samoista aineistoista valitsemia avainsanoja Fintossa? Mitä muita seikkoja nousee esiin ontologioiden sisältöä tutkittaessa?

Tutkimus toteutettiin kansalaisopiston elämäkertapiirin työskentelyn yhteydessä. Osallistujista viisi vapaaehtoista antoi katkelman omista elämäkertateksteistään, ja tutkija määritteli niille omasta näkökulmastaan sopivat avainsanat. Myös tekstien kirjoittajat olivat itse merkinneet tai kirjoittaneet, mitkä sanat he valitsevat tekstinsä avainsanoiksi.

Näillä avainsanalistoilla tehtiin hakuja internetissä olevaan Finto-ontologiakirjastoon ja näin selvitettiin, kuinka suuri osa avainsanoista sisältyi julkisiin ontologioihin. Samalla verrattiin myös tutkijan ja kirjoittajien tekemiä sanalistoja toisiinsa. Samoin selvitettiin, onko mitään eroja sen suhteen, kuinka suuri osuus sanoista sisältyy yleisesti saatavilla oleviin ontologioihin. Työskentelyn aikana kirjattiin ylös myös muita asiaan liittyviä huomioita ja niistä esitettiin lopuksi yhteenvedo.

9.2 Saadut tulokset

Työskentelyn aikana tehtiin seuraavia havaintoja.

Ei havaittu selkeää signaalia siitä, että käyttäjien määrittelemiä avainsanoja löytyisi hyvin Finto-palvelun ontologioista. Karkeasti ottaen noin puolet avainsanoista oli Fintos-

sa. Havaittiin myös lievä painotus, joka tukee olettamusta, että ontologiat sisältävät todennäköisemmin it-ammattilaisen valitsemia termejä. Tämä havainto oli yhdensuuntainen lähdeaineiston kanssa, mutta tulokseen tulee suhtautua suurella kriittisyydellä, sillä mittaaminen tapahtui niin pienellä aineistolla, että sillä ei ole todistusvoimaa.

Työn varsinaiset tulokset keskittyvät aineiston laadullisen käsittelyn puolelle. Sen yhteydessä voitiin kirjata seuraavat seikat:

- Avainsanojen valintaan vaikuttavat merkitsemiselle asetetut tavoitteet. Oli esim. eri asia merkitä avainsanoja nykypäivän kuin tulevaisuuden lukijoita varten.
- Avainsanojen merkitsemisen käytännöt kehittyvät vähitellen ja voivat muuttua työskentelyn edetessä. Eri aikoina tehdyt avainsanavalinnat on siis tehty eri ajatusmallien pohjalta.
- Käyttäjillä on eroja sen suhteen, kuinka paljon he merkitsevät avainsanoja: osa merkitsi runsaasti, osa vain muutamia.
- Käyttäjillä on erilaisia käytäntöjä esim. suhteessa sanojen merkitsemistapaan ja siihen muotoon, jossa sanat merkitään (yksikkö/monikko/eri taivutusmuodot). On hankala lähteä ohjeistamaan kaikkia webin käyttäjiä, sen sijaan on tärkeää huolehtia, että termit olisi mahdollista löytää ontologioista hyvin monessa eri taivutusmuodossa. Sen johdosta suositellaan, että ontologioihin sisällytettäisiin termit mahdollisuuksien mukaan kaikissa eri taivutusmuodoissa (tai ontologiapalvelussa olisi valinnaisena käytettävissä toiminto, joka tekee muunnoksen perusmuotoon).
- Erisnimiä tarvitaan ontologioihin lisää, jotta paikkoja ja henkilöitä koskevia tietoja ja aineistoja voidaan yhdistää.
- Sama käsite esiintyy useassa eri ontologiassa, mutta niillä on eri yläkäsitteitä ja eri alakäsitteitä. Tarvittaisiin nykyistä kattavampaa ontologioiden välistä linkittämistä eli siltaamista, jotta ontologiat todella auttaisivat aineistojen semanttisen yhteensopivuuden varmistamisessa. Ontologiakäsitteiden keskinäinen vastavuus/vastaamattomuus olisi hyvä esittää myös visuaalisesti.
- Murteet ja puhekieli ylipäättään vaikuttavat paljon käyttäjien ilmaisuihin. Tähän tulisi vastata sisällyttämällä ontologioihin myös murrelähttöisiä synonyymejä ja puhekielen ilmaisuja.
- Yhdyssanojen ja sanaliittojen tulkinta on vaikeaa huolimatta siitä, että koneellinen tekstin analysointi kehittyi myös koko ajan. Vaikeutta lisää omalta osaltaan käyttäjien tekemien yhdyssanavirheiden yleistyminen. Muutenkin tuli esiin esimerkkejä, joissa osoittautui, että inhimillinen tulkinta on edelleen monessa tilanteessa tärkeää.

- Vanhojen aineistojen tulkinnan kannalta on tärkeää säilyttää ontologioissa käsitteiden muutoshistoria.
- Mitä yleisemmän tason termi oli kyseessä, sitä todennäköisemmin se sisältyi Finto-ontologiakirjaston ontologioihin.
- Kun kyseessä ovat henkilökohtaisessa elämässä vaikeat asiat, niitä vastaavaa termiä ei usein löydy tekstistä. Tällaisiin sisältöihin pääsee parhaiten käsiksi vain avainsanamerkkintöjen avulla.

Yllä mainituista havainnoista osa oli yhteneväisiä lähdekirjallisuuden kanssa, osa oli täysin uusia omia huomioitani. Usealla tavalla nousi esiin se seikka, että suomen kielen yhdyssanojen kirjoitusasun vaikutus semantiikkaan tuottaa koneellisessa käsittelyssä joitakin lisähaasteita. Korostettakoon vielä, että omasta työstäni ei löytynyt sellaisia havaintoja, jotka olisivat olleet ristiriidassa jonkin aikaisemman tutkimuksen tulosten kanssa.

Kun otetaan huomioon, miten pieni aineisto oli kyseessä (viisi erilaista tekstinäytettä viideltä eri kirjoittajalta), oli tutkimus mielestäni yllättävän hedelmällinen. Se toi esille monia sellaisia seikkoja, joita olisi ollut vaikea hahmottaa pelkän teoreettisen päättelyn avulla. Nämä ovat myös asioita, jotka eivät ole kiinni suurten otosten tilastollisesta todistusvoimasta, vaan niitten kohdalla jo pieni aineistomäärä antaa tuloksen, jota voidaan yleistää.

Työssä on saatu uusia tuloksia, jotka ovat pääasiassa käytännöllisiä. Tehdyt uudet huomiot kertovat ontologioiden käytettävyydestä ja niihin sisältyy myös joitakin kehittämissuhteita. Tieteellisten ansioiden osalta työn anti keskittyy lähinnä aikaisempien tutkimusten tulosten testaamiseen. Niiltä osin ei ole mitään kiistettävää, oma tutkimukseni myötäilee hyvin muiden saamia tuloksia.

9.3 Uudet tutkimusongelmat

Seuraavassa esitetään tämän työn pohjalta esiin nousseita jatkotutkimuksen aiheita.

Kansalaisten itse tuottamien aineistojen hyödyntäminen

Kiinnostava jatkotutkimuksen aihe liittyy ontologioiden käyttöön, mikäli haluttaisiin yhdistää ”virallisia” terveystietoja kansalaisen omiin digitaalisiin (elämäkerta-)arkistoihin ja toisinpäin. Jo nyt kansalaiset voivat hakea omia tietojaan joistakin julkisista

tietokannoista, ja tulevaisuudessa tämä lisääntynee entisestään. Voisiko näihin palveluntarjoajien ja viranomaisten ylläpitämiin tietoihin liittää avainsanoja, joiden avulla niihin voisi yhdistää omien tietovarantojen sisältöä?

Esimerkiksi terveystiedoissa olisi avainsanoja, jotka viittaisivat ravitsemukseen, uneen, kotona tehtävien mittausten arvoihin jne. Kansalainen voi tallettaa itse näihin aiheisiin liittyviä tietoja ja merkintöjä ja liittää niihin kyseisiä avainsanoja. Näin ohjelma löytäisi henkilön omat merkinnät ja osaisi yhdistää ne asianmukaisesti ”virallisiin” terveystietoihin.

Olisi myös kiinnostavaa tutkia ontologioiden hyötyjä, kun halutaan laatia elämäkertaineisto muistisairaitten potilaitten hoitoa varten. Miten ontologinen jäsentely auttaa erityisesti silloin, kun tietojen jatkokäyttäjinä ovat esim. terveystietopalveluiden tuottajat? Myös näitä tilanteita varten voisi olla lista toivottavista avainsanoista, joilla potilaitten omia tietoarkistoja jäsenettäisiin. Tarkoituksena olisi hakea avainsanojen avulla aina kulloinkin kiinnostuksen kohteena olevaa asiaa.

Vastaavan kokeilun toteuttaminen eri aineistolla

Tutkimuksessa saatiin tietoa elämäkertatietojen jäsentämisestä. Tämän työn perusteella voimme alustavasti päätellä, miten hyvin käyttäjien laatimat folksonomiat ja asiantuntijoiden kokoamat ontologiat kohtaavat. Jää epäselväksi, ovatko tulokset laajennettavissa myös muita aihepiirejä käsitteleviin teksteihin. Tämän tarkistamiseksi vastaava kokeilu täytyisi toteuttaa muunlaisilla aineistoilla.

Yksi mahdollinen tutkimuskohde olisi toteuttaa vastaava vertailu myös nykyhetkessä eläville blogikirjoituksille, joissa ei erityisesti ole tarkoituksena säilöä arjen kuvausta tuleville polville. Niissä ovat yhtäläillä kyseessä yleensä harrastajakirjoittajien tekemät tekstit, joita tekijä jäsentää avainsanoilla ilman erityistä opastusta siihen. Nämä tekstit sijaitsevat internetissä ja niitä voidaan hakea avainsanojen avulla. Vertailua varten valittaisiin satunnaisesti esim. viisi blogia, joiden kirjoittaja on jäsentänyt tekstejään avainsanoilla. Sen jälkeen toteutettaisiin vastaavanlainen vertailu folksonomiajäsenysten ja ontologioiden välillä ja tavoitteena olisi verrata, ovat tulokset yhtenevät tämän tutkimuksen tulosten kanssa.

Vastaavan kokeilun toteuttaminen suuremmalla aineistomäärällä

Tämän tutkimuksen aineiston määrällisen käsittelyn yhteydessä todettiin, että se on valittavan suppea eikä sen tuloksilla ole validia tilastollista todistusvoimaa. Jotta saisimme paremmin tietoa ontologioiden kattavuudesta suhteessa käyttäjien tuottamiin aineis-

toihin, voitaisiin vastaava vertailu suorittaa suuremmalla aineistolla. Näin saataisiin tilastollisesti pätevää todistetta siitä, missä määrin ontologiat sisältämät käyttäjien itsensä käyttämää käsitteistöä.

Aineistojen laajentaminen semanttisten web-hakujen avulla

Käsitteiden määrittämisen jälkeen voitaisiin selvittää, onko jäsentämisestä hyötyä aineiston laajentamisen kannalta. Elämäkertaa voidaan laajentaa hakemalla semanttisen webin avulla täydentäviä sisältöjä. Hakujen avulla selvitetään, auttavatko avainkäsitteet löytämään internetistä samoja asioita käsitteleviä aineistoja.

Kun ontologinen jäsenitys siis on valmis, voitaisiin hyödyntää semanttisen webin hakuja ja hakea aineiston laatijalle tarjolle sellaisia avainsanoihin liittyviä sisältöjä, joita webissä tällä hetkellä on saatavilla. Tavoitteena on saada lisää tietoa ja aineistoa liittyen niihin asioihin, joita elämäkerrassa on kuvattu. Mikäli elämäkerran kirjoittaja hyväksyy saatuja hakutuloksia ja niitten kautta löydettyjä sisältöjä osaksi omaa aineistoaan tai katsoo niiden hyödyttävän omaa muistelutyötään, voitaisiin se tulkita merkiksi siitä, että haku oli tuottanut positiivisia tuloksia ja ontologinen luokittelu on tältä osin ollut hyödyllistä.

Auttavatko semanttisen webin tekniikat sisällön laajentamisessa ja rikastamisessa siinä määrin, että luokittelu ja hakujen tekeminen kannattavat? Semanttisten hakujen kohdalla ongelmaksi muodostuu se, että internetin web-sivuista suuri osa on talletettu ilman semanttista sisältöä määrittävää jäsenystä. Osa relevantista sisällöstä saattaa jäädä tästä syystä hakutulosten ulkopuolelle, eikä ongelma johdu siitä, että ontologinen luokittelu ei toimisi. Tehtyjä avainsanajäsennyksiä vastaavia sisältöjä ei vain yksinkertaisesti ole mahdollista löytää semanttisten hakutyökalujen avulla.

Hakuja tehtäisiin myös sellaisilla folksonomioihin sisältyvillä käsitteillä, joita ei ole Finto-palvelun ontologioissa, jolloin nähtäisiin, toimivatko ontologiakirjaston käsitteet paremmin sisältöjen hakutehtävissä kuin sellaiset käsitteet, jotka eivät sisälly yleiseen ontologiaan. Tämä kertoisi siis siitä, kuinka hyvin folksonomiat toimivat haettaessa sisältöjä internetistä.

Web-hakujen tuloksellisuuteen vaikuttaa myös se, kuinka paljon kyseistä hakusanaa vastaavaa materiaalia on internetissä. Toki tarkastelussa tärkein kriteeri onnistumiselle ei ole hakutulosten suuri määrä, vaan ylipäätään se, että onnistutaan yhdistämään oikeasti samaan aiheeseen liittyviä sisältöjä keskenään.

Elämäkertojen rikastamista kohtaan ovat myös muut osoittaneet kiinnostusta. Suomessa on ollut meneillään Semanttiseen kansallisbiografiaan liittyvä tutkimushanke¹, jossa on

¹ <https://seco.cs.aalto.fi/projects/biographies/>

tavoitteena kehittää menetelmiä ja rakentaa ympäristöä, jossa elämäkertoja voidaan esittää niin, että ne ovat laajennettavissa semanttisen webin teknologioilla. Omasta tutkimusideastani tämä eroaa selkeästi sen suhteen, että elämäkertatekstit eivät ole henkilöiden itsensä tuottamia.

Tutkimuksen lähdeaineistosta myös Van Damme ym. ovat tehneet käyttäjien valitsemien avainkäsitteiden avulla semanttisia hakuja. He ovat kuitenkin tehneet asiat eri järjestyksessä: ensin on tehty semanttisia hakuja, jotta on varmistuttu termin merkityksestä ja sen jälkeen avainsana on pyritty yhdistämään olemassa olevaan ontologiaan. [2007]

Elämäkertaontologian laatiminen

Tämän tutkimuksen perusteella näyttää, että arkielämästä kertovien tekstien sisällöt ja ammattikäyttöön tarkoitetut ontologiat ovat edelleen suhteellisen kaukana toisistaan. Tämä tuottaa tarpeen saada täydennystä nykyisiin ontologioihin, sillä muuten hyvin suuri osa webin sisällöistä jää semanttisten hakujen ulkopuolelle. Onhan yhä suurempi osa sisällöistä käyttäjien itsensä tuottamia sitä mukaa kuin web 2.0 -tyyliset ratkaisut leviävät.

Jatkotyön tavoitteena voisikin olla uusi ontologia, jota voi käyttää elämäkertatietojen jäsentämiseen. Suppeammassa muodossaan tuloksena voisi olla elämäkertatietoja koskeva täydennys johonkin vanhaan ontologiaan. Tällaiseen ontologioiden laatimis- tai täydentämistyöhön ei olisi ollut tämän tutkimuksen resurssien puitteissa mahdollista ryhtyä.

Lähteet

[Angel2007] Angeletou, Sofia, Sabou, Marta, Specia, Lucia and Motta, Enrico, Bridging the gap between folksonomies and the semantic web: an experience report. In: *The 4th European Semantic Web Conference 2007 (ESWC 2007)*, 3-7 Jun 2007, Innsbruck, Austria, 2007.

[Audac2015] Audacity-ohjelman kotisivut, <http://sourceforge.net/projects/audacity/>, luettu 23.6.2015.

[Barsa1983] Barsalou, Lawrence, Ad hoc categories. In *Memory & Cognition*, 11(3), 1983, ss. 211-227.

[Beel2009] Beel, Joeran, Gipp, Bela, Stiller, Jan-Olaf, Information Retrieval on Mind Maps – What could it be good for? In *Proceedings of the 5th international conference on collaborative computing: networking, applications and work sharing (collaborate-com'09)*, ss. 1-4, Washington (USA), Nov 2009.

[Berne2001] Berners-Lee, Tim, Hendler, James, and Lassila, Ora, The semantic web. In *Scientific American* 284, no. 5, May 1, 2001, ss. 28-37.

[Bratt2007] Bratt, Steve, Semantic Web and Other W3C Technologies to Watch, <http://www.w3.org/2007/Talks/0130-sb-W3CTechSemWeb/#%282%29>, luettu 18.11.2012.

[Braun2007] Braun, Simone, Schmidt, Andreas and Walter, Andreas, Ontology Matur-ing: A Collaborative Web 2.0 Approach to Ontology Engineering. In *Proceedings of the Workshop on Social and Collaborative Construction of Structured Knowledge (CKC) at the 16th International World Wide Web Conference (WWW 2007)*, Banff, Alberta, Canada, 2007.

[Bunge1998] Bunge, Mario, *Philosophy of science: from problem to theory*. Vol. 1. Transaction Publishers, 1998. (Uudelleenjulkaisu teoksesta Bunge, Mario, *Scientific research. 1. The search for system*, Springer, 1967.)

[D'Aqu2012] D'Aquin, Mathieu, Noy, Natalya F., Where to publish and find ontologies? A survey of ontology libraries. *Web Semantics: Science, Services and Agents on the World Wide Web* 11, 2012, ss. 96-111.

[Davis2000] Davis, Gordon B., Information systems conceptual foundations: looking backward and forward. In Baskerville, Richard, Stage, Jan and Degross, Janice I. (eds.) *Organizational and social perspectives on information technology: IFIP TC8 WG8.2 International Working Conference*, Springer US, 2000, ss. 61-82.

[Dubli2013] Dublin Core Metadata Initiative <http://dublincore.org/>, viitattu 7.8.2013.

[Dumas1999] Dumas, Joseph S., Redish, Janice, *A practical guide to usability testing*. Intellect Books, 1999.

[Echar2008] Echarte, F., Astrain, J.J., Cordoba, A., Villadangos, J., Self-adaptation of Ontologies to Folksonomies in Semantic Web. *World Academy of Science, Engineering and Technology* 43, 2008, ss. 325-331.

- [Eisen1989] Eisenhardt, K.M., Building theories from case study research, *Academy of Management Review* 14, No. 4, 1989, ss. 532-550.
- [Fense2001] Fensel, Dieter. *Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce*, Springer Berlin Heidelberg, 2001.
- [Finge2006-1] Fingerroos, Outi ja Haanpää, Riina, Muistitietotutkimuksen ydinkysymyksiä. Teoksessa Fingerroos, Outi, Haanpää, Riina, Heimo, Anne ja Peltonen, Ulla-Maija (toim.) *Muistitietotutkimus: metodologisia kysymyksiä*, Suomalaisen Kirjallisuuden Seura, 2006.
- [Finge2006-2] Fingerroos, Outi ja Peltonen, Ulla-Maija, Muistitieto ja tutkimus. Teoksessa Fingerroos, Outi, Haanpää, Riina, Heimo, Anne ja Peltonen, Ulla-Maija (toim.) *Muistitietotutkimus: metodologisia kysymyksiä*, Suomalaisen Kirjallisuuden Seura, 2006.
- [Finto2014] Suomalainen asiasanasto- ja ontologiapalvelu Finto, <http://finto.fi/fi/>, viitattu 25.6.2014.
- [Flyvb2011] Flyvbjerg, B., Case study, in Denzin and Lincoln (Eds.), *The Sage handbook on qualitative research*, Sage, Thousand Oaks, 2011, ss. 301-316.
- [Genda2007] Gendarmi, Domenico, Abbattista, Fabio and Lanubile, Filippo, Fostering Knowledge Evolution through Community-based Participation. In *Proceedings of the Workshop on Social and Collaborative Construction of Structured Knowledge (CKC) at the 16th International World Wide Web Conference (WWW 2007)*, Banff, Canada, 2007.
- [Gibso2007] Gibson, Andrew, Wolstencroft, Katy and Stevens, Robert, Promotion of Ontological Comprehension: Exposing Terms and Metadata with Web 2.0. In *Proceedings of the Workshop on Social and Collaborative Construction of Structured Knowledge (CKC 2007) at the 16th International World Wide Web Conference (WWW2007)* Banff, Canada, May 8, 2007.
- [Giunc2009] Giunchiglia, Fausto and Ilya Zaihrayeu, Lightweight ontologies. In *Encyclopedia of Database Systems*. Springer US, 2009, ss. 1613-1619.
- [Golde1993] Golden-Biddle, Karen, Locke, Karen, Appealing work: An investigation of how ethnographic texts convince, *Organization Science* 4, no. 4, 1993, ss. 595-616.
- [Golde2006] Golder, Scott A., Huberman, Bernardo A., Usage patterns of collaborative tagging systems (myös löydettävissä internetissä otsikolla "The Structure of Collaborative Tagging Systems"). *Journal of Information Science* 32.2, 2006, ss. 198-208.
- [GoodB2007] Good, B., Kawas, Edward and Wilkinson, Mark, Bridging the gap between social tagging and semantic annotation: ED the Entity Descriptor. In *Nature Precedings*, 2007. <http://www.researchgate.net/publication/36789076>
- [Grube1993] Gruber, Thomas R., A Translation Approach to Portable Ontology Specifications. In *Knowledge Acquisition* 5(2), 1993, ss. 199-220.
- [Grube2007] Gruber, Thomas, Ontology of Folksonomy: A Mash-up of Apples and Oranges. *International Journal on Semantic Web & Information Systems*, 3(2), 2007.
- [Grube1995] Gruber, Thomas, Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *International Journal Human-Computer Studies* Vol. 43, Issues 5-6, November, 1995, ss. 907-928.

[Grube2008] Gruber, Tom, Collective knowledge systems: Where the social web meets the semantic web. *Web semantics: science, services and agents on the World Wide Web* 6.1, 2008, ss. 4-13.

[Guari1998] Guarino, Nicola. Formal ontology in information systems. In *Proceedings of the first international conference (FOIS'98)*, June 6-8, Trento, Italy. Vol. 46. IOS press, 1998.

[Guari1995] Guarino, Nicola and Giaretta, Pierdaniele, Ontologies and Knowledge Bases: Towards a terminological Clarification. In N.J.I. Mars (ed.), *Towards Very Large Knowledge Bases*, IOS Press, 1995, ss. 25-32.

[Halpi2007] Halpin, Harry, Robu, Valentin and Shepherd, Hana, The complex dynamics of collaborative tagging. In *Proceedings of the 16th international conference on World Wide Web*, vol. 21, 2007, ss. 1-220.

[HeppM2007] Hepp, Martin. "Possible ontologies: How reality constrains the development of relevant ontologies." *Internet Computing, IEEE* 11.1, 2007, ss. 90-96.

[Hotho2006] Hotho, Andreas, Jäschke, Robert, Schmitz, Christoph and Stumme, Gerd, Information retrieval in folksonomies: Search and ranking. In *The semantic web: research and applications*, Springer Berlin Heidelberg, 2006, ss. 411-426.

[Huynh2008] Huynh-Kim-Bang, Benjamin, Dané, Eric and Grandbastien, Monique, Merging semantic and participative approaches for organising teachers' documents. In *Proceedings of ED-MEDIA 08: World Conference on Educational Multimedia, Hypermedia & Telecommunications*, 2008.

[Hyvön2018] Hyvönen, Eero, *Semanttinen web – linkitetyn avoimen tiedon käsikirja*, Gaudeamus, 2018.

[Hyvön2007] Hyvönen, Eero, Viljanen, Kim, Mäkelä, Eetu, Kauppinen, Tomi, Ruotsalo, Tuukka, Valkeapää, Onni, Seppälä, Katri, Suominen, Osma, Alm, Olli, Lindroos, Robin, Käsälä, Teppo, Henriksson, Riikka, Frosterus, Matias, Tuominen, Jouni, Sinkkilä, Reetta and Kurki, Jussi, Elements of a National Semantic Web Infrastructure - Case Study Finland on the Semantic Web. *Proceedings of the First International Semantic Computing Conference (IEEE ICSC 2007)*, Irvine, California, September, 2007. IEEE Press, 2007.

[Infim2015] Informaatiotutkimuksen ja interaktiivisen median laitos, Sanojen monimerkityksisyys - homonymia ja polysemia, Tampereen yliopisto, http://www.sis.uta.fi/infim/infim_2011/informaatiotutkimus/kurssisivut/a34/osajarjestelmat/semantiikka/homonymiaJaPolysemia.html, luettu 22.6.2015.

[Jaspe1999] Jasper, Robert, Uschold, Mike, A framework for understanding and classifying ontology applications. In *Proceedings 12th Int. Workshop on Knowledge Acquisition, Modelling and Management KAW*. Vol. 99. 1999.

[Järvi2011] Järvinen, Pertti ja Järvinen, Annikki, *Tutkimustyön metodeista*, Opinpajan kirja, Tampere, 2011.

[Järvi2012] Järvinen, Pertti, *On boundaries between field experiment, action research and design research*, Reports in Information Sciences 14, School of Information Sciences, University of Tampere, 2012.

[Jäsch2008] Jäschke, R., Hotho, A., Schmitz, C., Ganter, B. and Stumme, G., Discovering shared conceptualizations in folksonomies. In *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(1), 2008, ss. 38-53.

[Kalel2006] Kalela, Jorma, Muistitiedon näkökulma historiaan. Teoksessa Fingerroos, Outi, Haanpää, Riina, Heimo, Anne ja Peltonen, Ulla-Maija (toim.) *Muistitietotutkimus: metodologisia kysymyksiä*, Suomalaisen Kirjallisuuden Seura, 2006.

[Kieli2017] Kielitoimiston sanakirja. <http://www.kielitoimistonsanakirja.fi/>, viitattu 26.6.2017.

[KimYa2007] Kim, Hak-Lae, Yang, Sung-Kwon, Song, Seung-Jae, Breslin, John G., Kim, Hong-Gee, Tag Mediated Society with SCOT Ontology. In *Proceedings of the 5th Semantic Web Challenge at ISWC*, 2007.

[Knerr2006] Knerr, Torben, Tagging ontology - towards a common ontology for folksonomies, 2006. <https://tagont.googlecode.com/files/TagOntPaper.pdf> luettu 24.6.2014.

[Kompo2013] KompoZer-ohjelman web-sivut, <http://kompozer.net/>, luettu 17.2.2013.

[Korkia2006] Korkiakangas, Pirjo, Etnologisia näkökulmia muistiin ja muisteluun. Teoksessa Fingerroos, Outi, Haanpää, Riina, Heimo, Anne ja Peltonen, Ulla-Maija (toim.) *Muistitietotutkimus: metodologisia kysymyksiä*, Suomalaisen Kirjallisuuden Seura, 2006.

[Kotis2006] Kotis, Konstantinos and Vouros, George A., Human-centered ontology engineering: The HCOME methodology. In *Knowledge and Information Systems* 10.1, 2006, ss. 109-131.

[Kuosm2014] Kuosmanen, Päivi, Pekkarinen, Satu, Norrlin, Tuulikki, Minusta jää jälki - iloa ja voimaa tikutuksesta. Lappeenrannan teknillinen yliopisto, Lahti School of Innovation, Tutkimusraportit - Research reports no. 33, LUT Scientific and Expertise Publications, 2014.

http://www.lut.fi/documents/10633/30059/Minusta+j%C3%A4%C3%A4%20j%C3%A4lki_Tikutuksen+loppuraportti.pdf, viitattu 9.4.2016.

[LeeBa2012] Lee, Allen S., Baskerville, Richard L., Conceptualizing Generalizability: New contributions and a reply, In *MIS Quarterly* vol. 36, no 3, 2012, ss.749-761.

[Limpe2009] Limpens, F., Gandon, F., Buffa, M., *Linking folksonomies and ontologies for supporting knowledge sharing: a state of the art*. Technical report, EU Project, ISICIL, 2009.

[Linke2013] Linked Data. <http://linkeddata.org/>, viitattu 7.8.2013.

[Maedc2001] Maedche, Alexander, Staab, Steffen, Ontology learning for the semantic web, *IEEE Intelligent systems* 16.2, 2001, ss. 72-79.

[Mathe2004] Mathes, Adam, Folksonomies - cooperative classification and communication through shared metadata, In *Computer Mediated Communication* 47, no. 10, 2004.

[Merta2007] Mertanen, Virve, *Tietokirjoittajan käsikirja*, Tampere: Vastapaino, 2007.

[MikaP2011] Mika, Peter, Ontologies Are Us: A unified model of social networks and semantics. In *Web Semantics: Science, Services And Agents On The World Wide Web*, 5(1), 2011.

[Monto2008] Montonen, Timo, *Kirjoita tarinasi*, Otava, Keuruu, 2008.

[Nardi2007] Nardi, D. and Brachman, R. J., An Introduction to Description Logics. In Baader, Franz, Calvanese, Diego, McGuinness, Deborah, Nardi, Daniele and Patel-Schneider, Peter F., eds. *The Description Logic Handbook: Theory, Implementation, and Applications*, 2nd edition, Cambridge University Press, 2007, ss. 5-44.

[Niemi2014] Niemi-Laurila, Tanja, Muistitikut haltuun ja aineistoja keräämään, Mäntsälän Uutiset, 11.4.2014, <http://www.mantsalanuutiset.fi/artikkeli/290084-muistitikut-haltuun-ja-keruu-kayntiin>, luettu 9.4.2016.

[O'Rei2007] O'Reilly, Tim, What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. In *International Journal of Digital Economics* No. 65 (March 2007), 2007, ss. 17-37.

[OWL2012] W3C:n OWL-kielen määrittely. <http://www.w3.org/TR/owl2-overview/>, luettu 1.3.2013.

[Palom1991] Palomäki, Jari, Käsitteen käsite - ontologinen tarkastelu, teoksessa Marjomaa, Esko ja Vadén, Tere (toim.) *Ihmisen tiedonkäsittely, symbolien manipulointi ja konnektionismi*, Filosofisia tutkimuksia Tampereen yliopistosta XXII, Tampereen yliopiston jäljennepalvelu, 1991.

[Passa2008] Passant, Alexandre, Laublet, Philippe, Meaning Of A Tag: A collaborative approach to bridge the gap between tagging and Linked Data. In *Proceedings of the WWW 2008 Workshop Linked Data on the Web (LDOW2008)*, China, 2008.

[Passa2007] Passant, Alexandre, Using Ontologies to Strengthen Folksonomies and Enrich Information Retrieval in Weblogs. In *Proceedings of the First International Conference on Weblogs and Social Media ICWSM*, Boulder, Colorado, United States, March 2007.

[Peffe2007] Peffers, K., Tuunanen, T., Rothenberger, M.A. and Chatterjee, S., A design science research methodology for Information Systems research, *Journal of Management Information Systems* 24, No 3, 2007, ss. 45-77.

[Pelto2006] Peltonen, Ulla-Maija, Muistitieto folkloristiikassa. Teoksessa Fingerroos, Outi, Haanpää, Riina, Heimo, Anne ja Peltonen, Ulla-Maija (toim.) *Muistitietotutkimus: metodologia kysymyksiä*, Suomalaisen Kirjallisuuden Seura, 2006.

[Pelto2013] Peltonen, Ulla-Maija, Muistitiedon määritelmä Perinteentutkimuksen terminologia –verkkoaineistossa, toim. Apo, Satu ja Kinnunen, Eeva-Liisa, Helsingin yliopiston Filosofian, historian, kulttuurin ja taiteiden tutkimuksen laitos, <http://www.helsinki.fi/folkloristiikka/opiskelu/terminologia.htm#muistitieto>, luettu 9.8.2013.

[Picas2015] Picasa-ohjelman verkkosivut, <http://picasa.google.fi/>, luettu 18.6.2015.

[Poppe1979] Popper, Karl Raimund. *Objective knowledge: An evolutionary approach*. Oxford: Clarendon Press, 1979.

- [Porte2006] Portelli, Alessandro, Mikä tekee muistitietotutkimuksesta erityisen? suom. Salla Kivilaakso-Mendes ja Saana Viertomanner. Teoksessa Fingerroos, Outi, Haanpää, Riina, Heimo, Anne ja Peltonen, Ulla-Maija (toim.) *Muistitietotutkimus: metodologisia kysymyksiä*, Suomalaisen Kirjallisuuden Seura, 2006. Alkup. teksti What Makes Oral History Different? In Perks, Robert and Thomson, Alistair (eds.) *The Oral History Reader*, Routledge, 1998.
- [Pöysä2006] Pöysä, Jyrki, Kilpakirjoitukset muistitietotutkimuksessa. Teoksessa Fingerroos, Outi, Haanpää, Riina, Heimo, Anne ja Peltonen, Ulla-Maija (toim.) *Muistitietotutkimus: metodologisia kysymyksiä*, Suomalaisen Kirjallisuuden Seura, 2006.
- [RDF2004] W3C:n RDF-kielen määrittely, <http://www.w3.org/TR/rdf11-primer/>, luettu 1.3.2015.
- [Saart2009] Saarti, Jarmo, Folksonomiat ja kontrolloidut sanastot kaunokirjallisuuden kuvailussa, *Informaatiotutkimus* 28(1), 2009.
- [Sanas2018] Sanastokeskus TSK:n Tietotekniikan termitalkoot - sivusto, <http://www.tsk.fi/tsk/termitalkoot/fi>, viitattu 10.7.2018.
- [Schem2017] Schema.org-sivusto, <https://schema.org/>, viitattu 28.1.2017.
- [Schmi2006] Schmitz, Patrick, Inducing ontology from flickr tags. In *Collaborative Web Tagging Workshop at WWW2006*, Edinburgh, Scotland, vol. 50, 2006.
- [Seddo2012] Seddon, Peter B., Scheepers, Rens, Towards the improved treatment of generalization of knowledge claims in IS research: drawing general conclusions from samples, *European Journal of Information Systems* 21, 2012, ss. 6-21.
- [Seppä2014] Seppälä, Katri ja Hyvönen, Eero, Asiasanaston muuttaminen ontologiaksi, Yleinen suomalainen ontologia esimerkkinä FinnONTO-hankkeen mallista, Raportteja ja selvityksiä 3/2014, Kansalliskirjasto, 2014.
- [Smith2004] Smith, Barry, Ontology. In Floridi, Luciano (ed.), *The Blackwell guide to the philosophy of computing and information*, Blackwell Publishing, ss. 155-166, 2004.
- [Speci2007] Specia, Lucia, Motta, Enrico, Integrating folksonomies with the semantic web. In Franconi, E., Kifer, M., May, W. (eds.) *The Semantic Web: Research and Applications, ESWC 2007*. Lecture Notes in Computer Science, vol 4519. Springer, Berlin, Heidelberg, 2007, ss. 624-639.
- [Spite2005] Spiteri, Louise, Controlled Vocabularies and Folksonomies. Presentation at Canadian Metadata Forum, Ottawa, Ontario; 2005. Available at: <http://www.collectionscanada.gc.ca/obj/014005/f2/014005-05209-e-e.pdf>, luettu 19.6.2014.
- [Spyns2006] Spyns, Peter, De Moor, Aldo, Vandenbussche, Jan and Meersman, Robert, From folksonomies to ontologies: How the twain meet. In *On the move to meaningful internet systems 2006: CoopIS, DOA, GADA, and ODBASE*, Springer Berlin Heidelberg, 2006, ss. 738-755.
- [Stude1998] Studer, R., Benjamins, V. R., Fensel, D., Knowledge Engineering: Principles and Methods. In *IEEE Transactions on Knowledge and Data Engineering* 25(1-2), 1998, ss. 161-197.

- [Swoog2013] Swoogle – Semantic Web Search Engine, <http://swoogle.umbc.edu/>, viitattu 11.8.2013.
- [Tanas2007] Tanasescu, Vlad, Streibel, Olga, Extreme Tagging: Emergent Semantics through the Tagging of Tags. In Chen, Luke Liming et al. (eds.), *International Workshop on Emergent Semantics and Ontology Evolution*, 2007, ss. 84-94.
- [Tesco2008] Tesconi, Maurizio, Ronzano, Francesco, Marchetti, Andrea and Minutoli, Salvatore, Semantify del.icio.us: automatically turn your tags into senses. In *Proceedings of the First Social Data on the Web Workshop SDoW2008*, 2008.
- [Tikut2012] Tikutus-hankkeen web-sivut, <http://tikutushanke.blogspot.com/p/tietoa-hankkeesta.html>, luettu 18.03.2012.
- [Torni2008] Torniai, Carlo, Jovanovic, Jelena, Bateman, Scott, Gasevic, Dragan, Hatala, Marek, Leveraging folksonomies for ontology evolution in e-learning environments. In *Proceedings of the 2008 IEEE International Conference on Semantic Computing*. IEEE Computer Society, 2008, ss. 206-213.
- [Tuoma2006] Tuomaala, Saara, Sukupuolen kokemuksia muistitietohistoriaan. Teoksessa Fingerroos, Outi, Haanpää, Riina, Heimo, Anne ja Peltonen, Ulla-Maija (toim.) *Muistitietotutkimus: metodologisia kysymyksiä*, Suomalaisen Kirjallisuuden Seura, 2006.
- [Ukkon2000] Ukkonen, Taina, Menneisyyden tulkinta kertomalla: muistelupuhe oman historian ja kokemukset kertomusten tuottamisprosessina. Vol. 797. Suomalaisen Kirjallisuuden Seura, 2000.
- [Ukkon2006] Ukkonen, Taina, Yhteistyö, vuorovaikutus ja narratiivisuus muistitietotutkimuksessa. Teoksessa Fingerroos, Outi, Haanpää, Riina, Heimo, Anne ja Peltonen, Ulla-Maija (toim.) *Muistitietotutkimus: metodologisia kysymyksiä*, Suomalaisen Kirjallisuuden Seura, 2006.
- [Uscho1996] Uschold, Mike, Gruninger, Michael, Ontologies: Principles, methods and applications. *Knowledge engineering review* 11, no. 2, 1996, ss. 93-136.
- [Uscho1998] Uschold, Mike et al. "The enterprise ontology." *The knowledge engineering review* 13, no. 1, 1998, ss. 31-89.
- [W3C2017] The World Wide Web Consortium W3C <http://www.w3.org/>, luettu 29.1.2017.
- [VanDa2008] Van Damme, Céline, Coenen, Tanguy, and Vandijck, Eddy, Turning a corporate folksonomy into a lightweight corporate ontology. In *Business Information Systems*, Springer Berlin Heidelberg, 2008.
- [VanDa2007] Van Damme, Céline, Hepp, Martin, Siorpaes, Katharina, Folksonology: An integrated approach for turning folksonomies into ontologies, In *Workshop on Bridging the Gap between Semantic Web and Web 2.0 at WCSW-2007*, 2007, ss. 57-70.
- [Vande2005] Vander Wal, Thomas, *Explaining and Showing Broad and Narrow Folksonomies*, February 21, 2005.
<http://www.vanderwal.net/random/entrysel.php?blog=1635>, luettu 25.2.2013.
- [Vande2007] Vander Wal, Thomas, *Folksonomy*, February 2, 2007.
<http://vanderwal.net/folksonomy.html>, luettu 29.10.2012.

[Weber2008] Weber, Nicolas, The Evolution of Social Ontologies, In *Proceedings of the 3rd European Conference on Technology Enhanced Learning (EC-TEL)*, Maastricht, Netherlands, 2008.

[Welle2008] Weller, Katrin and Peters, Isabella, Seeding, weeding, fertilizing. Different tag gardening activities for folksonomy maintenance and enrichment. In *Proceedings of I-Semantics08, International Conference on Semantic Systems*. Graz, Austria, 2008.

[Veres2006] Veres, Csaba, The language of folksonomies: What tags reveal about user classification. In *Natural language processing and information systems*, Springer Berlin Heidelberg, 2006, ss. 58-69.

[Wierz1984] Wierzbicka, Anna, Apples are not a 'kind of fruit': the semantics of human categorization. In *American Ethnologist*, vol. 11, issue 2, 1984, ss. 313-328.

[Wikip2015] Wikipedia, avoin verkkosanakirja, <https://www.wikipedia.org/>, viitattu 19.6.2015.

[Virkk1951] Virkkunen, Henrik, *Teollisuuden kertakustannukset*. Niiden degressio sekä käsittely kustannuslaskennassa, Liiketaloustieteellisen Tutkimuslaitoksen julkaisuja 13, Helsinki, 1951.

[WordN2013] Wordnet – A Lexical Database for English. <http://wordnet.princeton.edu/>, viitattu 11.8.2013.

[WuZha2006] Wu, Xian, Zhang, Lei and Yu, Yong, Exploring social annotations for the semantic web. In *Proceedings of the 15th international conference on World Wide Web*, ACM, 2006.

Liite 1: Tutkimusaineistot

Kaikkien viiden aineiston avainsanat esitetään kukin omassa taulukossaan. Avainsanat on tallennettu siinä muodossa, kun aineiston laatija on ne itse merkinnyt. Listoista on jätetty pois avainsanaksi määritellyt erisnimet silloin, kun niiden ei voida olettaa olevan yleisesti tunnettuja.

Taulukon oikeanpuolimmaiseen sarakkeeseen on kirjattu tieto termin esiintymisestä Finto-ontologiakirjaston ontologioissa sekä muita hakutuloksiin liittyviä huomioita.

Aineisto 1

Kirjoittajan omista elämänvaiheista kertovasta aineistosta oli käytettävissä kolme tekstinäytettä. Niiden aiheina olivat kirjoittajan lapset, matkat ja työelämä. Aineiston pituus 5 sivua.

Avainsanat oli kirjoitettu käsin tekstisivujen reunoille.

Kirjoittajan valitse- mat avainsanat	Tutkijan valitseman avainsanat	Esiintyminen Finton ontologi- oissa
odotusaika	odotusaika	kyllä, mukana myös homonyymi, joka kuvaa yleistä odottamista
	raskausmyrkytys	kyllä
synnytys	synnytys	kyllä, useita homonyymejä
äitiysloma	äitiysloma	kyllä
Lastenlinna		kyllä
	varhaislapsuus	kyllä
	tukehtumisvaara	ei, tukehtuminen kyllä
Rinnekoti	Rinnekoti	kyllä
kehitysvamma	kehitysvamma	kyllä
CDC	CDC - cri du chat	kyllä
	kromosomi no 5	kyllä
	sosiaalihoitaja	sosiaalihoitajat kyllä
Oivala		ei
	päivätoimintakeskus	ei
Ensimmäinen ulkomaanmatka		ei, kyllä muodossa ”ulkomaanmatkailu”
Västerås	Västerås	ei
	Nordiska musikfesten 1959	ei

	vasemmanpuoleinen liikenne	ei
	korttipakka	ei
	pelikortit	kyllä
	raittiuslautakunta	ei, raittiuslautakuntatyöryhmä kyllä
toinen ulkomaanmatka		ei, kyllä muodossa ”ulkomaan- matkailu”
Leningrad	Leningrad	kyllä
Vaalimaa	Vaalimaa	ei
	Kasetti	kyllä
	Leningrad-hotelli, nykyisin St. Petersburg	ei
	Viipuri	kyllä
	Inturist	ei
	Polttoainekuponki	ei
	Neuvostoliitto	kyllä
	Inkerinmaa	kyllä
	M/S Georg Ots	ei
	Tallinna	kyllä
	Risteilymatka	risteily kyllä
	Viro	kyllä
Ensimmäinen etelän- matka		ei
	Etelänmatka	ei, seuramatka kyllä
Rhodos	Rhodos	kyllä
	Kreetta	kyllä
	Kreikka	kyllä
	Kypros	kyllä
	Bussimatka	ei, bussi kyllä
Moskova	Moskova	kyllä
	Silmäklänikka	ei
	Sukkahousut	kyllä
Sotshi	Sotshi	ei
Siperia	Siperia	kyllä
Venäjän kaukoitä		ei, kaukoitä kyllä
	Junamatka	kyllä
	Lentomatka	kyllä
	Omsk	kyllä
	Irkutsk	ei
	työskentely pankissa	ei, työskentely asiakastiloissa kyllä, pankki/pankit kyllä
	Hälytys	hälytykset kyllä

	Ryöstö	kyllä
	Työmaaruokailu	ei, työpaikkaruokailu kyllä!
	Pinaattikeitto	ei
	Jänis	kyllä
Hollolan säästöpankki	Hollolan säästöpankki	ei
	Lahti	kyllä (Lahti - Lahden kaupunki)
Suur-Lahden säästöpankki		ei
Salpausselän säästöpankki	Salpausselän säästöpankki	ei
	Päijät-Hämeen säästöpankki	ei
Rahanmuutos v.1963		ei
	Asiakkaat	kyllä
	Varallisuus	kyllä
	Ulkonäkö	kyllä
Jaakko Pulli	Jaakko Pulli	ei
Pankkilakko	Pankkilakko	ei
	Rikkuri	ei
25 kpl	61 kpl	38 kyllä, 32 ei

Aineisto 2

Aineisto kuvaa uusimaalaisen tilan vaiheita - sen asukkaita, tapahtumia ja toimintaa. Tekstin pituus 4,5 sivua.

Kirjoittaja oli valinnut avainsanat alleviivaamalla ne siinä muodossa kuin ne olivat tekstissä. Jos kaksi sanaa oli alleviivattu yhtenäisellä viivalla, tulkittiin ne yhdessä yhdeksi avainsanaksi. Mikäli peräkkäiset sanat oli alleviivattu erillisillä viivoilla, kirjattiin kyseiset sanat erikseen.

Kirjoittajan valitsemat avainsanat	Tutkijan valitsemat avainsanat	Esiintyminen Finton ontologioissa
Mustijoki	Mustijoki	kyllä
	Kirveskoski	ei
	Laukkosken kosket	ei
	Pornainen	kyllä
	joki	kyllä
koskea	koski	kyllä (koski, kosket) ei hakutuloksia muodolle ”koskea”
myllyjä	mylly	mylly kyllä, myllyjä ei
sahoja		”saha” kyllä, ”sahoja” ei, saha

		on merkityksessä, jossa tarkoittaa työkalua (yläkäsité ”konkreettinen eloton objekti”), suurempi saha oli ainoastaan nimikkeellä sahalaitokset
sähkölaitoksia		sähkölaitokset kyllä
vuolaat kosket		vuolaat ei, kosket kyllä, vuolasanaa ei tunnistettu missään muodossa
koskimylly		ei, mylly ja vesimylly kyllä
	tukki	tukit kyllä
uitettiin tukkeja		ei
tukinuitosta	tukinuitto	kyllä
uittoyhtiö		ei
	August Eklöf Oy	kyllä muodossa Aug. Eklöf
tukkisumia		ei
	suma	sumat kyllä
tukkilaiset	tukkilainen	tukkilaiset kyllä
pooshakaa, uittohaka, keksi	pooshaka, uittohaka, keksi	kyllä muodossa ”puoshaat, uittohaak, keksit” (Muotoa pooshaka ei tunnettu ollenkaan. Perusmuoto ”uittohaka” ei tuottanut hakutulokseksi sanaa ”uittohaak”.)
kiviarkkuja	kiviarkku	ei
sillan arkkujen		ei, arkut vain säilytykseen liittyvänä objektina
uppotukkeja	uppotukki	ei
tukinuittotyöt		ei
	laiivanrakennus	kyllä
	Hausjärvi	kyllä
	Tolkkinen	ei
	ranta	rannat kyllä
	joenranta	ei
karjan laitumina		ei, karjan kyllä, laitumet kyllä
rantalaidun	rantalaidun	ei, rannat kyllä, laitumet kyllä
maanvyörymän		ei löytynyt muodossa maanvyörymän, maanvyörymät kyllä
riippusillan	riippusilta	ei, riippusillat kyllä (ohjelma ei tunnistanut termiä riippusilta)
	pyykinpesu	kyllä
pyykkirantana		ei
piikkilanka-aita		ei
pyykkipaljuissa	pyykkipalju	ei, pyykkikartut kyllä, paljut kyllä

saaveissa		ei, saavit kyllä
muuripadassa	muuripata	ei, padat kyllä
pyykkinaruja	pyykkinaru	pyykkinaruja ei, pyykkinarut kyllä
pyykkikorissa	pyykkikori	ei
pyykkärit		ei
käsiopohjalla aloitettiin		ei
	uiminen	kyllä
	uimapaikka	ei, uimalat kyllä, uimarannat kyllä
pyykinpesukone		kyllä
	Hoover-pyykinpesukone	Hoover kyllä, pyykinpesukoneet kyllä
karjakeittiössä		ei, karjakeittiöt kyllä
muuripata		ei
virutettavaksi		ei
näkinkengille		ei
	onkiminen	kyllä, erityisesti muodossa ”onginta”
haavia vetämällä		ei
haavikalastukseen	haavikalastus	haavikalastukseen ei, haavikalastus kyllä
verkkohaavi		ei
	kalastaminen	ei, kalastus kyllä
rysillä	rysä	rysillä ei, rysät kyllä
katiskalla	katiska	katiskalla ei, katiskat kyllä
ruuhi		ruuhet kyllä
kirjolohtien	kirjolohtet	ei
kasvatusallas		ei
ravustuksessa, ravustus	ravustaminen	ravustuksessa ei, ravustus kyllä, ravustaminen ei
haavi		haavit kyllä
rapuhaaveja	rapuhaavi	ei
merta	merta	merrat kyllä
nuotion		nuotion ei, nuotio kyllä
sumppuun		ei
pärekoreihin		ei, lähinnä löytyvät päreet ja päretyöt
verannalla		ei löydy, veranta hakusanana tuo tulokseksi kuistit
kutomaan		ei, kutominen → kudonta kyllä
kävällä		ei, kävyt kyllä
verkkoa		ei varsinaisesti: tuo vastaukseksi

		nettiadressit, verkkoaineisto jne.
	mäenlasku	kyllä, sekä mäkihyppy että kelkalla laskeminen
nostettiin jäätä, jäiden nostoa	jäiden nostaminen	ei, nostaminen kyllä yleisessä merkityksessä
	jäiden säilyttäminen	ei, säilyttäminen kyllä
tallinsillan		ei
sahajauhokasassa		ei, sahajauho kyllä, kasa ei
	soutukilpailut	ei, soutu kyllä, kilpailut kyllä (Huom. Sulkavan suursoudut on mainittu tapahtumana, mutta sen luokka on ”liikuntatapahtumat, urheilukilpailut”)
sulkavalaisella		ei, Sulkava kyllä
kilpaveneellä		ei, kilpaveneet kyllä
	sulkavalainen kilpavene	ei
<i>yht. 59 kpl</i>	<i>yht. 44 kpl</i>	<i>46 kyllä, 35 ei</i>

Aineisto 3

Teksti on omaelämäkerrallista kuvausta lapsuuden vaiheista suomalaisessa pikkukaupungissa. Aineiston pituus 7 sivua.

Avainsanat oli merkitty alleiviivaamalla tekstiin, joten ne on kirjattu siinä muodossa kuin ne lauseessa olivat.

Kirjoittajan valitsemat avainsanat	Tutkijan valitsemat avainsanat	Esiintyminen Finton ontologioissa
	äiti	kyllä
	isä	kyllä
	muutto	kyllä
	Sysmä	kyllä
	Heinolan vanhainkoti	ei
	saari	muodossa ”saari” ainoastaan ikonologian ontologiassa, muissa ontologioissa ”saaret”
Koskensaari	Koskensaari	ei
	Heinola	kyllä
	Jyränkö	ei
	Koskensaarenkatu	ei
	Penikkasilta	ei
	talonmies	talonmiehet kyllä

	puusee ¹	ei, lähimmät termit ”käymälät”, ”kuivakäymälät” ja ”kompostikäymälät”
	saaren rakennukset	ei, rakennukset kyllä
	pihapiiri	ei, pihat kyllä
	uiminen	ei, uinti kyllä (löytyi hakemalla sanalla ”uiminen”)
	pyykinpesu	kyllä
koti		kyllä
	kodin keittiö	ei, keittiö kyllä, mutta vain kahdessa ontologiassa: ikonologian ontologiassa ja liiketoimintaontologiassa
	kammari	ei (ei myöskään muodossa ”kamari”)
	pystyuuni	kyllä
	radiopöytä	ei, pöydät kyllä
heteka	heteka	hetekat kyllä
	senkki	ei
vanhainkoti		vanhainkodit kyllä
	vapaa-aika	kyllä
	Lucia-juhla	ei, Lucian päivä kyllä
	naapurit	kyllä
	matka Helsinkiin	ei
	Salpausselän kisat 1958	ei, ”Salpausselän kisat” kyllä
	harrastukset	kyllä
uimakoulua	uimakoulu	uimakoulua ei, uimakoulut kyllä
	uimakandidaatin tutkinto	ei
	hiihtäminen	kyllä
	luisteleminen	ei, luistelu kyllä, ei löytynyt hakusanalla ”luisteleminen”
	voimistelu	kyllä
	tyttökerho	tyttökerhot kyllä
tyttöleirillä		ei
	kesäleiri	kesäleirit kyllä
	leirikaste	ei
	leiripuuhat	ei
vaatetus		kyllä
	vaatteet	kyllä
	ompeleminen	kyllä

¹ Huom kirjoitusasu!

	alusvaatteet	kyllä
	jalkineet	kyllä
	rannekello	kyllä
pikkusiskot ja -veli		ei
	pikkusisko	ei
	tuhkarokko	kyllä
	pikkuveli	ei
	nimen valinta	ei, nimet kyllä
	pulsaattorikone	ei
	kesätyö	kyllä
	lastenhoito	kyllä
	ruuanlaitto	kyllä (muodossa ruoanvalmistus, löytyi termillä ”ruuanlaitto”)
	marjanpoiminta	kyllä
	mieliruuat	ei
	kouluruokailu	kyllä
	vieraat	kyllä
	kyläily	kyllä
kuukautiset	kuukautiset	kyllä
	uusi vanhainkoti	ei, vanhainkodit kyllä
<i>yht. 9 kpl</i>	<i>yht. 58 kpl</i>	<i>37 kyllä, 26 ei</i>

Aineisto 4

Aineisto koostui lyhyistä, murteella kirjoitetuista ja usein humoristisista tarinoista, jotka kertoivat lapsuudesta ja nuoruudesta Pohjanmaalla. Mukana tarkastelussa oli 5 n. puolen sivun mittaista tarinaa.

Avainsanat oli kirjoitettu erikseen tarinoiden perään. Huom! Osa kirjoittajan itse määrittämistä avainsanoista oli murteellisessa muodossa (samoin kuin tarinatkin).

Kirjoittajan valitse- mat avainsanat	Tutkijan valitse- mat avainsanat	Esiintyminen Finton ontologi- oissa
evakko	evakko	kyllä, yleisimmin kuitenkin muodossa ”evakot”
	tupa	kyllä, yleisimmin muodossa ”tuvat”
kamari	kamari	ei, ainoastaan yhteydessä kamarimusiikkiin tms.
mummu		ei, lähin sana ”mummolat”
ahrasmielinen		ei, ei myöskään ”ahdasmielinen”

	piirakka	kyllä, ”piirakat”
	kiiski	kyllä (ei monikkomuodossa!)
	kalakeitto	ei, ”kala” kyllä, samoin ”keitot”
	torakka	kyllä
	lukutaidottomuus	kyllä
	vuokraisäntä	ei
elänlääkäri (huom. kirjoitusasu)	eläinlääkäri	”elänlääkäri” ei, ”eläinlääkärit” kyllä
	lestadiolainen	kyllä ”Lestadiolainen uusherätys”
käsilaukku	käsilaukku	käsilaukut kyllä
virsikirja		virsikirjat kyllä
	virsi laulaminen	ei, virret kyllä, laulaminen kyllä
	tanssit	kyllä, mutta taidemuotona ja tanssiurheiluna, tässä tekstissä kuvasi huvitapahtumaa
	levytanssit	ei, levy ei, äänilevy kyllä, tanssit: ks. edellinen sana
järjestöalo		ei
paperikuu		ei
	Olavi Virta	ei
raurooskaa		ei, raudat kyllä, mutta merkityksessä ”pyydykset”, tekstissä tarkoitti käsirautoja tms.
viuhka		viuhkat kyllä, mutta vain merkityksessä ”asusteet” tai ”aurat”, tekstissä tarkoitti suosiota tansseissa
	kommunistien juhlat	ei, kommunistit kyllä
	Kauhajoen yhteistalo	ei, Kauhajoki kyllä
pääsiäinen		kyllä
rokki		ei, rock kyllä
alushousut	alushousut	kyllä
häpeä		kyllä
	pultut	ei
vahtimestari	vahtimestari	vahtimestarit kyllä
	Aronkylän terveystalo	ei, Aronkylä kyllä, terveystalo ei
	puulämmitys	kyllä
pellit	pellit	kyllä
	häkä	kyllä
pesät		ei, ainoastaan merkityksessä linnunpesät tms. (eläintiede), tulipesät ainoastaan yhteydessä höyrykattiloihin, tekstissä tarkoitti

		tulipesiä
kattanto		ei, ei myöskään katsanto, katse kyllä
<i>yht. 19 kpl</i>	<i>yht. 25 kpl</i>	<i>18 kyllä, 19 ei</i>

Aineisto 5

Aineisto kuvasi kirjoittajan isän elämänvaiheita. Näytteen pituus oli n. 9 sivua.

Avainsanat oli merkitty käsin kirjoittamalla tekstisivujen marginaaleihin.

Kirjoittajan valitse- mat avainsanat	Tutkijan valitse- mat avainsanat	Esiintyminen Finton ontologi- oissa
	karjalaiset	kyllä
	evakko	kyllä
evakkomatka		ei
	Maanhankintalaki	ei
evakkojen asuttaminen		ei, evakot kyllä, asuttaminen ai- noastaan yhdyskuntarakenteen suunnittelun merkityksessä, ei uudelleenasuttamisena
	talon ostaminen	ei, talo antaa tulokseksi ”Talo Thorsbo” -käsitteen, talot tuo vastineeksi rakennuskäsitteen, ostaminen kyllä
	lehmä	kyllä
seurustelu	seurustelu	kyllä
	hää	kyllä
maan osto		kyllä
	pellon raivaaminen	ei, pellonraivaus kyllä
rakentaminen	rakentaminen	kyllä
	asuinrakennus	asuinrakennukset kyllä
tilat kodissa		ei, tilat kyllä, kodissa ei, koti kyllä
	tuvan kalusteet (vai ka- lustus?)	ei, ”tuvan”-hausta tulee tulok- seksi tuvan kieli, tuvat kyllä, ka- lusteet kyllä
kodin kalusteet		ei, kodinsisustus kyllä, kalusteet kyllä, koti kyllä
	kamari	kamarit kyllä
tilapäinen navetta		ei, tilapäinen ainoastaan muun- laisessa yhteydessä, esim. ”tila- päinen hoitovapaa”, navetat kyl-

		lä
	navetta	navetat kyllä
	sauna	kyllä
navetan rakentaminen	navetan rakentaminen	ei, navetan ei, navetat kyllä, rakentaminen kyllä
	kaivo	kyllä, haku muodossa ”kaivot” tuo paljon enemmän osumia, hakemalla muodossa ”kaivo” suurin osa tuloksista määrittää termiä ”kaivokset”
	vesipumppu	ei, vesi kyllä, pumpput kyllä, mutta pumppujen alalajeina vain lämpöpumput ja palopumput
navetan tilat		ei, navetan ei, navetat kyllä, tilat kyllä
	AIV-torni	ei, AIV kyllä, tornit kyllä (samoin rehutornit)
	raskaus	kyllä
lapset syntyvät		ei, lapset kyllä, syntyvät ei, syntyä ei, syntyminen ei, syntymä kyllä
	synnytys	kyllä
poika hukkuu		ei, pojat kyllä, hukkuu ei, hukkuminen kyllä
	hukkuminen	kyllä
	savikuoppa	ei, savikuopat kyllä
	hautajaiset	kyllä
anopin ajatuksia		ei, anopin -> anopinhammas, anopinjakkara jne., ajatuksia ei, ajatus -> Ajatus-niminen yhtiö, ajatukset ei (!?)
	anoppi	anopit kyllä
anoppi ja miniä		yhdistelmänä ei, anopit kyllä ja miniät kyllä
mummo ja tyttären lapset		ei, ainoastaan tyttäret kyllä, lapset kyllä
	mummo	ei, ainoastaan mummolat, Mummon kammari jne.
sukulaisvierailut		ei, sukulaiset kyllä, vierailut kyllä
sairaus		kyllä
	astma	kyllä
kuolema		kyllä
	mummon kuolema	ei, mummo ei, kuolema kyllä
luonteenpiirteitä		ei, luonteenpiirteet kyllä

	Kansanvalistusseuran kirjeopisto	kyllä
	maamieskoulu	maamieskoulut kyllä
	urheilu	kyllä
	puutyöt	kyllä
maanviljely		kyllä
	Maaseudun tulevaisuus	kyllä
	Pellervo	ei, Pellervon julkaisupalvelu kyllä
	AIV-rehu	kyllä
	heinätyöt	ei, heinä kyllä, työt ei (työtahti kyllä, työtaturma kyllä jne.)
lasten kasvatus		kyllä
	sokerijuurikas	kyllä
sähköt saadaan		ei
	sähköt	ei, sähkö kyllä (ei löydy, jos haetaan termillä ”sähköt”)
	lypsykone	lypsykoneet kyllä
traktori	traktori	kyllä
	Zetor	ei
hevosesta luopuminen		ei, hevonen kyllä (yksikkömuodossa!), luopuminen kyllä, mutta merkityksessä ”lopettava muuttaminen”, ei merkityksessä ”omistuksen siirtäminen”, mikä puolestaan löytyy ainoastaan hakusanoilla ”luovuttaminen” ja ”luovutus”
	hevonen	kyllä
kodin remontti		ei, koti kyllä, remontti kyllä
	leivinuuni	leivinuunit kyllä
	vesijohto	vesijohdot kyllä
	viemäri	viemärit kyllä
	keskuslämmitys	kyllä
	David Brown	ei
lato	lato	ladot kyllä
kuivuri	kuivuri	kuivurit kyllä
	puimuri	puimurit kyllä
	paalain	paalain ei, paalaimet kyllä
lehmät	lehmät	kyllä
sika	siat	kyllä
kanat	kanat	kyllä
kasvimaa	kasvimaa	kasvimaat kyllä

puutarha		kyllä, erityisesti muodossa ”puutarhat”
	marjat	kyllä
	mehumaija	mehumaijat kyllä
	omenapuut	kyllä
	leipominen	kyllä
arkiruoka		ei, arki kyllä, ruoka kyllä
	ruoka-ajat	ei, ruoka kyllä, ajat tuo tuloksiksi ajatelmat, ajattelu jne.
	ostokset	kyllä
kauppa-asiat		ei, kauppa kyllä, asiat ei
	myymäläauto	myymäläautot kyllä
vaatteet		kyllä
	vaatteiden ompeleminen	ei, vaatteet kyllä, ompelu (huom! muoto) kyllä
kodinkoneet		kyllä
	pesukone (pulsattori-)	pesukoneet kyllä (pulsattori- ei)
	sähkö	kyllä
	pakastin	pakastimet kyllä
puhelin	puhelin	kyllä
televisio	televisio	kyllä, televisio (viestintäväline) sekä televisiot (vastaotin)
	radio	kyllä, radio (viestintäväline) sekä radiot (vastaotinlaite)
auto	auto	kyllä
työssä jaksaminen		kyllä
	kutominen	kudonta kyllä
	kangaspuut	kyllä
	kulkuri (”kulkumies”)	kulkurit kyllä
vierailut		kyllä
	kylässä käyminen (”kyläreissut”)	ei, kyläily kyllä, käyminen ainoastaan aineenvaihduntaterminä
	ompeluseura	ompeluseurat kyllä
suvun vierailut		ei, suvun ei, vierailut kyllä
	sukulaisvierailut	ei
	perhejuhlat	kyllä
sairaudet		kyllä
	migreeni	kyllä
	verenluovutus	kyllä
metsänhoito	metsänhoito	kyllä
	Kirvu	kyllä
lehdet		kyllä

	Maalaisliitto	kyllä
tyttöjen opiskeluun kannustaminen		ei, tyttöjen ei, opiskeluun ei, kannustaminen kyllä, muodossa "kannustus"
	oppikoulu	kyllä
	sairaanhoitajakoulu	sairaanhoitajakoulutus kyllä
luopuminen tilasta		ei, luopuminen kyllä, mutta merkityksessä "lopettava muuttaminen", ei merkityksessä "omistuksen siirtäminen", tilasta ei, tila ei, tilat kyllä
	sukupolvenvaihdos	kyllä
	muutto	kyllä
omaa aikaa		ei, oma aika ei, joitain oma-alkuisia käsitteitä, kuten oma pääoma, aika kyllä
	eläkevuodet	ei, eläke kyllä, vuodet ei (monnikkomuotoa ei tunneta ollenkaan!), vuosi kyllä
	harrastukset	kyllä
lastenlasten hoito		ei, lapsenlapset kyllä, hoito kyllä, lastenhoito kyllä
	lapsenlapset	kyllä
ukkina		ei, ukki ei, ukit ei
ilo		kyllä
levon kaipuu		ei, lepo kyllä, kaipuu -> kaipaus kyllä
	vanhuus	kyllä
kuolema		kyllä
	Vilhon kuolema	ei, kuolema kyllä
<i>yht. 53 kpl</i>	<i>yht. 90 kpl</i>	<i>81 kyllä, 47 ei</i>