

Jari Salokangas

Moninäkömahaun hyödyntäminen verkkopalveluissa

Tietotekniikan pro gradu -tutkielma

11. marraskuuta 2014

Jyväskylän yliopisto

Tietotekniikan laitos

Tekijä: Jari Salokangas

Yhteystiedot: jari.p.t.salokangas@student.jyu.fi

Ohjaajat: Miika Nurminen ja Anneli Heimbürger

Työn nimi: Moninäkömäärän hyödyntäminen verkkopalveluissa

Title in English: The Utilization of Faceted Navigation in Web Services

Työ: Pro gradu -tutkielma

Suuntautumisvaihtoehto: Ohjelmisto- ja tietoliikennetekniikka

Sivumäärä: 98+0

Tiivistelmä: Tässä tutkimuksessa tutustutaan semanttisen webin maailmaan moninäkömäärän kautta sekä tutkitaan siihen liittyviä taustoja ja teorioita. Moninäkömäärä on eräs hakukäyttöliittymä, jolla käyttäjä voi hakea tietoa tutkivalla otteella. Moninäkömäärä tarjoaa hyvän tuen tavallisten sanahakukoneiden rinnalle, mutta sitä ei ole tuotu kovin näkyvästi esille ihmisten tietoon. Tutkimuksessa otetaan kantaa eri verkkopalveluiden moninäkömäärän sovellutuksiin. Tutkimuksen aikana tutkittiin seitsemää verkkopalvelua, jotka käyttävät moninäkömäärää kokonaan tai osana niiden tarjoamaa verkkopalvelua. Tutkimuksessa ilmeni, että osa sivustoista hyödynsi paremmin moninäkömäärän ominaisuuksia kuin muut, mutta mikään sivusto ei hyödyntänyt maksimaalisesti kaikkia moninäkömäärän tarjoamia ominaisuuksia. Tutkimus antaa tietoa moninäkömäärän soveltamisesta ja sen mahdollisista kehityskohteista.

Avainsanat: Moninäkömäärä, linkitetty data, ontologiat, semanttinen web, tagit, tiedonhaku, verkkopalvelu

Abstract: This research explores the world of semantic web through faceted navigation and discusses backgrounds and theories of it. Faceted navigation is one type of search interface, where user can search information with explorative edge. Faceted navigation offers good support together with traditional keyword-based search engines, but it hasn't yet been brought out very visibly to the people. This research addresses the applications of few web services

that use faceted search. During the research, seven different web services utilizing faceted search as a whole or as a part of its service were studied. In conclusion, some of the websites utilized faceted navigation better than others, but none of the websites utilized fully all the possible functionalities of a faceted navigation. This research gives insight into faceted search applications and how to improve them.

Keywords: Faceted navigation, information retrieval, linked data, ontologies, semantic web, tags, web service

Termiluettelo

Annotointi	metadatan lisäämisen prosessi, tavoitteena aineiston kuvaaminen, luokittelu ja jäsentely systemaattisella tavalla (engl. annotate).
API	rajapinta, joka määrittää ohjelmakomponenttien vuorovaikutuksen komponentin pohjalta tehtäville sovelluksille (engl. Application Programming Interface).
Assosiointi	objektien välille määritetty suhde, mikä sallii yhden objektin käskeä toista objektia tekemään jotakin sen puolesta (engl. associate).
Attribuutti	oliolle määritetty muuttuja, joka tallentaa jonkin olioon liittyvän tietoalkion (engl. attribute).
Data	bittiperäinen tieto, joka voi koostua luvuista, kirjaimista tai kuvista sellaisinaan, ilman erityistä merkitystä.
Entiteetti	olio, joka sisältää myös oliojoukot ja olioiden ominaisuudet (engl. entity).
HTTP	protokolla, jota selaimet ja WWW-palvelimet käyttävät tiedon siirtoon (engl. Hyper Text Transfer Protocol).
JSON	yksinkertainen avoimen standardin tiedostomuoto tiedonvälitykseen (engl. JavaScript Object Notation).
Käsitteellistäminen	ilmiön kuvaaminen käsitteiden avulla (engl. conceptualization).
Linkitetty data	RDF-kuvattua metatietoa, jossa WWW-resurssit kuvaavat asioita ja metatieto sisältää linkkejä (engl. Linked Data).
Linkitetty avoin data	yhteisöllinen projekti, joka tähtää semanttisen datan vapaaseen saatavuuteen (engl. Linked Open Data).
Linkittäminen	kahden tietoteknisen asian liittäminen yhteen hyperlinkillä (engl. linking, interlinking).
Luokka	olio itse tai olion tyyppi.
Metadata	tietoa metadatan kohteena olevasta tietolähteestä.
Mikroformaatit	web-pohjainen lähestymistapa semanttisille merkinnöille.

Moninäkömähaku	tekniikka tutkia tietokokoelmia käyttäen datajoukon metadatatista koostuvia näkymiä haun suodattajina (engl. faceted navigation).
Näkymä	tapa luokitella tuloksia / luokittelun tulos (engl. facet, view).
Näkymäluokittelu	organisoiva järjestelmä, joka sallii monta piirrettä tai attribuuttia jokaiselle objektille (engl. faceted classification).
Objekti	tietolähde.
Ontologia	tiedon esitys aihealueesta formaalissa käsittehierarkiassa, käyttäen jaettua sanastoa merkitsemään tyypit, ominaisuudet ja sisäiset suhteet käsitteille (engl. ontology).
OWL	kieli kuvaamaan ja jakamaan ontologioita WWW:ssä (engl. Web Ontology Language).
Ominaisuus	RDF-luokka, joka sisältää aihealueen ja arvojoukon (engl. property).
RDF	virallinen suositus semanttisen webin datamalleille (engl. Resource Description Framework).
RDFa	lisää joukon attribuuttitasoisia laajennuksia HTML, XHTML ja XML-tyyppisille dokumenteille mahdollistaen RDF-tyyliset kuvaukset osana web-dokumentteja (engl. Resource Description Framework in Attributes).
RDFS	tietämyksen esityskieli, jota käytetään RDF-kehyksessä (engl. Resource Description Framework Schema).
Resurssi	fyysinen tai virtuaalinen tietovaranto, jolla on rajoittunut saatavuus (engl. resource).
Semanttinen web	laajennos nykyiselle webille, mikä lisää semantiikkaa webmuotoiseen dataan (engl. Semantic web).
SPARQL	RDF-kyselykieli, jolla haetaan RDF-formaatissa olevaa tietoa (engl. SPARQL Protocol and RDF Query Language).
Tagi	avainsana tai -termi, joka voidaan liittää johonkin tietolähteeseen.
URI	merkkijono, jota käytetään tunnistamaan nimi tai resurssi (engl.

	Uniform Resource Identifier).
URL	merkkijono, joka sisältää viittaukseen resurssiin (engl. Uniform Resource Locator).
Web 2.0	WWW:stä käytetty nimitys, joka sisältää sivustoja, jotka käyttävät staattisista sivustoista paranneltuja teknologioita.
Web 3.0	WWW:stä käytetty nimitys, joka sisältää sivustoja, joissa on tietoa, jota koneetkin ymmärtävät.
Web Of Data	WWW:n laajennos, jossa verkkoa käytetään data-aineistojen jakamiseen linkitettyä dataa hyödyntäen. Synonyymi linkitettylle datalle.
XML	merkkäuskieli, joka kuvaa joukon sääntöjä koodaamaan dokumentteja formaatissa, joka on ihmis- ja koneluettava (engl. Extensible Markup Language).

Kuviot

Kuvio 1. Esimerkkikoodia RDF, SPARQL ja OWL-kielistä.	6
Kuvio 2. Esimerkki tageista. Kuvassa tagipilvi BibSonomy-sivustossa.	7
Kuvio 3. Esimerkki ontologiasta. (Noy ja Hafner 1997, s. 57).....	10
Kuvio 4. Esimerkki aloittelevan käyttäjän personomiasta Flickr-sivustossa.	14
Kuvio 5. Semanttisen webin teknologioita sisältävä kaavio. (Horrocks et al. 2005)	21
Kuvio 6. Esimerkki RDF-tripletistä. (Bizer, Heath ja Berners-Lee 2009, s. 4).....	26
Kuvio 7. Kuva linkitetystä data-avaruudesta. (<i>Linking Open Data Cloud Diagram</i> 2014) .	27
Kuvio 8. MediaFaces-järjestelmän järjestelmäarkkitehtuuri. (Zwol et al. 2010, s. 965)	33
Kuvio 9. Esimerkki näkymistä ja hakukategorioista Finna- ja Gigantti-sivustoissa.	35
Kuvio 10. Esimerkki moninäkömähakukäyttöliittymästä JYX-sivustossa.	42
Kuvio 11. Kuvankaappaus Finna-sivuston etusivusta.	56
Kuvio 12. Kuvankaappaus JYX-sivuston etusivusta.	57
Kuvio 13. Kuvankaappaus Yummly-sivuston etusivusta.	58
Kuvio 14. Kuvankaappaus Kulttuurisampo-sivuston etusivusta.	59
Kuvio 15. Kuvankaappaus Huuto-sivuston etusivusta.	60
Kuvio 16. Kuvankaappaus Gigantti-sivuston etusivusta.	61
Kuvio 17. Kuvankaappaus BibSonomy-sivuston etusivusta.	62
Kuvio 18. Kuva tutkittavien parametrien viitekehuksesta.	65
Kuvio 19. Esimerkit JYX, Huuto ja Bibsonomy-sivustojen hakuobjekteista.	70
Kuvio 20. Esimerkit Finna, Yummly, Gigantti ja Kulttuurisampo-sivustojen hakuobjekteista.	70

Taulukot

Taulukko 1. Tagityypit kootusti.	17
Taulukko 2. Tutkittavien sivustojen perustiedot.	66
Taulukko 3. Käyttöliittymän toiminnan tutkitut parametrit.	67
Taulukko 4. Käyttöliittymän näkymien tutkitut parametrit.	68
Taulukko 5. Käyttöliittymän hakuobjektin tutkitut parametrit.....	69
Taulukko 6. Käyttöliittymän hakemisprosessin tutkitut parametrit.	71
Taulukko 7. Lista sivustojen sosiaalisista ja semanttisista ominaisuuksista.	72

Sisältö

1	JOHDANTO	1
2	METADATA	4
2.1	Määrittely, luokittelu ja kuvaaminen	4
2.2	Tagit	6
2.3	Ontologiat	9
2.4	Folksonomiat	14
3	SOSIAALISET VERKKOPALVELUT JA SEMANTTINEN WEB	18
3.1	Sosiaaliset verkkopalvelut	18
3.2	Semanttinen web	19
3.3	Sosiaalinen semanttinen web	22
3.4	Linkitetty data	24
4	MONINÄKYMÄHAKU	31
4.1	Määrittely	32
4.2	Näkymät	34
4.2.1	Kontekstuaaliset näkymät	36
4.2.2	Näkymäluokittelu	37
4.3	Käyttöliittymä	41
4.3.1	Käyttäjäinteraktio	44
4.3.2	Käyttöliittymätyypit	46
4.4	Tiedonhaku	47
4.4.1	Tutkiva tiedonhaku	49
4.4.2	Hakemisen tukeminen	50
5	TUTKIMUKSEN EMPIIRISEN OSAN ASETELMA JA TOTEUTUS	53
5.1	Aikaisempi tutkimus	53
5.2	Tutkittavat verkkopalvelut	55
5.2.1	Finna	56
5.2.2	JYX	57
5.2.3	Yummly	58
5.2.4	Kulttuurisampo	59
5.2.5	Huuto	59
5.2.6	Gigantti	60
5.2.7	BibSonomy	61
5.3	Aineiston keruu- ja analyysimenetelmät	62
5.4	Tutkittavat parametrit	63
6	TUTKIMUKSEN TULOKSET	66
6.1	Sivustojen tiedot	66
6.2	Sivustojen käyttäjäinteraktio	67
6.3	Sivustojen näkymien ja objektien esitystapa	68

6.4	Sivustojen hakemisen prosessi	71
6.5	Sivustojen sosiaaliset ja semanttiset ominaisuudet	72
6.6	Yleiset kehitysehdotukset moninäkömahalle	73
6.7	Tagikäyttöliittymä moninäkömahakuna	74
7	YHTEENVETO.....	76
	LÄHTEET	79

1 Johdanto

Informaatioteknologian kehitys ja sen kautta tulleet innovaatiot ovat mullistaneet ihmisten jokapäivästä arkielämää. Elämme niin kutsutussa tietoyhteiskunnassa, jossa tieto kulkee tietokoneiden, mobiililaitteiden ja tietoliikenneverkkojen välillä ja sitä on saatavilla niille, joilla on mahdollisuus kytkeytyä Internetiin. Käyttäjälle avautuu netin kautta lähes äärettömältä tuntuva määrä tietoa, jota ihmiset ympäri maailmaa ovat sinne laittaneet tai tuottaneet. Nykyään monia eri tahoja kiinnostavat erilaiset tietoklusterit eri verkkopalveluissa, joihin on luotu ja kerätty valtava määrä dataa eri asioista tai henkilöistä. Tavallista nettiselaajaa kiinnostaa kuitenkin enemmän oikeanlaisen tiedon etsiminen, sen varmentaminen todeksi ja sen nopeahko löytäminen. Oikeanlaisen tiedon etsiminen ei kuitenkaan ole aina niin helppoa kuin luulisi. Jonkin asian hakeminen saattaa kestää tai ei tuota ollenkaan haluttua tulosta. Tämän lisäksi ihmisen saamaan tietoon ja tiedonkulkuun voi vaikuttaa monenlaisia tekijöitä, sillä hakuprosessi sinällään on läpinäkymätön. Esimerkiksi hakukoneiden ensimmäisen sivun osumat eivät välttämättä ole niitä absoluuttisia täysosumia, vaan jokin on voinut maksaa paikasta tai haun tulokset perustuvat sivuston sisältämiin linkkeihin. Tämän lisäksi haetun tiedon todentaminen oikeaksi voi olla hankalaa, jos tiedon lähteitä ei ole mainittu tai jos lähteeseen ei pääse käsiksi netin kautta.

Nyky-yhteiskunta tukeutuu vahvasti tiedon hakuun, keräämiseen, käsittelyyn, varastointiin ja sitä kautta uuden tiedon luomiseen. Monet ihmiset hakevat päivittäin tietoa hakukoneiden kautta erilaisista asioista, lukevat muiden henkilöiden tapahtumia ja mielipiteitä eri asioista sosiaalisen median tai uutissivustojen kautta sekä katsovat videoita eri asioista ja tilanteista. Tiedon kerääminen, analysointi ja uuden tiedon muodostaminen ei tule muuttumaan helpommaksi, ellei kehitetä laajenevalle datamäärälle sovelluksia tai laitteita, jotka pystyvät käsittelemään suurta määrää dataa nopeasti ja tehokkaasti. Nykyään suosituimmat palvelut Internetissä ovat järjestelmiä, joihin tietoa luovat pääasiallisesti verkkopalveluiden käyttäjät. Tästä seuraava askel olisi **semanttinen web**, jossa verkkoon luotua tai kerättyä tietoa ymmärtävät myös koneet semanttisella tasolla (Vandic, Dam ja Frasincar 2012, s. 427). Verkon tieto myös yhdistyisi tässä järjestelmässä niin, että erilaiset palvelut voivat hakea ja hyödyntää tietoa monesta eri lähteestä ja koota ne yhteen yhden verkkopalvelun alle. Tämä helpot-

taisi huomattavasti tiedon etsimistä, kun käyttäjä voisi hakea tietoa yhden verkkopalvelupor-
taalin kautta ja kerätä tietoa monesta eri lähteestä yhteen, samalla pystyen vertaamaan niitä
keskenään ja todentamaan niissä esiintyvän tiedon. Nykyään ihmiset hakevat tietoa käyttä-
mällä hakusanoja tai tageja tiedon hakemiseen. Tämänkaltaiset palvelut ovat jo muokanneet
käyttäjien tiedonhakemista niin, että itse haku koostuu useimmiten muutamasta avainsanasta
(Zwol et al. 2010, s. 961) ja tiedonhaku rajoittuu hyvin pinnalliseen määrään hakutuloksia.
Siispä toiveena olisi saada tiedon etsiminen Internetissä kehittymään seuraavalle tasolle, an-
tamalla sen käyttäjille parempia hakutyökaluja tiedonhakuun vanhojen vaihtoehtojen sijaan.

Moninäkömahaku on erilainen hakukäyttöliittymä, joka pyrkii helpottamaan käyttäjän tie-
donhakua keräämällä tietoa yhdestä tai useammasta eri lähteestä yhden käyttöliittymän tai
verkkopalvelun alle. Se kokoaa tietoa erilaisiksi hakunäkymiksi ja niitä voi selata monella
eri tavalla. Käyttäjälle avautuu mahdollisuus tarkentaa hakua metatietoon liittyviin tyyppiin tai
luokituksen perusteella, mihin hän aina kullakin kerralla etenee. Se helpottaa tiedonhakua
erityisesti silloin, kun käyttäjällä ei ole ennalta tietoa haettavasta aiheesta. Moninäkömahaka-
ku ei niinkään korvaa, vaan tukee perinteisiä sanahaku- tai tagihakukoneita. Moninäkömahaka-
haku kokoaa yhteen tiedon käsittelyyn liittyviä ongelmia (tiedon luokittelu, sen esittämi-
nen ja palauttaminen sekä käyttäjän tiedonhaku ja sen tukeminen) ja soveltuu käytettäväksi
myös semanttisen webin yhteydessä. Moninäkömahakukäyttöliittymiä on jo käytössä mo-
nissa eri verkkopalveluissa, kuten verkkokatalogeissa, kuvatalletuspaikoissa, treffipalveluis-
sa sekä liiketoimintasovelluksissa (Medynskiy, Dontcheva ja Drucker 2009, s. 2014), mutta
se ei ole vielä yleistynyt läheskään samalle tasolle, kuten Googlen kaltaiset sanahakukoneet
ovat käyttäjien keskuudessa. Tosin kovin monelle moninäkömahaku saattaa olla vielä tunte-
maton käsite, vaikka siihen olisikin aikaisemmin törmännyt: moninäkömahaku on ollut vain
toisenlainen hakukäyttöliittymä, jolla on voinut tuntua helpommalta löytää aihealueeseen
liittyvää tietoa.

Tämä tutkimus tutkii moninäkömahakua ja tarkastelee sen sovellutuksia erilaisissa verkko-
palveluissa. Teoriaosuudessa tutustutaan erilaisiin metadatan muotoihin, semanttiseen we-
biin sekä itse moninäkömahakuun. Tutkimuksessa tarkastellaan moninäkömahakua käyttä-
viä palveluita ja sitä, kuinka ne hyödyntävät moninäkömahakua käyttöliittymässään. Tut-
kimuksen tarkoituksena on antaa eräänlainen käsitys moninäkömahakun soveltumistavoista.

Tutkimuksessa vastataan kysymyksiin:

- Miten moninäkömähakua hyödynnetään eri verkkopalveluissa?
- Millainen on sivustojen käyttäjäinteraktio ja kuinka interaktiivinen käyttöliittymä on käyttäjälle?
- Miten hakemisen prosessi tapahtuu käyttöliittymässä ja mitä muita avustavia haku-komponentteja käyttöliittymä tarjoaa?
- Miten näkymät ja objektit esitetään käyttöliittymässä?
- Millaisia ovat sivustojen sosiaaliset ja semanttiset ominaisuudet?

Tutkimuksessa pohditaan myös saatujen tulosten perusteella, mitkä kulmakivet ovat edessä tämänkaltaisilla palveluilla, jotta moninäkömähaku tulisi enemmän käytetyksi ja tunnetuksi. Tämän lisäksi pohditaan, onko tageihin perustuva käyttöliittymä soveltuva moninäkömähakukäyttöliittymäksi. Tutkimusmenetelminä käytetään kirjallisuustutkimusta (kirjallisuuskat-saus) sekä empiiristä tutkimusta (verkkopalveluiden arviointi). Tutkimuksen aineistona on hyödynnetty aiheeseen liittyviä erilaisia tietolähteitä, kuten tutkimuksia, artikkeleita ja verkkosivuja.

Tutkielman rakenne koostuu seitsemästä kappaleesta. Luvussa 2 tarkastellaan erilaisia meta-datan muotoja. Luvussa 3 tarkastellaan semanttisen webin taustoja ja linkitettyä dataa. Luvussa 4 tarkastellaan moninäkömähakuun liittyviä taustoja ja teorioita sekä tarkastellaan tiedonhakua moninäkömähauun näkökulmasta. Luvussa 5 esitellään tutkimusasetelma sekä esitellään analysoitavat moninäkömähakua hyödyntävät verkkopalvelut. Luvussa 6 esitellään ja analysoidaan tutkimuksen tuloksia. Luvussa 7 luodaan yhteenveto tutkimuksesta. Tutkielman lopussa ovat lähteet ja liitteet.

2 Metadatan

Tässä luvussa esitellään erilaisia metadatan muotoja. Tutkimuksessa tarkasteltava metadatan on luokiteltu perinteiseen metadataan, semanttiseen metadataan (ontologiat) ja sosiaaliseen metadataan (tagit, folksonomiat). Metadatan perusteiden ymmärtäminen on tärkeää, sillä moninäkömähaku perustuu objekteihin liitettyyn metadataan (Holi 2010, s. 37). On tärkeää huomioida, että juurikin semanttisessa webissä metadatan on oltava eksplisiittistä, jotta datasta voidaan tehdä koneille ymmärrettävää (Hyvönen, Harjula ja Viljanen 2002, s. 49). Metadatan löytäminen, ymmärtäminen ja kokoaminen uudeksi tiedoksi ovat keskeisiä toimintoja semanttiselle webille ja moninäkömähäulle.

2.1 Määrittely, luokittelu ja kuvaaminen

Metadatan (metatieto) on tietoa tiedosta. Se on käsitteenä monimuotoinen, sille on laadittu runsaasti erilaisia luokituksia ja sen tarkka määrittely riippuu määrittelijästä. Esimerkiksi Gilliland-Swetland on määritellyt yleisen käsitelmän metadatalle, joka on hänen mukaan ”*kokonaisuus kaikista, mitä informaatio-objektista voidaan sanoa millä tahansa koostetasolla*”. Informaatio-objektilla tarkoitetaan tässä tapauksessa ”*mitä tahansa, mikä on ihmisen tai järjestelmän viitattavissa ja käsiteltävissä omana yksikkönään*”. Metadatan voidaan käyttää informaatioresurssien kuvailemiseen tai tunnistamiseen ja sitä on mahdollista hyödyntää tiedonhaussa hyödyntämällä sen rakenteisuutta ja hyperlinkkejä tiedon etsimiseen. (Gilliland-Swetland 2000, s. 1 ja 5) Metadatan voidaan myös lisätä toiseen dataan, jolloin tätä prosessia kutsutaan **annotoinniksi** (engl. *annotation*) (Häggström 2007, s. 37-38).

Metadatan voidaan luokitella implisiittisen ja eksplisiittisen luokittelun mukaan sekä käyttäen resurssi- ja asiakaspalvelinnäkökulmaa. **Implisiittinen metakuvaus** upottaa kielen rakenteen merkityksen algoritmeihin, jolloin ei synny erottavaa kuvausta siitä, mitä dokumentin elementit tarkoittavat. **Eksplisiittinen metakuvaus** on algoritmisesti tulkittava ja muokattava, joten se mahdollistaa dataan pääsyn, prosessoinnin ja jaon muiden ohjelmistojen kanssa. **Resurssinäkökulmassa** metadatan on joko ulkoista tai upotettua (metadatan suhde kuvattavaan resurssiin) ja **asiakaspalvelinnäkökulmassa** metadatan on joko keskitettyä tai

hajautettua (metadatan suhde itse ohjelmistoon). **Ulkoinen metadata** on uudelleenesitettyinä erikseen resursseista joita se kuvaa ja **upotettu metadata** on sisällytetty itse dokumenttiin. **Keskitetty metadata** on haettavissa yhdestä lähteestä ja **hajautettu metadata** on sijoittuneena eri paikkoihin, vaikkakin se voi asiakaspalvelimen näkökulmasta sijaita fyysisesti samassa paikassa. Resurssien näkökulmasta annotoinnit ovat hajautetussa metadatatassa ulkoisia. (Hyvönen, Harjula ja Viljanen 2002, s. 48-50) (Ahmed et al. 2001, s. 327-329)

McGuinness on jäsentänyt metadatan erilaisiin tasoihin. Yksinkertaisin käsitys metadatatista on **kontrolloitu sanasto**, toisin sanoen rajattu lista termejä (esim. luettelo). Seuraava määritelmä tästä olisi **termisanasto** (termit ja niiden selitykset luonnollisella kielellä). **Tesaurus** (engl. *thesaurus*) on seuraava taso, joka tarjoaa riippuvuussuhteita termien välille. **Taksonomia** (eksplisiittinen hierarkia) seuraa tämän jälkeen, mikä tarjoaa perustavanlaatuisen käsityksen yleistämisestä ja erikoistamisesta. Siitä seuraava taso olisi tarkat **aliluokkahierarkiat**. Seuraava taso sisältäisi kehykset, missä luokat sisältäisivät ominaisuuksia. Tätä ilmaisuvoimaisempi näkökanta olisi sisällyttää arvorajoitukset ontologiamäärittelyyn. Kun tämän jälkeen on tarve ilmaista enemmän tietoa, metatiedon täsmällisyyden ja pääteltävyyden vaatimukset kasvavat. (McGuinness 2003)

RDF, SPARQL ja OWL ovat keskeisiä kieliä metadatan kuvaamiselle webissä (Heath ja Bizer 2011). **RDF** on kieli, jota käytetään kuvaamaan informaatiota resursseista, jotka voivat olla mitä tahansa asioita (esim. dokumentteja, ihmisiä, fyysisiä objekteja tai abstrakteja konsepteja). RDF:ää voidaan käyttää web-datan julkaisemiseen ja linkittämiseen ja se on tarkoitettu tapauksiin, jossa webissä oleva tieto pitää käsitellä sovellusten avulla. RDF mahdollistaa tiedon jakamisen ilman, että menetetään tiedon merkitys. RDF:ssä resurssit esitetään tripleteinä (subjekti, predikaatti, ja objekti). (Schreiber ja Raimond 2014) **SPARQL** on kyselykieli, jota käytetään kyselykielenä RDF-datalle ja nimetyille graafeille (W3C 2013). **OWL** on kieli kuvaamaan ja jakamaan ontologioita webissä ja sitä käytetään sovelluksissa, joiden tarvitsee käsitellä tiedon sisältöä. Se tarjoaa RDF/XML:ää paremman tuen web-datan koneelliselle tulkinnalle tarjoamalla formaalin ontologian käsitteisiin liittyvien semanttisten rajoitteiden määrittelyyn. (McGuinness ja Harmelen 2004)

```

<rdf:Description rdf:about="http://www.w3.org/TR/rdf-syntax-grammar">
  <ex:editor>
    <rdf:Description>
      <ex:homePage>
        <rdf:Description rdf:about="http://purl.org/net/dajobe/">
        </rdf:Description>
      </ex:homePage>
    </rdf:Description>
  </ex:editor>
</rdf:Description>

```

```

PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX card: <http://www.w3.org/People/Berners-Lee/card#>
SELECT ?homepage
FROM <http://www.w3.org/People/Berners-Lee/card>
WHERE {
  card:i foaf:knows ?known .
  ?known foaf:homepage ?homepage .
}

```

```

<owlx:Annotation>
  <owlx:Documentation>An example OWL ontology</owlx:Documentation>
</owlx:Annotation>
<owlx:PriorVersion owlx:ontology="http://www.example.org/wine-112102.owl" />
<owlx:Imports owlx:ontology="http://www.example.org/food.owl"/>
<owlx:Annotation>
  <owlx:Label>Wine Ontology</owlx:Label>
</owlx:Annotation>

```

Kuvio 1: Esimerkkikoodia RDF, SPARQL ja OWL-kielistä.

2.2 Tagit

Tageilla tarkoitetaan informaatioteknologiassa avainsanaa tai -termiä, joka voidaan liittää johonkin informaatioon ja joita käytetään luokittelemaan tietoa informatiivisella tavalla. Tagit ovat epästrukturoitua metadattaa, jotka voivat koostua lauseista tai yksittäisistä sanoista eikä niiden tarvitse olla ennaltalaadittuja, vaan käyttäjät voivat luoda niitä itse hajautetusti. Tageissa käytetty tagien sanasto tai syy tagin valitsemiseen vaihtelee eri käyttäjillä. Tagien suosio perustuukin niiden kevyeen luonteeseen, monikäyttöisyyteen ja siihen, että kuka tahansa voi lisätä niitä ja ne ovat käyttäjien luettavissa ymmärrettävässä muodossa. Tämän takia monet verkkosivustot, kuten esimerkiksi Delicious ja Amazon.com hyödyntävät sivustoillaan tageja. (Treude ja Storey 2009, s. 14) (Liang et al. 2010, s. 51) (Rae, Sigurbjörnsson ja Zwol 2010, s. 92)

Tageja voidaan jakaa erilaisiin tyypeihin 1 tagin kuvauksen mukaan, sillä tagi itsessään voi sisältää laajan määrän semanttista tietoa (Iturrioz, Díaz ja Azpeitia 2011, s. 192). **Sisältöpohjaisella tagilla** voidaan kuvata lähteen todellinen sisältö (esim. *Ford, Kalevala*). **Kontekstipohjaisella tagilla** voidaan kuvata kohteen luonti- tai tallennushetki sijainnilla tai



Kuvio 2: Esimerkki tageista. Kuvassa tagipilvi BibSonomy-sivustossa.

aikamääreellä (esim. *Tampere*, *Näsinneula* ja *2014-06-20* yhtenä kokonaisuutena). **Attribuuttitagit** kuvaavat kohteen sisältöä ja sen piirteitä (esim. *hauska*, *pelottava*) tai kohteen tekijää/kirjoittajaa (esim. *Annan blogi*). **Omistajatagi** kuvaa kohteen oikeuksien hallitsijaa. **Subjektiiiviset tagit** kuvaavat käyttäjän mielipiteitä kohteesta (esim. *mahtava*, *cool*) ja niitä käytetään käyttäjälle kohdennettujen kohdesuositusten arvioinnissa. **Järjestelytageilla** voidaan kuvata henkilökohtaisia tietoja (esim. *laskut*, *työ*) tai muistuttaa tietyistä kohteisiin liittyvistä tehtävistä (esim. *toRead*, *toReview*). **Tarkoitukselliset tagit** ilmaisevat kohteeseen liittymättömiä toimintoja, jotka liittyvät käyttäjän tiedonetsimisprosessiin (esim. *lue tietoa kaloista*, *suomenna teksti*). **Tosiasialliset tagit** kuvaavat todellisia tietoja kohteesta, jotka ovat sellaisia, joita useimmat henkilöt liittäisivät tiettyyn kohteeseen. **Henkilökohtaisia tageja** käytetään järjestämään käyttäjän omia kohteita ja jonka kohdeyleisö on käyttäjä itse. **Itseviittaavat tagit** ovat lähteitä, jotka viittaavat itseensä. **Tagikimppu** (engl. *tag bundle*) on yksi tapaus hierarkkisista tageista ja se on metatagittamisen tulos, josta voidaan muodostaa hierarkkisia folksonomioita. (Gupta et al. 2010, s. 60)

Tageja voidaan luokitella myös käyttämällä tietoja siitä, että onko tagi käyttäjän vai koneen luettavissa sekä onko tagi kuvaileva vai määräävä. **Käyttäjäkuvaileva tagi** tai **standarditag** on tarkoitettu käyttäjiltä käyttäjille. **Konekuvaileva tagi** tai **konetagi** taas kuvaa kohdetta koneluettavassa muodossa ja se on myös oikea esitystapa tavanomaisille metatietokentille (esim. *päivämäärä*). Näitä tageja voidaan toimittaa niin käyttäjän kuin tagisivuston toimesta. Konetagit kuvataan usein tripletteinä, sisältäen kohteen nimiavaruuden, predikaatin ja arvon.

Konemäärävä tagi tai **toimintatagi** poikkeaa normaalista tagista niin, että siihen on lisätty jokin toimintoa kuvaava tagi (esim. *for:User*). **Käyttäjämäärävä tagi** tai **reaktiivinen tagi** on tagi muistuttamaan käyttäjän aikomuksista. Tällaiset tagit on määrätty käyttäjän toimesta ja niitä hallinnoi vain käyttäjä itse (esim. *toDo, toRead, toDownload*). (Iturrioz, Díaz ja Azpeitia 2011, s. 192-193)

Tagit voidaan jakaa luokkiin myös kielellisin perustein. **Funktionaaliset tagit** ovat tageja, jotka kuvaavat objektin toimintoja (esim. *ase*). **Funktionaaliset asematagit** ovat tageja, jotka sisältävät toiminnon lisäksi jonkin toimintoon liittyvän ajan tai paikan (esim. *kaluste, ruokailuväline*). **Alkuperäiset asematagit** kuvaavat, miksi jotkin asiat ovat yhdessä (esim. *roskat, astiat, likaiset astiat*). **Funktio- ja alkuperätagit** ovat tageja, jotka kuvaavat, miksi objekti on esillä, mikä on objektin tarkoitusperä tai mistä objekti on tullut (esim. *Michelangelo* ja *keskiaikainen Michelangelon maalaamassa maalauksessa*). **Taksonomiatagit** ovat tageja, jotka voivat auttaa objektin luokitteluksessa tiettyyn sopivaan luokkaan jotakin taksonomi-aa tai sanastoa mukaillen (esim. *eläinkunta* ja *nisäkäs* ihmisen kuvasta). **Adjektiivitagit** ovat tageja, jotka ilmaisevat objektia adjektiivein (esim. *sininen, iso*). **Verbitagit** ovat toimintasa- noja (esim. *etsiä rakentaa*). **Erisnimitagit** ovat tageja, jotka ilmaisevat kohteen nimen (esim. *Suomi, Helsingin kirkko*). (Gupta et al. 2010, s. 61)

Tagit voidaan myös jakaa tagin luomisprosessin perusteella manuaalisiin ja automaattisiin tageihin. **Manuaalisella tagilla** tarkoitetaan käyttäjän itse lisäämää tagia, joka voi olla liitetty mihin tahansa objektiin, tarkoituksena helpottaa etsimistä. Tällaisten tagien lisääminen voi olla aikaavievää sekä tagien nimeämiseen vaikuttaa käyttäjän kokemus tagaamisesta. **Automaattisella tagilla** tarkoitetaan tagia, joka tulee automaattisesti syötteenä ohjelman avulla. Automaattinen tagien luonti vaatii järkevästi toimiakseen ympärilleen sovellusympäristön, joka saatavien lisätietojen perusteella joko luo tai ehdottaa tageja käyttäjälle. (Koivisto 2010, s. 7-11)

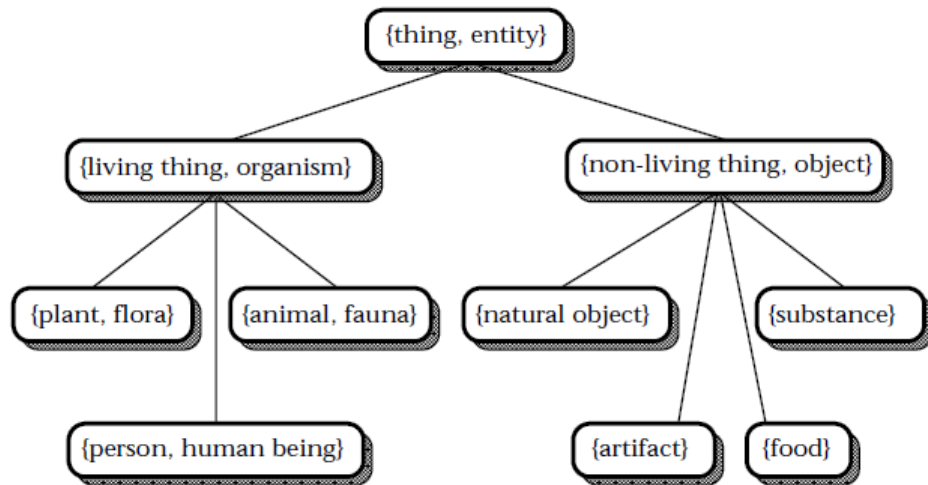
Vaikka tagit ovat hyödyllisiä, niihin kohdistuu myös muutamia ongelmia. **Tagispämmämisellä** tarkoitetaan tagien luomista kohteisiin vääränlaisella tavalla, tavoitteena ajaa toiminnal- laan omia intressejään eteenpäin. Tagispämmämisenä voidaan pitää ilmiötä, kuten yhden henkilön suuri kohteiden merkkäminen samalla tagilla tai tuhansien tagien luominen pienel- lä aikavälillä, jonka jälkeen tagien luontiin osallistuminen lähes loppuu. Tagispämmäjäillä on

tiettyjä erityispiirteitä, kuten korkea aktiivisuustaso, harvaan alueeseen kuuluvien kohteiden tagaaminen, korkea tagaamistaso kohdetta kohden ja massapostittaminen. Tagispämmäämistä voidaan havaita huomionhajaantumiskonseptilla (engl. *diffusion of attention*), jossa tagin saavuttama huomio määritellään tietyin aikaväleihin ja missä päähuomio kiinnitetään käyttäjiin, jotka käyttävät tagia ensimmäistä kertaa. Tällainen havainnointi auttaa vähentämään spämmin vaikutusalueetta ilman suodatuksen tarvetta. Muita tageihin liittyviä ongelmia on yhdenmukaistaminen (engl. *canonicalization*), moniselitteisyys (engl. *ambiguity*) (kts. 2.4) ja vähyys (engl. *sparsity*). (Gupta et al. 2010, s. 68-69)

Tutkijoiden Guyn ja Tonkinin toimesta on huomattu, että väärinkirjoittamisen ja huonon koodamisen takia tagijoukkoon päätyy myös turhia tageja. Tällaisia tageja ovat **yhdistesanaryhmit** (esim. *TimBernersLee*), tagit, jotka eivät seuraa yleisesti sovittuja sopimuksia (esim. numerointiin liittyen), **henkilökohtaiset tagit**, joilla ei ole suurempaa merkitystä muille (esim. *mynotebook*), tietokannassa yksittäisesti esiintyvät tagit (esim. päivämäärät ja monitulkintaiset tagit) sekä tageissa käytetyt symbolit. (Tonkin ja Guy 2006)

2.3 Ontologiat

Ontologia on tietty teoria olemassa olevasta (Noy ja Hafner 1997, s. 53) ja se tarjoaa potentiaalisia termejä kuvailemaan tietämystä kyseisestä aihealueesta (Chandrasekaran, Josephson ja Benjamins 1999, s. 20). Sen voidaan sanoa olevan ”*formaali, eksplisiittinen määrittely yhteisestä käsitteistöstä*” (Gruber 1993, s. 908), *joka määrittää käytetyt termit, joilla kuvataan ja esitetään alueen tieto ja tietämys*” (Waitelonis ja Sack 2011, s. 648) (Prieto-Diaz 2002, s. 458) (Devedzic 2002, s. 136) (Wache et al. 2001, s. 2). **Käsitteellistäminen** (engl. *conceptualization*) sisältää joukon objekteja, konsepteja ja muita entiteettejä, mistä tietämys ja riippuvuussuhteet ilmaistaan. Jokainen tietämysmalli on joko suorasti tai epäsuorasti sitoutunut johonkin käsitteellistämiseen. Muodollisesti ontologia sisältää termejä (jotka on organisoitu taksonomiaan), niiden määritelmiä ja niihin liittyviä aksioomia. (Noy ja Hafner 1997, s. 53) Sana ontologia on otettu filosofiasta, missä se tarkoittaa ”*järjestelmällistä selitystä olennosta*” (Corcho, Fernández-López ja Gómez-Pérez 2003, s. 42).



Kuvio 3: Esimerkki ontologiasta. (Noy ja Hafner 1997, s. 57)

Ontologioiden mahdollisuudet perustuvat niiden kykyyn jakaa tietämystä. Ilman ontologioita tai käsitteellistämistä aihealueella ei voi olla sanastoa, joka esittää tietämystä. Ontologia kapseloi itseensä sisäisen konseptuaalisen rakenteen mallinnettavasta aihealueesta, muodostaen aihealueen tietoesityksen ytimen. Tietämyksen jakaminen eri järjestelmien välillä vaatii kuitenkin tehokkaan ontologisen analyysin mallinnettavasta kohdealueesta. Analyysin tarkoituksena on selkeyttää tietämyksen rakennetta, sillä heikosti luotu ja analysoitu kohdealue johtaa hajanaiseen ja epäselvään tietämysperustaan. Analyysiin perustuva tietämyksen esityskieli tarvitsee kaksi asiaa: ontologiassa olevat termit tulee assosoida konseptien ja riippuvuussuhteiden kanssa sekä tiedon konsepteille ja riippuvuussuhteille tulee laatia koodisyntaksi. Onnistunutta tietämyksen esityskieltä voidaan jakaa muiden kanssa, joilla on samankaltaisia tarpeita tiedon esittämiseen kyseisessä aihealueessa. Se myös eliminoi tarpeen toistaa tiedon analysointiprosessi ja jaettuja ontologioita voidaan käyttää muodostamaan perusta aihealuekohtaisille tiedon esityskielille. Jaetut ontologiat mahdollistavat tiettyjä tilanteita kuvaavien tietämyskantojen rakentamisen, mikä lisää tietämyksen uudelleenkäyttöä. (Chandrasekaran, Josephson ja Benjamins 1999, s. 21)

Oikeanlaisen ontologian rakentaminen oikeaan tarpeeseen vaatii täsmällistä suunnittelua, sillä ontologian tarkoituksena ei ole esittää kaikkia olemassa olevia mahdollisuuksia, vaan tietämys koostuu järjestelmästä konsepteja ja niiden riippuvuussuhteista. Ontologian suunnittelu-

vaiheessa valitaan oleellisimmat aspektit todellisuudesta. Kaikki mallit noudattavat konseptisuhteiksi ja -aksioomiksi kutsuttuja periaatteita ja rajoituksia. Vaikka ontologioita voidaan esittää monella eri tavalla, suunnitteluvaiheessa useimmiten käytetään hierarkioiden mallintamista (vähintään käsitteellistämistasolla). Konseptihierarkiat ja taksonomiat esitetään kerroksina ja kaavioita käytetään parantamaan visuaalista esitystä. Kerrokset ontologian esittämisessä vaihtelevat ydinalueesta tehtäväkohtaisiin tai spesifeihin aihealueisiin. Tällainen ontologiasuunnittelu johtaa siihen, että ontologiat sisältävät tietoa asianmukaisesta hierarkisesta tai kerroksittaisesta malleista asiaankuuluvasta maailmasta. (Devedzic 2002, s. 138)

Ontologiat tarjoavat rakenteen mihin tietämuskantoja voidaan luoda. **Ontologiarakentaminen** sisältää joukon aktiviteetteja, jotka saatetaan ontologioiden käsitteellistämisen, suunnittelun, toteutuksen ja käyttöönoton aikana. Jotkin rakennusmenetelmät on spesifioitu tietyille ontologianrakentamisalueille, kuten ontologian uudelleenrakentaminen, ontologiaoppiminen, ontologia-arviointi, ontologiaevoluutio ja ontologioiden yhdistäminen. Rakentaminen antaa tietopohjan suunnittelulle perusteet ja sallii sille kurinalaisemman suunnittelun, auttaa määrittelemään olennaisimmat konseptit mallinnettavasta maailmasta ja mahdollistaa tiedon kokoamisen. (Devedzic 2002, s. 136-137) (Corcho, Fernández-López ja Gómez-Pérez 2003, s. 45) Sen tavoitteena on luoda looginen runko tietämykselle tai kehittää yleinen ymmärrys tiedonalasta. Asetetut tavoitteet ontologialle määräävät prosessin laajuuden ja monimutkaisuuden. Vain perusymmärryksen tarjoava ontologia voi tarvita vähemmän pyrkimystä kuin formaaleja loogisia argumentteja tukeva ontologia. Kysymykset, kuten ”*miksi me rakennamme ontologiaa*” ja ”*mihin me haluamme käyttää sitä*” ovat ensiaskelia ontologian rakentamisessa. (Prieto-Diaz 2002, s. 458) Ontologiat rakennetaan yleensä ryhmäyhteistyöllä, joissa jäsenet sijaitsevat eri kohteissa (Corcho, Fernández-López ja Gómez-Pérez 2003, s. 44).

Ontologioille on olemassa erilaisia rakennusmenetelmiä. Uscholdin ja Kingin 1990-luvulla testaama menetelmä sisälsi neljä eri aktiviteettia: ontologian tarkoituksen tunnistaminen, ontologian rakentaminen (kapselointi, koodaus ja integrointi), ontologian arviointi sekä ontologian dokumentointi. He myös esittivät kolmea eri strategiaa (top-down, bottom-up ja middle-out) ontologian pääkonseptien identifiointiin. (Corcho, Fernández-López ja Gómez-Pérez 2003, s. 45) (Prieto-Diaz 2002, s. 458-459) Tästä vähemmän tyypillistä rakennusmenetelmää käytettiin CYC-tietämuskannan rakentamisessa, joka koostui kolmesta vaiheesta.

Ensimmäinen vaihe sisälsi artikkelien ja tietopalasten manuaalisen koodaamisen, toisessa ja kolmannessa vaiheessa hankittiin uutta yleistietämystä käyttämällä luonnollisia kieliä tai koneoppimistyökaluja. Eräs formaalimpi ontologian rakentamismenetelmä on Grüningerin ja Foxin ehdottama menetelmä, jossa ensiksi tunnistetaan mahdolliset sovellukset, jossa kyseistä ontologiaa tullaan käyttämään, jonka jälkeen määritetään ontologian laajuus pätevyyskyselyiden avulla. Näitä kyselyitä ja niiden vastauksia käytetään ontologiaan liittyvien rakenteiden poimintaan. Ontologioita voidaan myös rakentaa uudelleenkäyttämällä osia ontologioista, rakentaakseen aihealuespesifejä ontologioita ja tietoperusteita. (Corcho, Fernández-López ja Gómez-Pérez 2003, s. 43 ja 45)

Ontologioita on kustannustehokkaampaa uudelleenkäyttää kuin luoda niitä tyhjästä. Uudelleenkäytössä halutun datan kuvaamiseen suositaan yleensä olemassa olevia, mielellään vakiintuneita tai hyvin testattuja ontologioita. Uudelleenkäyttö helpottaa merkittävästi myös ontologioiden yhteentoimivuutta, mahdollistaen datan kuvaamiseen samoja ontologioita käytävien datatarjoajien yhdistää dataa helpommin yhteen. Jotta ontologiat voivat helpottaa yhteentoimivuutta eri järjestelmien ja datajoukkojen kanssa, käyttäjien tulisi voida löytää olennaisia ontologioita nopeasti ja helposti. Heidän tulisi myös pystyä jakamaan ontologiansa tavalla, jolla toiset voivat helposti löytää, päästä käsiksi ja uudelleenkäyttää niitä. Ongelmana on kuitenkin ollut monivaiheinen prosessi, jossa käyttäjän pitää löytää, tutkia, valita ja testata haluttu ontologia. Sen tulisi myös olla jaettavissa muiden kanssa, jos käyttäjä laajentaa valittua ontologiaa. Ontologiakirjastojen onkin tunnistettu olevan tärkeä rakenne semanttisessa webissä: kasvava määrä ontologiakirjastoja on pureutunut tähän ongelmaan, mahdollistaen käyttäjän löytää, uudelleenkäyttää ja julkaista ontologioita, täten helpottaen datan yhteentoimivuutta semanttisessa webissä. (d'Aquin 2011, s. 1-3)

Ontologioiden keskeinen vahvuus on niiden suhteellinen itsenäisyys tiettyihin sovelluksiin. Ontologia koostuu suhteellisesti yleisestä tietämyksestä, jota voidaan uudelleenkäyttää monenkaltaisissa sovelluksissa sekä niitä voidaan jakaa sovellusten välillä. (Spyns, Meersman ja Jarrar 2002, s. 12) Ontologioita hyödynnetäänkin erityisesti ohjelmistojärjestelmissä, missä ontologioiden täytyy tietää kohteen yksilöistä ja niiden attribuuteista sekä niiden riippuvuussuhteista (Devedzic 2002, s. 143). Kyseisessä ympäristössä ontologiat tarkoittavatkin tietokonepohjaisia resursseja, jotka esittävät aihealueen hyväksytyjä semantiikkoja (Spyns,

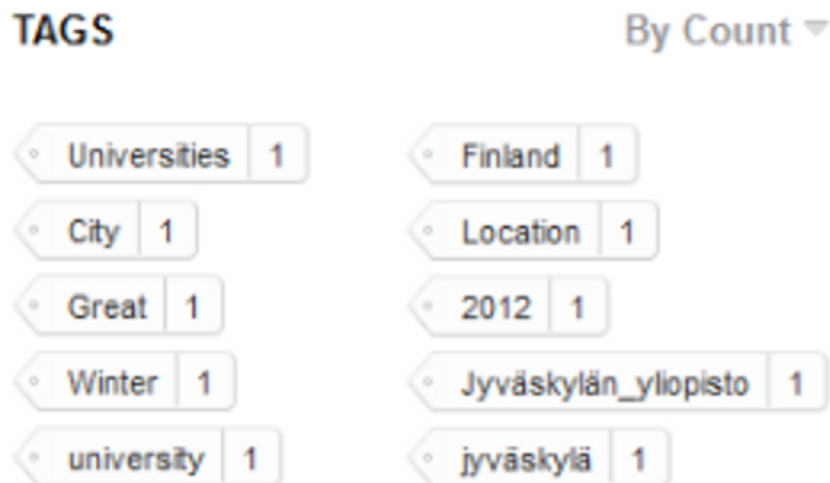
Meersman ja Jarrar 2002, s. 12). Kun informaatiolähteiden semantiikkoja kuvataan käyttämällä ontologioita, tekee se esitettävästä sisällöstä eksplisiittistä (Wache et al. 2001, s. 2), mikä sopii semanttisen webin kontekstiin. Useimmissa järjestelmissä ontologia ei kuitenkaan ole eksplisiittinen, mutta käyttämällä standardeja esittämiskieliä (kuten RDF tai OWL), ontologioita voidaan jakaa myös palveluiden välillä, jolloin dataa voidaan sovittaa yhteen jaettujen sovellusten välillä. Vasta tämän jälkeen ontologiat tarjoavat jaetun mallin esittää semanttisesti rikasta tietoa semanttisessa webissä. (Breslin, Passant ja Vrandečić 2011, s. 474)

Ontologioiden oleellisin ongelma on se, että eri ontologioilla voidaan kuvata samoja asioita ristiriitaisesti. Tämä johtaa ongelmaan, jossa ei tiedetä varmasti, mikä hierarkia nostetaan ensisijaisesti käyttäjälle esille. Monissa tapauksissa sovelluksen instanssit ja myös aihealueen konseptien merkityksiä vihjaavat säännöt (identiteetti, pakollisuus ja ankaruus) on sisällytetty ontologiaan. Nämä säännöt rajoittavat konseptien semantiikkoja ja konseptuaalisia riippuvuussuhteita tietyn sovelluskohdealueen tietyssä käsitteellistämisessä. Sääntöjä pitää noudattaa kaikkien niiden sovellusten, jotka haluavat käyttää jonkin ontologian tulkintaa. Ratkaisuna tähän ongelmaan on ontologioiden tupla-artikulointi, jolla erotetaan ontologiat datamallintamisesta ja mahdollistetaan uudelleenkäytettävien ontologioiden mallintaminen ja rakentaminen. (Spyns, Meersman ja Jarrar 2002, s. 12)

Ontologioiden käytön tekee haasteelliseksi se, että niiden kuvaamiseen käytettäviä kuvauskieliä on olemassa useita. Jotkin näistä kielistä voivat perustua ilmaisuvoimaltaan erilaisiin logiikkoihin ja olla siten epäsoivia päättelyn kannalta. Muutostilanteissa jopa samalla kielellä kuvattujen ontologioiden yhdistäminen voi olla työlästä. Kuitenkin semanttisen webin kontekstissa ontologioiden selkein etu on se, että eri käsitteille saadaan yksikäsitteiset tunnistet. Päättelyäkin pystytään tekemään riippumatta siitä, millaisessa tietämuskannassa kuvauksia säilytetään, sillä web on julkinen rajapinta näille tietämuskannoille. Joissakin yksittäistapauksissa olemassa olevan datan ”*ontologisointi*” voi onnistua suoraviivaisesti (esim. YSA-merkattujen asiansanojen tulkinta YSO-ontologiassa). Tosin ontologioiden varsinaiset hyödyt tulevat esille vasta sitten, kun ontologiaa noudattavaa metadataa ja sovelluksia on webissä käytössä enemmän.

2.4 Folksonomiat

Folksonomialle ei ole esitetty yhtä selkeää määritelmää, mutta joidenkin tutkijoiden mukaan termillä tarkoitetaan ihmisten resursseille asettamista tageista muodostuvaa kasvavaa rakennetta, jonka taustalla on käyttäjien palautteesta muovautuva yhtenäinen tagauskäytäntö. Folksonomia-termi tulee sanoista ”folk” (kansaa), ”taxi” (taksonomia) ja ”nomos” (hallinto). Folksonomia voidaan mallintaa kolmijakoisena hyperverkkona, joka muodostuu käyttäjien, resurssien ja tagien muodostamasta kokonaisuudesta. Toisin kuin formaaleissa taksonomioissa, folksonomioissa termien välillä ei ole ennalta määriteltyjä eksplisiittisiä suhteita, vaan kaikki termit sijaisevat samassa nimiavaruudessa ilman hierarkioita. (Gupta et al. 2010, s. 59) (Mika 2005, s. 524)



Kuvio 4: Esimerkki aloittelevan käyttäjän personomiasta Flickr-sivustossa.

Folksonomioita on kahdenlaisia määriteltyjen tagienluomisoikeuksien mukaan. **Kapeat folksonomiat** rajoittavat kohteiden tagaamista rajalliselle määrälle käyttäjiä. Kapean folksonomian esimerkisivusto on Flickr, jossa kohteet merkkää kohteen omistaja. **Leveässä folksonomiassa** kohteita voi tagata koko sivuston yhteisö niin, että yhden käyttäjän tagausaktiiviteetti samaistuu muiden käyttäjien tagausaktiiviteettiin. Leveän folksonomian esimerkisivustoja ovat Delicious ja BibSonomy. Tagaamisoikeuksilla oletetaan olevan vaikutuksia tagaamismotivaatioihin. Siinä missä kapean folksonomian käyttäjät käyttävät tavallisimpia tageja mainostaakseen omia kohteitaan, leveän folksonomian käyttäjillä on rajoitettu mielen-

kiinto tagaamisen sosiaalisiin tekijöihin, mutta he tagaavat kohteita mieleenpalauttamismielessä. Tagaamisen jatkuessa käyttäjille muodostuu oma tagisanasto, josta käytetään nimeä **personomiat**, jotka voivat olla hyvinkin erilaisia käyttäjien välillä. (Wetzker et al. 2010, s. 71) (Clements, Vries ja Reinders 2010, s. 8)

Folksonomioiden yhteydessä tärkeä perusasia on useissa erilaisten tietokokonaisuuksien tutkimuksissa huomattu havainto, että erilaisten tagien kokonaismäärä järjestelmässä ajan edetessä noudattaa **potenssilakia** (engl. *power law*), joka ilmaisee matemaattisen suhteen kahden muuttujan välillä. Tagikokonaisuuksia tutkittaessa verrattaviksi muuttujiksi valitaan esimerkiksi tagien esiintymä suhteessa tagien määrään tai tagien määrä suhteessa tagattuihin objekteihin. Jälkimmäisessä tapauksessa tageja sisältävien tietokokonaisuuksien yhteydessä potenssilain alkupää sisältää tageja, jotka ovat liian yleiskäyttöisiä ollakseen hyödyllisiä tagisuosituksina ja potenssilain loppupää taas sisältää käyttäjien käyttämiä suosituimpia tageja. Kyseisten tagien saavuttama suosio käyttäjien kesken vahvistaa sen, että käyttäjien välille on ajan mukaan syntynyt voimakas ymmärryssuhde ja hyväksyntä kyseisen sivuston yleisimmistä tageista. (Gupta et al. 2010, s. 63) (Heymann ja Garcia-Molina 2006, s. 3) (Sigurbjörnsson ja Zwol 2008, s. 329)

Folksonomioiden etuna on kyky vastata käyttäjien tarpeisiin ja kielellisyyteen tarkkuuden sijaan. Metadatan luominen manuaalisesti on ajallisesti ja taloudellisesti kallista sekä luotu metadatakirjasto voi olla täysin erilainen verrattuna esimerkiksi käyttäjien käyttämään sanastoon. Valmiiden taksonomioiden ja ontologioiden luonti voi olla hyödyllistä, mutta pysyvät taksonomiat ovat luonteeltaan kankeita ja keskitettyjä sekä objektit eivät välttämättä sovi vain yhteen kategoriaan. Hierarkkiset luokittelut taipuvat muuttumaan yhdeksi järjestelmälliseksi näkemykseksi eikä tällainen luokittelu helposti pysy uusien ja kehittyvien kielikoelmien perässä. Folksonomiat verrattuna formaaleihin taksonomioihin esiintyvät tasaisessa nimiavaruudessa eikä siten folksonomiassa esiintyvillä termeillä ole epäsuorasti muodostuneita suhteita toisiinsa nähden. Folksonomiat edustavat käyttäjien sanastoa ja niitä voidaan pitää yksinkertaisena iteratiivisena järjestelmänä, joka auttaa luomaan suosituimman tavan organisoida objekteja järjestelmässä. (Gupta et al. 2010, s. 58-59)

Yksi folksonomioiden ongelmista on **moniselitteisyys**, sillä eri käyttäjät merkkäävät termejä kohteisiin eri tavalla. Useiden sanojen alkukirjaimista muodostetut lyhenteet (akronyymit)

voivat johtaa moniselitteisyyteen. Nykyään tagit on määritelty joko yksittäis- tai yhdistesanoiksi, joten monen sanan sisältävä välimerkittömän tagin (esim. ”*vertigovideostillbbc*”) tieto voi hävitä tagaamisprosessissa. Yksittäissanaisten tagit hävittävät lauseen sanajärjestyksen tiedon. Folksonomioissa on muitakin ongelmia. Folksonomioissa ei ole synonyymien tai homonyymien kontrollointia. Niissä esiintyy usein sekä yksikkö- että monikkomuotoisia sanoja. Ne eivät myöskään tarjoa muodollista lähestymistapaa tagimuodon valintaan. Tagaamiseen osallistuvien erilaiset osaamisalueet ja tarkoitukset voivat johtaa tageihin, jotka kuvailevat kohdetta hyvin abstraktisti. (Gupta et al. 2010, s. 69)

Taulukko 1: Tagityypit kootusti.

Kuvauksen mukaan:	sisältöpohjainen kontekstipohjainen attribuutti subjektiivinen järjestely tarkoituksellinen henkilökohtainen itseviittaava
Käyttäjän mukaan:	käyttäjäkuvaileva konekuvaileva konemääräävä käyttäjämääräävä
Kielen mukaan:	funktionaalinen funktionaalinen asema alkuperäinen asema funktio tai alkuperä taksonomia adjektiivi verbi erisnimi
Luomisprosessin mukaan:	manuaalinen automaattinen
Virheiden mukaan:	yhdistesanaryhmät yleisesti sopimattomat muille merkityksettömät yksittäisesti esiintyvät symbolit

3 Sosiaaliset verkkopalvelut ja semanttinen web

Tässä luvussa tutustutaan webin sovellutuksiin, jotka hyödyntävät metadataa. Sosiaaliset verkkopalvelut (Web 2.0) ovat olleet paljon esillä eri medioissa ja ne ovat saavuttaneet käyttäjien suosion hyvin nopeasti. Katsaus semanttiseen webiin (Web 3.0) on tarkastelua seuraavalle webin murrokselle. Semanttinen web on tärkeä ymmärtää, sillä moninäkömähäun täysi potentiaali tulee esille, kun se hyödyntää semanttista webiä. Kun moninäkömähaku integroidaan ontologioiden ja semanttisen webin kanssa, tuloksena syntyy paradigma **semanttiselle moninäkömähäulle** (Holi 2010, s. 22). Semanttisen webin tarjoama linkitetty tietoavaruus tarjoaa moninäkömähäulle erinomaisen pääsyn dataan ja mahdollisuuden täydentää näytettävää dataa uusilla lähteillä. Useat tulevaisuuden sovellutukset, jossa yhdistyvät sosiaalisuus ja semanttisuus, odottavat vielä tekijäänsä.

3.1 Sosiaaliset verkkopalvelut

Sosiaaliset verkkopalvelut ovat sosiaalisia palveluita tarjoavia verkkosivustoja, jotka sallivat esimerkiksi käyttäjien tehdä muistiinpanoja verkostoon jaetuista resursseista (esim. web-sivuja, elokuvia tai kirjoja), hakea mielenkiintoista sisältöä käytettäväksi ja jaettavaksi sekä löytää samanhenkisiä ihmisiä verkoston sisältä. Tällaiset sivustot voivat tarjota esimerkiksi tagipalveluita käyttäjilleen, jolloin käyttäjät voivat tagata ja arvioida verkoston käyttäjien tuottamaa sisältöä sekä jakaa kyseisiä tageja verkoston yhteisön sisällä. Resurssien tallentaminen ja muistiinpaneminen huomioivien tagien avulla auttaa verkoston käyttäjiä ylläpitämään heidän mielenkiintoaan verkoston sisältämiin resursseihin. Verkoston sisältämien tagien sanotaan olevan sosiaalisia, kun käyttäjät jakavat niitä yhteisön kesken. (Swamynathan et al. 2008, s. 1) (Zubiaga, Körner ja Strohmaier 2011, s. 94)

Sosiaalisten verkkopalveluiden läpilyönti Internetin käyttäjien kesken (esim. *MySpace*, *Facebook*, *YouTube*) on muokannut käyttäjien verkossa tapahtuvaa toimintaa. Käyttäjien verkon ajankäytössä on lisääntynyt erityisesti vuorovaikutteinen ja yhteisöllinen toiminta. Käyttäjät käyttävät dynaamisia web-sivustoja videoiden ja kuvien jakamiseen, sisällön mielipiteiden ilmaisemiseen, löytääkseen mielenkiintoisia ihmisiä sekä ylläpitämään sosiaalisia suhteita.

Käyttäjien käyttäessä verkkopalvelun tarjoamia kommunikointityökaluja kontaktien ylläpitämiseen he rakentavat itselleen oman verkkoidentiteetin ja luovat ystävyysuhteita verkoston sisällä. Syntyneet sosiaaliset linkit todistavat, että linkin tehneiden käyttäjien välille on muodostunut luottamuksellinen suhde. Samalla, kun käyttäjät käyttävät verkkopalvelua, saavat sen ylläpitäjät itselleen käyttäjien toiminnan tuloksena suuren määrän dataa. (Clements, Vries ja Reinders 2010, s. 2) (Swamynathan et al. 2008, s. 1)

Sosiaalisen toiminnan tuloksena eri verkkopalvelut sisältävät valtavasti tietoa, joka voi koostua esimerkiksi itse käyttäjistä ja käyttäjien tekemisistä palvelun sisällä ja sen ulkopuolella. Useat eri tahot käyttävät tästä tietojoukosta termiä **big data**, joka koostuu heterogeenisistä ja itsenäisistä tietojärjestelmistä, jotka ovat jaettuja ja ei-keskitettyjä. Nämä piirteet, kuten myös se, että eri datatarjoajat käyttävät omia protokollia datan tallentamiseen, tekee haastelliseksi löytää hyödyllistä tietoa tästä datasta, vaikkakin siitä yritetään pyrkiä tutkimaan monimutkaisia ja muuttuvia riippuvuussuhteita. Kun big datan tietomäärä kasvaa, kasvaa myös siinä sisältyvän monimutkaisuuden ja riippuvuussuhteiden määrä. Keskeisin haaste onkin löytää hyödyllistä tietoa tai tietämystä tästä tietojoukosta. Tiedonkeräysprosessin on oltava hyvin tehokasta ja lähes reaaliaikaista, sillä kaiken tutkittavan datan tallentaminen nykyteknikoilla ei ole mahdollista. (Wu et al. 2014, s. 97-99) Vaikka big data nousee esille erityisesti sosiaalisten verkkopalveluiden kontekstissa, big data ei pelkästään koske sosiaalisia verkkopalveluita, vaan mitä tahansa laajoja ja toisiinsa kytkeytyviä tietoaaineistoja voidaan kutsua tällä nimellä.

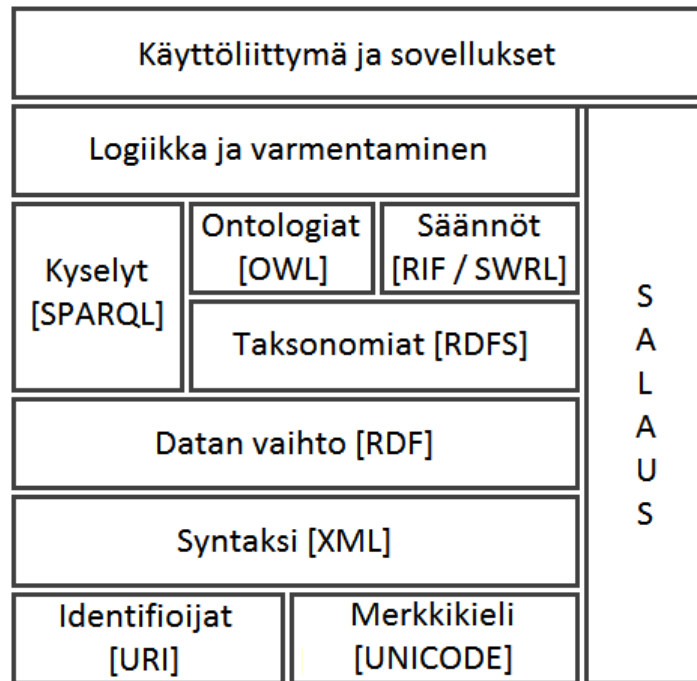
3.2 Semanttinen web

Semanttinen web on yksi keskeisimmistä käsitteistä informaatioteknologian tutkimuksen alueella. Sen idean kehitti vuonna 2001 Tim-Berners Lee, joka määritteli semanttisen webin ”uutena muotona web-sisällölle, joka on merkityksellistä myös koneille”. Päästäkseen tähän tavoitteeseen semanttisen webin pyrkimys on luoda merkityksellisiä linkkejä heterogeenisten tietolähteiden (RDF-metadata) välille. Tämä laajentaisi nykyisen webin mahdollisuuksia. Informaatio annettaisiin hyvin muotoillussa merkityksessä, mikä takaisi koneille ja henkilöille paremman saatavuuden tutkia ja työstää informaatiota yhteistoiminnassa. Linkittäminen taas mahdollistaisi navigoinnin yhdestä resurssista toiseen samantyyppiseen resurssiin eri tieto-

lähteestä, siten mahdollistaen henkilön löytää enemmän informaatiota kyseisestä aiheesta. (Waitelonis ja Sack 2011, s. 647-648) Täten semanttisen webin voi nähdä kokoelmana keskittämättömiä verkkoja, millä ei ole yksittäistä tietoperustaa ilmaistuna ja mihin kuka tahansa voi osallistua ilmaisujen luomiseen tuomalla ne julkisesti saataville. Näillä tietolähteillä ei välttämättä tarvitse olla mitään yhteistä, mutta käyttämällä jaettuja tunnisteita (URI:t) ja termejä, niiden tieto voidaan yhdistää tarjoamaan hyödyllisiä palveluita niin ihmisille kuin ohjelmistoasiakassovelluksille. (Oren et al. 2008, s. 1-2)

Semanttinen web hyödyntää useita metadatan kuvaamiseen käytettäviä kieliä (kts. 5). Se perustuu RDF-spesifikaatioon, jota käytetään webin tiedonsiirroissa standardimallina. RDF laajentaa webin linkitysrakennelmaa niin, että rakenteet käyttävät URI:a identifioimaan asioita ja nimeämään riippuvuussuhteita asioiden välille. (Waitelonis ja Sack 2011, s. 648) RDF-dokumentit itsessään ovat kokoelmia tripletejä, joista kaikki paitsi objektit ovat resursseja (objektit voivat olla joko kirjaimellisia merkkijonoja tai datatyyppiä). Kaikki RDF-tripletit muodostavat suoran nimetyn verkon, missä päätepisteet edustavat nimettyjä linkkejä (predikaatteja) kahden resurssin välillä. RDF:ssä jokainen resurssi on merkitty URI:lla tai useimmissa tapauksissa URL:lla. Resurssien riippuvuussuhteet ja ominaisuudet voidaan tarkentaa RDFS-sanastonkuvauskielellä, joka määrittää luokat ja ominaisuudet, joita voidaan käyttää kuvaamaan luokkia, ominaisuuksia ja muita resursseja. Suurempia web-sisältöjä tulkittaessa OWL-kieli on hyödyllisempi, tehden tulkittamisesta helpompaa tarjoamalla lisäsanaston formaalin semantiikan mukana. (Vandic, Dam ja Frasinca 2012, s. 427-428) (Waitelonis ja Sack 2011, s. 648)

Semanttisen webin idea, jossa koneet ymmärtävät web-sivuilla olevan tiedon, ei nykymuotoisilla web-sivuilla ole mahdollista. Tämä johtuu siitä, että nykymuotoiset web-sivut ovat ihmisille tarkoitettu lukumuodossa. Esimerkiksi ihminen ymmärtää sanan tai sanojen semanttisen merkityksen, mutta koneelle sana tai sanat ovat vain lista erilaisia merkkejä sovitettuna yhteen. Tämän lisäksi koneen on vaikea ymmärtää, mihin kontekstiin tietty sana kuuluu. Semanttinen web mahdollistaisi kuvata oikean maailman objekteja web-sivulla, mistä käytetään yleisesti nimitystä **konsepti**. Ne voivat olla muun muassa henkilöitä, yrityksiä tai tuotteita ja niiden välille on mahdollista asettaa riippuvuussuhteita, sallien koneiden ymmärtää näitä ja sitä kautta helpottaa automaattista tiedon koostamista resursseista. Esimerkiksi,



Kuvio 5: Semanttisen webin teknologioita sisältävä kaavio. (Horrocks et al. 2005)

jos kaksi eri web-sivua kuvaa samaa konseptia, niin silloin on mahdollista koota tämä informaatio yhteen yhdistämällä informaatiota kahdesta tai useammasta lähteestä, jotka kuvaavat tätä samaa konseptia. (Vandic, Dam ja Frasincar 2012, s. 427)

Nykyiset web-sivustot tarvitsevat lisävarustamista, jotta semanttiset hakukoneet voisivat koota ja yhdistää informaatiota edellä mainitulla tavalla. Tämä tapahtuu joko tekemällä lisämerkkauksia olemassa oleviin staattisiin tai dynaamisiin HTML-sivuihin tai sijoittamalla RDF-kuvaus omalle sivulleen. Web-sivujen varustamiseen on olemassa kaksi kieltä, joita ovat mikroformaatit ja RDFa. **Mikroformaatit** ovat sovittuja käytäntöjä noudattavia XHTML-merkkaukseen upotettua metadataa, jotka kuvaavat asioita kuten ihmisiä, tuotteita ja tapahtumia. Nämä formaatit ovat vähemmän joustavia kuin RDFa, mutta niitä on helpompi ymmärtää niiden henkilöiden, joilla ei ole aiempaa kokemusta semanttisesta webistä. **RDFa** taas on W3C-suositus, joka lisäämällä joukon attribuutti-tason laajennuksia XHTML:ään upottaa rikasta metadataa web-dokumentteihin. RDF-datamallikuvaus mahdollistaa RDFa:n käytön upottamalla RDF-triplettejä XHTML-dokumentteihin ja se myös mahdollistaa RDF-triplettien keräämisen käyttäjäagenttien avulla. (Vandic, Dam ja Frasincar 2012, s. 427)

Vaikka semanttisen webin rakenne on tarkoitettu koneille, on sen luominen ja ylläpitäminen sosiaalista toimintaa. Koneet helpottavat merkkien käsittelyä huomattavasti, mutta vain semanttisen webin käyttäjillä on tarvittavat tulkinnalliset ja assosiatiiiviset kyvyt luoda ja ylläpitää ontologioita ja erityisesti metadataa, joka käyttää ontologioita. On myös huomioitavaa, että semantiikan lisääminen web-arkkitehtuuriin tekee sen käyttäjistä tärkeän osan järjestelmää. Semanttinen web voidaankin nähdä kolmitasoisena näkymänä, jotka koostuvat yhteisön ja sen suhteet sisältävästä tasosta, semantiikkojen tasosta (ontologioista ja niiden suhteista) sekä objektien ja niiden suhteiden sisältävästä tasosta. Tämä näkymä voidaan myös ajatella kolmitasoisena mallina ontologioita, jotka sisältävät kolmea eri luokka-alkiota (aktorit, konseptit ja instanssit) ja hyperrajoja esittämään käyttäjien sitoutumista luokitteluun instansseja tiettyihin konsepteihin. Mallin tuloksena voidaan luoda esiintymäverkkoja ontologiaoppimiselle ja yhteisöllisille semanttisille verkoille. (Mika 2007, s. 14)

3.3 Sosiaalinen semanttinen web

Sosiaalinen semanttinen web on semanttisen webin teknologioiden ja sosiaalisen webin sovellusten suunnittelupohjien yhdistämisestä muodostettu kokonaisuus. Siinä tietyt formaalisesti esitetyt ihmisymmärrettävät osat voidaan koodata semanttisen webin työkaluilla ja muodostaa niille ymmärrys sekä niitä voidaan ylläpitää sosiaalisen webin tekniikoiden avulla. Sosiaalisessa semanttisessa webissä on mahdollista muodostaa sosiaalisesta toiminnasta syntyvälle yhteisölliselle tietämykselle ”älykkyyttä” semanttisen webin teknologioiden avulla sekä myös ylläpitää eri yhteisöjä tunnistavia piirteitä. Siinä yhdistyy semanttisen webin ilmaisuvoimainen tarkkuus ja ymmärrysvoima sekä sosiaalisen webin käyttäjäystävälliset yhteistyömekanismit. Se avaisi mahdollisuuden luoda uudenlaisia sovelluksia, jotka voisivat käyttää hyödyksi tiettyjen saatavilla olevien web-datojen välisiä semanttisia riippuvuussuhteita. (Greaves ja Mika 2008, s. 1-2)

Sosiaalinen semanttinen web tarjoaisi monia hyötyjä sen käyttäjille. Se voisi auttaa verkkopalvelun käyttäjiä jakamaan tietoa luomalla ihmisyyhteisöjä samanlaisten mielenkiinnon kohteiden ympärille. Semanttiset riippuvuussuhteet mahdollistaisivat informaation automaattisen paikantamisen ja yhdistämisen sekä tiedon esityksiä pystyttäisiin tukemaan ja muuntamaan erilaisten käyttäjien tarpeisiin (Greaves ja Mika 2008, s. 1). Se myös tarjoaisi tuen

datasiirroille ja yhdistämisoperaatioille sovellusten välillä, mahdollistaen yksilöiden ja yhteisöjen osallistumisen jaetun yhteenliitettävän tiedon luomiseen. Sosiaalista semanttista webiä voidaankin pitää verkostona, jossa on yhteenliitettyä ja semanttisesti rikasta tietoa. Se toisi yhteen sovelluksia ja sosiaalisia ominaisuuksia sosiaalisesta webistä sekä tietämyksen esittämisen kielet ja formaatit semanttisesta webistä. (Breslin, Passant ja Vrandečić 2011, s. 471)

Sosiaalinen semanttinen web voidaan rakentaa käyttämällä hyödyksi eri tekniikoita ja metadatan kuvauskieliä. RDF sallii semanttisten yhdistelmien ja integroitujen näkymien yksinkertaisen generoinnin niin alkuperäiseen kuin julkiseen informaatioon. Web-sivujen HTML-sisältö voidaan tehdä yhteensopivaksi RDF:n kanssa RDFa:n avulla sisällyttämällä RDF-annotaatiot XHTML-attribuutteihin. Tämä mahdollistaa tehokkaan semanttisen haun, eliminoiden tarpeen indeksoida erillisiä semanttisen kuvauksen sisältäviä sivuja. SPARQL-kyselykieltä voidaan käyttää etsimään avainsanojen lisäksi myös riippuvuussuhteita ihmisten ja objektien välillä koostetuista semanttisesta web-datasta. Käyttämällä RDF:ää ja SPARQL:a on mahdollista yhdistää monimuotoista informaatiota heterogeenisistä sosiaalisista verkkosivustoista, mikä mahdollistaa parannetun navigoinnin ja kyvyn hakukyselyille. (Breslin, Passant ja Vrandečić 2011, s. 499)

Sosiaalisen semanttisen webin toteuttaminen kohtaa haasteita. Koska semanttisen webin pyrkimyksenä olisi luoda yhteenliittymiä sosiaalisten sivustojen välille, niin niiden tämänhetkinen eristäytyneisyys muista samankaltaisista sivustoista on iso rajoite tavoitteen toteutumiselle. Erilaiset sosiaaliset sivustot voivat sisältää täydentävää tietoa, aihealueita tai osapaloja henkilön etsimästä vastauksesta, mutta vain yhden web-sivuston toimintaan osallistuvilla henkilöillä ei ole valmista pääsyä asiaankuuluvaan tietoon muista paikoista. Esimerkiksi henkilö voi etsiä tietoa jostakin asiasta, löytäen ripoteltua tietoa useista eri sosiaalisen median verkkosivustoista, mutta ei saa esille näkymää, jossa kaikki tämä informaatio on koostettu yhteen eri palveluista. Kun uusia sivustoja, yhteisöjä ja palveluita tulee webiin saataville, tulee niiden keskinäinen vuorovaikutuksen puute hyvin selkeästi esille. (Breslin, Passant ja Vrandečić 2011, s. 470)

Merkittävä syy miksi sosiaaliset verkkopalvelut pysyvät eristäytyneinä on se, että palvelut voivat täten sitoa käyttäjät sivustoonsa. Toinen syy liittyy itse sosiaalisen webin luonteeseen: se luo yksittäisiä tietovarastoja, jotka eivät pysty toimimaan keskenään toistensa kanssa. Tä-

män lisäksi sosiaalisessa webissä synergiat ovat kalliita hyväksikäyttää sekä uudelleenkäyttö ja datan yhteenliittäminen on vaikeaa ja hankalaa. Koska käyttäjä ei voi liikkua sosiaalisesta verkkopalvelusta toiseen menettämättä ennen lisättyä ja ylläpidettyä informaatiota ja historiaa, sosiaalinen verkkopalvelu voi lukita sen käyttäjät rajoittamalla yhteenliittämismahdollisuuksia muista palveluista. Tosin sivustot voivat käyttää avoimia rajapintoja jakamaan sivuston tietoja muille taloudellisista syistä (lisääntynyt hakuliikenne ja mainontapotentiali) tai käyttäjäkeskeisistä syistä (henkilökohtaisen tiedon siirto muihin palveluihin). On myös huomioitava, että monirelaatioidatayhdistelmien tekeminen vapaasti saataville tuo esille niin hyötyjä (tutkimustyö) kuin myös ongelmia (yksityisyydensuoja). (Breslin, Passant ja Vrandečić 2011, s. 470 ja 499)

Semanttisen webin teknologioita voidaan käyttää myös parantamaan bloggausympäristöä. Tästä syntyvää tulosta nimitetään **semanttiseksi bloggaukseksi**, joka tähtää perinteisen bloggaamisen rikastamiseen käyttämällä metadatan rakenteesta (mikä liittyy mihin ja kuinka) ja sisällöstä (mistä tämä viesti kertoo). Tämä parantaisi minkä tahansa saatavilla olevan blogiviestin rakennetta parempaan linkittämiseen ja uudelleenkäyttöön. Steve Cayzerin mukaan bloggausympäristö paranisi myös kahdella muulla tavalla: jaettuja ontologioita hyödyntämällä saadaan niin blogiviestin metadataan kuin sen aihealueisiin rikkaampi rakenne sekä kyselyiden navigoiminen ja tiedon löytäminen parantuu. Laajentamalla dataa lisätyillä rakenteilla ja sisältöön liittyvällä metadatatalla voidaan mahdollistaa uusia tapoja tehdä kyselyitä ja navigoida blogidataa. (Breslin, Passant ja Vrandečić 2011, s. 484) (Cayzer 2004)

3.4 Linkitetty data

Linkitetty data tarkoittaa webiin julkaistavaa dataa, joka on ”*koneluettava, sen merkitys on selkeästi kuvailtu, se on linkitetty muihin ulkopuolisiin datajoukkoihin ja siihen voidaan tehdä linkitys muista ulkopuolisista datajoukoista*”. Sen perusideana on luoda webiin tyytetyjä linkkejä eri lähteissä olevien datojen välille. Linkkejä voidaan tehdä esimerkiksi kahden fyysisesti eri sijainneissa olevien datalähteiden välille tai myös heterogeenisten järjestelmien välille, missä kyseiset datajoukot eivät toimi keskenään kovin hyvin. Linkitetyn data-avaruuden keskeinen ominaisuus on sen avoimuus: kuka tahansa voi julkaista sinne dataa ja datan julkaisijoita ei ole rajoitettu esitettävän datan sanastovalinnoissa. Kun nykyiset

HTML-dokumentit on linkitetty ei-tyypitettyillä hyperlinkeillä, linkitettyssä datassa dokumentit sisältävät dataa RDF-muodossa. RDF:ää käytetään luomaan tyypitettyjä lausekkeita, jotka linkittävät vapaavalintaisesti asioita maailmassa. Tuloksena syntyvästä kokonaisuudesta käytetään myös nimeä **Web of Data**, jota voidaan pitää yhtenä mahdollisena toteutumana semanttiselle webille. (Bizer, Heath ja Berners-Lee 2009, s. 2 ja 4)

Berners-Lee on esittänyt joukon sääntöjä strukturoidun datan julkaisemiseen ja yhdistämiseen webissä. Niistä käytetään nimeä **linkitetyn datan periaatteet** ja ne tarjoavat perusteet datan julkaisuun ja yhdistämiseen käyttäen webin infrastruktuuria, samalla noudattaen sen arkkitehtuuria ja standardeja. Sääntöjen tarkoituksena on tuoda esille tapaa, jossa kaikki julkaistu data tulee osaksi yhtä globaalia data-avaruutta. Sääntöjä ovat: ”käytä URI:a nimenä asioille”, ”käytä HTTP:n URI:a, jotta ihmiset voivat tarkastella niitä”, ”kun joku katsoo URI:a, tarjoa hyödyllistä tietoa käyttäen RDF- ja SPARQL-standardeja” sekä ”sisällytä linkkejä toisiin URI:in, jotta on mahdollista löytää lisää asioita”. Linkitetyn datan periaatteet voidaan myös tiivistää seuraavasti: yhdistävä datamalli, standardoitu datan pääsymekanismi, hyperlinkkipohjainen datan löytäminen sekä itsekuvaava data. Terminä ”linkitetyn datan” voidaan sanoa myös viittaavan näihin edellä mainittuihin käytänteisiin. (Bizer, Heath ja Berners-Lee 2009, s. 1-2) (Heath ja Bizer 2011) Kaiken kaikkiaan linkitetyn datan periaatteet rajoittavat RDF:n ilmaisuvoimaa ja sen tarjoamia mahdollisuuksia, mutta helpottavat linkitettyä dataa hyödyntävien sovellusten rakentamista.

Linkitetty data-avaruus on yleisen web-arkkitehtuurin päälle rakennettu kokonaisuus, jolla voidaan jakaa rakenteista dataa globaalilla mittakaavalla (Hausenblas et al. 2012, s. 1). Se nojaa kahteen keskeiseen web-teknologiaan, jotka ovat **URI** ja **HTTP**. URI tarjoaa geneerisen tavan tunnistaa maailmassa esiintyviä entiteettejä. URI:lla identifioituja ja HTTP-skeemaa käyttäviä entiteettejä on mahdollista tarkastella lähdeviittaamalla URI HTTP-protokollan yli. Tällä tavalla HTTP-protokolla tarjoaa yksinkertaisen, universaalien mekanismin resursien tarjoamiseen. Kun URI ja HTTP täydennetään RDF-datateknologialla, HTML tarjoaa välineet rakentaa ja linkittää dokumentteja webissä ja RDF tarjoaa geneerisen, graafipohjaisen datamallin webissä olevan datan rakentamiseen ja linkittämiseen. RDF-malli koodaa dataa tripletteinä ja URI:a käytetään kunkin tripletin osan esittämiseen: subjekti ja objekti identifioivat lähteen tai URI:n ja merkkijonon, kun taas predikaatti tarkentaa, kuinka sub-

jekti ja objekti ovat samaistuneet. Kun käytetään HTTP:n URI:ja tunnistamaan resurssit, HTTP-protokollaa hakumekanismina ja RDF-datamallia esittämään resurssikuvauksia, linkitetty data rakentuu suoraan yleiseen webin arkkitehtuuriin. (Bizer, Heath ja Berners-Lee 2009, s. 3-4) Koska datan linkittäminen jaettuna webiin tarvitsee standardin mekanismin spesifioimaan objektien yhteyksien olemassaolon ja tarkoituksen kuvattuna datassa, on RDF oiva mekanismi linkittämään asioita myös siksi, että RDF-linkit ovat tyyppitettyjä (Heath ja Bizer 2011).

```
Subject: http://dig.csail.mit.edu/data#DIG
Predicate: http://xmlns.com/foaf/0.1/member
Object: http://www.w3.org/People/Berners-Lee/card#i

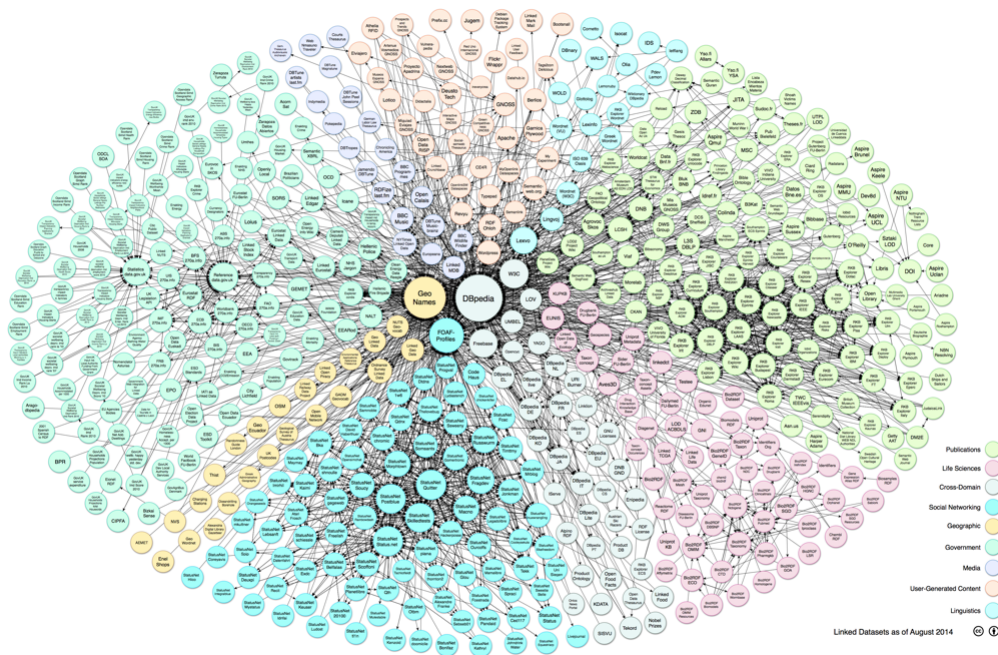
Subject: http://data.linkedmdb.org/resource/film/77
Predicate: http://www.w3.org/2002/07/owl#sameAs
Object: http://dbpedia.org/resource/Pulp_Fiction_%28film%29
```

Kuvio 6: Esimerkki RDF-tripletistä. (Bizer, Heath ja Berners-Lee 2009, s. 4)

Linkitetty data-avaruus hyödyntää myös muitakin teknologioita. SPARQL-kyselykieltä käytetään rakennettuihin kyselyihin (Grimnes et al. 2012, s. 107) sekä RDFS ja OWL tarjoavat perustuksen luoda **sanastoja**. Ne ovat luokkia ja ominaisuuksia sisältäviä kokoelmia ja niitä voidaan käyttää kuvaamaan entiteettejä maailmassa ja kuinka ne liittyvät toisiinsa. Sanastot itsessään on ilmaistu RDF-muodossa ja ne käyttävät termejä RDFS:stä ja OWL:stä, kumpikin tarjoten vaihtelevia asteita ilmaista kohteena olevia aihealueita. Sanastoja voidaan yhdistää RDF-triplettien avulla, jotka linkittävät luokkia ja ominaisuuksia yhdestä sanastosta muihin, täten kuvaten verkkoa samaistuneiden sanastojen välille. Sanastoja voi julkaista vapaasti linkitettyyn data-avaruuteen. (Bizer, Heath ja Berners-Lee 2009, s. 4)

Linkitetylle datalle on perustettu projekti, joka tähtää semanttisen datan vapaaseen saatavuuteen kaikille. Tästä projektista käytetään nimeä **Linked Open Data**. Sen tarkoituksena on myös tarjota aloituspisteitä riippuvuussuhteiden poimimiseen informaatioresurssien keskuudesta. (Waitelonis ja Sack 2011, s. 648) Projektin idean voidaan sanoa myös olevan ”*yhteisöllinen pyrkimys webissä olevien jäsenettyjen datajoukkojen julkaisuun ja linkitykseen*” (Grimnes et al. 2012, s. 107). Projektin seurauksena webiin on jo julkaistu suuri määrä Linked Open Dataa, sisältäen dataa erilaisista kohdealueista kuten media, maantiede, julkaisut,

hallinto ja tiede (Haase, Schmidt ja Schwarte 2011, s. 1). Kun datan määrä kasvaa, syntyy kasvava tarve tuottaa ihmiskäyttäjille tehokkaita mekanismeja datan hakemiseen, tutkimiseen ja kuluttamiseen (Bozzon et al. 2011, s. 109). Kasvaessaan kokoa ja aihealueiden saatavuutta, tämä data tulee yhä kiinnostavaksi myös innovatiivisten sovellusten rakentamiseen, jotka integroivat heterogeenistä dataa eri lähteistä (Haase, Schmidt ja Schwarte 2011, s. 1).



Kuvio 7: Kuva linkitetystä data-avaruudesta.

(*Linking Open Data Cloud Diagram 2014*)

Linkitetty data mahdollistaa uudentyyppisten sovellusten luomisen. Tällaisia ovat esimerkiksi generiset linkitetyn datan selaimet ja hakukoneet sekä kohdealueespesifit sovellukset. Sovelluskehitysnäkökulmasta linkitetyllä datalla on mielenkiintoisia piirteitä: data on tiukasti eroteltu alusta- ja esitysaspekteista, data on itsekuvailevaa sekä HTTP:n ja RDF:än käyttö yksinkertaistaa dataan pääsyä verrattuna webin API:hin. Linkitetty data on myös avointa, joten sovelluksia ei tarvitse toteuttaa kiinteissä datalähteissä, vaan ne voivat löytää uusia datalähteitä ajon aikana RDF-linkkejä seuraamalla. Rajattoman globaalien data-avaruuden päällä toimiminen mahdollistaa linkitetyn datan sovellusten tarjota enemmän valmiimpia vastauksia uusien datalähteiden ilmestyessä webiin. (Bizer, Heath ja Berners-Lee 2009, s. 1, 2 ja 4) Muihin webissä oleviin strukturoituihin datoihin verrattuna taas linkitetty data tarjoaa yksit-

täisen standardoidun pääsymekanismiin. Tämä parantaa sen yhteensopivuutta verrattuna monimuotoisiin käyttöliittymiin ja tulosformaatteihin, mikä tekee linkitetystä datasta korkeasti yhteensopivan. (Waitelonis ja Sack 2011, s. 648)

Linkitetyn datan kehitys avaa uusia mahdollisuuksia tutkimustyölle ja yritystoiminnalle. Se tarjoaa yrityksille mahdollisuuksia rakentaa innovatiivisia sovelluksia tuomalla yhteen yrityksen sisäiset datalähteet ulkoisen datan kanssa, kasvattaakseen ja rikastaakseen sisäisiä tietovarantoja (Haase, Schmidt ja Schwarte 2011, s. 1). Se myös motivoi palveluiden ja linkitetyn datan välisten riippuvuussuhteiden tutkimista (Taheriyani et al. 2012, s. 1). Sille on myös todettu olevan taloudellista potentiaalia, vaikkakin sitä ei ole vielä kunnolla tutkittu (Rey et al. 2012, s. 3). Tutkimustyötä on jo tehty näyttämään linkitetyn datan sovellusten potentiaalia teknisestä näkökulmasta sekä myös tiedonhaun etsimisen vaatimuksien analysointia on tutkittu käyttäjän näkökulmasta (Bozzoni et al. 2011, s. 109). Avoin data mahdollistaisi hallintojen tietoaaineistojen avaamisen ja niiden rikastamisen RDF-muotoon. Kun data on strukturoitua ja riippuvuussuhteet muihin dataresursseihin on eksplisiittisesti selitetty, linkitetty data sallii uuden tiedon löytämisen helposti linkkejä seuraamalla (Sah ja Wade 2012, s. 103). On kuitenkin huomioitava, että linkitetyn datan tutkimuksen huomio on kasvavassa määrin vaihtumassa datan organisoimisesta ja esittämisestä datan linkitykseen ja yhdistämiseen (Bianchini ja Antonellis 2012, s. 283).

Linkitetyn datan kehitys kohtaa haasteita. Jotta linkitetty data voisi nousta suureen suosioon, palveluja kehitettäessä on huomioitava sellaisia asioita, kuten adoptio, uudelleenkäyttö ja datan laatu. Kulutuspuolta paradigmasta ei ole vielä täysin hyväksikäytetty, vaikkakin kansallista kehitystä on jo käynnistetty (*Linkity-seminaari* 2014). Lisäksi täysillä dataan pureutuvien generisten sovellusten vaisuun määrään on osasyynä epäkelpo datan laatu, vaikkakin tätä on osittain pyritty ratkaisemaan. Epäkelpo data yleensä kärsii keskinkertaisesta laadusta erityisesti, kun iso määrä kyseisestä datasta on generoitu automaattisesti web-pohjaisista teksteistä. Tyypilliset ongelmat sisältävät vanhentunutta, epätarkkaa, keskeneräistä, yhteenkuulumatonta tai epämääräistä dataa. **Datan uusiokäyttöön valmistaminen** (engl. *data curation*) on hyvin tärkeä asia linkitetystä datassa, sillä se on osittain välttämätön prosessissa, jossa julkaistaan linkitettyä dataa. Ratkaisuna epäkelpoon dataan olisi joko havaita ongelmallinen data alkuperäisistä datajoukoista kyselyitä luodessa tai sisällyttää loppukäyttäjä

kyselyiden osallistumiseen ja ottamalla vastaan niistä saatu palaute. Tosin on huomioitava, että kyselyiden teko linkitetyssä datassa ei ole mahdollista ilman syvällistä ymmärrystä sen rakenteesta ja kaikkien mukana olevien ominaisuuksien ja luokkien nimistä ja semantiikasta. (Rey et al. 2012, s. 3) Tähän ratkaisuksi voidaan olettaa, että loppukäyttäjille näkyvissä sovelluksissa käyttäjät eivät tiedosta käyttävänsä linkitettyyn dataan pohjautuvaa sovellusta, geneeristen sovellusten sopiessa vain asiantuntijoille.

Linkitettyä dataa hyödyntävien spesifisten sovellusten kehittäminen on ongelmallista, sillä se on aikaa ja kuluja vievää. Kehittäjät kohtaavat ensiksi liudan uusia dataformaatteja ja kyselykieliä (kuten esim. RDF, OWL ja SPARQL). He myös kamppailevat heterogeenisyydestä datatasolla, mikä vaatii monenlaisia uusia tietokantajärjestelmiä ja työkaluja datan varastointiin, prosessointiin ja saatavuuteen. Kun asiaankuuluva data on tunnistettu ja integroitu järjestelmään, linkitetyn datan järjestelmät vaativat uusia dataintegraatioparadigmoja hoitaakseen tietyt haasteet ja mahdollisuudet, kuten skeemajoustavuuden ja datasemantiikat. Loppukäyttäjälle suunnatuissa linkitetyn datan ohjelmistojen käyttöliittymissä on tärkeinä huomioida geneerinen visualisointi, tutkiminen ja interaktioparadigmat linkitetylle datalle. (Haase, Schmidt ja Schwarte 2011, s. 1) Lisäksi, linkitetyn datan prosessoinnin piirissä on havaittu lisävaatimuksia pakollisten vaatimusten (suorittaminen, suoritusteho, skaalautuvuus, joustavuus) lisäksi, kuten URI:t identifioijina, RDF-tuki, käyttöliittymä, levyn osiointi, päivittäminen, indeksointi, häirintä, rikkaan datan prosessointi ja tehokas graafiprosessointi (Hausenblas et al. 2012, s. 2-3).

Linked Open Data -projekti ei myöskään ole ongelmaton. Kun linkitettyä dataa tukevien avointen datajoukkojen määrä kasvaa Linked Open Data -projektin myötä, nykyisistä linkitetyn datan hakukoneista on tulossa entistä tärkeämpiä välineitä löytämään olennaista dataa myöhempää tutkimista varten. Kuitenkin nykyiset hakumekanismit ovat enemmän keskittyneitä tarjoamaan automatisoitua informaatiopääsyä palveluihin ja yksinkertaisia hakutulost listoja käyttäjille. Esimerkiksi ne esittävät hakutulokset vähenevässä tärkeysjärjestyksessä, perustuen johonkin kriteeriin (olennaisuus luokan nimeen). Tuloslista ei kuitenkaan tarjoa tehokkaita sisäänpääsykeinoja linkitetuille dataresursseille, sillä URI:t ja otsikot Linked Open Data -resursseissa eivät ole kovin informatiivisia. Tehokas pääsy ja hyvät etsimismekanismit dataan ovat tärkeitä linkitetylle datalle. Ongelmaa on tutkittu, mutta tarkastelu on asettunut

vasta eri lähestymistapojen tutkimiseen. (Sah ja Wade 2012, s. 103-104)

Linkitettyssä data-avaruudessa dokumenttien lisäksi myös raakadata on altistettu ja linkitetty vapaiden standardien mukaisesti. Tällöin datan edellytetään olevan vapaasti saatavilla ja semanttisesti yhteenkytkettynä. Tällä tavalla julkaistu data sallii datan jakamisen ja uudelleenkäytön valtavalla skaalalla, tarjoten mahdollisuuksia kehittää uusia lisäarvoa tuottavia palveluita ja sovelluksia linkitettyjen datajoukkojen päälle. (Rey et al. 2012, s. 1) Dataa pystytään myös uudelleenkäyttämään, mutta sen edellytyksenä datan on oltava tiettyyn pisteeseen hyvin strukturoitu. Mitä enemmän säännöllinen ja hyvin muotoilu datan rakenne on, sitä helpommin voidaan luoda työkaluja luotettavasti kyseisen datan prosessointiin ja uudelleenkäyttöön. HTML on kuitenkin enemmän orientoitunut tekstuaalisen dokumentin kuin datan rakentamiseen. Kun dataa on ripoteltu ympäröivään tekstiin, ohjelman on vaikeaa poimia rakenteisen datan kappaleita HTML-sivuista. Ratkaisuna tähän ongelmaan on käyttää mikroformaatteja tai web API:ja (tarjoaa tuloksia rakenteisena datamuotona, kuten RDFa, XML ja JSON). (Heath ja Bizer 2011)

Linkitetyn datan päämäärä johtaa globaaliin data-avaruuteen (Bizer, Heath ja Berners-Lee 2009, s. 1) (Rey et al. 2012, s. 1). Siinä resurssit tunnustetaan URI:en kautta, ne ovat itsekuvaltuja (URI-lähdejulkaisumekanismien kautta) ja ne on globaalisti yhdistetty RDF-linkkien avulla (Bianchini ja Antonellis 2012, s. 283). Se ylittää eri datalähteisiin ja mahdollistaa uusien datalähteiden löytämisen sekä sisältää tietoa monimuotoisista aihealueista (Bizer, Heath ja Berners-Lee 2009, s. 2). Linkitetty data on ajanut eteenpäin uusia tapoja paljastaa, jakaa ja yhdistää dataa, informaatiota ja tietoa semanttisessa webissä (perustuen URI:in ja RDF:n). Koska suuri määrä dataa on saatavilla linkitettyssä datapilvessä, kykenevyys pakottaa tekniikoita tehokkaaseen etsintään, tutkimiseen ja visualisoimiseen on tulossa entistä ratkaisemmavaksi. (Castano, Ferrara ja Montanelli 2012, s. 157)

4 Moninäkömähaku

Tässä luvussa kuvataan ja tarkastellaan moninäkömähakuun liittyviä taustoja ja teorioita. Näkömähakuparadigma juontaa juurensa näkömähakuteoriasta, jonka esitteli S. R. Ranganathan jo 1930-luvulla (Colon Classification) (Holi 2010, s. 22) (Peter J. Wild 2009, s. 423). Vuonna 1955 Ranganathanin esittelyn pohjalta, Classification Research Group -ryhmällä oli toiveena nähdä näkömähakuteoriasta perustana kaikelle tiedonhaulle (Fagan 2010, s. 58). 1970-luvusta lähtien näkömähakuteoriaa on sovellettu tiedonhaku tutkimuksissa (Holi 2010, s. 22). Moninäkömähaku tuli osaksi webiä tiedonhaku sovelluksissa 1990-luvun loppupuolella, joista ensimmäinen kehitetty sovellus oli nimeltään HiBrowse, joka kehitettiin suurten lääketieteellisten tekstien kokoelman hakuun. HiBrowseen jälkeen moninäkömähakuteorian ajatusta on hyödynnetty useissa järjestelmissä, kuten Flamenco, Relation Browser ja mSpace. (Hägström 2007, s. 34) (Holi 2010, s. 22) Näkömähakuteoriasta on hyödynnetty myös käyttöliittymäsuunnittelussa (Allen 1995), tietokantasuunnittelussa (Kashyap 2003) ja monimutkaisen informaatioavaruuden mallintamisessa (Crystal 2007) (Slavic 2008, s. 258). Moninäkömähakua on hyödynnetty useilla alueilla, kuten tietojenkäsittelytiede, tiedonhaku, verkkokauppasivustot, tavarakaupat ja kuvakokoelmasivustot. Viime vuosien aikana kasvava määrä ohjelmistoja on alkanut käyttämään moninäkömähakuteorian tarjoamia etuja hakukoneissaan. (Komamizu, Amagasa ja Kitagawa 2012, s. 350-351) (Kules ja Capra 2012, s. 114)

Tämä tutkimus keskittyy olemassa olevien moninäkömähakupalveluiden tarkasteluun. On kuitenkin hyvä tietää, että moninäkömähakuteorialle on olemassa avoimia alustoja, joiden avulla voidaan rakentaa moninäkömähakupalveluita. Näitä ovat esimerkiksi Solr ja ElasticSearch. **Solr**¹ on Apachen kehittämä avoin ja hyvin yleinen moninäkömähakuteorian web-sovelluksissa mahdollistama hakukone. Sen ominaisuuksiin kuuluu muun muassa sanahaku, osuimien korostaminen, moninäkömähaku, reaaliaikainen indeksointi sekä tietokantaintegraatio. **ElasticSearch**² on taas toinen yleistyvä hakukone, joka perustuu samaan indeksointi- ja hakukirjastoon kuin Solr (Apache Lucene). Se hyödyntää REST-API:a ja JSON:ia kyselyissään ja on siten soveltuva monelle ohjelmointikielelle.

1. <http://lucene.apache.org/solr/>

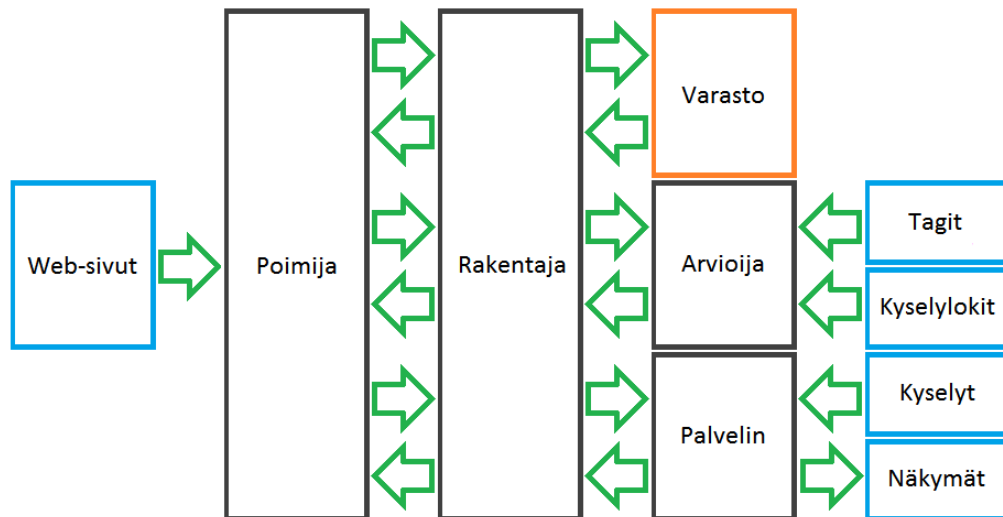
2. <http://www.elasticsearch.org/overview/>

4.1 Määrittely

Moninäkömähauulla tarkoitetaan tekniikkaa tutkia tietokokoelmia käyttäen datajoukon metadatasta koostuvia näkymiä haun suodattajina. Näkymiä käytetään tässä järjestelmässä tarkentamaan haun kysely haluttuun kohteeseen (Koren, Zhang ja Liu 2008, s. 477). Tämänkaltaisen hakutekniikka tarjoaa iteratiivisen tavan tarkentaa hakutuloksia moninäkömäisessä taksonomiassa. Ominaisuudet, kuten hakutuloksen monimuotoisuus, ennalta tarvittavan tiedon tarpeettomuus ja hakutuloksen löytyminen jokaisella kerralla, vähentävät merkittävästi informaatioylivuormaa. (Wei et al. 2013, s. 41) Moninäkömähausta onkin tullut kriittinen ominaisuus parantamaan löytämistä ja käyttäjän hakukokemusta monenkaltaisille hakusovelluksille (esim. verkkokaupat). Se on suosittu ja intuitiivinen interaktioparadigma myös tutkimiseen ja tiedonlouhintaan, sillä se antaa käyttäjälle vahvan työkalun tarkastella, analysoida ja navigoida moniulotteista dataa. (Ben-Yitzhak et al. 2008, s. 33) Moninäkömähakua voidaan myös soveltaa hakuagenttina käyttäjälle, missä agentti päätelee kyselyehdotukset kyselylokianalyysin avulla (Zwol et al. 2010, s. 961).

Van Zwol ja Sigurbjörnssonin tutkimuksessa kuvaillaan MediaFaces-moninäkömähakujärjestelmän järjestelmäarkkitehtuuri. Järjestelmä koostuu viidestä pääkomponentista, jotka ovat näkömävarasto, näkömäpoimija, näkömäärvioija, näkömäpalvelin ja näkömärakentaja. **Näkömävaraston** tehtävänä on hallita purettuja objekteja ja näkymiä sekä objektiparien välisiä riippuvuussuhteita. **Näkömäpoimijan** tehtävänä on kerätä objekteja ja näkymiä luotettavista lähteistä. Näkömäpoimija aktivoituu, kun uusi datalähde tulee saataville tai vanha datalähde päivittyy. Näkömäpoimija käsittelee sisääntulodatan, poimii objektit ja lähettää ne näkömärakentajamoduulille, joka varastoi datan näkömäpalvelimelle. **Näkömäärvioijamoduuli** rankkaa kandidaattinäkömät jaksottaisesti kullekin objektille. Kaikki olemassaolevat näkömät rankataan käyttämällä uusinta tietokannan kyselylokia. **Näkömäpalvelimen** vastuulla on palvelimen ja sovelluksen välinen liikenne. **Näkömärakentajamoduuli** on kommunikointimoduuli näkömävaraston ja kaikkien muiden komponenttien välillä. Kun näkömäpoimija käsittelee uuden lähteen, se lähettää poimitut näkömät näkömärakentajalle, jonka vastuulla on tallentaa ne näkömävarastoon. Kun näkömäärvioijamoduuli aktivoituu, kysyy se näkömärakentajalta listaa näkömistä rankattavaksi. Tämän jälkeen arvioijamoduuli lähettää rankatut näkömät takaisin rakentajamoduulille, joka päivittää näkömien arvot nä-

kymävarastossa. Kun näkymäpalvelin saa syötekyselyn, se kysyy näkymärakentajalta listaa näkymäkandidaateista, joita se voi tarjota käyttäjälle. (Zwol et al. 2010, s. 964-968) Muita moninäkömähaun avainteknologioita edellä mainittujen lisäksi ovat hierarkiarakentaminen ja yhdistelmätermien generointi. (Wei et al. 2013, s. 57-58)



Kuvio 8: MediaFaces-järjestelmän järjestelmäarkkitehtuuri. (Zwol et al. 2010, s. 965)

Moninäkömähaun kautta etsittävä datajoukko koostuu objekteista, jotka voivat olla esimerkiksi kirjoja, dokumentteja tai tavaroita. Näkymät ovat attribuutteja kohdeobjekteille, mitkä ilmaisevat erilaisia ominaisuuksia kuvatusa kohdejoukosta. Moninäkömähakukäyttöliittymän kautta käyttäjät voivat tutkia objekteja interaktiivisella tavalla näkymien avulla. Näkömäkeskeinen navigointijärjestelmä näyttää objektit ja näkymät, jonka jälkeen käyttäjä voi valita näkymän käyttöliittymästä. Moninäkömähakukäyttöliittymä noutaa objekteja, jotka täsmäävät kokoelman kanssa ja näyttävät ne objektit ja näkymät, jotka assosioivat tulosobjektien kanssa. Jos käyttäjä ei ole tyytyväinen hakutulokseensa, käyttäjä voi valita uuden näkymän tai kumota nykyisen näkymän, jolloin käyttöliittymä päivittää itsensä valittujen uusien arvojen perusteella. Tämä prosessi toistetaan, kunnes käyttäjä on löytänyt tarvittavan hakutulokokoelman haetusta aiheesta. (Komamizu, Amagasa ja Kitagawa 2012, s. 350)

4.2 Näkymät

Moninäkömauksessa ortogonaalisten tai itsenäisten taksonomioiden luokittelun dimensioita kutsutaan **näkymiksi** (Oren, Delbru ja Decker 2006, s. 560) (Holi 2010, s. 22). Hearst (Hearst 2008, s. 1) luokittelee näkymät ”joukoksi tärkeitä lipukkeita järjesteltynä sellaisella tavalla, joka peilaa aihealueeseen liittyviä konsepteja”. LaBarre (La Barre 2007, s. 82) taas määrittelee näkymät esittämään ”kategorioita, ominaisuuksia, attribuutteja, tunnusmerkkejä, riippuvuussuhteita, funktioita tai konsepteja, jotka ovat keskeisiä tälle organisoidulle joukolle dokumentteja ja entiteettejä, mitkä ovat erityisen mielenkiintoisia tälle käyttäjäryhmälle”. (Fagan 2010, s. 58) Näköma sisältää niin kutsuttuja hakukategorioita: itsenäisiä konsepteja, jotka kuuluvat näkömiin. Näkömien organisointi ja rakenne taas paljastavat riippuvuussuhteita eri konseptien ja valitun aiheen välillä. (Holi 2010, s. 22) Näkömille on olemassa useita eri esitystapoja (kuten hierarkia, yksitasoinen listä, kartta tai jana) (Hägström 2007, s. 33) sekä olemassa olevia näkömämalleja (joukkoteoriapohjainen, FCA (*Formal Concept Analysis*) ja kevytrakenteinen ontologia) (Wei et al. 2013, s. 46). Joitakin näkömän hakukategorioita voidaan esittää liukusäätiminä, jossa käyttäjä määrittää näkömälle halutun numeerisen arvon.

Näkömät edustavat useaa funktiota moninäkömauksen käyttöliittymässä. McGuinnessin (McGuinness 2003) mukaan ne toimivat sanastonkontrolloijina, sivuston navigoijina ja tukijoina, kokonaisnäkömän tarjoajina ja ennakkoluulojen asettajana, selaustukena, hakutukena sekä selvennys- ja täsmennystukena. Edellä mainitut funktiot tarjoavat tärkeitä etuja käyttäjälle. Ne käyttävät kategorisointijärjestelmiä, jotka ovat yhtenäisiä ja valmiita sekä ne ovat ennustettavia. Ne näyttävät esikatseluita seuraavista hakupoluista sekä näyttävät, kuinka palata takaisin edelliseen hakutulospäkömään. Ne myös ehdottavat loogisia vaihtoehtotuloksia ja helpottavat käyttäjää välttämään nolihakutuloksia kaventamalla hakua. Toisaalta näkömisissä on myös joitakin haittapuolia. Käyttäjä näkee näkömistä etukäteen mielenkiintoisimmat kategoriat, kun taas tärkeitä trendejä ei välttämättä näytetä käyttäjälle. Näkömien kategoriarakenteiden automaattinen rakentaminen voi olla vain osittain onnistunut, jolloin kategoriarakenteet pitää rakentaa manuaalisesti. (Fagan 2010, s. 58)

Näkömiä voidaan luoda monella eri tavalla. Niitä voidaan luoda käyttämällä olemassa olevaa sanastoa tai tesaurusta. Jos sopiva sanasto on olemassa, sitä voidaan käyttää näkömien

Rajaa hakua

Verkossa saatavilla

Toimiala ▾

Organisaatio ▾

Aineistotyyppi ▾

- ▶ Kirja (92419)
- ▶ Lehti/Artikkeli (72409)
- ▶ Opinnäyte (14773)
- ▶ Äänite (10918)
- ▶ Kuva (10179)
- ▶ Asiakirja (7380)
- ▶ Esine (5583)
- ▶ Muu (2750)
- ▶ Video (2611)
- ▶ Nuotti (1724)
- ▶ Paikka (448)
- ▶ Taideteos (390)
- ▶ Kartta (44)

Aihe ▾

Tekijä ▾

Valmistusvuosi TT ▾

Kavenna hakua

Kategoria

Elokuvat

- [Lasten elokuvat ja sarjat](#) 5
- [Leffaboksit / TV-sarjat](#) 4
- [DVD-elokuvat](#) 1
- [3D Blu-ray elokuvat](#) 1
- [Blu-ray-elokuvat](#) 1

Kamerat

- [Muistikortit kameroille](#) 9
- [Video ja kuvanmuokkaus](#) 6
- [Kameratarvikkeet](#) 3
- [Järjestelmäkamerat](#) 1
- [Action-kameroiden tarvikkeet](#) 1

Näytä lisää ↻

Hinta

Alle 25	173
25 - 49	307
50 - 99	241
100 - 249	278
250 - 499	208
Yli 500	325

Hinta

↻

Kuvio 9: Esimerkki näkymistä ja hakukategorioista Finna- ja Gigantti-sivustoissa.

luomiseen tai niiden osien luomiseen. Jos sopivaa sanastoa ei ole olemassa, näkymäsanasto voidaan luoda manuaalisesti erityistä tarvetta varten. Monessa tapauksessa, korttijärjestelyä tai vastaavaa menetelmää käytetään näkymien luomiseen. Olemassa oleva metadata analysoidaan ja objektien sisältö asetetaan luokitukseen. Jotkin näkymät voidaan luoda automaattisesti käyttämällä olemassa olevaa objektien metadataa. Tämä koskee näkymiä, jotka perustuvat metadataan, joka on liitetty objekteihin standardissa formaatissa (esim. Dublin Core -skeeman mukaisilla metadatakentillä). Kun näkymät on luotu, hakuobjektit indeksoidaan ja annotoidaan näiden näkymien mukaan. Tämä indeksointi voidaan kokonaan tehdä automaattisesti (indeksointi luomispäivän mukaan), mutta pieni osa joudutaan aina tekemään manuaalisesti (indeksointi manuaalisen sanaston mukaan). (Holi 2010, s. 38)

Semanttisessa moninäkömahaussa näkymiä projisoidaan ontologisista rakenteista kaksiosaisessa prosessissa. Ensimmäisessä vaiheessa projisoidaan hierarkkinen kategoriapuu, joka tehdään seuraamalla graafia tiettyjen sääntöjen perusteella. Graafista poimitaan talteen vastaan tulevia oleellisia käsitteitä, jonka jälkeen ne liitetään kategoriapuuhun tietämyskannan suhteiden mukaisesti. Useimmiten suhteet ovat **hyponymioita** (merkitys sisältyy kokonaan toisen käsitteen merkitykseen) tai **meronymioita** (käsitteet merkitsevät tarkasteltavan kä-

sitteen osaa). Toisessa vaiheessa dokumentit yhdistetään projisoituihin kategorioihin. Tämä tehdään määrittelemällä ominaisuus (predikaatti), jolla dokumentin on viitattava kategoriana käytettyyn resurssiin (tehdään sääntöjen perusteella). Tämä prosessi ei oleellisesti muuta hakua verrattuna tavalliseen moninäkömähakuun, vaan haku edelleen tapahtuu rajaamalla tulosjoukkoa kategoriavalintoja yhdistelemällä. (Häggström 2007, s. 34-35)

4.2.1 Kontekstuaaliset näkymät

Kontekstuaaliset näkymät (engl. *contextual facets*) ovat erilainen interaktioteknikka moninäkömähauille. Ne sallivat käyttäjän katsoa ja valita vaihtoehtoisia arvoja tiettyyn osaan web-sivustoa. Ne myös muuntavat staattiset web-sivuelementit käyttöliittymäkomponenteiksi, joita voidaan käyttää suodattamaan ja noutamaan samankaltaisia web-sivuja ilman erillistä hakukyselykäyttöliittymää. Kontekstuaaliset näkymät käyttävät hyödyksi olemassa olevan, tutunoloisen objektin yksityiskohtasivun mallia. Ne tarjoavat käyttäjille helpon käyttöliittymän, jolla on pääsy laajaan valikoimaan metadatakategorioita. Valittaessa arvon kontekstuaalisesta näkymästä web-sivu suorittaa kyselyn, joka sisältää valitun arvon. Esimerkiksi kun käyttäjä tarkastelee objektia, klikatessaan objektin tiettyä metadataspesifikaatiota näkee hän muiden saatavilla olevien objektien vastaavan metadatan. (Medynskiy, Dontcheva ja Drucker 2009, s. 2014-2015)

Kontekstuaaliset näkymät tarjoavat muutaman avainhyödyn olemassaoleviin moninäkömähakukäyttöliittymiin. Navigointikäyttöliittymä on upotettu yksityiskohtasivulle, jolloin käyttäjän ei tarvitse oppia erillistä käyttöliittymää objektien selaamiseen tai hakemiseen. Käyttäjät voivat heti selata tai suodattaa heille persoonallisesti olennaisia kategorioilta, kun he huomaavat ne objektien yksityiskohtasivuissa. Potentiaalisesti käyttäjät voivat myös katsoa vähemmän suosittumista kategorioista joko piilossa olevista hakuelementeistä tai kehittyneemmistä hakukäyttöliittymistä. Ne tarjoavat myös tehokkaan tavan navigoida web-sivua käyttäjille, joka päätyy suoraan objektin yksityiskohtasivulle: käyttäjällä on mahdollisuus heti aloittaa navigointi muihin sivuston vaihtoehtoihin. (Medynskiy, Dontcheva ja Drucker 2009, s. 2014-2015)

Kontekstuaalisista näkymistä on myös muitakin hyötyjä. Hutchinsin, Hollansin ja Norma-

nin tutkimuksen mukaan kontekstuaaliset näkymät vähentävät kognitiivista taakkaa, mitä tarvitaan käyttäjältä siirtämään hänen tavoitteensa syötteeksi moninäkömahakukäyttöliittymälle. Kun kyselynrakentamiskäyttöliittymä yhdistetään objektin yksityiskohtakäyttöliittymään, se helpottaa käyttäjää kaventamaan toteutukseen tarvittavaa kuilua. Käyttäjät voivat myös nopeasti arvioida, onko objektista näytettävä attribuutti näytettynä yksityiskohtasivulla otettavissa kyselyyn mukaan ja siten mahdollista käyttää sitä. Tämä auttaa käyttäjää muodostamaan käsityksen, auttaako kyseinen käyttöliittymäelementti häntä pääsemään haluttuun maaliin ja minimoi järjestelmältä tarvittavaa työpanosta ilmaista tämä tavoite. (Medynskiy, Dontcheva ja Drucker 2009, s. 2015-2016)

Kontekstuaalisten näkymien potentiaalia moninäkömahakukäyttöliittymän osana on arvioitu laadullisessa tutkimuksessa (Medynskiy, Dontcheva ja Drucker 2009, s. 2017-2018). Sen tuloksissa todettiin, että kontekstuaalisten näkymien käyttö paransi keskiverroksi testajia tekemään parempia päätöksiä ja nopeammin. Tosin kaikilla testattavilla oli jotakin hämmennyksiä kontekstuaalisten näkymien käytöstä, vaikkakin suurin osa käyttäjistä onnistui käyttämään niitä onnistuneesti tiedonsuodattamisessa, ja useimmat käyttivät sitä pääkäyttöliittymänä. Testattavat pitivät käyttöliittymästä siksi, että se tarjosi paremman käyttöliittymän selaamiseen ja hakemiseen verrattuna tutkimuksessa tutkitun ohjelmiston natiiviin käyttöliittymään. Se salli heidän hakea suoraan yksityiskohtasivulta ilman, että tarvitsi selailta yksityiskohtasivujen ja hakutulossivujen välillä. Se laittoi myös käyttäjät tutkimaan useita eri vaihtoehtoja nopeammin sekä se sisälsi muodoiltun ja informatiivisen esikatselun sivupalkissa. Vaikka käyttöliittymästä pidettiin, kolmasosa käyttäjistä ei nähnyt mitään lisäarvoa käyttöliittymässä eikä kontekstuaalisissa näkymissä.

4.2.2 Näköluokittelu

Moninäkömahaku perustuu metadataan, joka on liitetty haettuihin objekteihin. Metadata järjestetään ortogonaalisista kategoriajoukoista muodostettujen näkymien mukaan. Tätä metadatan järjestelyä kutsutaan **näköluokitteluksi** (engl. *faceted classification*). Näköluokittelussa haetut objektit luokitellaan jokaisen ortogonaalisen näkymän mukaisesti. (Holi 2010, s. 37-38) Jokainen objekti kuuluu useampaan kuin yhteen näkymän kategoriaan, tosin jotkut näkymät saattavat sallia vain yhden kategorian objektia kohden. Näköluokittelu

eroaa huomattavasti esimerkiksi Dewey'sin tai YKL:n (*Kirjastot.fi* 2014) luokitteluista, joissa kaikki objektit luokitellaan yhtenäisen taksonomian mukaan ja jokainen objekti kuuluu tasan yhteen kategoriaan taksonomiassa. (Holi 2010, s. 38) Luokittelun tuloksena syntyvän näkymäkokoelman pääpilarit ovat näkymähierarkia ja näkymissä olevien dokumenttien kuvaus (Ben-Yitzhak et al. 2008, s. 34). Näkymiä voidaan myös luokitella rajatusti (engl. *strict facet classification*), jolloin objektit on luokiteltu näkymän sisällä tasan yhteen luokkaan (Wilson 2007, s. 5).

Moninäköinen luokittelutapa perustuu dokumenttien ja olioiden useisiin dimensioihin. Tämä tarkoittaa sitä, että taksonomian käyttö tämänkaltaisissa järjestelmissä on mahdoton, sillä taksonomiassa dokumenteilla ja olioilla on tasan yksi paikka kiinteässä hierarkiassa. (Häggström 2007, s. 33) Luokittelun rakentamista ovat muokanneet teoriat tietämyksen rakenteesta, kuten tavanomainen joukko pääluokkia (ensimmäinen askel yleisen luokittelun hierarkiassa) tai perustavanlaatuinen joukko kategorioita (perusjoukko piirteitä, joita käytetään). Näkymien luokittelun rakenne voi sisältää seuraavat asiat: koko määrittelyalue, lista tai hierarkia aihealueista (subjektit), näkymien erittely, lista tai hierarkia termeistä, alinäkymien erittely sekä lista tai hierarkia alinäkymien termeistä. On kuitenkin tiedettävä, että dokumentin päättäminen, sen subjektit, sen ymmärtäminen sekä sen kuvaileminen muutamalla sanalla ovat alustavia tekoja näkymäluokittelun käytölle. (Vickery 2008, s. 147 ja 156)

Näkymät ovat voimakas tapa ilmaista informaation kompleksisuutta. Niitä voidaan (kuten muitakin konsepteja ja metodeja) käyttää monella eri tavalla erilaisten implisiittisten ja eksplisiittisten filosofisten olettamusten kanssa. (Giess, Wild ja McMahon 2008, s. 389) Niitä voidaan myös järjestää eri tavoin: yksinkertaisesta monimutkaiseen, spatiaalinen tai kronologinen järjestys sekä merkittävyyden, esiintymistiheyden tai aakkosten mukaan (Vickery 2008, s. 155). Luokittelutapoja on olemassa näkymäluokittelun lisäksi muun muassa enumeratiivinen (top-down) luokittelu ja hierarkkinen luokittelu. Nykyään elektroniset laitteet mahdollistavat näkymäluokittelua käyttäville sovelluksille vähemmän sijaintirajoitteita, mitä voidaan pitää etuna, sillä aikanaan kirjastoissa olleet näkymäluokittelusovellukset olettivat, että objektit sijaitsevat fyysisesti vain yhdessä paikassa. (Peter J. Wild 2009, s. 422 ja 426)

Näkymäanalyysissä kompleksinen aihealue jaetaan useampaan osaan. Kun otetaan joukko objekteja tutkittavaksi joltakin tietyltä aihealueelta tai tieteenalalta, tulokseksi saadaan

joukko sanoja, lauseita tai yhdyssanoja, joita voidaan pitää potentiaalisina näkyminä. Kerätyistä yhdyssanoista ja myös käyttäjäsyötteistä saadaan selville, kuinka sanat eri näkymissä voidaan yhdistää. Näkymien luokittelussa pyritään vastavuoroisesti rajattuihin ja itsenäisesti yhdistettäviin näkymiin. Yleensä analyysi tulee aloittaa termeistä ja riippuvuussuhteista, jotka löytyvät subjekteista ja mitkä ovat avuksi käyttäjille. Luokittelua voidaan myös tehdä niin, että on ydinnäkymiä ja näkymiä, jotka ottavat huomioon käyttäjän tarpeet. Kun subjekti-joukon analysointi on tehty ja tarvittava määrä näkymiä on luotu, tulee seuraavaksi käydä läpi koko tähän asti kertynyt sanasto ja lisätä jokainen kerätty termi näkymään. Tämän seurauksena voi tulla ongelmia, kuten synonyymisanat, homonyymisanat, operaatioiden ja entiteettien erottelu sekä tilanne, jossa tiettyä näkymää voidaan käyttää eri konteksteissa tarkoittamaan eri asioita. (Vickery 2008, s. 150-153) Näkymäanalyysi johtaa systemaattiseen löytämiseen, sidesanaisen ja semanttisen rakenteen kokoamiseen. Rakenteen riippuvuussuhteet taas johtavat indeksoijat ja käyttäjät sanaston ympärille ja edistävät osumia kysely- ja objektikuvauksen välille. Nämä suhteet ovat yleensä jaettu laajasti ekvivalenssiin, assosiativisiin ja hierarkkisiin, vaikkakin eri tyyppisiä ja alityyppejä on olemassa satakunta. (Schwartz 2008, s. 833-834)

Moninäkömaailmassa **luokittelulla** tarkoitetaan symbolisten termien hierarkkista rakennetta. Symbolit viittaavat tai edustavat oikean tai kuvitteellisen maailman entiteettejä. Entiteetit tai dokumentit voivat olla tekstejä, kuvia tai ääniä, jotka on kuvailtu teksteillä. Rakennetta taas voidaan käyttää etsimisen apuna tai tiedonpalauttamistyökaluna (osoitetaan haluttuun entiteettiin). Dokumentteja luokitellaan seuraavien vaiheiden avulla. Ensiksi kaikkien dokumenttien avaruudesta valitaan tietty kasa dokumentteja. Dokumentit tai otos niistä tutkitaan, jonka jälkeen kootaan lista niistä piirteistä, joita pidetään merkittävänä luokittelun käyttäjille. Seuraavaksi tehdään päätöksiä siitä, mitä piirteitä käytetään muodostamaan dokumenteista hierarkia. Tämän jälkeen jokainen dokumentti avaruudessa luokitellaan merkittävistä piirteistä ryhmiin asetettavaksi. Dokumenttien kuvaelma rakennetaan, sisältäen symbolisten termien kombinaation (ryhmää, johon se on asetettu). Lopuksi lista dokumenttien piirteistä suljetaan ja jatkuvia laajennus- ja tarkistusprosesseja tehdään jatkossa myöhemmin. (Vickery 2008, s. 145-146)

Luokitteluun on otettava mukaan hyvin valitut sanat käytettäväksi termeiksi. Yksi keino lisä-

tä hakujen onnistumista on kuvata subjektit niin dokumenteista kuin tiedustelijasta yleisiksi termeiksi. Luokittelijan on tunnistettava termit, millä ilmaistaan subjektien ensimmäinen laatiminen. Kun nämä on tunnistettu, seuraa dokumenttien subjektien lisääminen. Menettelytapa on yleisesti samanlainen kuin lisättäessä termejä, alinäkyviä ja alinäkyvien termejä. Sille on myös välttämätöntä olla mekanismi, joka osaa lisätä luokitteluun lisättävälle termille sen ylempitasoiset termit luokitteluketjuun. Web-ympäristössä on myös huomioitava, että jokaista dokumenttia on kohdeltava erillisinä dokumentteina, jotta ristikkäisyhdistelmiä ei tapahtuisi. Uusien dokumenttien luokittelu tuo jatkuvasti esille uusia synonyymejä, termejä ja jopa näkymiä, joita voi käyttää luokittelussa. Uudet mahdolliset termilisäykset laiteaan joko listaan tai lisätään aliluokaksi hierarkiaan. On myös tärkeää, että paikallinen versio näkymäluokittelusta kykenee reagoimaan nopeasti uusiin piirteisiin aihealueesta, jossa sitä käytetään: tulisi siis olla mahdollisuus ilmoittaa uusista termeistä paikalliselle vastuussa olevalle toimijalle, joka päivittäisi skeeman. (Vickery 2008, s. 156-157)

Kun yksittäinen luokka on tehty, se voidaan uudelleenesittää **ekspressiivisen notaation** kautta. Siinä luokan kuvaileva nimi (notaatio) voidaan erottaa sen varsinaisesta tunnisteesta ja rakenteesta luokituksesta, mikä on luontevaa ontologioiden ja linkitetyn datan kontekstissa. Kun kaikki elementit skeemarakenteessa ovat koneluettavia, näyttö ja luokittelun käyttö tulevat itsenäiseksi notaatiosta. Teoriassa notaatiojärjestelmän voi ottaa pois ja korvata tai linkittää sen rinnakkaisjärjestelmään. Ekspressiivinen notaatio voi olla myös hyödyllinen työkalu uusia luokitteluja suunniteltaessa, koska sille on helppo tapa tallettaa rakenteellisia sääntöjä ja pakottaa jäykkyyttä (engl. *rigidity*) dataesityksiin. Tietokannassa tätä voidaan myös hyödyntää täydentävänä työkaluna ohjaukselle ja validoinnille tai kun luodaan koneavustettu indeksointityökalu. Ekspressiivistä notaatiojärjestelmää voidaan käyttää digitaalisten resursien indeksoinnissa, missä notaation pituutta ei pidetä heikkoutena. Tällaisessa järjestelmässä ihmisen ei tule käsitellä notaatiota millään tavalla. (Slavic 2008, s. 263-264)

Koska entiteeteillä on luontaisesti monia piirteitä, tärkeiden ja merkittävien piirteiden määrittäminen ja arvottaminen voi olla ongelmallista. Lisäksi määritettyjen piirteiden tulisi olla käyttäjälle avuksi. Yksi keino ratkaista ongelma on eri tieteenalojen tekijöiden yhdessä päättetyt piirteet, joita he ovat itsenäisesti pitäneet merkittävänä ja saaneet hyväksynnän termeistä muilta alan tekijöiltä. Ei-akateemisilla aloilla yleensä dokumenttien tekijät ilmaisevat ja

perustelevat lukijalle tiettyjen termien tärkeyden. Tosin näitä näkökulmia voi muokata käyttäjäkeskeinen lähestymistapa, missä luokittelu on suunniteltu tietyille käyttäjäryhmälle. Heidän kiinnostustensa tutkiminen voisi määrittellä, mitä termejä tulisi lisätä luokitteluun ja luokitukseen. (Vickery 2008, s. 147) Kiinnostusten tutkiminen onnistuisi mahdollisesti folksonomioiden avulla, folksonomioiden määrittäessä käyttäjien käyttämän sanaston. Tämä avaisi tietä käyttäjäkeskeiselle luokittelulle, missä käyttäjille todella esitettäisiin avustavia piirteitä sivuston entiteeteistä.

Suunniteltaessa luokittelua dokumenteille tulee niissä kuvatut ilmiöt ottaa monipuolisesti huomioon. Monien dokumenttien kanssa joudutaan käsittelemään koostesubjekteja ja jotkin subjektit voivat olla hyvin monimutkaisia, sisältäen usean attribuutin yhdistelmän. Esimerkiksi ”*ekologinen tutkimus aavikkoeläinten ihonväreistä*” sisältää useita eroteltavia attribuutteja. Silloin kun luokitellaan objektista kirjoitettua asiaa, joudutaan ottamaan huomioon muun muassa sisällytettyjen termien ainesosa, sen käyttäytyminen, sen interaktiot muiden objektien kanssa, sille toteutettavat operaatiot ja sen ympäristö. Esimerkiksi termi ”*lääkitys*” voidaan kuvata ”*henkilön toimintoina terveyden ylläpitoon tai sairauden parantamiseen*”, missä subjektin määritelmä johtaa itse aiheeseen ja kategoriat (henkilön tyyppi, anatominen osa, prosessit, operaatiot, operaation osatekijät) tarkentavat sitä. Näiden kaikkien termien yhdistelmä antaa kyvyn täsmätä subjekti yksittäiseen dokumenttiin. Haun aikana käyttäjälle on erotettava tietty näkökulma dokumenteista, mikä esitetään hänelle luokitteluna. Nämä erikseen listatut näkökulmat tunnetaan siis näkyminä. (Vickery 2008, s. 147-148)

4.3 Käyttöliittymä

Moninäkömahaun käyttöliittymä on yksi sen tunnistavista tekijöistä käyttäjälle (kts. 10). Moninäkömähakukäyttöliittymä sisältää yleensä sellaisia komponentteja, kuten listan eri näkymistä ja niiden rajoittuvuusarvoista, hakutulosten listan ja kyseistä hakutulospolkua ilmaisevan ”*leivänmurun*” (Hearst 2008). Käyttöliittymä näyttää kokoelman objekteja ja monia näkymiä datasta tai metadatasta sekä sallii käyttäjien parantaa tai laajentaa hakukyselyä esitettävien kategorioiden mukaisesti (Medynskiy, Dontcheva ja Drucker 2009, s. 2014). Käyttöliittymä näyttää myös kunkin kategorian kohdalla valinnasta seuraavan hakutulosten määrän, joten käyttäjälle selviää hyvin, kuinka paljon käyttäjän haku rajautuu kunkin hakupolun

yhteydessä. Moninäkömähakuun liittyvät ideat ja inspiraatiot tulevat dynaamisista hakukäyttöliittymistä. (Suominen 2008, s. 3-4)

hakupolku tai "leivänmuru"

JYX — Etusivu → Haku JYX:stä

Lisää hakuehto

Nimeke

Add

Valitse hakuehto

Tekija
 Jyväskylän yliopiston ylioppilaskunta (12)

Asiasana
 opiskelijalehdet (12)
 Jytkkari (9)
 Jyväskylän kasvatustieteellinen korkeakoulu (8)
 Jyväskylän yliopisto (6)
 Jyväskylän yliopiston ylioppilaskunta (8)
 Jyväskylän ylioppilaslehti (6)
 Niitelmä (8)
 Ivikkari (8)
 Jyväskylän kasvatustieteellinen korkeakoulu (8)
 Jyväskylän yliopisto (8)
 Jyväskylän yliopiston ylioppilaskunta (8)
 Jyväskylän ylioppilaslehti (8)
 Niitelmä (8)
 Ylioppilaskunnat (8)
 Jyväskylä (4)
 ...Lisää

Julkaisuaika
 2000 - 2010 (4)
 1970 - 1979 (6)
 1960 - 1969 (2)

Hakutulokset

Hakusana(t): ohjelmointi **Hae**

Valitut hakuehdot:
 Tekijä: Jyväskylän yliopiston ylioppilaskunta Asiasana: opiskelijalehdet

Tuloksia per sivu: 10 Järjestysperuste: relevanssi Järjestysuunta: laskeva

Huulla ohjelmointi löytyi 12 osua(a).

Näytetään nimekkeet 1-10 / 12 1 2 Seuraava sivu

Jyväskylän ylioppilaslehti 04/1977
 Jyväskylän yliopiston ylioppilaskunta (1977)

Jyväskylän ylioppilaslehti 03/2000
 Jyväskylän yliopiston ylioppilaskunta (2000)

Jyväskylän ylioppilaslehti 11/1968
 Jyväskylän yliopiston ylioppilaskunta (1968)

Jyväskylän ylioppilaslehti 11/1969
 Jyväskylän yliopiston ylioppilaskunta (1969)

Jyväskylän ylioppilaslehti 02-03/1972
 Jyväskylän yliopiston ylioppilaskunta (1972)

Jyväskylän ylioppilaslehti 08-09/1975
 Jyväskylän yliopiston ylioppilaskunta (1975)

Jyväskylän ylioppilaslehti 09/1976
 Jyväskylän yliopiston ylioppilaskunta (1976)

Jyväskylän ylioppilaslehti 20/1978
 Jyväskylän yliopiston ylioppilaskunta (1978)

Jyväskylän ylioppilaslehti 05/1979
 Jyväskylän yliopiston ylioppilaskunta (1979)

näkymiä ja niiden rajoittuvuusarvoja

hakutulostista

Kuvio 10: Esimerkki moninäkömähakukäyttöliittymästä JYX-sivustossa.

Moninäkömähakukäyttöliittymä omaa tietynlaisia ominaisuuksia. Se esittää käyttäjälle näkymiä, jotka mahdollistavat indeksoitujen sisältöjen selaamisen minkä tahansa valitun sisällön kautta. Näkymien organisointi ja rakenne taas paljastavat riippuvuussuhteita eri konseptien ja valitun aiheen välillä. (Holi 2010, s. 22) Moninäkömähauksen yksi keskeinen ominaisuus on se, että käyttöliittymän kautta käyttäjä ei voi päätyä tulossivulle, joka ei sisällä yhtään hakutulosta. Yleensä tällaiset hakutulokset piilotetaan käyttäjän näkyvistä. (Suominen 2008, s. 4) Tosin on huomioitava, että jos käyttäjä sisällyttää hakuunsa vapaavalintaisia hakusanoja, tällöin haku ei välttämättä tuota tulosta. Tästäkin huolimatta moninäkömähaku soveltuu erityisesti hyvin analyysiin ja tukee hyvin käyttäjän tutkivaa tiedonhakua, sillä se mahdollistaa aineiston selauksen monesta eri näkökulmasta ja paljastaa erilaisia riippuvuussuhteita aineiston objektien välillä (Suominen 2008, s. 1).

Moninäkömähauksen käyttöliittymän suunnittelua voidaan lähestyä kolmivaiheisen, käyttäjän-

keskeisen käyttöliittymäsuunnitelun kautta. Nämä kolme vaihetta ovat analyysi-, suunnittelu- ja arviointivaihe. Analyysivaiheessa on tärkeää ymmärtää järjestelmän käyttäjien tavoitteet ja käyttökonteksti. Nämä saadaan selville luomalla käyttäjäpersoonia ja erilaisia käyttökäytännöitä. Käyttöliittymän suunnitteluvaihe koostuu yleensä inkrementaalisiin keinoihin toteutuvasta prototyypisuunnittelusta. Samalla jokaisen version yhteydessä kokeillaan ja arvioidaan toteutettuja vaihtoehtoja. Paperi- tai sovellusprototyypit ovat keinoja, joita voidaan käyttää järjestelmän suunnittelussa. Käyttöliittymän arviointia voidaan toteuttaa käyttäjä- tai asiantuntijatestauksella. Asiantuntija-arvioinnissa käytetään keinoja, kuten heuristista arviointia, kognitiivista arviointia ja simulointitestausta. Näitä menetelmiä voidaan käyttää jo varhaisessa suunnitteluvaiheessa kohdentamatta suurta resurssimäärää prosessiin. Käyttäjättestausmenetelmiä ovat käytettävyydestit, ja niihin voidaan sisällyttää vertailutestejä sekä ymmärrettävyydesteitä. (Suominen 2008, s. 6-7)

Perinteisiin moninäkömähakukäyttöliittymiin on kehitetty parannuksia. Wilson, Andre ja Schraefel esittelivät takaisinpaluun korostuksen, joka helpottaa käyttäjiä pitämään muistissa näkymien välistä assosiointia suunnatussa moninäkömähakussa. Huynh, Miller ja Karger taas kehittivät kaksi järjestelmää, jotka automaattisesti generoivat kustomoituja moninäkömähakukäyttöliittymiä puolirakenteisen datan selaamiseen. Puolirakenteiselle datalle on myös tehty järjestelmiä, jotka tarjoavat innovatiivista interaktiivisuutta web-sivuille. Näitä ovat Zoetrope³, missä käyttäjät vuorovaikuttavat web-sivuihin niiden muuttuessa ajan mittaan ”*linssien*” avulla (jotka ovat samankaltaisia kontekstuaalisille näkymille), ja MashMaker⁴, joka tarjoaa työkaluja datan kuvaamiseen ja uudelleenmiksaamiseen. (Medynskiy, Dontcheva ja Drucker 2009, s. 2015)

Moninäkömähakukäyttöliittymissä on havaittu myös puutteita. Ohjatut navigointikäyttöliittymät hyötyisivät suuresti, jos ne tarjoaisivat rikkaamman metadatatiedon haettavasta datasta. Kyky nähdä joustavia ja dynaamisia koosteita näkymädatan yli (joita esiintyy business intelligence -sovelluksissa, jotka käsittelevät strukturoitua dataa) antaisi käyttäjien tehdä enemmän informatiivisia tihennys- ja yleistysvalintoja, mitkä taas tukisivat heitä parempaan päätöksentekoon. Toinen puute moninäkömähakussa on se, että sen perusdatamalli, jossa do-

3. <http://www.cond.org/zoetrope.html>

4. <https://software.intel.com/en-us/articles/intel-mash-maker-mashups-for-the-masses>

kumentit on assosioitu joukolla arvoja usean itsenäisen (ortogonaalisen) näkymähierarkian läpi, on liian rajoittava, jos dokumentin mahdolliset valittavat näkymät ovat korreloivia. Esimerkiksi, jos kuvataan tiettyä vaatetta sinisenä ja punaisena sekä isona ja pienenä kokonaan niin, että vain pieni koko tulee punaisena ja suuri sinisenä, siten näkymät ”väri” ja ”koko” ovat korreloivia eikä itsenäisiä. Yksi tapa korjata nämä kaksi puutetta on palauttaa tulosedokumenttien määrän lisäksi rikkaampi kooste metadatatista, joka tukisi parempaa päätöksentekoa. (Ben-Yitzhak et al. 2008, s. 33-34)

4.3.1 Käyttäjäinteraktio

Moninäkymähakusovellukset pyrkivät hyvään käyttäjäinteraktioon rakentamalla sen ympärille intuitiivisen ja helppokäyttöisen käyttöliittymän (Holi 2010, s. 22). Tyypillisesti käyttäjäinteraktio koostuu useista vaiheista, missä käyttäjä voi joko kirjoittaa tai parantaa hakukyselyä tai navigoida näkymähierarkian dataa joko tihennys- tai yleistysoperaatioilla. Kun käyttäjä valitsee tietyt näkymäarvot sen hetkessä hakukontekstissa, näkymäsovellus näyttää mahdollisia tihennyksiä niistä näkymistä alikategorioihin sekä hakutulosten määrän eri hakukategorioissa. Nämä luvut näyttävät käyttäjälle määrällisen yleiskatsauksen datamäärästä, mikä antaa käyttäjälle tiedon tehdä päätöksiä mahdollisille parannuksille ja tarkentaakseen haun siihen kohdetietoon, jota käyttäjä tarvitsee. (Ben-Yitzhak et al. 2008, s. 33) (Komamizu, Amagasa ja Kitagawa 2012, s. 349) Kun käyttäjä valitsee hakuobjektin tuloslistasta, vie se käyttäjän yksityiskohtaiselle kuvaussivulle, joka listaa kaiken saatavilla olevan tiedon valitusta objektista. Tältä sivulta käyttäjä voi palata takaisin tuloslistalle tai aloittaa uuden hakukyselyn klikkaamalla jotakin objektin attribuuttia. (Medynskiy, Dontcheva ja Drucker 2009, s. 2014)

Moninäkymähaun perusmallina on näyttää luokittelu ja antaa käyttäjän työstää tiensä syvemmälle kunnes sopiva kuvaus vastaa hänen kyselyään. Hakukyselyn syvyys voi ulottua aihealueesta aina alitermeihin asti. Hakemisen ja selaamisen prosessin tulisi antaa käyttäjälle mahdollisuus liikkua ylös, alas ja sivuille mistä tahansa pisteestä luokittelussa. Tämä on helppo saavuttaa web-ympäristössä. Luokittelussa kaikkien yhtäaikainen näyttö ruudulla voi olla hankalaa, sillä suuri määrä objekteja ja dokumentteja johtaa suureen määrään näkymiä. Tähän on useita ratkaisuja, jotka perustuvat näkymien osittaiseen näyttämiseen. Käyttäjälle

voidaan näyttää vain näkymien nimet ja kun valitaan näkymä, niin näytetään termisisältö ja toistetaan prosessi jokaiselle valitulle näkymälle. Käyttäjälle voidaan myös näyttää kaikki näkymät ja vain osa ensiksi listatuista termeistä: kun näkymät valitaan, näytetään koko sisältö ja toistetaan prosessi jokaiselle näkymälle. Käyttäjälle voidaan myös näyttää vain kaikista tärkeimmät näkymät ja niiden termit: kun termivalinta näistä on valmis, näytetään toinen joukko näkymiä ja niin edelleen kunnes kaikki on näytetty. Ongelmia syntyy jokaisessa edellä mainituista tapauksista, jos käsitellään suurta tai hierarkkista sisältömäärää. (Vickery 2008, s. 157-158)

Moninäkömahaku on visuaalinen kyselyparadigma, jonka tavoitteena on rajoittaa hakuavaruutta joukkoon olennaisia lähteitä. Tämä tapahtuu rakentamalla valintakyselyä selaamalla tai lisäämällä rajoitteita siihen. Jokainen askel käyttöliittymässä on askel kyselyn rakentamisessa ja käyttäjä näkee välituloksia sekä mahdollisesti tulevia askelia hänen rakentaessaan kyselyä. (Oren, Delbru ja Decker 2006, s. 561) Kokonaisuutena käyttäjä toteuttaa monivaiheisen resurssien hakemisprosessin, joka alkaa jostakin tutkittavasta alkukokoelmasta (joka sisältää yhden rajoitteen), jonka jälkeen käyttäjä jatkaa orientoitumis- ja täsmennysaskelilla lopputuloskokoelmaan, jonka hän tutkii tarkasti. Tämän prosessin aikana käyttäjän valitsemat näkömäärojen tulokset ovat joko konjuktiivisia tai disjuktiivisia kyselyitä (näkömien luonteen ja sovelluksen suunnittelun mukaan), joita toteutetaan resurssikokoelmalle. Koska moninäkömahauksen ytimessä sijaitsee tasaisia tai hierarkkisia näkömiä ja jokainen näkömä sisältää yhden tai useamman näkömääroksen sekä jokainen objekti voidaan assosoida alikokoelmalla näitä arvoja, tarvitsee tämä navigointiparadigma yleensä rikkaan metadatan kuvaamaan riippuvuussuhteita näkömien ja resurssien välillä. (Thai, Rouille ja Handschuh 2012, s. 2)

Perinteistä interaktiokaavaa seuraavat moninäkömahakukäyttöliittymät kärsivät useasta haittapuolesta. Ensiksi, hakukyselyn teon ja kuvaussivuston välinen ero voi tehdä käyttäjälle hankalaksi päättää, mitkä objektin attribuutit voidaan ottaa käyttöön osaksi hakua. Toiseksi, käyttöliittymässä näytetään usein oletuksena vain kaikista yleisimmät olennaisista kategorioista ja loput piilotetaan erilliseen avattavaan käyttöliittymään, sillä valtava lista kaikista kategorioista tekisi hakukäyttöliittymästä hankalan. Lopuksi, moninäkömahakukäyttöliittymät olettavat käyttäjien aloittavan selaamisen aloitussivusta, vaikka saapumiskeinoja mihin

tahansa listaobjektiryhmään on olemassa monia. Tällöin käyttäjä päätyy Yee et al. mukaan "*loppupeliin*", ohittaen alun ja keskikohdan hakuprosessista. Tämä ei anna selvää kuvaa käyttäjälle siitä, kuinka hän voisi noutaa samankaltaisia objekteja samasta web-sivustosta. (Medynskiy, Dontcheva ja Drucker 2009, s. 2015) Pysyvillä hakulinkeillä on kuitenkin etunsa, esimerkiksi jos käyttäjä haluaa tallentaa tehdyn haun. Tällöin pysyvät hakulinkit jäisivät käyttäjien haltuun ja niitä voisi mahdollisesti jakaa muiden käyttäjien kesken, tehden niistä sosiaalisia hakutulolinkkejä.

4.3.2 Käyttöliittymätyypit

Moninäkömahaulle on olemassa useita käyttöliittymätyyppejä. Yleinen **moninäkömahakukäyttöliittymä** antaa käyttäjän valita näkymiä useammasta näkömääkategorista kerrallaan. Monta näkömää kerralla valittuna antaa käyttäjälle laajan kriteerikokoelman, joka voi antaa laajemman kokoelman objekteja, jotka täsmäävät yhteen tai useampaan näkömäärvoon. Tällöin käyttäjän tarvitsee katsoa objekteja tarkemmin selvittääkseen, mitä näkömän arvoja kukin objekti vastaa. Tästä erikoistapaus on **yksittäisnäkömahakukäyttöliittymä**, joka antaa käyttäjän selata vain yhtä näkömää kerrallaan, käyttäjän joutuessa rajaamaan omaa hakuaan askel kerrallaan. Jos tässä käyttöliittymässä usean valitun näkömän jälkeen päädytään eihaluun hakutulokseen, käyttäjä joutuu palaamaan takaisin lähtöasetelmaansa. Kummassakin mainituissa käyttöliittymissä käyttäjän valitessa objektia joutuu hän vertailemaan useita objekteja monen näkömän kokoelmasta yhden sijaan. (Holi 2010, s. 22) (Thai, Rouille ja Handschuh 2012, s. 22)

Rinnakkaismoninäkömahaku tarjoaa mielenkiintoisen lähestystavan hakemiseen. Se on hakuparadigmasta jalostettu moninäkömahakukäyttöliittymä, joka sallii käyttäjän helposti rakentaa ja katsoa useamman hakukyselyn tulokset yhtäaikaaisesti. Tavallisilla moninäkömahakukäyttöliittymillä on rajoitteena se, että käyttäjä näkee kerralla vain yhden hakukyselyjonon tulokset. Rinnakkaismoninäkömahaku näyttää myös käyttäjälle jokaisen hakutuloksen väliset riippuvuussuhteet toisiinsa nähden. Rinnakkaismoninäkömahaku on yleistys perinteisestä moninäkömahasta, joten potentiaalisesti sen tarjoamat hyödyt on saavutettavissa niissä sovellusalueissa, joihin perinteinen moninäkömahaku on lisätty onnistuneesti. (Buschbeck et al. 2013, s. 3023-3026)

Automatisoitu moninäkömahaku tarjoaa käyttäjälle tavan selata yhtä tai useampaa tekstiä eri näkymistä ilman mitään syötteitä. Se perustuu semanttisiin yksikköihin, joka voi olla lauseke tai sekvenssi lausekkeita. Semanttiset yksiköt yhdessä aiheessa voivat olla eri sijainnista ja semanttisilla yksiköillä on etäisyydet. Semanttiset yksiköt yhdessä muodostavat semanttisen näkymän, joka on joukko semanttisesti läheisiä semanttisia yksiköitä, jotka edustavat yhtä näkökulmaa merkityssä tekstissä. Tällainen moninäkömahakumeکانismi osaa hakea semanttisia yksiköitä tekstistä, rakentaa hierarkioita semanttisista näkymistä ja osaa navigoida hierarkiassa. Tämä on hyödyllistä käyttäjälle, sillä hän voi navigoida semanttisissa näkymissä ilman syötteitä ja saada mielenkiintoista sisältöä tekstistä. Tämä mekanismi ottaa huomioon ihmisen lukemisprosessin ja kokeet osoittavat tämän mekanismin tehokkuuden. (Xu ja Zhuge 2014, s. 1)

Myös **tagipohjaisia hakupalveluita** voidaan pitää eräänlaisina erikoistapauksina moninäkömahasta. Tässä tapauksessa tageja voidaan pitää näkyminä tai hakukategorioina, joita valitsemalla pääsee aina tiettyyn hakutulokseen. Tagit toimivat haun rajoittajina, suodattaen hakutuloksia, joissa valittu tagi ei esiinny. Jos jokaisen valinnan jälkeen tagiavaruutta rajoitetaan niin, että sellaisia tageja ei näytetä, mitkä eivät tuota hakutulosta, tällöin yksi moninäkömahauksen perusominaisuuksista (ei päädytä tyhjiin hakutulokseen), toteutuu myös tagihakukäyttöliittymällä. Toisaalta tagit eivät paljasta samanlaista informaatiota kuin näkymien hakukategoriat tekevät moninäkömahassa. Tageilta puuttuu tieto, kuinka paljon kyseisellä tagilla avautuu hakutuloksia. Tageja ei myöskään ole yleensä ryhmitelty erilaisiin kategorioihin, kuten näkymät ja hakukategoriat on moninäkömahassa. Esimerkiksi tagipilvessä tagit esiintyvät yksittäisinä sanoina ja ne on järjestetty yleensä aakkosten mukaan. Täten voidaan sanoa, että tagit voivat toimia sinällään vain osittain näkyminä ja hakukategorioina.

4.4 Tiedonhaku

Moninäkömahakua voidaan käyttää tietokokoelmien tutkimiseen ja tiedonhakuun isoille datajoukoille (Oren, Delbru ja Decker 2006, s. 571) (Thai, Rouille ja Handschuh 2012, s. 1). Tarkasteltavat tietokokoelmat voivat koostua monenlaisista objekteista, esimerkiksi tekstidokumenteista. Vaikka objektit sisältävät paljon tietoa, niitä ei yleensä kohdella viitekehysten mukaisesti moninäkömahakäyttöliittymässä: tarvittava metadata ei aina ole saatavilla tai

ei vastaa käyttäjän mielenkiinnon kohteita. Tämän lisäksi objektien lineaarinen listaus tekee tulosten näyttämisen hankalaksi, kun käyttäjä selaa tai vertaa useiden näkymien arvoja niiden ollessa eri tärkeysjärjestyksessä. Näille ongelmille on ehdotettu ratkaisuksi moniulotteista näyttötapaa, joka tukee objektien järjestämistä valittujen näkymäarvojen mukaan. (Thai, Rouille ja Handschuh 2012, s. 1)

Semanttinen tiedonhaku eroaa merkittävästi tavanomaisesta sanahausta, jossa hakurobotit keräävät tietoa web-sivustoista suorittamalla hakusanapohjaisia hakuja. Kun käyttäjä hakee tietoa tietyllä hakulausekkeella, kaikki web-sivut, jotka täsmäävät tähän lausekkeeseen palautetaan hakutuloksena. Hakukone ei kuitenkaan ymmärrä sanojen merkitystä, vaan ne ovat sille vain merkkijonosekvenssejä. Semanttisen haun avulla on mahdollista etsiä tietynlaisia asioita, jotka täyttävät tiettyjä kriteerejä. Samalla voidaan esittää muita hakutuloksia muiden kriteerien perusteella. Tämä mahdollistaa käyttäjän selvästi ilmaista hakukyselyn rajoitteet. Esimerkiksi semanttisessa haussa käyttäjä voi rajoittaa hakunsa henkilön julkaisemiin papereihin, missä taas sanahaun tapauksessa haettaisiin vain tuloksia, joissa esiintyy kirjoittajan nimi ja sukunimi. (Vandic, Dam ja Frasinca 2012, s. 427) Tällainen toiminta kuitenkin edellyttää, että hakukohteena olevilla sivustoilla on käytetty semanttista merkkäystä.

Tiedonhaku moninäkömähäussa tapahtuu valitsemalla näkymistä kiinnostavia kategorioita. Moninäkömähaku tarjoaa käyttäjälle tukun alikategorioita, joista voidaan nähdä kokonaistulos hakutuloksista, jonka jälkeen käyttäjä voi kaventaa hausta syntyvää tuloslistaa. Jokainen valinta rajaa tulosjoukkoa niin, että tulosjoukko sisältää ne dokumentit, jotka kuuluvat valittuihin kategorioihin tai valittujen kategorioiden alikategorioihin. (Fagan 2010, s. 58) (Hägström 2007, s. 33) (Suominen 2008, s. 4) Tiedonhaku voi myös sisältää interaktioita muiden käyttöliittymässä olevien komponenttien kanssa, kuten avainsanahaku tai suora hakulinkki johonkin hakutulokseen.

Moninäkömähaku tarvitsee kaksi tärkeää toiminnallisuutta, jotta se voisi tukea tehokasta hakemista. Nämä ovat valintaoperaatio ja avainsanapohjainen valintaoperaatio, joka tukee avainsanahakua moninäkömähäun kontekstin aikana. **Valintaoperaatio** sallii käyttäjän valita näkymän tutkiakseen valintaa ja silloin järjestelmä kaventaa hakuavaruutta valittuun näkymään ja arvoon. **Avainsanapohjainen valintaoperaatio** taas on valinnainen toiminto, mutta se sallii käyttäjän antaa hänen tarpeitaan edustavia avainsanoja, joiden jälkeen järjestelmä

noutaa kyseiset objektit, jotka sisältävät annetut avainsanat. (Komamizu, Amagasa ja Kitagawa 2012, s. 351)

4.4.1 Tutkiva tiedonhaku

Tutkiva tiedonhaku (engl. *exploratory search*) on yksi menetelmä hakea tietoa aiheesta, vaikkakin kaikenlaista tiedonhakemista voidaan pitää tutkivana. Sitä voidaan myös kutsua selaamiseksi (engl. *browsing*). Tutkivassa tiedonhaussa käyttäjä tutkii tietojoukossa olevien objektien riippuvuussuhteita. Yleisin esimerkki tutkivan tiedonhaun soveltamisesta on löytää oikea objekti tai objektin etsimismenetelmä. Käyttäjälle annetaan objekteja ja niiden ominaisuuksia sisältävä tietojoukko, jonka jälkeen käyttäjä valitsee mielestään hänelle parhaiten sopivimman objektin tavoitteisiinsa nähden (esim. paras ja halvin vaihtoehto). Jotta tutkiva hakeminen olisi tehokasta, käyttäjän tarvitsee löytää kaikki relevantit ominaisuudet ja keskittyä niihin sekä tutkia kaikki muut ominaisuudet, jotka liittyvät valittuihin ominaisuuksiin. (Sacco ja Tzitzikas 2009, s. 1-3)

Kun käyttäjä haluaa harjoittaa tutkivaa tiedonhakua, ei tavallinen sanahaku taivu käyttäjän haluille. Käyttäjän etsiessä tietoa uudesta aihealueesta ei hän välttämättä ole varma mitä hän etsii, mutta haluaa kuitenkin saada ymmärryksen aiheesta, sen sisäisistä riippuvuussuhteista sekä sen riippuvuussuhteista muihin aihealueisiin. Tällaista hakemista tulisi tukea navigaatioelementeillä, kuten hakemistoilla, aihealueeseen läheisesti liittyvillä termeillä sekä suosituksilla. (Holi 2010, s. 21-22) Tutkiva tiedonhaku verrattuna sanahakuun paljastaa käyttäjälle täydemmän kokonaisnäkökulman aiheesta: hakuprosessi on luonteeltaan avoin, moninäkömäinen ja iteratiivinen. Se on yleisesti käytössä tieteellisessä tutkimisessa, oppimisessa ja ymmärryksen luomisessa. Tutkiva tiedonhaku vaatii enemmän käyttäjältään, mutta palkitsee hänet syvemmällä tietämyksellä aihealueesta. (Marchionini 2006) (Dimitrova et al. 2013, s. 1)

Moninäkömähaku on näkömäteorioihin perustava tutkiva tekniikka strukturoiduille datajoukoille (Oren, Delbru ja Decker 2006, s. 560). Se tukee erilaista lähestymistapaa tiedonhaakuun, ollen esimerkki tutkivan tiedonhaun toteuttavasta hakutekniikasta. (Holi 2010, s. 22) (Wei et al. 2013, s. 42). Moninäkömähaussa riippuvuussuhteiden tutkiminen onnistuu nä-

kymien avulla: näkymillä on monia rajoittuvuusarvoja ja käyttäjän valitessa eri näkymiä valitsee hän arvoja rajoittaakseen tietoavaruuden häntä kiinnostaviin olennaisiin objekteihin. Näkymäteoria on sovellettavissa puolirakenteisen RDF-datan navigointiin: informaatioelementit ovat RDF-subjekteja, näkymät ovat RDF-predikaatteja ja rajoittuvuusarvot ovat RDF-objekteja. (Oren, Delbru ja Decker 2006, s. 559-560)

Moninäkymähakukäyttöliittymät tukevat niin tutkivaa kuin suoraa hakemista monessa erilaisessa kohdealueessa (Medynskiy, Dontcheva ja Drucker 2009, s. 2014). Erilaiseksi moninäkymähaun muihin hakumuotoihin tekee sen, että haettavat objektit sijaitsevat dynaamisessa taksonomiassa, toisin sanoen useammassa kategoriassa. Tämä eroaa suuresti avainsanahausta, jossa etsitään tuloksia annettujen avainsanojen perusteella, tai lomake- ja kansiohausta, jossa objektit esiintyvät vain yhdessä kategoriassa. (Wei et al. 2013, s. 44) Tutkivat käyttöliittymät sallivat käyttäjien löytää informaatiota ilman aikaisempaa tietämystä skeemasta. Tutkiva hakutekniikka on tarpeellinen erityisesti silloin, kun datan skeema tai rakenne on käyttäjälle tuntematon. (Oren, Delbru ja Decker 2006, s. 560)

Moninäkymähakukäyttöliittymät tukevat tutkivaa hakemista monella eri tavalla. Käyttäjä voi rakentaa ja muokata monimutkaisia hakukyselyitä, jotka jättävät turhat tulokset pois. Näkymien lisäys ja poisto onnistuu nopeasti yhdellä hiiren klikkauksella. Näkymät ottavat käyttöön interaktiivista metadataa osana hakukäyttöliittymää ja näyttävät yksityiskohtia siitä, mitkä näkymät, kategoriat ja arvot ovat saatavilla mistäkin tietojoukosta. Näkymät myös näyttävät, kuinka monta objektia on kullakin näkymällä. Moninäkymähaku antaa myös tuen joustavalle tiedonhakustrategialle, antaen käyttäjälle mahdollisuuden yhdistää selaus- ja hakustrategioita. (Kules ja Capra 2012, s. 114 ja 116)

4.4.2 Hakemisen tukeminen

Ihmiset käyttävät ja hakevat hyvin paljon tietoa hakukoneiden kautta. Tiedonhakua näiden avulla on perinteisesti harjoitettu erilaisten vapaasti valittavien sanojen tai termien avulla. Tällainen hakukone on esimerkiksi Google, joka on tällä hetkellä lähes integroitunut ihmisten päättiedonhakumenetelmäksi. Tiedon hakeminen sanahakukoneiden kautta toimii erityisen hyvin silloin, kun käyttäjällä on selkeästi selvillä ongelma, johon hän haluaa vastauksia

ja haettava asia voidaan kuvata yhdellä sanalla tai lausekkeella (esim. henkilö tai organisaatio). (Holi 2010, s. 21-22)

Useassa tiedonhaku tutkimuksessa on saatu Novotнын (Novotny 2004) tutkimuksen kaltaisia tuloksia, että käyttäjät ovat omaksuneet vaikutteita hakemiseensa eri web-hakukoneista (kuten Google, Facebook, Twitter, Amazon.com ja Wikipedia). Käyttäjät eivät suunnittele huolellisesti hakuprosessejaan, vaan mielummin antavat muutaman hakutermin (yleensä yhdestä kolmeen sanaa), näkevät ensimmäiset tulokset ja suodattavat hakutulostaan muutamalla yksinkertaisella interaktiolla. Erityisesti aloittelevat käyttäjät käyttävät omia termejään ilman riittävää tietämyksen tasoa haettavasta tiedosta. (Ramdeen ja Hemminger 2012, s. 703) (Zwol et al. 2010, s. 961) Tämän lisäksi kun käyttäjän käyttäytymistä on tutkittu eri web-hakujärjestelmissä, on huomattu, että kolmasosa käyttäjän interaktioista on hakutulossivulla, kun loput ajasta käytetään selaamisaktiviteetin kautta hankituilla sivuilla (Castano, Ferrara ja Montanelli 2012, s. 163).

Monet nykyiset sanahakukoneet hyödyntävät avainsana-pohjaisia hakualgoritmeja tiedonhakuun. Tämä metodi on nopea ja helppo, mutta kärsii korkeasta hakutulosten palautusmäärästä ja haun tarkka raja-alue metadatan perusteella on hankalaa. Ongelmat selittyvät sillä, että tämänkaltaiset hakukoneet hakevat yleensä avainsanoja sisältäviä hakutuloksia. (Sah ja Wade 2012, s. 105) Käyttäjille olisikin hyödyllistä päästä käsiksi työkaluihin, jotka antavat mahdollisuuden aloittaa haun yksinkertaisilla avainsanoilla ja jatkaen hakua mahdollisilla tarkennuksilla, varsinkin jos haettava aihealue ei ole käyttäjälle ennalta tuttu. Moninäkömähäussa onnistuu niin laajien hakukyselyiden kaventaminen ja täsmentäminen kuin haun aloittaminen sanahaulla. Näkymät moninäkömähäussa voivat myös hakea itseensä monenlaista metadatan monesta eri lähteestä (esim. tekijän nimi, aihe, genre tai kieli). (Kules ja Capra 2012, s. 117)

Moninäkömähakukäyttöliittymällä on useita etuja ja eroja avainsanahakuun tai eksplisiittisiin, kyselykielellä muotoiltuihin kyselyihin. Se on visuaalinen käyttöliittymä, joka poistaa tarpeen kirjoittaa eksplisiittisiä kyselyitä. Se tarjoaa vain rajoittuvuusarvoja, jotka eivät johda tyhjiin hakutuloksiin. Rajoittuvuusarvojen antaminen joka hakuaskeleella sallii tutkimisen ennalta tuntemattomasta datajoukosta. Se sallii myös parametrusten hakujen suorittamisen ja spesifioinnin valittuihin attribuuteihin (näkymät). Nämä kaksi hakuvaihtoehtoa eivät

kuitenkaan sulje toisiaan pois, vaan moninäkömahakuun voidaan integroida avainsanahaku mukaan. Tämä lähestymistapa onkin tullut hyvin suosituksi tietojoukkojen etsimiseen ja navigointiin. (Oren, Delbru ja Decker 2006, s. 560) (Vandic, Dam ja Frasincar 2012, s. 427)

Hakemisen tukemiseen on otettava huomioon myös käyttöliittymä, sillä tutkimuksissa on osoitettu moninäkömahakukäyttöliittymän vaikuttavan ja muuttavan käyttäjän hakukäyttäytymistä. Käyttöliittymän kapasiteetti ja siten myös käyttöliittymän saavutettavuus ja ulkoasu vaikuttavat käyttäjän toimintaan, taktiikoihin ja strategioihin, hänen joko hakiessa, katsoessa tai valitessa hakutuloksia. Tutkimuksissa on huomattu, että hakijat voivat käyttää näkymiä eri tavalla eri tilanteissa yhden hakusession aikana. Esimerkiksi Kules, Capra, Banta ja Sierra (Kules et al. 2009) raportoivat että hakijat katsoivat näkymiä enemmän ensimmäisen sivun hakusivulla kuin haun alisivuilla. Näkymien huomattiin myös olevan yhtä tai jopa enemmän tärkeitä käyttäjälle kuin itse hakutulos. (Kules ja Capra 2012, s. 115) (Ramdeen ja Hemminger 2012, s. 704) Täten voidaan todeta, että näkymillä on tärkeä rooli käyttäjän hakemisen tukemisessa.

5 Tutkimuksen empiirisen osan asetelma ja toteutus

Tässä luvussa esitellään tutkimuksen kohteena olevat verkkopalvelut ja tutkimusaineiston keruu- ja analyysimenetelmät. Tutkimuksessa tutkitaan, kuinka moninäkömähakua hyödynnetään eri verkkopalveluissa: millainen on sivustojen käyttäjäinteraktio ja kuinka interaktiivinen käyttöliittymä on käyttäjälle, miten hakemisen prosessi tapahtuu käyttöliittymässä ja mitä muita avustavia hakukomponentteja käyttöliittymä tarjoaa, miten näkymät ja objektit esitetään käyttöliittymässä sekä millaisia ovat sivustojen sosiaaliset ja semanttiset ominaisuudet. Tässä luvussa myös tarkastellaan erilaisia aikaisempia tutkimuksia moninäkömähakusta.

5.1 Aikaisempi tutkimus

Moninäkömähakututkimuksia on julkaistu muun muassa itse moninäkömähakusta, moninäkömähakukäyttöliittymästä sekä tutkimuksen aikana toteutetuista moninäkömähakupalveluista. Moninäkömähakututkimus on eniten keskittynyt alueisiin, kuten vertailututkimukset (muihin hakuparadigmoihin) sekä käyttöliittymän käytettävyyden tutkiminen hakemisessa. Tutkimuksia on myös tehty katsauksina, joissa on saatu tärkeää tietoa moninäkömähakuun liittyvistä tutkimuksista: on kartoitettu erilaisia moninäkömähakujärjestelmiä eri verkkopalveluissa, kartoitettu mahdollisia ongelmakohtia sekä tulevaisuuden tutkimuskohteita ja -suuntauksia. Moninäkömähakukäyttöliittymiä on tutkittu myös näillä tavoilla: suurimittainen lokianalyysi, vertaileva käyttäjätestaus sekä silmänliikkeiden seuraukset (Kules ja Capra 2012, s. 117).

Tutkimuksia, joissa on vertailtu moninäkömähakua muihin hakukäyttöliittymiin on saatu seuraavia tuloksia. Prattin, Hearstin ja Faganin (Pratt, Hearst ja Fagan 1999) tutkimuksessa lähes kaikki testattavat pitivät moninäkömähakutyökalua parempana kuin rankkaus- ja klusterointityökalua. Tutkimukseen osallistuneiden mielestä moninäkömähaku oli järkeenkäypä, avulias, hyödyllinen ja selkeä sekä he olisivat valmiita käyttämään moninäkömähakua uudelleen hakemiseen. Uddinin ja Janecekin (Uddin ja Janecek 2007) tutkimuksessa huomattiin, että käyttäjät olivat nopeampia käyttäessään moninäkömähakukäyttöliittymää ja

kahdesta kolmessa tehtävässä ero oli huomattava. Moninäkömähauulla onnistui paremmin löytää aiheeseen liittyvää olennaista tietoa tavalliseen hakuun verrattuna. Käyttäjät myös arvioivat moninäkömähauun paremmaksi käyttää. Kules ja Shneiderman (Kules ja Shneiderman 2008) vertailivat tutkimuksessaan tavallista hakukäyttöliittymää ja moninäkömähakukäyttöliittymää. Moninäkömähakua käyttäessä käyttäjät tutkivat heidän hakuaan laajemmalla alueelta samalla tuntien olevansa enemmän organisoituneempia hauissaan. Käyttäjät myös mielsivät moninäkömähauun parantaneen heidän hakutuloksiaan. Tutkijat myös havaitsivat, että käyttäjät hakiessaan tietoa omaksuivat ja käyttivät eri hakutaktiikoita. Yee et al. (Yee et al. 2003) tutkimuksessa verrattiin Flamenco-nimistä moninäkömähakukäyttöliittymää perinteiseen hakukäyttöliittymään. Suurempi osa tutkittavista pystyi löytämään olennaista tietoa aiheesta moninäkömähauun avulla verrattuna tavalliseen hakukäyttöliittymään. Tutkimuksen tulokset paljastivat myös, että kategoriapohjainen lähestymistapa on onnistunut tapa tarjota pääsy kuvakokoelmiin ja niiden tarkasteluun. (Fagan 2010, s. 59-60) (Kules ja Capra 2012, s. 117-118) (Peter J. Wild 2009, s. 426)

Tutkimuksia, joissa on tutkittu moninäkömähakukäyttöliittymää on saatu seuraavia tuloksia. Toisessa Flamenco-käyttöliittymään (English et al. 2002) liittyneessä tutkimuksessa testattiin kahta erilaista käyttöliittymää. Tutkimus osoitti, että käyttäjät suosivat monipuolisempaa hakukäyttöliittymää etsimiseen ja selaamiseen, mutta kokivat yksinkertaisemman käyttöliittymän paremmaksi täsmähakea tiettyä kuvaa. Kummassakin käyttöliittymässä havaittiin, että käyttäjät aloittivat tiedonhaun selaamisesta varsinaisen hakemisen sijaan. Capra et al. (Capra et al. 2007) tutkimuksessa verrattiin kolmea käyttöliittymää. Tutkimuksessa huomattiin, että käyttäjät pitivät arvossa perinteisen käyttöliittymän tuttuutta (hyperlinkit ja avainsanahaku), mutta listasivat näkymät kahdessa kokeellisessa käyttöliittymässä niiden parhaiksi ominaisuuksiksi. Tutkijat tulivat siihen lopputulokseen, että hyvin suunniteltu hierarkkinen organisointi informaatioavaruudelle on suositeltavaa. Zhang ja Marchionini (Zhang ja Marchionini 2005) analysoivat tutkimuksessaan yhden käyttöliittymän käytettävyyttä, joka käytti näkymiä hakutulosten kategorisointiin. Näkymiä hyödyntävä käyttöliittymä havaittiin tilastollisesti paremmaksi miellyttävyydessä. Hearst (Hearst 2008) totesi tutkimuksessaan, että tekemien johtopäätöksien mukaan näkymät ovat ”*todistettu tekniikka tukemaan tutkivaa hakemista ja löytämistä*”. (Fagan 2010, s. 60-61) (Kules ja Capra 2012, s. 117)

Wei, Liu ja Zheng (Wei et al. 2013, s. 41-58) tekivät katsauksen tutkimuksen moninäkömahakupalveluista, missä tarkasteltiin palveluiden näkymien hakumalleja sekä siihen liittyviä metodeja ja tekniikoita. Tutkimuksessa tuotiin esille pääpiirteet olemassa olevista moninäkömahakupalveluista. Moninäkömähaussa näkömämalleja on kolmenlaisia: joukkoteoriapohjainen, FCA (*Formal Concept Analysis*) tai kevytrakenteinen ontologia. Moninäkömähaun avainteknologioihin taas kuuluu näkömätermien poiminta, hierarkiarakentaminen, yhdistelmätermien generointi ja näkömärankkaus.

Moninäkömähaululle on myös ehdotettu kehityskohteita. Hearst (Hearst 2008) mainitsi tutkimuksessaan kehityskohteiksi näkömiin lisättäväksi isoja aihealuekeskeisiä kategoriajärjestelmiä, kehittyneitä älyominaisuuksia (automaattinen täydennys) sekä näkymien visuaalisen ilmiannon parantamista. Hän mainitsi myös, että syötettyjen avainsanojen tulisi vaikuttaa esitettävien näkymien järjestykseen. Näkymien kehityksen tulisi myös huomioida mobiilikäyttöliittymät. (Fagan 2010, s. 61) Kules ja Capra mainitsevat tutkimuksessaan, että silmänliikeseuraustutkimuksia on vuodesta 2009 lähtien tehty todella vähän (Kules ja Capra 2012, s. 117). Wei, Liu ja Zheng (Wei et al. 2013, s. 56-57) tarkastelivat tutkimuksessaan myös tulevaisuuden tutkimusaiheita moninäkömähausta, kuten automaattinen taksonomioiden generointi moninäkömähaululle, näkömäluokittelu ja -navigointi monimutkaisille avoimille datajoukoille, käyttöliittymän ja visualisoinnin ongelma-alueet (visuaalinen sotku ja muutossokeus) sekä olennaisuuden arviointi hakutuloksista.

Edellä mainitut tutkijat ovat olleet merkittäviä tekijöitä moninäkömahakututkimuksessa. Muita mainittavia tekijöitä moninäkömähaun tutkijoina ovat Vickery, Broughton, Slavic sekä Beghtol (Peter J. Wild 2009, s. 420).

5.2 Tutkittavat verkkopalvelut

Tutkittavat verkkopalvelut sisältävät julkaisusivustoja, semanttisia sivustoja ja kaupallisia sivustoja. Bibsonomya tarkastellaan erikoistapauksena, sillä se ei ole suora sovellutus moninäkömähausta. Vaikka vastaavan tyylinen katsaus on tehty aikaisemmin Wein, Liun ja Zhengin toimesta (Wei et al. 2013), se keskittyi enemmän sivustojen järjestelmällisiin ominaisuuksiin. Tämä tutkimus eroaa myös Wein, Liun ja Zhengin tutkimuksesta sen suhteen, että tutkitta-

via parametreja on enemmän, ja että tutkimuksen painopisteenä on suomalaiset, uudehkot ja erityyppiset verkkopalvelut.

5.2.1 Finna

Finna on portaali kymmenien arkistojen, kirjastojen ja museoiden aineistoihin ja palveluihin. Se perustuu avoimen lähdekoodin VuFind-ohjelmistoon. Sivuston kautta voi etsiä tietoa kirjastojen ja arkistojen kokoelmista. Käyttäjä voi esimerkiksi katsella kuvia museoesineistä ja taideteoksista sekä ladata erilaisia kirjoja. Käyttäjä voi portaalin kautta myös lainata kirjoja ja tarkastella organisaatiokohtaisia näkymiä (JYKDOK). Hakupalvelun luotettavuuden takaa asiantuntijaorganisaatioiden tuottama sisältö. (Finna 2014) Finna valittiin tutkimukseen erityisesti sen sisältämän laajan ja monipuolisen aineiston takia.



Kuvio 11: Kuvankaappaus Finna-sivuston etusivusta.

Sivuston etusivulta on mahdollisuus aloittaa tiedonhaku monella eri tavalla. Käyttöliittymässä on mahdollisuus tutkia koko tietokantaa sanahaualla, jota voi täsmentää hakemaan tietoa tietystä paikasta tai tiedon esiintymismuodosta (esim. kuva, lehtiartikkeli). Haun voi myös aloittaa suoraan tarkennetulla haulla. Etusivu myös esittää pysyviä hakutulostenlinkkejä, joi-

hin käyttäjä voi hypätä klikkaamalla niitä. Tietoa voi myös hakea aikajanarajauksella tietyltä aikaväliltä. Käyttäjillä on mahdollisuus selata istuntokohtaista hakuhistoriaansa ja rekistroytimällä sivustoon hakuhistorian voi tallentaa omaan käyttäjätiliin.

5.2.2 JYX

JYX on Jyväskylän yliopiston julkaisuarkisto, jonne tallennetaan aineistoa digitaalisessa muodossa. Se perustuu avoimen lähdekoodin DSpace-ohjelmistoon. Sinne tallennetaan Jyväskylän yliopistossa tuotettuja tai sen kokoelmista digitoituja aineistoja, kuten väitöskirjoja, artikkeleita, e-kirjoja, verkkolehtiä, historiallisia karttoja sekä musiikkia. Yliopiston tutkijoilla on mahdollisuus rinnakaistallentaa artikkelinsa JYX:iin avoimesti saataville. JYX:n tavoitteita on edistää avointa pääsyä tieteelliseen tutkimukseen ja edistää digitaalisen tiedon pitkäaikaissäilyvyyttä. (JYX 2014) JYX sopi tutkimukseen hyvin, sillä se on Jyväskylän yliopiston omakehittämä palvelu.

The screenshot shows the JYX website homepage. At the top, there is a header with the JYX logo and navigation links: 'Kirjautu', 'Ota yhteyttä', 'Ohjeet', 'Tietoja JYXistä', and language options 'en' and 'fi'. Below the header, the main content area is divided into several sections:

- Jyväskylän yliopiston julkaisuarkisto**: A welcome message and a brief description of the archive.
- Hae**: A search bar with a 'Go' button and links for 'Laajennettu haku' and 'Ohjeet'.
- Selaa**: A list of filters for browsing, including 'Kaikki kokoelmat', 'Tekiä', 'Nimeke', 'Asiasana', and 'Julkaisuajka'.
- Pääkokoelmat**: A list of main collections, such as 'Artikkelit [842]', 'Blogi-kirjoitukset [6]', 'Data-aineistot [1]', 'Elektroniset kirjat [94]', 'Esitelmät ja posterit [27]', 'Historialliset kartat [5822]', 'Julkaisusarjat [554]', 'Kandidaatintutkielmat [918]', 'Kehityshankkeet [0]', 'Konferenssit ja seminaarit [247]', 'Lehdet [2428]', 'Lisensiaattintyöt [397]', 'Nuottikokoelmat [5]', 'Opinto-oppaat [23]', 'Oppimateriaalit [19]', 'Pro gradu -tutkielmat [11774]', 'Seminaarit [56]', 'Suomen Kansan eSävelmät [8613]', 'Tutkimusraportit [3]', 'Valokuvat [530]', 'Vuosikertomukset [6]', and 'Väitöskirjat [1339]'.
- Uusimmat julkaisut**: A list of the most recent publications, including articles on social media, education, and biology.

At the bottom of the page, there is a footer with navigation links: 'Jyväskylän yliopisto | Yliopiston kirjasto | JYX-julkaisuarkisto | JYKDOK-tietokanta'.

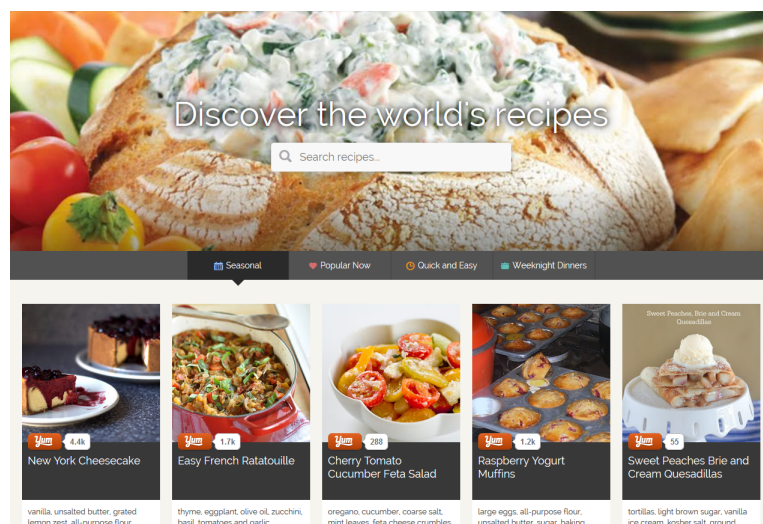
Kuvio 12: Kuvankaappaus JYX-sivuston etusivusta.

Sivuston etusivulla on mahdollisuus hakea, selata ja tutkia uusimpia julkaisuja tai pääkokoelmia. Tavallisen sanahaun voi myös tehdä tarkennettuna hakuna. Selaamisen voi aloittaa

kaikista kokoelmista tai täsmentää sen tiettyihin attributteihin (esim. tekijä, julkaisuaika). Pääkokoelmat kertovat linkin kautta avautuvien hakutulosten määrän. Uusimmat julkaisut voi saada myös RSS-syötteenä, jolloin niihin pääsee käsiksi ilman sivustolle menemistä.

5.2.3 Yummly

Yummly on yhdysvaltalainen semanttinen ruoanlaiton portaali ruokien ja reseptien hakemiseen. Palveluun voi rekisteröityä käyttäjäksi, jolloin palvelu oppii käyttäjän mieltymyksistä ja inhoista, jonka jälkeen se pystyy tarkasti luokittelemaan haetut ruoat ja antamaan reseptisuosituksia käyttäjälle. Yummly on nopeimmiten kasvava ruokasivusto maailmassa ja sillä on 15 miljoonaa käyttäjää kuukausittain. Yummly hyödyntää yksinoikeudellisia teknologioita semanttisissa hauissaan. (Yummly 2014) Yummlyn valinta tutkimukseen oli ehdoton, sillä se oli ainoa löydetty sosiaalinen semanttinen verkkopalvelu.

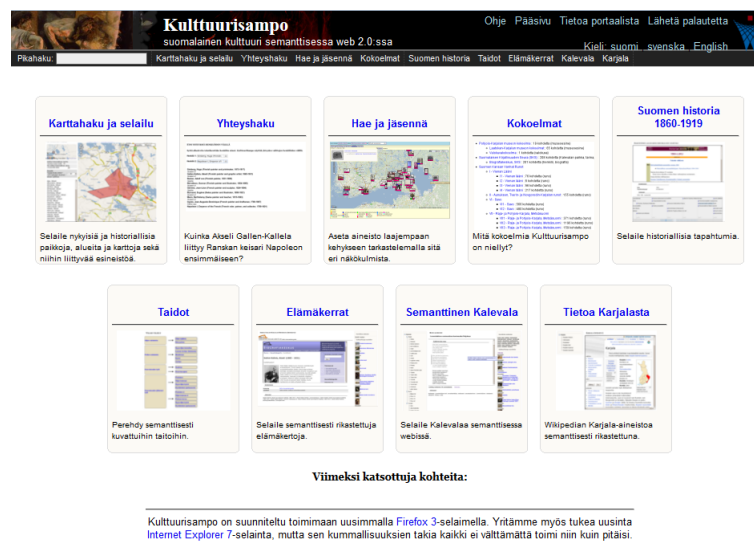


Kuvio 13: Kuvankaappaus Yummly-sivuston etusivusta.

Sivuston etusivulla on paljon näytettävää: käyttäjälle näytetään selvästi mahdollisuus sanahakuun, mutta selatessa sivustoa alas avautuu muitakin hakupolkuja. Käyttäjä voi selata reseptejä tiettyjen kategorioiden mukaan (vuodenaika, nyt suosittu, nopea ja helppo sekä arkipäivän ruoka). Käyttäjällä on myös mahdollisuus selata useita suosittuja hakuja, niiden ollessa hakulinkkejä, joista löytyy hakuun liittyviä reseptejä. Reseptejä voi myös hakea sivuston kärkijulkaisijoiden mukaan sekä rekisteröityneillä käyttäjillä on mahdollisuus tarkastella aikaisempia tallennettuja hakuja ja reseptejä.

5.2.4 Kulttuurisampo

Kulttuurisampo on kansallinen ja yhteisöllinen julkaisukanava, jonka kautta voidaan julkaista erilaista sisältöä semanttisessa webissä. Järjestelmä sisältää 20:n suomalaisen museon, kirjaston, arkiston ja muun organisaation kokoelmia sekä yhteisöllisiä Web 2.0 -verkkoaineistoja. Sivuston erityispiirteinä ovat temaattiset näkökulmat, jotka ovat erilaisia tapoja hakea tietoa sen sisällön tai rakenteen avulla. (KulttuuriSampo 2014) Kulttuurisampo valittiin vertailevana pohjana muihin tutkimukseen valittuihin sivustoihin, sillä se on yksi aikaisimmista suomalaisista pyrkimyksistä toteuttaa semanttinen tiedonhaku sivusto.



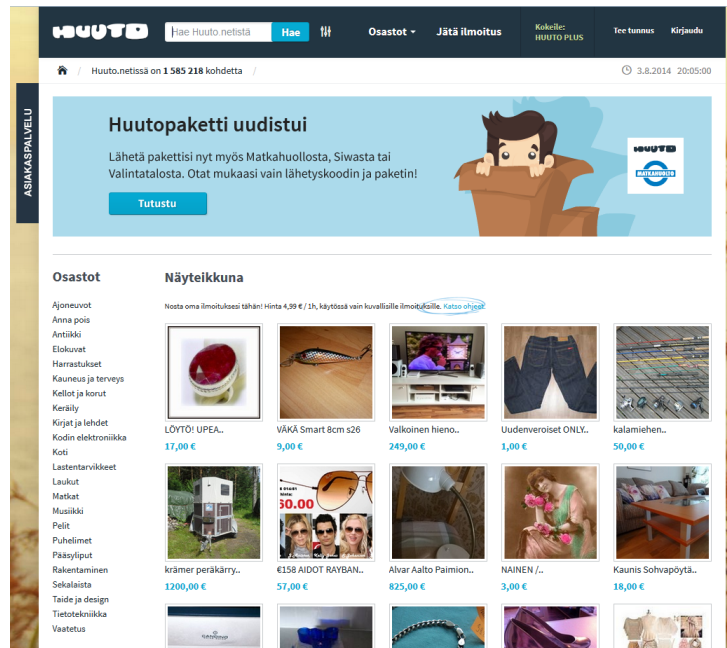
Kuvio 14: Kuvankaappaus Kulttuurisampo-sivuston etusivusta.

Sivuston etusivu tarjoaa muutaman mahdollisuuden aloittaa hakeminen. Pikahauulla voi etsiä tietoa avainsanojen avulla ja karttahaulla voi hakea tietoa paikan perusteella. Yhteyshauulla voi hakea tietoa, jotka liittyvät kahteen annettuun henkilöön. Hakuaan voi myös jäsentää ja haun voi aloittaa tutkimalla suoraan tietokokoelmia. Kun käyttäjä valitsee haluamansa hakuvaihtoehdon, avautuu hänelle siihen tarkoitukseen räätälöity käyttöliittymä.

5.2.5 Huuto

Huuto on suomalainen verkkopalvelu, jossa ihmiset voivat huutokaupata erilaisia tavaroita ja ostaa niitä. Sivusto tarjoaa myös maksullisia palveluita, kuten Plus-palvelu (sisältää hintatavaraukset ja yksityisviestit) sekä näyteikkuna (käyttäjä saa oman ilmoituksensa näkyville

sivuston etusivulle rajatuksi ajaksi). Sivusto on sosiaalinen, sillä palveluun laitettavat tavara-ilmotukset ovat muiden ihmisten tekemiä sekä itse kaupankäynti tapahtuu käyttäjien kesken. Huuto valittiin tutkimukseen siksi, että ensivaikutelmaltaan sivusto vaikutti siltä, että se saattaisi sisältää puutteita moninäkömahaun toteutuksesta.



Kuvio 15: Kuvankaappaus Huuto-sivuston etusivusta.

Sivuston etusivu tarjoaa monipuolisen valikoiman vaihtoehtoja hakemiselle. Tavallisen sana-haun ohella haun voi aloittaa tarkennettuna tai osastokohtaisesti. Etusivu tarjoaa myös useita mahdollisuuksia hypätä tiettyyn kohteeseen, joko näyteikkunan tai alhaalla olevien kategorioiden kautta. Käyttäjälle näyttäytyy myös sivuston suosituimmat hakusanat tageina, joita klikkaamalla käyttäjä pääsee itse hakutulokseen.

5.2.6 Gigantti

Gigantti on yritys, joka on osa Norjassa toimivaa Elkjøp-konsernia, joka kuuluu englantilaisen kodinelektronikkajätin, DSG International Oyj:n omistukseen (Gigantti 2014). Sen toimialana on kodinelektronikka sekä siihen liittyvät tele- ja datatuotteet (kuten tietokoneet). Gigantti palvelee asiakkaitaan tavarataloissa ja verkossa. Erityisesti sen verkkosivuilta on mahdollista hakea hyvinkin tarkasti tuotteita moninäkömähakukäyttöliittymän avulla.

Gigantti valittiin tutkimukseen siksi, että se on hyvin yleinen ja tunnettu yritys ostaa kodin tekniikkaa Suomessa.

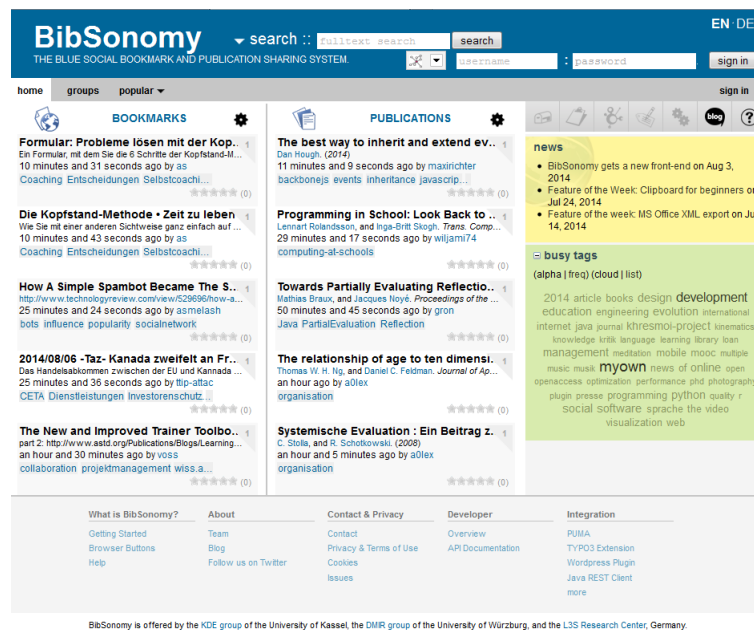


Kuvio 16: Kuvankaappaus Gigantti-sivuston etusivusta.

Sivuston etusivu tarjoaa muutaman hakumahdollisuuden käyttäjälle, vaikkakin ensisilmäyksellä niitä luulisi olevan enemmän. Sanahaun lisäksi tuotteiden selaamisen voi tehdä tuoteryhmäkohtaisesti aina alituoteryhmiin asti. Paljon tilaa on varattu erilaisille tuotetarjouksille, joita klikkaamalla pääsee itse tuotteeseen. Tuotetarjousryhmiä on myös mahdollista selata yksitellen tai näyttää kaikki kerralla.

5.2.7 BibSonomy

BibSonomy on sosiaalinen kirjanmerkkaukserkkopalvelu, missä käyttäjät keräävät ja merkkauvat URL-osoitteita ja julkaisuja (Wetzker et al. 2010, s. 73). Sivuston tarjoamat työkalut kirjanmerkkaukseen ovat helppokäyttöisiä ja palkitsevia käyttää eikä sivuston toiminnan osallistumiseen vaadita mitään erityistaitoja. Aluksi Bibsonomyn sivuston tarkoituksena oli organisoida BibTeX-pohjaisia lähdeviitetietoja. BibTeX on LaTeX-ladontajärjestelmään kuuluva työkalu kirjallisuusviitteiden ylläpitoon, mitä monet tutkijat käyttävät tutkimusten kirjoittamisessa. Myöhemmin BibTeX-merkkauksista tehtiin toinen resurssi sivustolle kirjanmerkkien lisäksi. (Benz et al. 2009, s. 323-324) BibSonomy valittiin tutkimukseen, sillä sivusto oli tullut tutuksi yhdestä aikaisemmasta tutkimuksesta.



Kuvio 17: Kuvankaappaus BibSonomy-sivuston etusivusta.

Sivuston sivu on jaettu neljään osaan: otsikko näyttää käyttäjälle informaatiota, kuten nykyinen sivu ja polku, navigointilinkit ja hakulaatikko, kaksi listaa: toinen kirjanmerkeille ja toinen julkaisuille, sekä listan tageja, jotka ovat merkityksellisiä listoissa oleville objekteille. Näkymä mahdollistaa sivuston folksonomiassa navigoinnin mihin suuntaan tahansa. Sivusto hyödyntää tagipilviä ilmaisemaan sivuston sisältöä sivustossa käytettyjen tagien avulla. (Benz et al. 2009, s. 323-324) Sivuston käyttäjillä on myös mahdollisuus hierarkkisten tagien käyttöön.

5.3 Aineiston keruu- ja analyysimenetelmät

Tutkimuksen aineiston keräämiseen käytetään eräänlaista ohjaavaa sivuston läpikäyntirunkoa, tarkoituksena kerätä tutkittavat parametrit yleisesti samanlaisen kaavan mukaan. On tosin huomioita, että sivustoissa voi esiintyä eroja, jolloin läpikäyntiä on sovellettava ja tutkittavaa sivustoa pitää lähestyä toisella tavalla. Aineiston keruu tapahtuu itsenäisesti ja aineisto analysoidaan taulukkomuodossa vertailemalla saatuja arvoja keskenään sekä tekemällä niistä havaintoja ja johtopäätöksiä.

Aineiston keräämiseen ja analysoimiseen haettiin ideoita muista aikaisemmista tutkimuk-

sista, mutta loppujen lopuksi päädyttiin hyvin käytännönläheiseen lähestymistapaan. Aikaisemmissa tutkimuksissa käytetyt menetelmät koostuivat muun muassa aikaan perustuvista arvioinneista sekä useat tutkimukset vertasivat kahta tai useaa hyvin erityyppistä käyttöliittymää keskenään tai aineistonkeruumenetelminä oli käytettävyydestejä. Joissain tutkimuksissa ei kuvailtu tarpeeksi tarkasti käytettyä tutkimusmenetelmää, jotta siitä olisi voinut hyödyntää lähestymistapoja tähän tutkimukseen.

Käyttöliittymän läpikäyntirunko on seuraavanlainen. Ensiksi asetutaan hakukäyttöliittymään aloituspisteeseen tai etusivuun, josta kerätään asioita visuaalisesti havainnoimalla. Sen jälkeen etsitään kahta tai kolmea eri asiaa, tarkoituksena saada hieman erilaisia hakutuloksia (jos se on mahdollista) ja analysoida niitä. Hakutuloksen lopuksi ollaan päädytty johonkin tulokseen, jolloin itse tulosobjekti on tarkastuksessa. Tämän jälkeen havainnoidaan mahdolliset vaihtoehdot navigoinnille: takaisinpaluu, haun tallentaminen sekä ehdotukset uusille hakupoluille ja -tuloksille. Tällä tavalla tutkimus voidaan osittain toistaa, tosin silloin on otettava huomioon muutamia asioita: sivuston toiminnan muutokset voivat johtaa erilaisiin toimenpiteisiin sekä havainnoijan kokemukset voivat johtaa erilaisiin johtopäätöksiin.

5.4 Tutkittavat parametrit

Tutkimuksen aineistoon kerätään kustakin tutkittavasta sivustosta erilaisia parametreja. Ne on luokiteltu tutkimuksen aikana suunniteltuun viitekehukseen tutkimuskysymysten perusteella ja niiden merkitys on selitetty tarkemmin. Viitekehukseen valittujen tarkasteltavien ominaisuuksien inspiraatio on tullut tarkastelemalla Wein, Liun ja Zhengin (Wei et al. 2013) tutkimusta, jonka jälkeen ideoita tarkasteltaviksi ominaisuuksiksi on ideoitu teoriaosuudesta ja osa on keksitty itse. Kyseistä viitekehystä voidaan myös pitää yhtenä tutkimustyön tuloksena.

Ensimmäinen parametriryhmä koostuu sivuston perustiedoista: **sivuston nimi**, **verkkosivuston tyyppi**, **verkkosivuston tyyli** ja **datakohdejoukko**. Sivuston tyyppi voi olla muun muassa hakuportaali, sosiaalinen verkkopalvelu, verkkokauppa tai yhdistelmä edellisistä. Verkkosivuston tyyli määritellään kahdella eri arvolla: rento tai ammattimainen. Datakohdejoukko voi vaihdella sivustojen välillä ja yksi sivusto voi sisältää useita eri objekteja, kuten

kirjoja, julkaisuja tai tekstidokumentteja.

Toinen parametriryhmä koostuu käyttöliittymän käyttäjäinteraktiosta: **käyttöliittymän tyyppi, käyttäjäinteraktio ja käyttöliittymän interaktiivisuus**. Käyttöliittymän tyyppi voi olla yksittäinen (käyttäjä voi selata käyttämällä vain yhtä näkymää), moninäköinen (käyttäjä voi valita monta näkymää kerrallaan), rinnakkainen (tukee rinnakkaista hakuvalintaa useiden näkymien välillä) tai automatisoitu (tarjoaa joukon navigaatio-operaatioita käyttäjälle selata yhtä tai useampaa tekstiä eri näkymistä ilman mitään syötteitä). Käyttäjäinteraktiolla selvitetään, kuinka hyvin käyttöliittymä noudattaa moninäkömaun käyttäjäinteraktiota sekä miten käyttäjä on vuorovaikutuksessa käyttöliittymän kanssa. Käyttöliittymän interaktiivisuus käsittää koko käyttöliittymän toiminnan muun muassa tarkastelemalla, ovatko vuorovaikutteiset käyttöliittymäelementit (hakukenttä, objektit) staattisia vai dynaamisia.

Kolmas parametriryhmä koostuu näkymiin sekä hakuobjektiin liittyviin asioihin: **näkymien luokittelu, näkymien esitystapa, objektien rankkaus, objektin esitystapa ja objektin sisältämä metadata**. Näkymien luokittelu voi olla joko enumeratiivinen (top-down) luokittelu, näköluokittelu tai hierarkkinen luokittelu. Näkymien esitystapa voi olla muun muassa hierarkkinen, lista, kartta tai jana. Objektien rankkaus kuvaa, kuinka käyttäjä voi järjestää hakutuloksensa, esimerkiksi käyttäjä voisi järjestää hakutulosobjektin kirjaimen, vuosiluvun tai objektille annetun arvostelun mukaan. Objektin esitystapa ja sisältämä metadata liittyvät hakutuloksena oleviin objekteihin.

Neljäs parametriryhmä syventyy tarkemmin käyttöliittymän hakemisprosessin toimintaan: **navigointi, tukevat hakukomponentit ja hakuehto- tai -tulosten tallennettavuus**. Navigointi keskittyy siihen, miten sujuvasti päästään rajaamaan ehtoja ja kuinka palataan takaisin alkupisteeseen (esimerkiksi klikkauksia laskemalla). Avustaviin hakukomponentteihin voi kuulua sanahaku tai tagihaku. Hakuehto- ja tulosten tallennettavuus tarkastelee, onko kyseistä tehtyä hakutulosta mahdollisuus käyttää uudelleen (tallennus omaan käyttäjätiliin tai pysyvä url, jolla pääsee takaisin samaan hakuun).

Viimeisessä parametriryhmässä pohditaan sivustojen **sosiaalisia ja semanttisia ominaisuuksia**. Sivustoista tarkastellaan sosiaalisina ominaisuuksina, voiko loppukäyttäjä vaikuttaa hakutuloksiin tai osaan datasta (onko suunnattu yhdelle käyttäjälle vai käyttäjäryhmälle, suo-

situkset, muiden merkkeamat tiedot) ja mitä mahdollisuuksia käyttäjällä on vuorovaikuttaa muihin palvelun käyttäjiin sekä semanttisina ominaisuuksina, onko havaittavissa tukea ontologioille, RDF-esityksille tai onko sivusto indeksoitu jossakin semanttisessa hakukoneessa. Sosiaalisista ominaisuuksista jätetään tarkastelun ulkopuolelle muiden sosiaalisten verkkopalveluiden käyttö osana käyttöliittymää (esim. Facebook ja Twitter).

	Objektiiviset parametrit	Subjektiiiset parametrit
Sivuston perustiedot	<ul style="list-style-type: none"> Nimi Tyyppi Datakohdejoukko 	<ul style="list-style-type: none"> Tyyli
Käyttöliittymä	<ul style="list-style-type: none"> Tyyppi 	<ul style="list-style-type: none"> Käyttäjäinteraktio Interaktiivisuus
Näkymät ja hakuobjektit	<ul style="list-style-type: none"> Luokittelu ja esitystapa <hr/> <ul style="list-style-type: none"> Rankkaus ja esitystapa Metadatasältö 	
Hakemisen prosessi	<ul style="list-style-type: none"> Tukevut hakukomponentit Hakuehtojen ja tulosten tallennettavuus 	<ul style="list-style-type: none"> Navigointi
Ominaisuudet	<ul style="list-style-type: none"> Sosiaaliset ominaisuudet Semanttiset ominaisuudet 	

Kuvio 18: Kuva tutkittavien parametrien viitekehyksestä.

6 Tutkimuksen tulokset

Tässä luvussa tarkastellaan tutkimuksen tuloksia. Tutkimusta toteutettiin 25.07 - 05.08.2014 välisenä aikana. Tutkimuksessa tutkittavat sivustot käytiin läpi luvussa 5.4 tutkimuksen aikana suunnitellun läpikäyntirungon avulla. Tulosten yksinkertaiseen esittämiseen käytetään taulukoita ja tarkempi kuvaus taulukoiden arvoista kuvaillaan parametriryhmittäin jaetuissa kappaleissa. Tutkimuksen tulokset tuovat esille, että moninäkömähaku soveltuu hyvin monenlaisten sivustojen hakumenetelmäksi.

6.1 Sivustojen tiedot

Taulukko 2: Tutkittavien sivustojen perustiedot.

	Tyyppi	Tyyli	Datakohdejoukko
Finna:	hakuportaali	rento	useita
JYX:	hakuportaali	ammattimainen	tutkimusaineisto
Yummly:	sos. verkkopalvelu	rento	ruokaresepti
Kulttuurisampo:	hakuportaali	ammattimainen	useita
Huuto:	verkkokauppa	rento	tavara
Gigantti:	verkkokauppa	rento	tuote
BibSonomy:	sos. verkkopalvelu	ammattimainen	julkaisu, kirjanmerkki

Tutkimukseen valittiin tasaisesti sosiaalisia, semanttisia ja kaupallisia verkkopalveluita. Jokainen sivusto voitaisiin luokitella tavalla tai toisella hakuportaaleiksi niiden hakumahdollisuuksien takia, mutta kaikki sivustot eivät ole puhtaita hakukoneita tai hakukäyttöliittymä on vain yksi osa sivuston tarjoamaa palvelua. Sivustojen tyylit jakautuvat tasaisesti rentoihin ja ammattimaisiin ulkoasuihin. Suurimmassa osassa sivustoissa haettava datakohdejoukko oli selvästi määritelty. Finna ja Kulttuurisampo-sivustojen kautta pystyy hakemaan erityyppisiä kohteita, näiden sivustojen datakohdejoukon koostuen useammasta vaihtoehdosta, kuten äänitteistä tai kirjoista (Finna) ja erilaisista esineistä (Kulttuurisampo). Gigantti ja Huuto-sivustoissa on myös mahdollista hakea monentyyppisiä kohteita, mutta ne voidaan koota yhden asiatyypin ympärille.

6.2 Sivustojen käyttäjäinteraktio

Taulukko 3: Käyttöliittymän toiminnan tutkitut parametrit.

	Käyttöliittymän tyyppi	Käyttäjä- interaktio	Käyttöliittymän interaktiivisuus
Finna:	moninäkömä	korkea	keskitaso
JYX:	moninäkömä	keskitaso	alhainen
Yummyly:	moninäkömä	korkea	korkea
Kulttuurisampo:	yksittäinen	alhainen	alhainen
Huuto:	moninäkömä	alhainen	keskitaso
Gigantti:	moninäkömä	korkea	keskitaso
BibSonomy:	yksittäinen	alhainen	keskitaso

Sivustojen käyttöliittymätyypeiksi on valikoitunut yksittäinen tai moninäkömäinen haku-käyttöliittymä. Kulttuurisampo-sivustossa näkymistä pystyi valitsemaan vain yhden vaihtoehdon kerrallaan. BibSonomy-sivuston tapauksessa tagien valinta aiheutti kahdenlaista toimintaa. Jos käyttäjä oli aloittanut hakemisen sanahauulla, tagi lisättiin sanahaun jatkeeksi, kun taas jos käyttäjä aloitti haun tagista, toista tagia ei ollut mahdollista lisätä hakuun, vaan uuden tagin valitseminen aiheutti uuden tagihaun. Muissa sivustoissa näkymiä pystyi valitsemaan useampia samaan hakuun sekä poistamaan niitä haun aikana.

Käyttäjäinteraktio onnistui parhaiten niillä sivustoilla, joissa käyttöliittymä oli selkeä ja näkymät oli helppo tunnistaa omana kokonaisuutena. Käyttöliittymän interaktio oli parempi, jos näkymät olivat ymmärrettäviä ja niiden funktio haun kaventajana oli ymmärrettävissä. Kulttuurisampo-sivustossa interaktio ontui, sillä käyttäjä pystyy valitsemaan vain yhden näkymän kerralla. Huuto-sivustossa haittaavaksi tekijäksi osoittautui se, että valittujen näkymien poistaminen oli tehty erittäin epäintuitiiviseksi (klikkaamalla uudestaan valittua kategoriaa), jonka lisäksi käyttäjää ei opasteta millään tavalla tähän käyttöliittymän toimintoon (esim. tooltip tai visuaalinen ohje). Huuto- ja BibSonomy-sivustot taas eivät tarjonneet samankaltaisia objekteja hakutuloksen päätyttyä, joten selaamista ei voinut jatkaa hyppäämällä muihin samankaltaisiin objekteihin. BibSonomy-sivustossa näkyminä toimivien tagien valinta aiheutti kahta erilaista toimintoa, mikä häytti käyttöliittymän interaktiota. Lisäksi

hämmennystä voi aiheuttaa myös tageilla ja konsepteilla (alitagit mukana) hakemisen ymmärtäminen. Näkymien ymmärrettävyys joissakin sivustoissa ei ollut aina selkeää: Gigantti- ja Huuto-sivuston tapauksessa ”Osasto”-näkyvä voi aiheuttaa ensikertalaiselle hämmennystä siitä, mikä vaikutus tämän näkymän valinnalla on hakuun.

Sivustojen käyttöliittymien interaktiivisuus oli suurimmaksi osaksi hyvin yksinkertaista. Monet käyttöliittymäkomponentit olivat staattisia, eivätkä ne reagoineet käyttäjien toimintoihin. Esimerkiksi kaikissa sivustoissa esiintynyt sanahakukenttä ei kaikissa sivustoissa tarjonnut automaattisesti hakusanavaihtoehtoja. Muutamassa sivustossa oli pieniä toimintoja, jotka reagoivat käyttäjän viedessä hiiren tiettyyn kohtaan. Gigantti-sivustossa esiintyi navigointia helpottava toiminto: hakuobjekteja pystyi selaamaan automaattisesti ilman hiiren klikkauksia vierittämällä sivua alaspäin. Kaikissa sivustoissa näkymän valinta päivitti sivun uudelleen. Staattisin tapaus tutkitavista sivustoista oli JYX, jossa kaikki sivuston objektit olivat suurimmaksi osaksi hyperlinkkejä.

6.3 Sivustojen näkymien ja objektien esitystapa

Taulukko 4: Käyttöliittymän näkymien tutkitut parametrit.

	Näkymien luokittelu	Näkymien esitystapa
Finna:	näkyvä	lista
JYX:	näkyvä	lista
Yummyly:	näkyvä	lista
Kulttuurisampo:	ei	lista
Huuto:	näkyvä	lista
Gigantti:	näkyvä	lista
BibSonomy:	ei	lista, pilvi

Näkymien luokittelussa sivustot käyttivät ilmeisimmin näkymäluokittelua, tosin tarkkaa tietoa käyttöliittymään valituista näkymistä ei ole. Gigantti ja JYX olivat ainoat sivustot, joissa osa näkymistä vaihtui haettavan kohteen mukaan (JYX:ssä käytössä vain muutamilla kokoelmissa). Esimerkiksi Gigantti-sivustossa haettaessa tietokonetta näkymistä sai valita ydinten

lukumäärän, kun taas älypuhelinta etsittäessä näkymistä sai valita SIM-kortin tyyppin. Muissa sivustoissa näkymät eivät vaihtuneet haettavan asian kontekstin mukaisesti, vaan pysyivät samoina. Kulttuurisampo-sivuston ”Hae ja jäsenä”-sivun käyttöliittymä ei tarjonnut näkömäloukittelua, vaan näkymät oli järjestetty esiintyvyyden mukaan. Bibsonomy-sivuston näkymät toimivat samalla tavalla, mutta käyttäjä pystyi valitsemaan joko aakkos- ja esiintyvyyjärjestyksen näille näkymille.

Kaikissa sivustoissa näkymien esitystavaksi oli valittu lista ja näkymä esitettiin tekstinä. Vain Bibsonomy-sivustossa näkymiä saattoi näyttää listan lisäksi jollakin muulla tavalla. Jos näkömävaihtoehtoja on paljon, näkymien näyttäminen vertikaalisena listana on järkeenkäypää ajatus. Jos näkymiä on valita vähän (kuten esimerkiksi Huuto-sivustossa) ja ne eivät peitä merkittävästi osia sisällöstä, näkymiä voisi näyttää myös horisontaalisena janana. Näkymien hakukategorioiden esittämisessä esiintyi vaihtelua. Osa sivustoista näytti ensiksi vain osan kategorioista ja painamalla erillistä painiketta avautui näkyville muut hakukategoriat. Painamisen jälkeen kaikki mahdolliset hakukategoriat näytettiin käyttäjälle. Finna-sivusto piilotti eniten tarjottavia kategorioita, näyttäen vain yhden näkymän kategorioita avattuna. Joitakin kategoriavaihtoehtoja tarjottiin vapaavalintaisena arvona tekstikenttänä, jolloin nämä näkymät luokiteltiin tietotyypin sisällyttämiseksi hakuun. Yummly oli ainoa sivusto, joka tarjosi näkymän hakukategorioiden arvoja liukusäätimenä.

Taulukko 5: Käyttöliittymän hakuobjektin tutkitut parametrit.

	Objektien rankkaus	Objektin esitystapa	Objektin metadata
Finna:	useita	kuva ja hyperlinkki	paljon
JYX:	useita	hyperlinkki	vähän
Yummly:	useita	kuva ja hyperlinkki	paljon
Kulttuurisampo:	ei	kuva ja hyperlinkki	keskitaso
Huuto:	useita	kuva ja hyperlinkki	keskitaso
Gigantti:	useita	kuva ja hyperlinkki	paljon
BibSonomy:	useita	kuva ja hyperlinkki	keskitaso

Hakuobjekteja pystyi järjestelemään lähes kaikissa sivustoissa hyvin monipuolisesti. Valittavina vaihtoehtoina useimmiten esiintyi vastaavuus, muita järjestämismahdollisuuksia oli ajanjakso (uusimmasta vanhimpaan), hinta (halvimmasta kalleimpaan) ja määrä (pienemmästä suurempaan). Objektien esitystavassa kukin sivusto näytti objekteja hyvin pelkistetyksi, objektien koostuen kuvasta ja hyperlinkistä, joka vei itse hakuobjektisivulle. Objektien metadataesityksessä esiintyi vaihtelevia arvoja ja joissakin sivustoissa objektin esittämä metadata vaihteli objektin tyyppin mukaan (mitä on näytettävä). Objektien metadatan esittämisessä löytyi myös eroja sivustojen väliltä, joidenkin sivustojen tyytyessä näyttämään vähemmän tietoa objektista. Kaikissa tapauksissa esitettävä metadata vastasi vain osaa kaikesta objektiin liitetystä metadatasta.



Kuvio 19: Esimerkit JYX, Huuto ja Bibsonomy-sivustojen hakuobjekteista.



Kuvio 20: Esimerkit Finna, Yummly, Gigantti ja Kulttuurisampo-sivustojen hakuobjekteista.

6.4 Sivustojen hakemisen prosessi

Taulukko 6: Käyttöliittymän hakemisprosessin tutkitut parametrit.

	Navigointi	Tukevat haku- komponentit	Haun tallennettavuus
Finna:	keskitaso	sana	haku, objekti, URL
JYX:	keskitaso	sana	objektiin pysyvä URL
Yummyly:	korkea	sana, tagi	objekti
Kulttuurisampo:	alhainen	sana	ei
Huuto:	alhainen	sana	objekti, URL
Gigantti:	keskitaso	sana	objekti
BibSonomy:	keskitaso	sana, tagi	käyttäjä, objekti

Hakemisen prosessissa ei havaittu suuria eroja sivustojen välillä. Kaikissa sivustoissa hakuprosessi ja tietojoukossa navigointi tapahtui ensiksi asettamalla ensimmäinen hakusana tai näkymä, jonka jälkeen lisätään haku-ehdot (joko uusia hakusanoja, tageja tai valitsemalla näkymiä). Toistettaessa edellistä toimintaa tarpeeksi kauan päädytään johonkin hakutulosityhmään, josta valitaan haluttu objekti, jota käyttäjä voi tarkastella erillisessä sivussa. Navigointia pystyi myös harjoittamaan leivänmurun tai navigointipainikkeiden kautta. Navigoimalla seuraavalle sivulle käyttäjä saattoi valita, montako seuraavaa hakutulosta sivulla näytetään. Leivänmuru edusti useimmissa sivustoissa käyttäjän nykyisen haun historiaa. Finna-sivustossa leivänmuru oli kavennettu näyttämään hakupolun sijaan vain viimeisin muutos hakuun. JYX-sivustossa taas leivänmuru edusti hierarkkista listausta sivuston kokoelmista (sivusto \rightarrow polku \rightarrow ... \rightarrow objekti).

Tutkimuksen tulosten perusteella kaikki sivustot käyttävät sanahakua tukevana hakukomponenttina. Tälle on ymmärrettävä selitys. Sanahaku on jokaisessa sivustossa sijoitettu sivuston etusivulle, jolloin hakukomponentti avaa nopean ja yksinkertaisen pääsyn haulle. Sanahaku on myös hyvin tuttu lähes kaikille webin käyttäjille, joten hakemiseen hakeutuminen onnistuu ymmärrettävästi. Toisaalta moninäkömahaku ei ensimmäisenä avaudu käyttäjälle, vaan siihen pääsee käsiksi joko aloittamalla haun sanahauulla tai jonkun muun linkin kautta.

Haun tallennettavuudessa sivustojen välillä oli vaihtelua. Suurimmassa osassa sivustoissa

hakuobjekteja pystyi tallentamaan käyttäjätiliin. JYX-sivusto oli ainoa, jossa hakuobjektin pystyi tallentamaan pysyvällä URL:lla käyttäjäliittymän kautta. Huuto- ja Finna-sivustoissa hakuehdot näkyivät sessiosta riippumatta myös URL:ssa, tosin Finna-sivuston haku-URL oli hankalahko tyhjien HTTP-parametrien takia. Pysyvä URL on siitä hyödyllinen, että se mahdollistaa käyttäjien tallentaa hakutuloksia ilman rekisteröitymistä palveluun. Finna-sivusto oli ainutlaatuinen sen suhteen, että itse hakutuloksen (sisältäen useita hakuobjekteja) pystyi tallentamaan käyttäjätiliin ja tarkastelemaan sitä myöhemmin. Joissakin sivustoissa hakuobjektit pystyi ryhmittämään haluamaansa ryhmiin (esim. Finna ja Yummlly).

6.5 Sivustojen sosiaaliset ja semanttiset ominaisuudet

Taulukko 7: Listaus sivustojen sosiaalisista ja semanttisista ominaisuuksista.

	Sosiaaliset ominaisuudet	Semanttiset ominaisuudet
Finna:	objektien arvostelu käyttäjätili vaikuttaa hakuihin	yhdistetty JYX:n tietokantaan
JYX:	-	RefWorks
Yummlly:	objektien lisäys, suosittelu ja jako käyttäjätili vaikuttaa hakuihin	Food ja Recipe API
Kulttuurisampo:	objektien kommentointi	sisällöntuontantoprosessi
Huuto:	objektien lisäys ja arvostelu käyttäjätili vaikuttaa hakuihin	-
Gigantti:	objektien arvostelu	-
BibSonomy:	objektien lisäys ja arvostelu käyttäjätili vaikuttaa hakuihin	-

Sosiaalisia ominaisuuksia löytyi lähes kaikista sivustoista. Suuressa osassa nämä ominaisuudet rajoittuivat vain hakutulokset kommentointiin tai arviointiin. Käyttäjätilin hyödyntämistä hakutulosten esittämisessä niin, että samalla perushaulla tehdyt tulokset voivat vaihdella eri käyttäjillä, hyödynsi muutama sivusto. Näistä sivustoista Yummlly-sivusto hyödynsi käyttäjätilin tietoja näkyvimmin, mahdollistaen kytkeä käyttäjätilin tiedot itse hakuun täyttä-

mällä käyttäjätiliin omat makumieltymyksensä, jotka liittyivät esitäyteyiksi näkymien kategoriavaihtoehtoiksi automaattisesti kaikkiin myöhempiin hakuihin. Finna-sivustossa käyttäjätilin tietoina vaikuttivat tallennetut hakutulokset ja lainaukset, BibSonomy-sivustossa omat kirjanmerkkaukset (tosin tulosten rankkaukseen nämä eivät vaikuttaneet muuten kuin yhden hakuparametrin oletusarvoina) sekä Huuto-sivustossa käyttäjätiedot vaikuttivat omien myytävien kohteiden hallintaan. Voidaan siis todeta, että tutkimuksen otoksen mukaan käyttäjätilin tietoja tai hakuhistoriaa hyödynnetään vaihtelevasti moninäkömähakukäyttöliittymissä.

Sivustojen semanttisia ominaisuuksia oli vaikea havaita itse hakukäyttöliittymässä, jolloin tarvittava tieto haettiin muualta. JYX-sivuston selauskäyttöliittymä tarjoaa tietojen viemisen RefWorks:n avulla. Kulttuurisampo-sivuston mukaan hakutuloksen sisällöntuotto tapahtuu semanttisesti, vaikkakin tätä oli vaikea havaita (haussa ”liitos ontologiakohteeseen”-kohta). Yummly-sivusto taas tarjoaa rajapinnan hyödyntää Yummlyä itse kehitetyissä sovelluksissa sekä sivusto hakee hakuobjekteja muista sivustoista ja blogeista.

6.6 Yleiset kehitysehdotukset moninäkömähauille

Tutkimusten tulosten perusteella moninäkömähauun ominaisuuksia hyödynnetään kokonaisuudessaan hyvin, mutta vaihtelevasti. Myös joidenkin perusominaisuuksien puuttuminen (kuten näkymien intuitiivinen poisto) on huomioitava asia. Täten yleisenä kehitysehdotuksena voidaan pitää moninäkömähauun perusteiden valjastamista hakukäyttöliittymään. Kukin sivusto voisi myös parantaa sivustojaan tekemällä joitakin toimintoja automaattisiksi. Näkymien tarjoaminen haettavan tuotteen mukaan ja automaattinen selaus seuraaville hakuobjekteille voisi parantaa käyttöliittymän interaktiota. Hakuobjektien tarkastelu ilman sivun päivittämistä sekä monen objektin tarkastelu yhtäaikaaisesti voisi tarjota parempia vertailumahdollisuuksia käyttäjälle. Tosin tässä tulee huomioida asia, että sivuston selaaminen tehdään yleensä ”Google-yhteensopivaksi”, jotta sivuston ulkopuoliset käyttäjät voivat päästä käsiksi erinäisiin tietoihin ulkopuolisten linkkien kautta. Sivustot, joiden haettava tietoavaruus on suuri (esim. Finna) voisivat käyttää rinnakkaista tai automaattista moninäkömähakua käyttöliittymässä parantaakseen käyttäjän tiedonhakua.

Käyttäjätilin tietojen hyödyntämistä hauissa hyödynnettiin tutkittavissa sivustoissa vaihtelevasti.

vasti. Käyttäjätilin hyödyntäminen voisi tarjota paljon hyödyllisiä ominaisuuksia käyttäjälle. Esimerkiksi metadatan esittäminen (näyttää käyttäjää kiinnostavaa metadataa kaiken mahdollisen sijaan) ja objektien rankkaus käyttäjän mieltymysten mukaan voisi tuoda paremmin esille häntä kiinnostavat objektit. Myös yksi hyödyllinen ominaisuus voisi olla personoitujen tulosten saaminen itselleen (esim. RSS-syötteen kautta) käyttäjätiliin, mitä esimerkiksi Yummys-sivusto tekee jo, tarjoamalla käyttäjälle uusia reseptisuosituksia käyttäjätiliin. Toisaalta käyttäjätilin tiedot tuovat esille myös yksityisyyteen liittyviä asioita: haluaako käyttäjä, että hänen mieltymyksensä ja hakuhistoriansa on hänelle näkyvästi tallennettuna, vaikkakin haku sinällään on jo tallentunut palvelun käyttöhistoriaan.

6.7 Tagikäyttöliittymä moninäkömähakuna

Tutkittavista sivustoista BibSonomy ei ollut niinkään moninäkömähakukäyttöliittymä, vaan sivusto hyödyntää tageja hakutoiminnoissaan. Tutkimuksen aikana tarkasteltiin kyseisen käyttöliittymän mahdollista soveltuvuutta moninäkömähakuksi.

Tällä hetkellä BibSonomy-sivuston ei voida todeta olevan moninäkömähakukäyttöliittymä, mutta tutkimuksen perusteella tagit voisivat tarjota muutamalla toiminnon muuntamisella samankaltaisia ominaisuuksia, mitä moninäkömähaku tarjoaa. Tagit tekevät jo osaa moninäkömähäun toiminnoista: ne ovat kontekstiriippuvaisia ja ne rajoittavat hakua. Tagien jonkinlainen ryhmittely sekä näyttämällä jokaisen tagilisäyksen jälkeinen hakutulospääte veisi tageja lähemmäksi moninäkömähakua. Toisaalta ryhmittämisessä on oma ongelmansa: tagit ovat käyttäjien itse määrittämiä vapaavalintaisia merkkijonoja ilman erityistä merkitystä ja tagien osittainen arvo perustuu juurikin niiden henkilökohtaisuuteen. Siispä niille sivustoille, joiden tagit ovat määräytyneet käyttäjien toiminnan tuloksena muokkaantuisi omanlainen näkömähäköelma, joka ei riipu sivustoon lisättävistä objekteista, vaan käyttäjien luomista näkömähäköelma. Luultavasti sivustolle luotavaa folksonomian perustaa pitäisi tukea jollakin tavalla, esimerkiksi olemassa olevalla ontologialla.

Toisaalta jos BibSonomy-sivuston käyttäjällä on käytössä hierarkkisia tageja, niin silloin omien lähteiden selaus BibSonomyssa on rajoittuneessa mielessä jo moninäkömähakua. Toisaalta tulosten yleistettävyyttä monelle käyttäjälle voidaan pitää todellisena ongelmana, mut-

ta tätä asiaa voidaan tarkoituksenmukaisesti kyseenalaistaa, jos moninäkömähakua ajatellaan apuna tietojen tai viitteiden mieleen palauttamiseen. Tätä voidaan pitää täysin olennaisena ongelmana, jos kirjanmerkkejä on kertynyt käyttäjälle useampia satoja. Tällöin organisoitongelman painopisteeksi siirtyy olemassa olevien metatietojen mahdollisimman helppo massaeditointi. Tähän ratkaisuksi voisi ajatella näkymien automaattista luomista ”lennossa” sekä synkronointi- ja koostamismahdollisuuksien lisäämistä.

7 Yhteenveto

Tutkimuksessa tutkittiin seitsemää eri verkkopalvelua, jotka käyttivät moninäkömahakukäyttöliittymää osana tai kokonaan palvelussaan. Tutkimuksen tulosten perusteella moninäkömahauksen ominaisuuksia hyödynnettiin vaihtelevasti. Osa sivustoista hyödynsi paremmin ominaisuuksia kuin muut sivustot, mutta mikään sivusto ei hyödyntänyt maksimaalisesti kaikkia moninäkömahauksen tarjoamia ominaisuuksia. Hyödynnettyjä ominaisuuksia oli näkymien luokittelu, näkymien muuttuminen haettavan objektin mukaan ja semanttinen haku muista sivustoista. Suurimpana puutteena voidaan mainita rinnakkaisen tai automaattisen haun puute sekä käyttäjän datan liittäminen hakuun. Tutkimuksen tulokset antavat eräänlaisen kuvan, kuinka moninäkömahakua hyödynnetään eri verkkopalveluissa ja mihin asioihin sen hyödyntäminen keskittyy. Tutkittavista sivustoista moninäkömahakua hyödynsi parhaiten Yummys-sivusto, joka tarjosi parhaimman tuen semanttiselle haulle ja joka myös hyödynsi sosiaalisena ominaisuutena selkeimmin käyttäjän mieltyämyksiä ja hakuhistoriaa hauissa. Sivustoa voidaan pitää tämän tutkimuksen kontekstissa eräänlaisena mittarina moninäkömahaulle. Sivustojen välillä havaittiin myös huomattavia eroja hakemisen laadukkuudessa ja käyttäjäystävällisyydessä.

Millainen on sivustojen käyttäjäinteraktio ja kuinka interaktiivinen käyttöliittymä on käyttäjälle?

Sivustojen hakukäyttöliittymän käyttäjäinteraktiossa oli vaihtelua. Käyttäjäinteraktio onnistui parhaiten niillä sivustoilla, joissa käyttöliittymä oli selkeä ja näkymät oli helppo tunnistaa omana kokonaisuutena. Käyttöliittymän interaktio oli parempi, jos näkymät olivat ymmärrettäviä ja niiden funktio haun kaventajana oli ymmärrettävissä. Käyttäjäinteraktiota heikensi, jos käyttäjä pystyi valitsemaan vain yhden näkymän kerrallaan, valittuja näkymiä ei pystynyt selkeästi poistamaan haun aikana käyttöliittymän kautta, selaamista ei voinut jatkaa hyppäämällä muihin samankaltaisiin objekteihin, näkymän valinta aiheutti kahta erilaista toimintoa tai jos näkymän valinnan vaikutusta hakuun ei ymmärretty. Sivustojen käyttöliittymien interaktiivisuus oli suurimmaksi osaksi hyvin yksinkertaista ja monet käyttöliittymäkomponentit olivat staattisia. Muutamassa sivustossa oli pieniä toimintoja, jotka reagoivat käyttäjän vihdessä hiiren niiden päälle. Yhdessä sivustossa hakuobjekteja pystyi selaamaan automaatti-

sesti ilman hiiren klikkauksia. Kaikissa sivustoissa näkymän valinta päivitti sivun uudelleen.

Miten hakemisen prosessi tapahtuu käyttöliittymässä ja mitä muita avustavia hakukomponentteja käyttöliittymä tarjoaa?

Hakemisen prosessissa ei havaittu suuria eroja sivustojen välillä. Kaikissa sivustoissa hakuprosessi ja tietojoukossa navigointi tapahtui ensiksi asettamalla ensimmäinen hakusana tai näkymä, jonka jälkeen lisättiin hakuehtoja. Toistettaessa edellistä toimintaa tarpeeksi kauan päädyttiin johonkin hakutulosityhmään, josta valittiin haluttu objekti. Navigointia pystyi myös harjoittamaan ”leivänmurun” tai navigointipainikkeiden kautta. Käyttäjä sai myös valita, montako seuraavaa hakutulosta sivulla näytetään. Leivänmuru edusti useimmissa sivustoissa käyttäjän nykyisen haun historiaa ja yhdessä sivustossa se edusti hierarkkista listausta sivuston objekteista. Avustavina hakukomponentteina kaikki sivustot käyttivät sanahakua ja kaksi sivustoa hyödynsi tageja.

Miten näkymät ja objektit esitetään käyttöliittymässä?

Näkymien luokittelussa viisi sivustoa käytti näkymäluokittelua, tosin tarkkaa tietoa käyttöliittymään valituista näkymistä ei ole. Vain kahdessa sivustossa osa näkymistä vaihtui haettavan kohteen mukaan. Muissa sivustoissa näkymät eivät vaihtuneet haettavan asian kontekstin mukaisesti, vaan pysyivät samoina. Tutkittavista sivustoista kaksi ei tarjonnut ollenkaan näkymäluokittelua, vaan näkymät järjestettiin esiintyvyyden tai aakkosjärjestyksen mukaan. Kaikissa sivustoissa näkymien esitystavaksi oli valittu lista ja yhdessä sivustossa näkymiä saattoi näyttää myös pilvenä. Näkymien hakukategorioiden esittämisessä esiintyi vaihtelua. Osa sivustoista näytti ensiksi vain osan kategorioista ja painamalla erillistä painiketta avautui näkyville muut hakukategoriat. Joillekin hakukategorioille saattoi valita arvon vapaavalintaisesti. Kaikissa sivustoissa näkymä esitettiin tekstinä.

Hakuobjektien yleisin esitystapa oli kuva ja hyperlinkki sekä niitä pystyi järjestelemään lähes kaikissa sivustoissa hyvin monipuolisesti. Objektien metadataesityksessä esiintyi vaihtelevia arvoja ja joissakin sivustoissa objektin esittämä metadata vaihteli objektin tyyppin mukaan. Objektien metadatan esittämisessä löytyi myös eroja, joidenkin sivustojen tyytyessä näyttämään vähemmän tietoa objektista. Kaikissa tapauksissa esitettävä metadata vastasi vain osaa kaikesta objektiin liitetystä metadatasta.

Millaisia ovat sivustojen sosiaaliset ja semanttiset ominaisuudet?

Sosiaalisia ominaisuuksia löytyi lähes kaikista sivustoista. Suuressa osassa nämä ominaisuudet rajoittuivat vain hakutulosoikeuksien kommentointiin tai arviointiin. Vain yhdessä sivustossa oli mahdollista kytkeä selkeästi käyttäjätilin tiedot itse hakuun ja kolmessa muussa sivustossa käyttäjätilin tiedot vaikuttivat käyttäjän hakuihin jollakin tavalla. Sivustojen semanttisia ominaisuuksia oli vaikea havaita itse hakukäyttöliittymässä. Sivustojen semanttisia ominaisuuksia olivat tietojen vieminen muihin järjestelmiin, haun semanttinen sisällöntuotto, semanttinen tiedonhaku muista sivustoista sekä rajapinnan hyödyntäminen itse kehitetyissä sovelluksissa.

Lopuksi, millainen on semanttisten hakukoneiden ja moninäköhaun tulevaisuus?

On tiedostettava, että tiedon käsittelemiseen, luomiseen ja hakemiseen kuluu paljon aikaa. Erilaisten tutkimusten kautta löytyvien teorioiden ja sovellusten löytäminen auttaa yhteiskuntaa tekemään parempia sovelluksia tiedon hakemiseen ja käsittelyyn. Tulevaisuudessa on mahdollista käyttämällä metadatojen sovelluksia luoda järjestelmiä, joissa tietokoneet ymmärtävät tietoa ja luovat itse uutta tietoa olemassa olevasta tiedosta. Semanttisia hakukoneita on jo nyt, mutta näiden tämänhetkinen kattavuus on aivan toinen asia. Kaikkia tekstiaineistoja ei varmaankaan tulla muuttamaan semanttisesti, ellei luonnollisen kielen käsittely edisty merkittävästi. Toisaalta taas yksittäisten tietoa tuottavien organisaatioiden osalta semanttisten tietojoukkojen kontrolloitu tuottaminen kuulostaa realistisemmalta tavoitteelta, oikean yritysideoita löytymisen jälkeen. Linkitetty avoin dataprojekti taas on luonut paljon linkitettyjä data-aineistoja, jonka perusteella voidaan sanoa, että datajoukot eivät enää ole ainakaan kokonaan itsenäisiä, mutta niitä käyttävät sovellukset ja palvelut ehkä ovat. Ehkäpä tulevaisuudessa saatamme kuitenkin nähdä Googlen kaltaisen semanttisen hakupalvelun, joka hakee yleisesti tietoa aiheesta kuin aiheesta (tästä viitteitä antaa jo sivusto DBpedia⁵). Tämä askel tietäisi suurta mullistavaa muutosta ihmisten yleiselle tutkivalle tiedonhauille.

5. <http://dbpedia.org/About>

Lähteet

Ahmed, Kal, Daniel Rivers-Moore, Joshua Lubell, Andrew Watt, Mark Birbeck, Jay Cousins, Rob Worden, Miloslav Nic, Danny Ayers ja Ann Wrightson. 2001. *Professional XML meta data*.

Ben-Yitzhak, Ori, Nadav Golbandi, Nadav Har'El, Ronny Lempel, Andreas Neumann, Shila Ofek-Koifman, Dafna Sheinwald, Eugene J. Shekita, Benjamin Sznajder ja Sivan Yogev. 2008. "Beyond basic faceted search." Teoksessa *Wisdom (WSDM)*, toimittanut Marc Najork, Andrei Z. Broder ja Soumen Chakrabarti, 33–44. ACM. doi:10.1145/1341531.1341539.

Benz, Dominik, Folke Eisterlehner, Andreas Hotho, Robert Jäschke, Beate Krause ja Stum Gerd. 2009. "Managing publications and bookmarks with BibSonomy". Teoksessa *Proceedings of the 20th ACM conference on Hypertext and hypermedia*, 323–324. HT '09. Torino, Italy: ACM. ISBN: 978-1-60558-486-7. doi:10.1145/1557914.1557969.

Bianchini, Devis, ja Valeria Antonellis. 2012. "Linked Data Services and Semantics-Enabled Mashup". Teoksessa *Semantic Search over the Web*, toimittanut Roberto De Virgilio, Francesco Guerra ja Yannis Velegrakis, 283–307. Data-Centric Systems and Applications. Springer Berlin Heidelberg. ISBN: 978-3-642-25007-1. doi:10.1007/978-3-642-25008-8_11.

Bizer, Christian, Tom Heath ja Tim Berners-Lee. 2009. "Linked data-the story so far". Toimittanut IGI Global. *International Journal on Semantic Web and Information Systems (IJSWIS)* 5 (3): 1–22. <http://www.igi-global.com/article/linked-data-story-far/37496>.

Bozzon, Alessandro, Marco Brambilla, Emanuele Della Valle, Piero Fraternali ja Chiara Pasini. 2011. "A Conceptual Framework for Linked Data Exploration." Teoksessa *The International Conference on Web Engineering (ICWE) Workshops*, toimittanut Andreas Harth ja Nora Koch, 7059:109–118. Lecture Notes in Computer Science. Springer. ISBN: 978-3-642-27996-6. doi:10.1007/978-3-642-27997-3_10.

- Breslin, JohnG., Alexandre Passant ja Denny Vrandečić. 2011. "Social Semantic Web". Teoksessa *Handbook of Semantic Web Technologies*, toimittanut John Domingue, Dieter Fensel ja James A. Hendler, 467–506. ISBN: 978-3-540-92912-3. doi:10.1007/978-3-540-92913-0_12.
- Buschbeck, Sven, Anthony Jameson, Adrian Spirescu, Tanja Schneeberger, Raphaël Troncy, Houda Khrouf, Osmo Suominen ja Eero Hyvönen. 2013. "Parallel faceted browsing". Teoksessa *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, 3023–3026. Paris, France: ACM. ISBN: 978-1-4503-1952-2. doi:10.1145/2468356.2479601.
- Capra, Robert, Gary Marchionini, Jung Sun Oh, Fred Stutzman ja Yan Zhang. 2007. "Effects of structure and interaction style on distinct search tasks". Teoksessa *JCDL '07: Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries*, 442–451. ACM. ISBN: 978-1-59593-644-8. doi:10.1145/1255175.1255267.
- Castano, Silvana, Alfio Ferrara ja Stefano Montanelli. 2012. "Thematic Clustering and Exploration of Linked Data." Teoksessa *SeCO Book*, toimittanut Stefano Ceri ja Marco Brambilla, 7538:157–175. Lecture Notes in Computer Science. Springer. ISBN: 978-3-642-34212-7. doi:10.1007/978-3-642-34213-4_11.
- Cayzer, Steve. 2004. "Semantic blogging and decentralized knowledge management". *Commun. ACM* 47, numero 12 (joulukuu): 47–52.
- Chandrasekaran, B., John R. Josephson ja V. Richard Benjamins. 1999. "What are Ontologies, and Why Do We Need Them?" *IEEE Intelligent Systems*:20–26.
- Clements, Maarten, Arjen P. de Vries ja Marcel J. T. Reinders. 2010. "The task-dependent effect of tags and ratings on social media access." *ACM Trans. Inf. Syst.* 28 (4): 21. doi:10.1145/1852102.1852107.
- Corcho, Óscar, Mariano Fernández-López ja Asunción Gómez-Pérez. 2003. "Methodologies, tools and languages for building ontologies: Where is their meeting point?" *Data Knowledge Engineering* 46 (1): 41–64.

- d'Aquin, Mathieu. 2011. "Where to Publish and Find Ontologies? A Survey of Ontology Libraries", nide 11. <http://www.websemanticsjournal.org/index.php/ps/article/view/217>.
- Devedzic, Vladan. 2002. "Understanding ontological engineering." *Commun. ACM* 45 (4): 136–144. doi:10.1145/505248.506002.
- Dimitrova, Vania, Lydia Lau, Dhavalkumar Thakker, Fan Yang-Turner ja Dimoklis Despotakis. 2013. "Exploring Exploratory Search: A User Study with Linked Semantic Data". Teoksessa *Proceedings of the 2nd International Workshop on Intelligent Exploration of Semantic Data*, 2:1–2:8. Intelligent Exploration of Semantic Data (IESD) '13. Paris, France: ACM. ISBN: 978-1-4503-2006-1. doi:10.1145/2462197.2462199.
- English, Jennifer, Marti Hearst, Rashmi Sinha, Kirsten Swearingen ja Ka-Ping Yee. 2002. "Flexible Search and Navigation using Faceted Metadata". https://oc.ac.ge/pluginfile.php/451/mod_resource/content/0/_2_English_et_al_2003_Flexible_Search_and_Navigation_using_Faceted_Metadata.pdf.
- Fagan, Jody Condit. 2010. "Usability Studies of Faceted Browsing: A Literature Review." doi:10.6017/ital.v29i2.3144.
- Finna. 2014. <https://www.finna.fi/Content/about>. [Online; noudettu 26. elokuuta 2014].
- Giess, M.D., P.J. Wild ja C.A. McMahon. 2008. *International Journal of Information Management* 28 (5): 379–390. doi:10.1016/j.ijinfomgt.2007.10.001.
- Gigantti. 2014. <http://www.gigantti.fi/cms/gigantti-fi/tietoa-gigantista/>. [Online; noudettu 26. elokuuta 2014].
- Gilliland-Swetland, Anne J. 2000. "Setting the Stage". Teoksessa *Introduction to Metadata: Pathways to Digital Information*, toimittanut Murtha Baca. Getty Research Institute. http://www.getty.edu/research/publications/electronic_publications/intrometadata/setting.html.
- Greaves, Mark, ja Peter Mika. 2008. "Semantic Web and Web 2.0". *Semantic Web and Web 2.0, Web Semantics: Science, Services and Agents on the World Wide Web* 6 (1): 1–3.

- Grimnes, Gunnar Aastrand, Malte Kiesel, Mohammed Abufouda ja Armin Schröder. 2012. “Semantic Integration through Linked Data in the iGreen project.” Teoksessa *GIL Jahres-tagung*, toimittanut Michael Clasen, Georg Fröhlich, Heinz Bernhardt, Knut Hildebrand ja Brigitte Theuvsen, 194:107–110. ISBN: 978-3-88579-288-8.
- Gruber, Thomas R. 1993. “Toward Principles for the Design of Ontologies Used for Knowledge Sharing”. *Toward Principles for the Design of Ontologies Used for Knowledge Sharing* 43:907–928. doi:10.1006/ijhc.1995.1081.
- Gupta, Manish, Rui Li, Zhijun Yin ja Jiawei Han. 2010. “Survey on Social Tagging Techniques”, *ACM SIGKDD Explorations Newsletter* (New York, NY, USA) 12 (1): 58–72. doi:10.1145/1882471.1882480.
- Haase, Peter, Michael Schmidt ja Andreas Schwarte. 2011. “The Information Workbench as a Self-Service Platform for Linked Data Applications.” Teoksessa *Consuming Linked Data (COLD)*, toimittanut Olaf Hartig, Andreas Harth ja Juan Sequeda, nide 782. Sun SITE Central Europe (CEUR) Workshop Proceedings. CEUR-WS.org.
- Häggström, Thomas. 2007. “Toimintakeskeisen semanttisen moninäkömäähaun toteutus ja evaluointi kulttuurialan portaalisovelluksessa”. Thesis Required for a Diploma, Teknillinen Korkeakoulu, Otaniemi, Helsinki.
- Hausenblas, Michael, Robert Grossman, Andreas Harth ja Philippe Cudré-Mauroux. 2012. “Large-scale Linked Data Processing - Cloud Computing to the Rescue?.” Teoksessa *CLOSER*, toimittanut Frank Leymann, Ivan Ivanov, Marten van Sinderen ja Tony Shan, 246–251. SciTePress. ISBN: 978-989-8565-05-1.
- Hearst, Marti A. 2008. “UIs for Faceted Navigation: Recent Advances and Remaining Open Problems”.
- Heath, Tom, ja Chris Bizer. 2011. *Linked Data: Evolving the Web into a Global Data Space*. <http://linkeddatabook.com/editions/1.0/>.
- Heymann, Paul, ja Hector Garcia-Molina. 2006. *Collaborative Creation of Communal Hierarchical Taxonomies in Social Tagging Systems*. Tekninen raportti. Computer Science Department, Stanford University. <http://dbpubs.stanford.edu:8090/pub/2006-10>.

- Holi, Markus. 2010. “Crisp, Fuzzy and Probabilistic Faceted Semantic Search”. Doctoral Dissertation, Aalto University.
- Horrocks, Ian, Bijan Parsia, Peter Patel-Schneider ja James Hendler. 2005. “Semantic Web Architecture: Stack or Two Towers?” Teoksessa *Principles and Practice of Semantic Web Reasoning*, toimittanut François Fages ja Sylvain Soliman, 3703:37–41. Lecture Notes in Computer Science. Springer Berlin Heidelberg. ISBN: 978-3-540-28793-3. doi:10.1007/11552222_4.
- Hyvönen, E., toimittanut. 2002. *Representing Metadata about Web Resources*. HIIT Publications 2002.
- Iturrioz, Jon, Oscar Díaz ja Iker Azpeitia. 2011. “Reactive tags: associating behaviour to prescriptive tags.” Teoksessa *The 22nd ACM Hypertext Conference (HT)*, toimittanut Paul De Bra ja Kaj Grønbaek, 191–200. ACM. ISBN: 978-1-4503-0256-2. doi:10.1145/1995966.1995993.
- JYX. 2014. <http://openaccess.jyu.fi/jyx>. [Online; noudettu 26. elokuuta 2014].
- Kirjastot.fi. 2014. <http://www.kirjastot.fi>. [Online; noudettu 7. lokakuuta 2014].
- Koivisto, Antti. 2010. “Kontekstitietoon perustuva automaattinen tagien luonti mobiililaitteilla”. Thesis Required for a Diploma, Tampere University of Technology.
- Komamizu, Takahiro, Toshiyuki Amagasa ja Hiroyuki Kitagawa. 2012. “Faceted Navigation Framework for XML Data.” *International Journal of Web Information Systems (IJWIS)* 8 (4): 348–370. doi:10.1108/17440081211282865.
- Koren, Jonathan, Yi Zhang ja Xue Liu. 2008. “Personalized Interactive Faceted Search”. Teoksessa *Proceedings of the 17th International Conference on World Wide Web (WWW)*, 477–486. WWW '08. ACM. ISBN: 978-1-60558-085-2. doi:10.1145/1367497.1367562.
- Kules, Bill, ja Robert Capra. 2012. “Influence of training and stage of search on gaze behavior in a library catalog faceted search interface.” *Journal of the Association for Information Science and Technology (JASIST)* 63 (1): 114–138. doi:10.1002/asi.21647.

Kules, Bill, Robert Capra, Matthew Banta ja Tito Sierra. 2009. "What do exploratory searchers look at in a faceted search interface?" Teoksessa *JCDL '09: Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries*, 313–322. New York, NY, USA: ACM. ISBN: 978-1-60558-322-8. doi:10.1145/1555400.1555452.

Kules, Bill, ja Ben Shneiderman. 2008. "Users can change their web search tactics: Design guidelines for categorized overviews." *Inf. Process. Manage.* 44 (2): 463–484. doi:10.1016/j.ipm.2007.07.014.

KulttuuriSampo. 2014. <http://www.kulttuurisampo.fi/about.shtml>. [Online; noudettu 26. elokuuta 2014].

La Barre, Kathryn. 2007. "Faceted Navigation and Browsing Features in New OPACS: Robust Support for Scholarly Information Seeking?" *Knowledge Organization* 34 (2): 78–90.

Liang, Huizhi, Yue Xu, Yuefeng Li, Richi Nayak ja Xiaohui Tao. 2010. "Connecting users and items with weighted tags for personalized item recommendations." Teoksessa *The 21st ACM Hypertext Conference (HT)*, toimittanut Mark H. Chignell ja Elaine Toms, 51–60. ACM. ISBN: 978-1-4503-0041-4. doi:10.1145/1810617.1810628.

Linkity-seminaari. 2014. <http://www.kansalliskirjasto.fi/kirjastoala/tapahtumat/1403000585819.html>. [Online; noudettu 7. lokakuuta 2014].

Marchionini, Gary. 2006. "Exploratory Search: From Finding to Understanding". *Commun. ACM* 49 (4): 41–46. doi:10.1145/1121949.1121979.

Linking Open Data Cloud Diagram. 2014. [Online; noudettu 20. lokakuuta 2014], 30. elokuuta. <http://lod-cloud.net/>.

McGuinness, Deborah L. 2003. "Ontologies Come of Age". Luku 6 teoksessa *Spinning the Semantic Web*, toimittanut Dieter Fensel, James Hendler, Henry Lieberman ja Wolfgang Wahlster, 171–196. MIT Press. [http://www-ksl.stanford.edu/people/dlm/papers/ontologies-come-of-age-mit-press-\(with-citation\).htm](http://www-ksl.stanford.edu/people/dlm/papers/ontologies-come-of-age-mit-press-(with-citation).htm).

- Medynskiy, Yevgeniy Eugene, Mira Dontcheva ja Steven M. Drucker. 2009. "Exploring websites through contextual facets." Teoksessa *CHI*, toimittanut Dan R. Olsen Jr., Richard B. Arthur, Ken Hinckley, Meredith Ringel Morris, Scott E. Hudson ja Saul Greenberg, 2013–2022. ACM. ISBN: 978-1-60558-246-7. doi:10.1145/1518701.1519007.
- Mika, Peter. 2005. "Ontologies Are Us: A Unified Model of Social Networks and Semantics", 522–536.
- . 2007. "Ontologies are us: A unified model of social networks and semantics". *Web Semantics: Science, Services and Agents on the World Wide Web* 5 (1): 5–15.
- Novotny, E. 2004. *I don't think I click: A protocol analysis study of use of a library online catalog in the Internet age*. <http://crl.acrl.org/content/65/6/525.full.pdf>.
- Noy, N.F., ja C.D. Hafner. 1997. "The state of the art in ontology design: A survey and comparative review". *AI magazine* 18 (3): 53.
- Oren, Eyal, Renaud Delbru, Michele Catasta, Richard Cyganiak, Holger Stenzhorn ja Giovanni Tummarello. 2008. "Sindice.com: a document-oriented lookup index for open linked data." *International Journal of Metadata, Semantics and Ontologies (IJMSO)* 3 (1): 37–52. doi:10.1504/IJMSO.2008.021204.
- Oren, Eyal, Renaud Delbru ja Stefan Decker. 2006. "Extending faceted navigation for RDF data". Teoksessa *5th International Semantic Web Conference, Athens, GA, USA, November 5-9, 2006*.
- McGuinness, Deborah L., ja Frank van Harmelen. 2004. *OWL Web Ontology Language Overview*. Tekninen raportti. "[Online; noudettu 26. elokuuta 2014]".
- Peter J. Wild, Chris A. McMahon, Matt D. Giess. 2009. "Describing engineering documents with faceted approaches: Observations and reflections." doi:10.1108/00220410910952410.
- Pratt, Wanda, Marti A. Hearst ja Lawrence M. Fagan. 1999. "A Knowledge-Based Approach to Organizing Retrieved Documents", 80–85.
- Prieto-Diaz, Rubén. 2002. "A Faceted Approach to Building Ontologies". Teoksessa *International Conference on Conceptual Modeling (submitted)*. Tampere, Finland. doi:10.1109/IRI.2003.1251451.

- Rae, Adam, Börkur Sigurbjörnsson ja Roelof van Zwol. 2010. "Improving tag recommendation using social networks". Teoksessa *Adaptivity, Personalization and Fusion of Heterogeneous Information*, 92–99. RIAO '10. Paris, France: Le Centre De Hautes Etudes Internationales d'Informatique Documentaire. <http://portal.acm.org/citation.cfm?id=1937055.1937077>.
- Ramdeen, Sarah, ja Bradley M. Hemminger. 2012. "A tale of two interfaces: How facets affect the library catalog search." *Journal of the Association for Information Science and Technology (JASIST)* 63 (4): 702–715. doi:10.1002/asi.21689.
- Schreiber, Guus, ja Yves Raimond. 2014. *RDF 1.1 Primer*. Tekninen raportti. [Online; noudeutu 26. elokuuta 2014].
- Rey, Guillermo Álvaro, Irene Celino, Panos Alexopoulos, Danica Damljanovic, Mariana Damova, Ning Li ja Vladan Devedzic. 2012. *Semi-automatic generation of quizzes and learning artifacts from Linked Data*. <http://lile2012.linkededucation.org/>.
- Sacco, Giovanni Maria, ja Yannis Tzitzikas. 2009. *Dynamic Taxonomies and Faceted Search Theory, Practice, and Experience*. New York: Springer. ISBN: 978-3-642-02358-3.
- Sah, Melike, ja Vincent Wade. 2012. "A Novel Concept-Based Search for the Web of Data Using UMBEL and a Fuzzy Retrieval Model." Teoksessa *ESWC*, toimittanut Elena Simperl, Philipp Cimiano, Axel Polleres, Óscar Corcho ja Valentina Presutti, 7295:103–118. *Lecture Notes in Computer Science*. Springer. ISBN: 978-3-642-30283-1. doi:10.1007/978-3-642-30284-8_14.
- Schwartz, Candy. 2008. "Thesauri and Facets and Tags, Oh My! A Look at Three Decades in Subject Analysis." *Library Trends* 56 (4): 830–842. doi:10.1353/lib.0.0014.
- Sigurbjörnsson, Börkur, ja Roelof van Zwol. 2008. "Flickr Tag Recommendation based on Collective Knowledge". Teoksessa *17th International World Wide Web Conference, WWW 2008, Beijing, China – April 21-25, 2008, Proceedings*, 327–336. ACM. ISBN: 978-1-60558-085-2. doi:10.1145/1367497.1367542.
- Slavic, Aida. 2008. "Faceted Classification: Management and Use". *Axiomathes* 18 (2): 257–271. doi:10.1007/s10516-007-9030-z.

W3C. 2013. *SPARQL 1.1 Overview*. Tekninen raportti. [Online; noudettu 26. elokuuta 2014].

Spyns, P., R. Meersman ja M. Jarrar. 2002. "Data Modelling versus Ontology Engineering". *SIGMOD Record* 31 (4): 12–17.

Suominen, Osma. 2008. "Käyttäjakeskeinen moninäkömähaku semanttisessa portaalissa". Master's Thesis, Helsingin Yliopisto.

Swamynathan, Gayatri, Christo Wilson, Bryce Boe, Kevin Almeroth ja Ben Y. Zhao. 2008. "Do social networks improve e-commerce?: a study on social marketplaces". Teoksessa *Proceedings of the first workshop on Online social networks*, 1–6. 1st Workshop on Online Social Networks (WOSN) '08. Seattle, WA, USA: ACM. ISBN: 978-1-60558-182-8. doi:10.1145/1397735.1397737.

Taheriyan, Mohsen, Craig A. Knoblock, Pedro Szekely ja Jose Luis Ambite. 2012. "Semi-Automatically Modeling Web APIs to Create Linked APIs". Teoksessa *Proceedings of the First Linked APIs workshop at the Ninth Extended Semantic Web Conference*. <http://lapis2012.linkedservices.org/papers/2.pdf>.

Thai, VinhTuan, Pierre-Yves Rouille ja Siegfried Handschuh. 2012. "Visual Abstraction and Ordering in Faceted Browsing of Text Collections." *ACM Transactions on Intelligent Systems and Technology (TIST)* 3 (2): 21. doi:10.1145/2089094.2089097.

Tonkin, Emma, ja Marieke Guy. 2006. "Folksonomies: Tidying Up Tags?" *D-Lib* volume 12(1). <http://webdoc.sub.gwdg.de/edoc/aw/d-lib/dlib/january06/guy/01guy.html>.

Treude, Christoph, ja Margaret-Anne Storey. 2009. "How tagging helps bridge the gap between social and technical aspects in software development". Teoksessa *ICSE '09: Proceedings of the 2009 IEEE 31st International Conference on Software Engineering*, 12–22. IEEE Computer Society. ISBN: 978-1-4244-3453-4. doi:10.1109/ICSE.2009.5070504.

Uddin, Mohammed Nasir, ja Paul Janecek. 2007. "Performance and Usability Testing of Multidimensional Taxonomy in Web Site Search and Navigation". *Performance Measurement and Metrics* 8 (1): 18–33.

- Vandic, Damir, Jan-Willem van Dam ja Flavius Frasincar. 2012. “Faceted product search powered by the Semantic Web.” *Decision Support Systems* 53 (3): 425–437. doi:10.1016/j.dss.2012.02.010.
- Vickery, Brian. 2008. “Faceted Classification for the Web”. *Axiomathes* 18 (2): 145–160. doi:10.1007/s10516-007-9025-9.
- Wache, H., T. Vögele, U. Visser, H. Stuckenschmidt, G. Schuster, H. Neumann ja S. Hübner. 2001. “Ontology-Based Integration of Information - A Survey of Existing Approaches”, 108–117.
- Waitelonis, Jörg, ja Harald Sack. 2011. “Towards exploratory video search using linked data”. *Multimedia Tools and Applications*: 1–28. ISSN: 1380–7501. doi:10.1007/s11042-011-0733-1.
- Wei, Bifan, Jun Liu, Qinghua Zheng, Wei Zhang, Xiaoyu Fu ja Boqin Feng. 2013. “A Survey of Faceted Search.” *J. Web Eng.* 12 (12): 41–64.
- Wetzker, Robert, Carsten Zimmermann, Christian Bauckhage ja Sahin Albayrak. 2010. “I tag, you tag: translating tags for advanced user models”. Teoksessa *Proceedings of the third ACM international conference on Web search and data mining*, 71–80. Wisdom (WSDM) '10. New York, New York, USA: ACM. ISBN: 978-1-60558-889-6. doi:10.1145/1718487.1718497.
- Wilson, T. 2007. *Strict faceted classification*. [Online; noudettu 20. lokakuuta 2014]. http://www.facetmap.com/pub/strict_faceted_classification.pdf.
- Wu, Xindong, Xingquan Zhu, Gong-Qing Wu ja Wei Ding. 2014. “Data Mining with Big Data”. *IEEE Trans. on Knowl. and Data Eng.* 26 (1): 97–107. doi:10.1109/TKDE.2013.109.
- Xu, Bei, ja Hai Zhuge. 2014. “Automatic faceted navigation.” *Future Generation Comp. Syst.* 32:187–197. doi:10.1016/j.future.2012.12.003.
- Yee, Ka P., Kirsten Swearingen, Kevin Li ja Marti Hearst. 2003. “Faceted metadata for image search and browsing”. Teoksessa *Proceedings of the SIGCHI conference on Human factors in computing systems*.
- Yummly*. 2014. <http://www.yummly.com/about/>. [Online; noudettu 26. elokuuta 2014].

Zhang, Junliang, ja Gary Marchionini. 2005. "Evaluation and evolution of a browse and search interface: relation browser". Teoksessa *dg.o2005: Proceedings of the 2005 national conference on Digital government research*, 179–188. Digital Government Research Center. <http://portal.acm.org/citation.cfm?id=1065279>.

Zubiaga, Arkaitz, Christian Körner ja Markus Strohmaier. 2011. "Tags vs shelves: from social tagging to social classification". Teoksessa *Proceedings of the 22nd ACM conference on Hypertext and hypermedia*, 93–102. HT '11. New York, NY, USA: ACM. ISBN: 978-1-4503-0256-2. doi:10.1145/1995966.1995981.

Zwol, Roelof van, Börkur Sigurbjörnsson, Ramu Adapala, Lluís Garcia Pueyo, Abhinav Katiyar, Kaushal Kurapati, Mridul Muralidharan et al. 2010. "Faceted exploration of image search results." Teoksessa *Proceedings of the 19th international conference on World Wide Web (WWW)*, toimittanut Michael Rappa, Paul Jones, Juliana Freire ja Soumen Chakrabarti, 961–970. ACM. ISBN: 978-1-60558-799-8. doi:10.1145/1772690.1772788.