

Jari Salokangas

**Tagien käyttö ja hyödyntäminen sosiaalisissa
verkkopalveluissa**

Tietotekniikan kandidaatintutkielma

8. maaliskuuta 2013

Jyväskylän yliopisto

Tietotekniikan laitos

Tekijä: Jari Salokangas

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

Ohjaaja: Miika Nurminen

Työn nimi: Tagien käyttö ja hyödyntäminen sosiaalisissa verkkopalveluissa

Title in English: The Use And Invocation Of Tags In Social Network Services

Työ: Kandidaatintutkielma

Suuntautumisvaihtoehto: Ohjelmisto- ja tietoliikennetekniikka

Sivumäärä: 43+0

Tiivistelmä: Tämä tutkimus keskittyi tutkimaan tagien käyttöä muutamassa tagipalveluihin erikoistuneessa sosiaalisessa verkkosivustossa. Tutkimuksen aikana huomattiin, että tageja hyödynnetään hyvin monipuolisesti näissä sivustoissa osana käyttäjälle tarjottavaa palvelukokonaisuutta: tärkeimpinä palveluina tarjottiin palveluun lisättävien objektien tagaamisen lisäksi tagisuosituksia, tagihakuja ja sosiaalista verkkotoimintaa palvelun käyttäjien kesken. Tutkimuksen lisäksi tämä tutkimus pohjustaa tageihin liittyvää teoreettista taustaa. Tämän ajateltiin olevan tärkeää, sillä tageista ei ole paljoa olemassa tutkimuksia, missä tageihin liittyviä teorioita ja taustoja olisi koottuna yhteen kattavasti.

Avainsanat: Tagi, Tagaaminen, Tagihierarkiat, Sosiaaliset verkkopalvelut

Abstract: This study focused on the use and invocation of tags in a couple of social network services that specialize delivering tag services to their users. During the research, it was found that the sites use tags in a very varied ways as part of the service package offered to the user: the most important services offered in addition of tagging objects included tag recommendations, using tags to search information and social networking between the users of the service. In addition of the research this study performed, this study also lays the theoretical background of tags. This was thought to be important, because there isn't much research found that aggregates the theory and context of tags in an extensive way.

Keywords: Tag, Tagging, Tag Hierarchies, Social Network Services

Kuviot

Kuvio 1. Algoritmi hierarkkisten taksonomioiden generoimiseen sosiaalisista tagausjärjestelmistä.	14
Kuvio 2. Kuvankaappaus Delicious-sivuston etusivusta.	19
Kuvio 3. Kuvankaappaus Bibsonomy-sivuston etusivusta.	20
Kuvio 4. Kuvankaappaus Flickr-sivuston etusivusta.	22
Kuvio 5. Esimerkki käyttäjille tarjottavista tagisuosituksista Delicious-sivustossa.	25
Kuvio 6. Esimerkki kohteiden rankkauksesta Bibsonomy-sivustossa.	26
Kuvio 7. Esimerkki tagihausta Delicious-sivustossa.	28
Kuvio 8. Esimerkki tagiklusterihausta Flickr-sivustossa.	29
Kuvio 9. Esimerkki tagipilvestä Flickr-sivustossa.	29

Taulukot

Taulukko 1. Tutkittavien verkkosivustojen perustietoja.	31
Taulukko 2. Tutkittavissa verkkosivustoissa tutkittuja parametreja.	31

Sisältö

1	JOHDANTO	1
2	TAGIPOHJAINEN METATIETO	3
	2.1 Tagit	3
	2.2 Tagien luonti	5
	2.3 Folksonomiat	8
3	TAGIRAKENTEITA	11
	3.1 Tagipilvet	11
	3.2 Tagihierarkiat	13
	3.3 Tagihiutaleet	16
4	SOSIAALISET VERKKOPALVELUT	18
	4.1 Delicious	19
	4.2 BibSonomy	20
	4.3 Flickr	22
5	TAGIEN HYÖDYNTÄMINEN SOSIAALISISSA VERKKOPALVELUISSA	23
	5.1 Vertailtavat arvot ja tutkimusmenetelmät	23
	5.2 Tagisuositukset	23
	5.3 Kohteiden arviointi tagien avulla	26
	5.4 Tagien avulla hakeminen	27
	5.5 Käyttäjän profilointi tagien avulla	30
	5.6 Tutkimuksen tulokset	31
6	YHTEENVETO	33
	LÄHTEET	34

1 Johdanto

Informaatioteknologiassa erilaisista tietolähteistä muodostettavaa metadataa on tutkittu paljon, mutta sitä ei ole tutkittu ilman hyvää syytä. Koska elämme tietoyhteiskunnassa tietokoneiden ja tietoliikenneverkkojen ympäröimänä, on tiedon määrä sekä sen käsittely ja varastoiminen kasvanut huomasti. Kotikoneen käyttäjälle avautuu internetin kautta lähes ääretön määrä tietoa, jota ihmiset ympäri maailmaa ovat sinne tuottaneet. 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 muiden hakijoiden tuloksiin tai sivuston sisältämiin linkkeihin.

Hajautetun metadatan piiriin kuuluvat **tagit** ovat keino käsitellä ja järjestellä tietoja. Useimmat nykypäivän netin käyttäjäympäristöistä hyödyntävät tageja jollakin tavalla. Tagien luonti ja tagien kautta hakeminen auttaa käyttäjiä etsimään ja löytämään haluamaansa tietoa käyttämästä järjestelmästä. Samalla voidaan myös sanoa, että käyttäjien luomien tagien kautta muodostuva verkosto antaa järjestelmän ylläpitäjille tietoa sen käyttäjistä. Tämän päivän tietoyhteiskuntaa voidaankin pitää kaksiteräisenä miekkana, jossa käyttäjän hakiessa tietoa saa käyttäjä tiedon halutusta aiheesta itselleen, samalla antaen tietoa itsestään järjestelmälle tiedon kustannuksella. Tämä asia ei tosin ole pelkästään tageista kiinni, vaan siihen vaikuttaa muitakin asioita.

Käyttäjän tekemää tagien luomista kutsutaan **tagaamiseksi** (engl. *tagging*), joka on kevytkäyttöinen sosiaalinen tietojenkäsittelymekanismi ja jonka konsepti tulee kyseiseltä alalta. Tällaisia sosiaalisia tietojenkäsittelyteknologioita kutsutaan yleisesti nimellä **Web 2.0:ksi**. Nykyiset suursivustot, kuten Facebook, Youtube ja Flickr ovat netin Web 2.0 -ilmiön tuotoksia. Tagaamista käytetään yleisesti näissä käyttöjärjestelmissä, joissa sitä kutsutaan nimellä **sosiaalinen tagaaminen** (engl. *social tagging*). Käyttäjät asettavat tageja haluamilleen kohteilleen tarjotakseen tietoa kohteesta sekä kategorisoidakseen niitä. Tagien ja tagaamisen suosio perustuu siihen, että tagien ei tarvitse olla ennaltamääriteltyjä ja käyttäjä voi itse va-

lita tagin nimeämisperusteet sekä tagien määrän kohdetta kohden. Näiden edellä mainittujen perusteella tageja käytetään luokittelemaan tietoa informatiivisella tavalla sekä käyttäjien hajautetusti muodostaman luokittelun ansiosta ovat vastakohtainen näkökulma formaalille ylhäältä ohjatulle luokittelulle. (Treude ja Storey 2009, s. 12 ja 14)

Erilaiset sosiaaliset verkkopalvelut hyödyntävät tageja ja niistä muodostuvaa metadataa monella eri tavalla. Käyttäjän tagatessa objekteja voidaan hänelle sivuston toimesta antaa tagisuosituksia kyseiselle tagattavalle objektille. Tagiarvioilla käyttäjät voivat arvioida kohteita liittämällä tagin oheen kuvaavan arvion. Tagien avulla voidaan myös hakea tehokkaasti tietoa, kun tagit ja kohteet muodostavat verkkomaisen kokonaisuuden. Tagien avulla haettaessa voidaan sulkea paljon epärelevanttia tietoa, jota käyttäjä ei olisi edes halunnut hakea. Sosiaaliset verkkopalvelut ovat myös siitä mielenkiintoinen tutkimuksen kohde, että ne ovat avanneet käyttäjille uuden tavan löytää ja luoda uutta tietoa. Samalla tapahtuu myös vuorovaikutteista toimintaa käyttäjien kesken.

Tämän tutkimuksen tarkoituksena on tutkia sosiaalisissa verkkopalveluissa käytettäviä tageja käytettävissä olevien tagirakenteiden tai käyttötapojen osalta. Tutkimuksessa selvitetään tagit ja tagien luomiseen liittyviä asioita sekä käydään läpi erilaisia tagirakenteita, joita käytetään sosiaalisissa verkkopalveluissa. Tutkimuksessa tutkitaan muutaman yleisen tagisivuston tagien käyttöä ja verrataan valittuja parametrejä valittujen sivustojen kesken. Tutkittavien verkkopalveluiden määrän vähyden takia tutkimus ei anna kokonaiskuvaa tagien käytöstä. Tutkimuksessa pohditaan myös sitä, miksi sosiaaliset verkkopalvelut ovat ottaneet tagit ja tagaamisen käyttöönsä tutkien tageihin, hierarkisiin tageihin sekä muihin samaan ympäristöön liittyviä tutkimuksia. Tutkimuksen aineistona on hyödynnetty erilaisia tutkimusartikkeleita, jotka liittyvät tageihin ja sosiaalisiin verkkopalveluihin.

Luvussa 2 esitellään tagipohjaiseen metatietoon liittyvää teoreettista taustaa. Luvussa 3 tarkastellaan erilaisia tageista muodostettuja tietorakenteita. Luvussa 4 tarkastellaan tutkimukseen valittuja sosiaalisia verkkopalveluita. Luvussa 5 tarkastellaan valittujen sosiaalisten verkkopalveluiden tagien käyttöä ja tutkimuksen tuloksia. Luvussa 6 luodaan yhteenveto tutkimuksesta.

2 Tagipohjainen metatieto

Metatieto tai **metadata** on tietoa tiedosta. Se on käsitteenä monimuotoinen ja sille on laadittu runsaasti erilaisia luokituksia. Sen tarkka määrittely riippuu määrittelijästä. Esimerkiksi Gilliland-Swetland on määritellyt yleisen käsitelmän metadatalle, joka on hänen mukaan ”*kokonaisuus kaikesta, 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.*” Metadataa on mahdollista hyödyntää tiedonhaussa hyödyntämällä sen rakenteisuutta ja hyperlinkkejä tiedon etsimiseen. (Nurminen 2005, s. 31-32)

2.1 Tagit

Tageilla tarkoitetaan informaatioteknologiassa avainsanaa tai -termiä, joka voidaan liittää johonkin informaatioon ja niitä käytetään luokittelemaan tietoa informatiivisella tavalla. Tagit ovat epästrukturoitua metadataa, jotka voivat koostua lauseista tai yksittäisistä sanoista eikä niiden tarvitse olla ennaltalaadittuja. Käyttäjän valitsemien tagien sanasto tai syy tagin valitsemiseen vaihtelee käyttäjästä toiseen. Tagien suosio perustuukin niiden kevyeen luonteeseen ja monikäyttöisyyteen ja siihen, että 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) (Koivisto 2010, s. 7-8) (Liang et al. 2010, s. 51) (Rae, Sigurbjörnsson ja Zwol 2010, s. 92)

Tageja voidaan jakaa erilaisiin tyypeihin 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. *Honda, Odyssey*). **Kontekstipohjaisella tagilla** voidaan kuvata kohteen luonti- tai tallennushetki sijainnilla tai aikamäärällä (esim. *San Fransisco, Golden Gate Bridge ja 2005-10-19*). **Attribuuttitagit** kuvaavat kohteen sisältöä ja sen piirteitä (esim. *hauska, pelottava*) tai kohteen tekijää/kirjoittajaa (esim. *Jeremy's Blog*). **Omistajatagi** kuvaa kohteen oikeuksien hallitsijaa. **Subjektiiiset tagit** kuvaavat käyttäjän mielipiteitä kohteesta (esim. *hauska, cool*) ja niitä käytetään käyttä-

jälle kohdennettujen kohdesuosituksen arvioinnissa. **Järjestelytageilla** voidaan kuvata henkilökohtaisia tietoja (esim. *myPaper*, *myWork*) tai muistuttaa tietyistä tehtävistä kohteeseen liittyen (esim. *to-read*, *to-Review*). **Tarkoituseräiset tagit** ilmaisevat kohteeseen liittymättömiä toimintoja jotka liittyvät käyttäjän tiedonetsimisprosessiin (esim. *lue tietoa LaTeX:sta*, *käännä 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** on tagien tagittamista jonka tuloksena muodostuu 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 **standarditagi** on tarkoitettu käyttäjiltä käyttäjille ja sillä on monenlaisia eri käyttötarkoituksia. **Konekuvaileva tagi** tai **konetagi** kuvaa kohdetta koneluettavassa muodossa. Näitä tageja voidaan toimittaa niin käyttäjän kuin tagisivuston toimesta. Konetagit kuvataan usein ”tripleteinä”, sisältäen kohteen nimiavaruuden, tilan ja arvon. **Konemääräävä tagi** tai **toimintatagi** poikkeaa normaalista tagista niin, että siihen on lisätty jokin toimintoa kuvaava tagi (esim. *for:User*, *for:U2*). **Käyttäjämääräävä 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 ajan tai paikan toimintoon liittyen (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 tarkoituserä tai mistä objekti on tullut (esim. ”*Michelangelo*” ja ”*keskiaikainen*” Michelangelon maalaamassa maalauksessa). **Taksonomiatagit** ovat tageja, jotka auttavat objektin luokittelemisessa tiettyyn sopivaan luokkaan (esim. ”*eläinkunta*” ja ”*selkäjännteiset*” liitettynä haikaraan). **Adjektiivitagit** ovat tageja, jotka ilmaisevat objektia adjektiivein (esim. ”*punainen*”, ”*hieno*”). **Verbitagit** ovat toimintasanvoja (esim. ”*tutkia*” ”*tehdä*”). **Erisnimitagit** ovat tageja, jotka ilmaisevat kohteen nimen (esim.

”Uusi-Seelanti”, ”Manhattanin silta”). (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 mediaan, 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)

Tutkijoiden Guyn ja Tonkinin toimesta on huomattu, että väärinkirjoittamisen ja huonon koodamisen takia tagijoukkoon päätyy myös turhia tageja. Tällaisia tageja ovat **yhdistesannaryhmät** (esim. ”TimBernersLee”), tagit, jotka eivät seuraa yleisesti sovittuja sopimuksia (esim. numerointiin liittyen), henkilökohtaiset tagit, joilla ei ole suurempaa merkitystä muille (esim. ”mydog”), tietokannassa yksittäisesti esiintyvät tagit sekä tageissa käytetyt symbolit. (Gupta et al. 2010, s. 69)

2.2 Tagien luonti

Tagaajat voidaan jakaa kategorisoijiin ja kuvailijoihin. **Kategorisoijat** lisäävät tageja objekteihin, jotta ne on helpompi löytää tulevaa käyttöä varten. **Kuvailijat** tagaavat objekteja, jotta muut löytäisivät objektit helpommin. Kuvailijan sanavarasto verrattuna keskinkertaisiin kategorisoijiin on laajempi, kun taas kategorisoijan sanavarasto on henkilökohtainen ja tagit ovat subjektiivisia. Kuvailijoilla verrattuna kategorisoijiin on enemmän kerran käytettyjä tageja ja heidän tagimääränsä kasvaa nopeammin kuin kategorisoijien tagimäärä. Kategorisoijat taas pyrkivät mahdollisimman alhaiseen entropiaan tagien suhteen. (Gupta et al. 2010, s. 60) (Körner et al. 2010, s. 3) (Zubiaga, Körner ja Strohmaier 2011, s. 93)

Käyttäjien tagaamiseen liittyy erilaisia motivaatio- ja tavoitetiljoja, joiden takia käyttäjät tagaavat objekteja. Gupta, Li, Yin ja Han kokosivat tekemässään tutkimuksessaan aikaisempien tutkimusten tuloksia ja luokittelivat tagaamisen motivaatiot yksityiskohtaiseksi listaksi. **Mieleepalauttamistavoite** motivoi käyttäjää tagaamaan objekteja tavoitteena helpottaa nii-

den löytämistä tulevaisuudessa tai liittää jokin tulevaisuuden toiminto liittyen objektiin joko itsensä tekemänä tai hakemana tai muiden toimesta tekemänä. **Tuki- ja jakotavoite** motivoi käyttäjää tagaamaan objekteja tavoitteena luoda objekteille lisäarvoa tunnetuille tai tuntemattomille joukoille. **Mielenkiintotavoite** motivoi käyttäjää tagaamaan objektejaan tavoitteena käyttää suosittuja tageja järjestelmässä ajamaan käyttäjiä katsomaan käyttäjän itse luomia objekteja saadakseen niille enemmän näkijöitä. **Pelaamisessa ja kilpailemisessa** käyttäjien tagien muodostamisen motivaatio perustuu sisäisiin tai ulkoisiin sääntöihin, esimerkiksi tavoitteena tagata tageja mitä muut käyttäjät aikovat tagata. **Itsensä näyttämisen halu** motivoi käyttäjää niin, että käyttäjän tekemillä tageilla voidaan luoda järjestelmään oma identiteetti ja jättää oma identiteettijälki muihin objekteihin, tavoitteena henkilökohtaistaa jokin objekti itseensä. **Mielipiteen ilmaiseminen** motivoi käyttäjää niin, että käyttäjä haluaa tuoda ilmi arvoja tietystä objektista jakamalla omia arvojaan käyttäjien kesken liittyen tiettyyn objektiin. Joskus käyttäjä voi toimia näin halutessaan kerätä mainetta yhteisön sisällä. **Tehtävien järjesteleminen** motivoi käyttäjää tagaamaan objektejaan tulevien työtehtävien organisoimiseen. **Sosiaalinen signalointi** motivoi käyttäjää niin, että käyttäjä kommunikoi tagien välityksellä antaen tietoja objektista toisille käyttäjille. **Rahanteko** motivoi käyttäjää tagaamaan objekteja, koska käyttäjä saa objektien tagaamisesta rahaa. **Teknologinen helppous** motivoi joitakin käyttäjiä tagaamaan objekteja, koska käyttäjän käyttämä teknologia tukee objektien tagaamista ja tekee tagaamisen helpoksi käyttäjälle. (Gupta et al. 2010, s. 59-60)

Delicious-sivuston käyttäjien tagaamisen toiminta opetti sivuston ylläpitäjille, että käyttäjien toteuttama tagaaminen ei aina toimi optimististen mielikuvien mukaisesti. Opetuksen pääidea on, että henkilökohtaiset arvot ylittävät sivuston arvot. Tällä tarkoitetaan sitä, että käyttäjien on ensiksi löydettävä arvoa itselleen ennenkuin hän voi avustaa parantamaan sivuston arvoa. Delicious:n tapauksessa seurasi, että käyttäjät löysivät arvoa siitä, että he säilyttivät omia henkilökohtaisia tagejaan eivätkä jakaneet niitä muille, kaiken muun käytön ollessa toisarvoista. Tagien käytöstä tulikin henkilökohtaista toimintaa, jolla käyttäjä merkkasi omia suosikkejaan myöhemmän löytämisen varalle. Sivuston tagit tarjosivat sosiaalisen arvon sijaan henkilökohtaista arvoa käyttäjilleen aina, kun joku muu käyttäjä käytti häntä tai hänen luomaansa tagia tiedon palauttamiseen. (Porter 2006)

Logistisella sekaregressiomallilla (engl. *logistic mixed model regression*) voidaan arvioida

käyttäjän tagien valintaan vaikuttavan kolme eri asiaa. **Imitoimisessa** käyttäjään vaikuttaa aikaisempien käyttäjien verkkopalveluun laittamat tagit, joita käyttäjä imitoi luodessaan tageja kohteilleen. **Organisomisessa** käyttäjä uudelleenkäyttää tageja, joita on luotu verkkopalveluun muiden käyttäjien toimesta. **Suosituksessa** käyttäjään vaikuttaa verkkopalvelun tarjoamat tagisuositukset, joista hän valitsee tageja, kun hän tagaa kohteita. (Rader ja Wash 2008, s. 241)

Vaikka tagaamisen motivaation tutkiminen on vaikeaa ilman suoraa vuorovaikutusta tagaajaan, on kuitenkin olemassa muutamia suhdelukuja, joilla voidaan mitata tagaamista. **Sanastokoko** kuvaa tagien lukumäärää käyttäjän tagisanastossa. Kategorisoijilla tämä on pienempi kuin kuvailijoilla, koska he pyrkivät pitämään heidän sanastonsa pienenä. **Tagiresurssisuhde** kuvaa käyttäjän tagien määrää suhteessa muistiinpantuihin kohteisiin. Kategorisoijien luulisi saavan pienempiä arvoja verrattuna kuvailijoihin, sillä kuvailijoiden sanavarastojen oletetaan olevan rajattoman suuria. **Tagien keskiarvo kohdetta kohden** kuvaa, kuinka monta tagia käyttäjä tagaa kohdetta kohden. Tagaajat, jotka yleensä tagaavat paljon tageja kohdetta kohden saavat suurempia keskiarvoja verrattuna vähemmän tageja kohteisiin laittaviin. **Orposuhde** kuvaa käyttäjän suhdetta luoda orpotageja. Ne ovat tageja, joita käyttäjä tagaa vain pieneen määrään kohteita. Orposuhde kuvaa myös sitä käyttäjän sanaston määrää, jotka esiintyvät orvoissa tageissa. Orposuhteen tulosväli on nolasta yhteen, missä suurempi arvo kuvaa käyttäjiä, joilla on paljon orvoiksi jääneitä tageja. Kuvailijoiden oletetaan lähestyvän enemmän ykköstä ja kategorisoijien enemmän nollaa. Edellä mainitut mittasuhteet kuvaavat vain tagaamisen käyttömalleja ja ovat riippumattomia semantiikoista. (Körner et al. 2010, s. 3-4) (Zubiaga, Körner ja Strohmaier 2011, s. 95-96)

Vaikka tagit ovat hyödyllisiä, niihin kohdistuu myös muutamia ongelmia. **Tagispämmäämisellä** tarkoitetaan tagien luomista kohteisiin vääränlaisella tavalla tavoitteena ajaa toiminnallaan omia intressejään eteenpäin. Tagispämmäämisestä voidaan pitää ilmiötä, kuten yhden henkilön suuri kohteiden merkkäminen samalla tagilla tai tuhansien tagien luominen pienellä aikavälillä, jonka jälkeen tagien luontiin osallistuminen lähes loppuu. Tagispämmääjillä on sellaisia 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 vaikutusalueita ilman suodatuksen tarvetta. Muita tageihin liittyviä ongelmia on yhdenmukaistaminen (engl. *canonicalization*), moniselitteisyys (engl. *ambiguity*) (kts. folksonomiat) ja vähyys (engl. *sparsity*). (Gupta et al. 2010, s. 68-69)

Koska tagien sanavalintojen valitsemisen suhteen ei ole rajoittavia tekijöitä, käyttäjien luomat tagit ovat vapaamuotoisia, sisältävät monitulkintaisia sanoja (yhellä sanalla on useampi merkitys) sekä tagisynonyymejä (monella sanalla on yksi merkitys). Tämän lisäksi käyttäjän luomista tageista kolme viidestä on henkilökohtaisia, vain yhden käyttäjän käyttämiä tageja. Nämä haittapuolet tuovat lisähaasteita, kun käytetään tageja kuvailemaan käyttäjän mieltymyksiä käyttämässään järjestelmässä. (Liang et al. 2010, s. 51)

2.3 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 yhteinäinen 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) (Koskinen 2007, s. 10-11)

Folksonomioita on kahdenlaisia määriteltyjen tagienluomisoikeuksien mukaan. **Kapeat folksonomiat** rajoittavat kohteiden tagaamista rajalliselle määrälle käyttäjiä. Kapean folksonomian esimerkkisivusto 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 esimerkkisivustoja 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 mielenkiinto 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)

Folksonomiat ja sosiaalinen tagaaminen yhdessä auttavat ohittamaan objektien käsittelyyn liittyviä ongelmia. Folksonomioiden etuna on kyky vastata käyttäjien tarpeisiin ja kielellisyyteen tarkkuuden sijaan. Metadatan luominen manuaalisesti on ajallisesti ja taloudellisesti kallista sekä luotu metadatarikasto voi olla täysin erilainen verrattuna esim. 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. Hiearkkiset luokittelut taipuvat muuttumaan yhdeksi järjestelmälliseksi näkemykseksi eikä tällainen luokittelu helposti pysy uusien ja kehittyvien kielikokoelmien perässä. Folksonomiat verrattuna formaaleihin taksonomioihin esiintyvät tasaisessa nimiavaruudessa eikä siten folksonomiassa esiintyvillä termeillä ole epäsuorasti muodostuneita suhteita toisiinsa nähden. Koska käyttäjät tagaavat objekteja, folksonomiat edustavat käyttäjien sanastoa. Tagaaminen on hyvin yksinkertaista ja se auttaa käyttäjiä saamaan suoraa palautetta toisilta käyttäjiltä. Tällainen palautekierre johtaa epäsymmetriseen

kommunikointiin käyttäjien ja metadatan välillä. Folksonomioita voidaan pitää yksinkertaisena iteratiivisena järjestelmänä, joka auttaa luomaan suosituimman tavan organisoida objekteja järjestelmässä. (Gupta et al. 2010, s. 58-59)

Moniselitteisyys on folksonomioissa ongelma, koska eri käyttäjät merkkäävät termejä kohteisiin eri tavalla. Useiden sanojen alkukirjaimista muodostetut, ns. akronyymit voivat johtaa moniselitteisyyteen. Nykyään tagit on määritelty joko yksittäis- tai yhdistesanoiksi, joten monen sanan sisältävä välimerkittömmän tagin (esim. ”*vertigovideostillbbc*”) tieto voi hävitä tagaamisprosessissa. Yksittäissanaisten tagit hävittävät lauseen sanajärjestyksen tiedon. 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)

3 Tagirakenteita

3.1 Tagipilvet

Tagipilvellä tarkoitetaan yleensä sosiaalisissa verkkosovelluksissa olevaa visuaalista rakennetta, joka kuvaa sovelluksen käyttäjien luomien tagien osakokonaisuutta. Tagipilvien tarkoituksena on helpottaa käyttäjää tagien selaamis- ja etsintäprosessissa sekä luoda katselmus sovelluksessa olevista tiedoista tagien avulla. (Gupta et al. 2010, s. 65) (Hearst ja Rosner 2008, s. 1)

Tagipilvelle on yleensä rajattu tietty tila, jolle mahdutetaan parhaiten kuvaavia ja suosituimpia tageja. Tagipilveen valitut tagit perustuvat niiden hyödyllisyyteen, jotka määritellään Hassan-Monteron ja Herrero-Solanan kaksiulotteisessa näyttötavassa kolmen asian avulla. Ensiksi määritellään tagin kapasiteetti esittää kukin kohde verrattuna muihin samaan kohteeseen osoitettujen tagien kanssa. Toiseksi verrataan tagien osoitekohteita verrattuna muihin tageihin ja kolmanneksi verrataan tagien kapasiteettia esittää kohteet, jotka ovat vähemmän esitettyinä muissa kohteisiin liitetyissä tageissa. (Gupta et al. 2010, s. 65)

Tagipilvellä on käyttäjän näkökulman mukaan muutama käyttäjää helpottava ominaisuus verrattuna tavalliseen hakusanaetsintään. Sinclairin ja Cardew-Hallin toteuttamassa tutkimuksessa havaittiin, että käyttäjät etsivät tagipilven kautta tietoa, jos tiedonetsintäprosessi oli enemmän tavanomaisempi. Tutkimuksessa havaittiin myös, että tagipilvi antaa helposti ideoita käyttäjälle siitä, mistä aloittaa tiedonetsintä, antaa käyttäjälle visuaalisen katselmuksen sovelluksen tiedoista, tutustuttaa käyttäjää palveluun sekä etsitty tieto auttaa käyttäjää keskittämään etsintöjensä. Tagipilven selaamisen on myös huomattu tarvitsevan vähemmän kognitiivista taakkaa kuin hakusanojen formuloiminen. (Gupta et al. 2010, s. 65) (Hearst ja Rosner 2008, s. 6-9)

Tagipilvillä on myös omat ongelmansa. Tagipilvet vinouttavat käyttäjän hakemista niin, että hakuprosessi suuntautuu suosituimpien tagien kautta. Tällöin tagipilvet sulkevat pois hyödyllistä informaatiota käyttäjältä. Tagipilvien kautta hakeminen tarvitsee enemmän hakuja kuin suoralla hakusanalla hakukoneen kautta hakeminen. Tagipilvi ei myöskään anna mah-

dollisuutta kapeuttaa hakemisprosessia käyttäjän haluamassa ajankohdassa. Tagipilvien ulkopuolelle on arvioitu jäävän keskimäärin puolet sivuston tietokannan kohteista. Vaikka jotkin tagisysteemit yrittävät vähentää tätä ongelmaa näyttämällä vähemmän suosittuja tageja käyttäjän selatessa artikkeleita, se ei välttämättä ole kovin hyödyllistä, jos käyttäjä etsii tiettyä tietoa. (Gupta et al. 2010, s. 65) (Hearst ja Rosner 2008, s. 6-9)

Tagipilviin liittyviä tutkimuksia on tehty paljon viimeisten vuosien aikana. Bohmannin ja Schrammelin tekemät tutkimukset paljastivat, että aakkosjärjestykseen lajitellut tagipilvet pärjäsivät parhaiten ja olivat käyttäjien mielestä mielekkäimpiä, kun käyttäjä etsi erityisiä tageja. Batemannin tekemässä tutkimuksessa todettiin, että suurempien fonttikokojen havaittiin vaikuttavan kaikista johdonmukaisemmin tagivalintoihin. Fontin painon (fontin kirjaimen paksuus suhteessa fonttikokoon) todettiin olevan toiseksi suurin vaikuttava tekijä tagivalintoja tehdessä. Fontin värillä ei ollut merkitystä tagivalintoja tehtäessä. Näiden tulosten oletetaan vaikuttavan tagipilvien ulkoasun suunnitteluun. (Christie, Lueg ja Baghaei 2010, s. 360)

Hearstin ja Roshenin toteuttamassa tutkimuksessa havaittiin, että tagipilvien yhtenä päähyötynä on niiden kyky näyttää tagien käyttöön liittyviä trendejä. **Trendeillä** tarkoitetaan ajan kanssa muuttuvia toimintamalleja, jotka liittyvät johonkin asiaan. Tagien käyttöön liittyvien trendien seurantaan käytettyjen työkalujen ajatellaan olevan tehokkaita ja ennustavia työkaluja esimerkiksi tuotteiden markkinoinnissa, tuotekehittelyssä ja innovaatioiden löytämisessä. Tutkimuksessa myös havaittiin, että tagipilviä hyödyntämällä pystytään kuvaamaan yksittäisen ihmisen tai ryhmän käyttäytymismalleja. (Hearst ja Rosner 2008, s. 8)

Hearstin ja Roshenin tutkimuksessa tuotiin myös ilmi, että suurin hyöty tagipilvestä saattaa olla käyttäjien toiminnan näyttäminen muille käyttäjille. Niinpä tagipilvien hyödyllisin käytötapa voi olla tagipilvien käyttö sosiaalisena viestijänä. Tällaisia sosiaalisia tagipilviä kutsutaan **henkilökohtaisiksi tagipilviksi**, jotka näyttävät listan tiettyyn käyttäjään liittyviä tageja. Henkilökohtaiset tagipilvet näyttävät käyttäjän itse luomaa tietoa muiden sivulla olevien käyttäjien luomien tietojen sijaan, antavat käyttäjän itse muokata tagipilvessä olevia tageja niin, että näkyvillä on hyödyllistä ja mielenkiintoista tietoa sekä näyttää luettua tietoa postitettujen tagien lisäksi. Hearstin ja Roshnerin mielestä henkilökohtaiset tagipilvet ovat hyviä itsereflektioon sekä ne antavat hyvän kuvan siitä, mitä tietty käyttäjä lukee ja

ajattelee. Tämän kaltainen henkilökohtainen tagipilvi eroaa aikaisemmin toteutetuista henkilötagipilvitutkimuksista sillä, että aikaisemmissa tutkimuksissa henkilökohtaisia tagipilviä ovat taganneet myös muut käyttäjät, jolloin tagipilvi kuvasikin käyttäjien näkökulmaa tietystä käyttäjästä. (Christie, Lueg ja Baghaei 2010, s. 361) (Hearst ja Rosner 2008)

3.2 Tagihierarkiat

Tagihierarkioilla tarkoitetaan tageista luotua rakennepuuta, joka kuvaa rakenteeseen valittujen tagien hierarkkisen taksonomian. Tagihierarkioiden yhtenä etuna on tiedonhaun keskittäminen muodostuneessa hierarkkisessa rakennepuussa niin, että toisiinsa linkitetyt tagit ovat toisiaan sisällöllisesti lähellä. Hierarkkisessa rakenteessa on helpompaa navigoida verrattuna esimerkiksi tagipilven kautta tapahtuvaan tiedonhakuun. (Gupta et al. 2010, s. 65-66) (Heymann ja Garcia-Molina 2006, s. 1)

Heymannin ja Garcia-Molanen rakentama tagihierarkioiden luomiseen käytettävä algoritmi käyttää hyödyksi käyttäjien objekteihin liitetyissä tiedoissa epäsuorasti esiintyviä samankaltaisuuksia ja käyttäjien enemmistön selityksiä objekteista. Tagihierarkioiden luomiseen käytettävä algoritmi olettaa, että tagaamisjärjestelmä koostuu kohteista, käyttäjistä ja tageista. Järjestelmän tieto koostuu selitetyistä kohteista, joista jokaista selitettyä kohdetta kohden on yksi käyttäjä, yksi kohde sekä yksi tai useampi siihen liitetty tagi. Tämän lisäksi kohteen selitys voi sisältää muutakin tietoa, kuten päivämääriä tai käyttäjien huomautuksia. Tagit koostetaan tagivektoreiksi, jonka indeksi vastaa sillä liitettyjen kohteiden määrää. Tagien samankaltaisuuskaavio kuvaa tagien samankaltaisuuksia valitussa tietokohteessa, jossa jokainen tagi edustaa yhtä kärkeä. Kaksi kärkeä on yhdistetty särmällä, jos samankaltaisuuden kuvaamiseen käytettävä perusarvo ylittyy. Samankaltaisuus lasketaan soveltaen kosinien samankaltaisuuskaavaa tagivektorien välille. (Heymann ja Garcia-Molina 2006, s. 1-2)

Algoritmi aloittaa hierarkian rakentamisen yhdestä juurialkiosta, joka edustaa rakenteen huipua. Tämän jälkeen tageja lisätään pienuusjärjestyksessä rakenteeseen tagien samankaltaisuuskaavioon perustuvien arvojen mukaan. Tulevan tagin paikka lasketaan jo rakenteessa olevien tagien kanssa. Lisättävä tagi lisätään joko kaikista samankaltaisimman tagin lapsialkioksi tai juurialkioon, jos lisättävälle tagille ei löydy hyvää sukulaisalkiota. (Heymann ja

Algorithm 1 An extensible greedy algorithm for hierarchical taxonomy generation from social tagging systems using graph centrality in a similarity graph of tags.

Require: $L_{generality}$ is a list of tags t_i, \dots, t_j in descending order of their centrality in the similarity graph.

Require: Several functions are assumed: $s(t_i, t_j)$ computes the similarity (using cosine similarity, for example) between t_i and t_j . $getVertices(G)$ returns all vertices in the given graph, G .

Require: $taxThreshold$ is a parameter for the threshold at which a tag becomes a child of a related parent rather than of the root.

```

1:  $G_{taxonomy} \leftarrow \langle \emptyset, root \rangle$ 
2: for  $i = 1 \dots |L_{generality}|$  do
3:    $t_i \leftarrow L_{generality}[i]$ 
4:    $maxCandidateVal \leftarrow 0$ 
5:   for all  $t_j \in getVertices(G_{taxonomy})$  do
6:     if  $s(t_i, t_j) > maxCandidateVal$  then
7:        $maxCandidateVal \leftarrow s(t_i, t_j)$ 
8:        $maxCandidate \leftarrow t_j$ 
9:     end if
10:  end for
11:  if  $maxCandidateVal > taxThreshold$  then
12:     $G_{taxonomy} \leftarrow G_{taxonomy} \cup \langle maxCandidate, t_i \rangle$ 
13:  else
14:     $G_{taxonomy} \leftarrow G_{taxonomy} \cup \langle root, t_i \rangle$ 
15:  end if
16: end for

```

Kuvio 1: Algoritmi hierarkkisten taksonomioiden generoimiseen sosiaalisista tagausjärjestelmistä.

Garcia-Molina 2006, s. 3)

Tagihierarkioiden luomiseen käytettävän tiedon tulee sisältää luonnollisia hierarkiasuhteita. Luonnolliset hierarkiat kuitenkin sisältyvät tietoihin, sillä käyttäjät luonnostaan tagaavat kohteita omien henkilökohtaisten taksonomioiden kautta sekä kuvailevat kohteitaan eri kuvailemistasoilla. Sivustojen tiedon hierarkisoiminen on vaikeampaa, jos sivustolla oleva tieto omaa matalan **tiheysluvun** (luku, joka kuvaa käyttäjien tagaustaaajuutta) sekä matalan **päällekkäisyysluvun** (luku, joka kuvaa käyttäjien samojen kohteiden tagaustaaajuutta). Vaikka algoritmeja voidaan luoda soveltumaan alhaisille tiheys- ja päällekkäisyysluville, näiden lukujen alhaisten arvojen on yleisesti huomattu vaikeuttavan tagihierarkioiden luomisia rakenteettomasta tiedosta. (Heymann ja Garcia-Molina 2006, s. 3-4)

Hierarkkisia suhteita voidaan rakentaa myös tagien välille niin, että luodaan tagiryhmiä, joissa jokaisessa ryhmässä on tageja, jotka edustavat samaa konseptia. Zhou ja Kingin tutkimuksen kolmivaiheisessa lähestymistavassa ensiksi muodostettiin tagiryhmiä, jotka sisälsivät semanttisesti samaistuvia tageja. Tämän jälkeen muodostetaan tagiklustereita, jotka jakavat saman konseptin ryhmässä olevien tagien kanssa. Viimeiseksi luodaan tagihierarkioita ryhmien suhteen niin, että tagiryhmien väliset hierarkiasuhteet perustuvat konseptien samantaisuuksiin. Kahden tagin semanttinen samaistuvuus tutkitaan käyttämällä kattavuussuhdetta, joka tutkii, kuinka paljon ensimmäisenä verrattavan tagin resursseista voidaan ilmaista toisen vertailtavan tagin avulla. (Zhou ja King 2009, s. 25-31)

Tageista voidaan muodostaa lauseyhteydellisiä folksonomioita ja konseptihierarkioita käyttäen apuna FCA-metodia (engl. *Formal Concept Analysis*). **Lauseyhteydellinen folksonomia** viittaa jaettuun tagivarastoon, joka on kerätty keskitetyn sosiaalisen verkon sijasta useista samankaltaisista ryhmistä yhdeksi kokoelmaksi. Tagien keräyskohteena voi toimia esimerkiksi **blogosfääri**, jolla tarkoitetaan kaikkien olemassa olevien blogien kokonaisuutta. Lauseyhteydellinen folksonomia tarjoaa lauseyhteydellisen tagausympäristön ryhmityille yhteisöille. FCA-metodia on pääasiassa käytetty epäsuorasti annetun tiedon tutkimiseen ja analysoimiseen. FCA alkaa formaalilla kontekstilla, joka käsittää objekteja, attribuutteja ja relaation joka kuvaa, mitkä attribuutit liittyvät mihinkin objekteihin. Kontekstissa olevilla objekteilla on keskenään yhteisiä yleisattribuutteja. Keskinäinen notaatio FCA:ssa on konsepti, jossa on ryhmä objekteja ja attribuutteja, jotka kuvaavat toisiaan parhaiten.

FCA:ta voidaan käyttää myös tiedossa, jossa attribuutit sisältävät arvoja. (Kim, Hwang ja Kim 2007, s. 1340-1342)

Folksonomioita voidaan rakentaa käyttäjien personomioiden kautta, jolloin personomioista muodostetaan yhtenäinen hierarkkinen rakenne. Personomiat sinällään sisältävät matalan tason hierarkioita, sillä monet sosiaaliset verkkopalvelut antavat mahdollisuuden kerätä tai ryhmittää tageja ja objekteja yhteen, jolloin matalan tason hierarkioita pystyy muodostumaan. Rakenteisia personomioita ilmaistaan samankaltaisuuksien ja sidosrajotteiden kautta. Samankaltaisuuksien tutkimiseen käytetään hyväksi *Affinity Propagation* -algoritmia, joka tutkii rakenteisesta tiedosta esimerkkipisteitä, jotka pystyvät esittämään tietopaketin kaikki pisteet. Esimerkkipisteissä on tietopisteitä, jotka yhdessä muodostavat yhden rypäsrakenteen. Syntyneet klusterit voidaan liittää yhteen samalla algoritmilla, mutta jotta vältetään silmukoilta ja oikoreiteiltä, algoritmia laajennetaan sisältämään rajoitteita, jotka varmistavat, että syntynyt folksonomia ei sisällä vältettäviä asioita sekä muodostaa hierarkian. (Plangprasopchok, Lerman ja Getoor 2010, s. 555-559)

3.3 Tagihiutaleet

Tagihierarkioita voidaan rakentaa myös tagipilvissä olevista tageista. Tällaisia tagipilvistä koostettuja tagihierarkioita kutsutaan **tagihiutaleiksi**. Koska tagihiutaleet muodostetaan tagipilvistä, tagihiutaleet näyttävät sivuston avainsanoja sekä missä kontekstissa niitä on käytetty. Tagihiutale on myös hyödyllinen näyttämään aihepiirien kehityksiä sosiaalisessa verkossa sen visuaalisen ilmiannon avulla, joka näyttää tagien väliset aikajanat. (Candan, Caro ja Sapino 2008, s. 75-78)

TMine-”analyysimetodilla” voidaan sosiaalisissa verkkopalveluissa olevista tagipilvistä koostaa rakenteisia kokonaisuuksia, jotka kapseloivat tagien väliset riippuvuussuhteet. TMine:ssä käsitteet ovat enemmän dominoivampia, jos ne sijaitsevat lähempänä juurta ja samankaltaiset käsitteet ovat rakentuneet saman alapuolen klusteriin. TMine käyttää käsitteiden samankaltaisuuden ja dominoituvuuden kartoitukseen LSA-metodia (engl. *Latent Semantic Analysis*), jonka jälkeen analysoidaan tilassa olevien tagien väliset asiayhteydelliset suhteet käyttämällä hyväksi totuusarvotulkintaa tagiavaruudessa. Tagien arvioinnissa käytettävä LSA-

pohjainen tagiavaruuden koostaminen pohjautuu SVD-matriisihajotelmatekniikkaan, jonka syötearvoksi laitetaan $m \times n$ -muotoinen dokumentti-terminen esiintymämatriisi, jossa m ilmaisee dokumentteja ja n ilmaisee niihin liitettyjä termejä. Tagiavaruudessa olevien tagien riippuvuustiedot ja sijainnit arvioidaan, jonka jälkeen TMine tiivistää tagipilven hierarkkiseksi hiutaleeksi. Koska hierarkioiden muodostamiseen ei käytetä kielenprosessointitekniikoita, on se kielineutraali ja siksi hyvin tehokas. (Candan, Caro ja Sapino 2008, s. 75-78)

Tagipilvien muodostamiseen käytetään myös tagFlakea, joka käyttää hyväkseen TMinea. TMine tekee tarvittavat tagihierarkiat oman algoritminsa avulla ja tarjoaa tagFlakelle mekanismin, joka mahdollistaa tagiavaruudessa navigoimisen ja dokumenttien luokittelun luotujen tagien hierarkiasuhteiden pohjalta. TagFlake analysoi tagipilviä ja muodostaa dokumenteista kerätyistä tageista hierarkioita ilman, että sen tarvitsee käyttää korkealottuvuuksellisia kuvainnointimekaniikkoja tai että sen tarvitsisi rajoittaa ulottuvuuksien määrää kahteen. TagFlake on hyödyllinen työkalu dokumenttien organisoimiseen ja se antaa käyttäjälle mahdollisuuden navigoida moniulotteisessa tagiavaruudessa. Kuten TMine, on tagFlake myös kielineutraali. (Caro, Candan ja Sapino 2008, s. 1069-1070)

TMinen ja tagFlaken yhteisessä systeemissä TMine kerää merkittävät tagit annetuista dokumenteista ja kokoaa ne semanttiseksi hierarkiaksi. Tämän jälkeen tagFlake käyttää hierarkiaa asiapohjaisena navigointivälineenä tagi- ja dokumenttiavaruudessa. Tekstialkiot kiinnitetään sitä paremmin kuvaaviin tageihin ja artikkeleista kerättyjen hierarkioiden ajanjaksoja arvioidaan ja vertaillaan, jotta saadaan kuvattua aihepiirien kehityksiä ajassa. Tämän avulla tagFlake pystyy tarjoamaan aiheen kehittymisen seurantamoduulin, joka pystyy kuvaamaan dokumentissa olevien aiheiden kehittymistä dokumenttien tagihierarkioissa. TagFlaken käyttöliittymä pystyy näyttämään käyttäjälle yksittäisen tagin suosion kehityksen ja käytetyt kontekstit annetulla aikavälillä. (Caro, Candan ja Sapino 2008, s. 1071-1072)

4 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 tarjoavat usein tagipalveluita käyttäjilleen, jolloin käyttäjät voivat tagata ja arvioida verkoston käyttäjien tuottamaa sisältöä sekä jakaa kyseisiä tajeja 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)

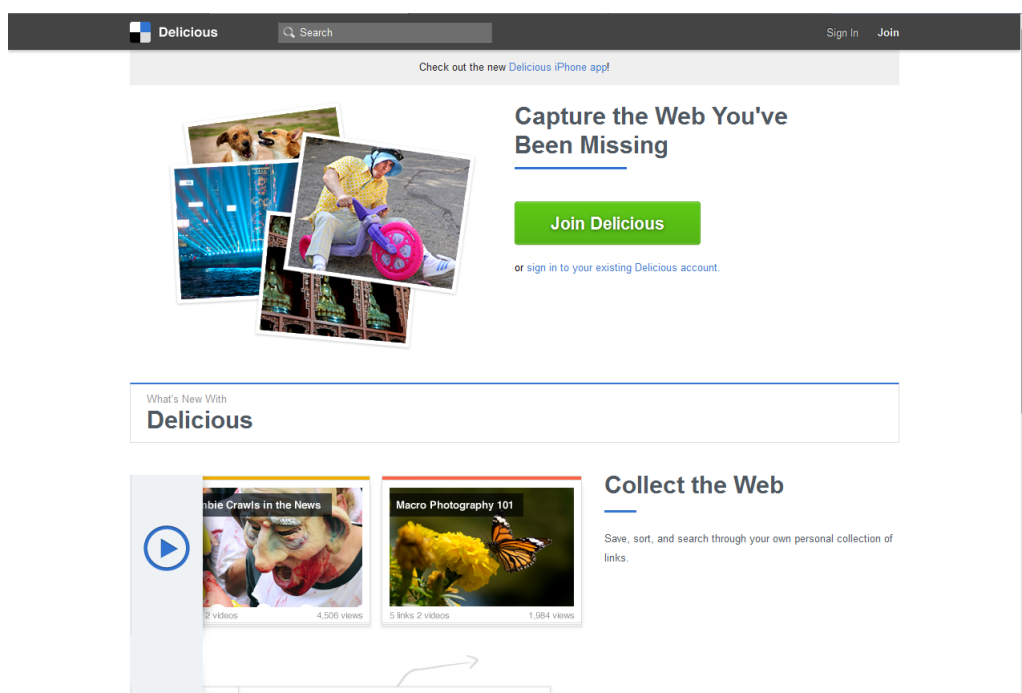
Viime vuosina tapahtunut sosiaalisten verkkopalveluiden läpilyönti Internetin käyttäjien kesken (esim. *MySpace*, *Facebook*, *YouTube*) on muokannut käyttäjien verkossa tapahtuvaa toimintaa. Käyttäjien verkon ajankäytössä on lisääntynyt erityisesti vuorovaikutteinen ja yhteisöllinen toiminta. Käyttäjät 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 ystävyyssuhteita 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)

Monet sosiaaliset verkkosivustot tarjoavat sisällön jakamista niin, että sisältö itsessään ei sisällä kuvausta itsestään. Tällaisissa järjestelmissä monet käyttäjät ovat halukkaita osallistumaan hankalien tiedostojen tagaamiseen. Käyttäjät ovat myös halukkaita tagaamaan olemassaolevaa sisältöä sekä ilmaisemaan mielipiteensä antamalla arvioita. Vaikka suurin osa käyttäjistä hyödyntää tajeja henkilökohtaisiin tarkoituksiin, on kuitenkin osoitettu, että sosiaalisen tagauksen tuloksia voidaan käyttää sisällön palauttamiseen koko verkoston toimesta.

(Clements, Vries ja Reinders 2010, s. 2)

4.1 Delicious

Delicious oli ensimmäinen sosiaalinen kirjanmerkkiverkkosivusto, joka keräsi paljon suosiota netin käyttäjien keskuudessa. Sen käyttäjät voivat tallentaa ja tagata heidän ylläpitämäänsä omia kirjanmerkkikokoelmiaan netissä, jolloin ne ovat saatavilla mistä tahansa nettiin liitetystä tietokoneesta. Se perustettiin vuonna 2003 ja on sittemmin kerännyt itselleen valtavan kokoelman käyttäjien kirjanmerkkejä, tehden siitä suurimman ihmisten tekemän web-linkkidatakokoelman. Sivuston sisältämät kirjanmerkki- ja tagihistoriat ovat julkisia ja siten kaikkien saatavilla ja analysoitavissa. (Wetzker et al. 2010, s. 73) (Benz et al. 2010, s. 872) (Rader ja Wash 2008, s. 239)



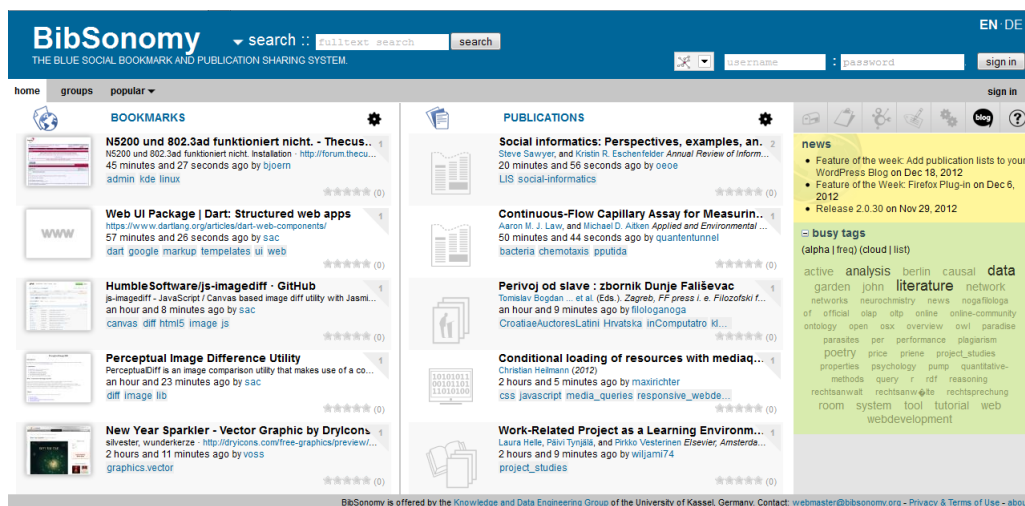
Kuvio 2: Kuvankaappaus Delicious-sivuston etusivusta.

Jokainen luotu uusi kirjanmerkki sivustossa sisältää siihen liitettyä assosioivaa metadataa, kuten kirjanmerkkiä tallentavan käyttäjän nimen, käyttäjän valitsemat tagit ja kirjanmerkin luomispäivämäärän. Käyttäjien selatessa sivuston kirjanmerkkejä näkyy hänelle osajoukkoja kirjanmerkeistä, joilla on rajattu metadata. Esimerkiksi painamalla tagikirjastoa, käyttäjälle

näytetään kaikki käyttäjien toimesta kirjanmerkatut web-sivut tai klikatessa käyttäjänimeä näytetään käyttäjälle klikatun käyttäjän kirjanmerkaamat web-sivut. Web-sivun metadatan voi laittaa myös näkymään ja näyttämään jokaisen käyttäjän, joka on kirjanmerkannut sivuston ja kaikki siihen liittyvät tagit. Kun käyttäjä luo uuden kirjanmerkin, käyttöliittymässä näytetään hänelle tagisuosituksia, käyttäjän tageja ja suosittuja tageja kyseiseen sivuun liityen. (Rader ja Wash 2008, s. 240)

4.2 BibSonomy

BibSonomy on sosiaalinen kirjanmerkkauverkkopalvelu, missä käyttäjät keräävät ja merkaavat URL-osoitteita ja julkaisuja (Wetzker et al. 2010, s. 73). Se on yksi sosiaalisista verkostovustoista, joka on saavuttanut suuren suosion Internet-käyttäjien keskuudessa. Suosion syynä on pidetty sitä, että osallistumiseen ei vaadita mitään erikoistaitoja sekä sivuston tarjoamat työkalut kirjanmerkkaukseen ovat helppokäyttöisiä ja palkitsevia käyttää. Bibsonomyn sivuston tarkoituksena oli organisoida BibTeX-pohjaisia lähdeviitetietoja. BibTeX on LaTeX-ladontajärjestelmään kuuluva työkalu kirjallisuusviitteiden ylläpitoon, mitä monet tutkijat käyttävät tutkimusten kirjoittamisessa. (Benz et al. 2009, s. 323-324)



Kuvio 3: 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 myös tagipilviä ilmaisemaan sivuston sisältöä sivustossa käytettyjen tagien avulla. (Benz et al. 2009, s. 323-324)

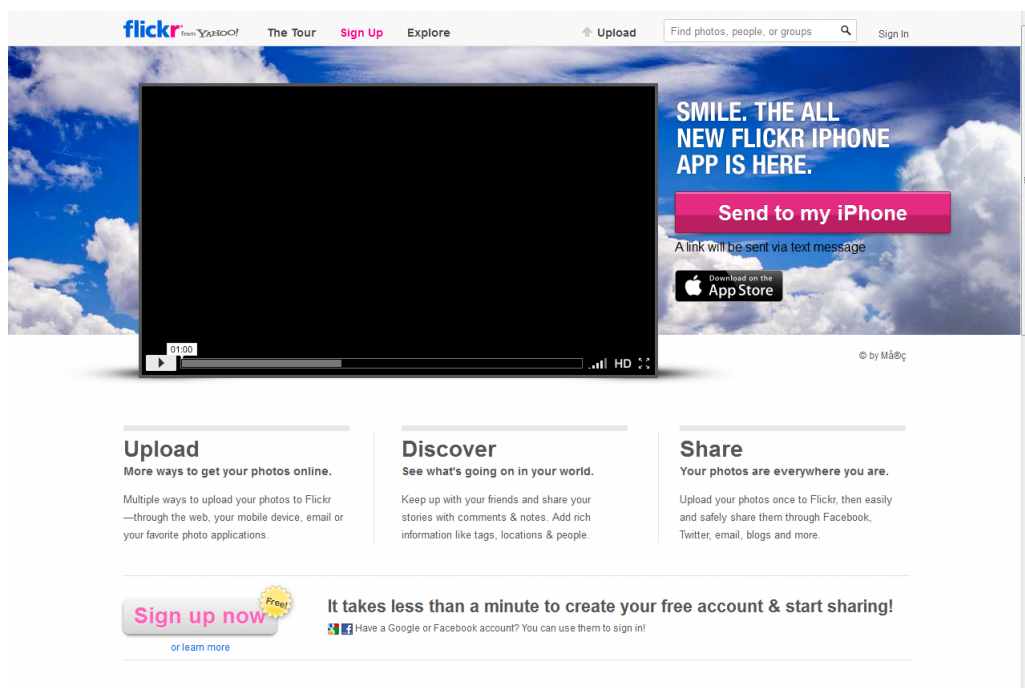
Sosiaalisissa kirjanmerkkijärjestelmissä folksonomiarakenne käännetään suoraan hyperlinkeiksi. Jokainen resurssin nimi, tagi tai käyttäjänimi on hyperlinkki, joka osoittaa sitä vastaavaan kohteeseen. Käyttäjälle linkitetty rakenne mahdollistaa monimuotoisen selaamisen halutusta aihealueesta. Käyttäjä voi esimerkiksi aloittaa hakemisen tagista ja päätyä selaaamaan jotakin mielenkiintoista käyttäjää, joka on kirjanmerkannut hakijan mielenkiintoa herättäviä resursseja. (Benz et al. 2010, s. 855)

Käyttäjän tallentaessa julkaisua resurssiin hänen täytyy asettaa julkaisua vastaavaa metadataa, kuten esimerkiksi otsikko ja lista tageja. Bibsonomy sisältää erilaisia tiedonkerääjämoduuleja (engl. ”scrapers”) jotka etsivät ja keräävät automaattisesti metatietoja käyttäjän vierailemilta web-sivuilta tai digitaalisista kirjastoista (kuten *SpringerLink* ja *IEEE Xplore*) kootakseen yksiselitteisiä lähdemerkintöjä julkaisuista. Bibsonomy myös tarjoaa tagisuosituksia helpottaakseen käyttäjän tagaamista. (Benz et al. 2010, s. 851 ja 867)

Käyttäjän merkatessa resursseja, hän samalla osallistuu Bibsonomy-sivuston tietokannan tukemiseen, sillä hänen merkkaamansa muutokset ovat näkyvissä muille. Hakiessa ja selatessa Bibsonomy ottaa huomioon muiden käyttäjien merkkaukset, kun käyttäjä hakee tai selaa palvelun resursseja. Bibsonomy hyödyntää olemassa oleviin datarakenteisiin FolkRank-algoritmia. Kyseisen algoritmin idea on samantyyppinen kuin PageRank-algoritmin: web-sivu on tärkeä, jos siihen on linkittyneenä monia muita sivuja ja jos kyseiset sivut ovat itsessään merkittäviä. FolkRank:ssa edellä mainittu avainidea on valjastettu tageihin: resurssi johon on tagattu merkittäviä tageja merkittävilta käyttäjiltä tekee resurssista itsessään merkittävän. Sama pätee symmetrisesti niin käyttäjiin kuin tageihin. Rankkaussysteemi harkitsee myös resursseja jotka eivät sisällä annettuja tageja. (Benz et al. 2010, s. 851 ja 859)

4.3 Flickr

Flickr on sosiaalinen verkkopalvelu, jossa käyttäjät voivat vuorovaikuttaa sivuston sisältämään mediaan ja käyttäjiin monella eri tavalla. Käyttäjät voivat jakaa sinne laittamiaan valokuvia ja tagata niitä. Käyttäjät voivat myös muodostaa sosiaalisia suhteita vuorovaikutamalla suorasti tai epäsuorasti palvelun käyttäjäkuntaan. Sivuston käyttäjät voivat ylläpitää suhteitaan toisiin käyttäjiin, jotka voidaan myöhemmin tunnistaa käyttäjän toimesta olemaan joko kaveri, perheenjäsen tai joku muu kontakti. Käyttäjät voivat myös allekirjoittaa itsensä ryhmiin, jonka jäsenyys ilmaisee käyttäjän olevan kiinnostunut kyseisen ryhmän aihealueesta ja sen sisällä olevista käyttäjistä. Käyttäjä voi myös olla kuulumatta mihinkään ryhmään sekä omata kontakteja, joilla on omat kuvakokoelmansa ja ryhmät, joihin he kuuluvat. (Rae, Sigurbjörnsson ja Zwol 2010, s. 92 ja 94) (Rader ja Wash 2008, s. 239)



Kuvio 4: Kuvankaappaus Flickr-sivuston etusivusta.

Käyttäjä voi tagata omia objektejaan käyttämällä sivuston tarjoamia tagisuosituksia. Kun käyttäjä haluaa lisätä uuden tagin haluamaansa kuvaan, sivusto ehdottaa hänelle tageja, jotka perustuvat tageihin, joita joko hän tai muut häneen liitetyt kontaktit tai hänen kuulumissaan ryhmissään olevat käyttäjät ovat aikaisemmin käyttäneet. Suositellut tagit päivittyvät dynaamisesti jokaisen tagilisäyksen jälkeen. (Rae, Sigurbjörnsson ja Zwol 2010, s. 94)

5 Tagien hyödyntäminen sosiaalisissa verkkopalveluissa

5.1 Vertailtavat arvot ja tutkimusmenetelmät

Edellisessä kappaleessa mainittujen sosiaalisten verkkopalveluiden tutkittavia ja tarkasteltavia parametreja ovat tagisuositusten hyödyntäminen, kohteiden arviointi, arviointi tagien avulla, tagien hyödyntäminen tiedonhaussa sekä miten ja kuinka käyttäjän käyttämät ja muodostamat tagit näkyvät käyttäjän identiteetin mallintamisessa.

Tutkimusta toteutettiin päivinä 20.12.2012 ja 7.1.2013. Tagisuositusten hyödyntämistä tutkitaan lisäämällä sivustoon muutama objekti. Kohteiden arviointia, tagien hyödyntämistä tiedonhaussa ja tagien suhdetta käyttäjän identiteettiin tutkitaan tarkastelemalla sivuston ominaisuuksia kyseisistä asioista. Tutkimuksessa keskitytään lähinnä yksittäisen käyttäjän kokemuksiin sivustoista, eikä tutkimusta varten luoda sosiaalisia ryhmiä, jotka yhdessä keräävät dataa sekä käyttävät ja tutkivat sivustoja, eikä myöskään luoda isoa testimateriaalia, jolla voisi esimerkiksi tagisuosituksia testata paremmin.

5.2 Tagisuositukset

Sosiaaliset verkkopalvelut, jotka hyödyntävät tageja ja tagaamista, voivat tarjota **tagisuosituksia** käyttäjälle hänen tagatessaan verkkopalvelun objekteja. Tagisuositusjärjestelmä voi tällöin suositella käyttäjälle erilaisia tageja, joista käyttäjä voi valita joitain tai olla valitsematta. Tagisuositukset tarjoavat käyttäjille parempaa tagauspalvelua ja ylläpitävät parhaiden ja soveltuvien tagien suosiota käyttäjien keskuudessa. Tageja voidaan suositella joko niiden laatuun, yhteisesiintymän, yhtenäisen tiedon (esim. sanojen, objektien ja tagien kesken) tai objektin ominaisuuksien perusteella. (Gupta et al. 2010, s. 67)

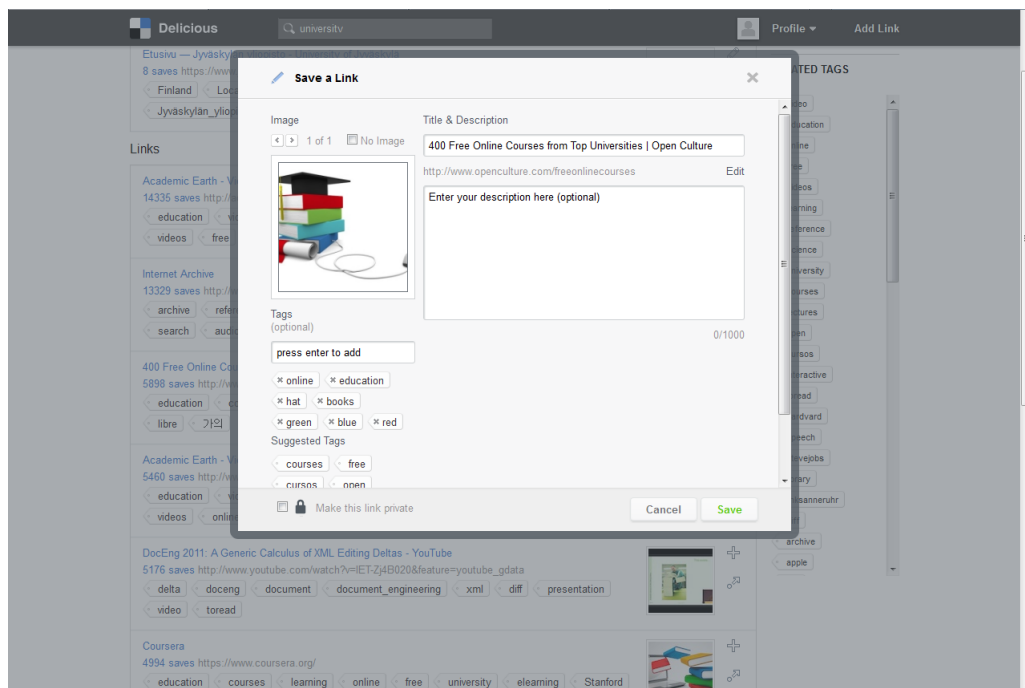
Tagisuositusten tarjoamiseen on esitetty erilaisia tapoja monissa eri tutkimus-artikkeleissa. **Sisältöpohjaista suosittelua** tarjotaan yleensä uusille käyttäjille ja silloin, kun aktiivinen käyttäjä tai resurssi on uusi (Zhang, Zhang ja Tang 2009, s. 10). **Graafipohjaisessa suosituksessa** hyödynnetään olemassa olevan käyttäjän tagaushistoriaa ja resurssiin liitettyjä tageja kumpienkin ollessa jo olemassa olevia itse verkkopalvelussa (Zhang, Zhang ja Tang

2009, s. 10). **Tagipohjaiset suositusjärjestelmät** sisältävät ne järjestelmät, jotka käyttävät tageista kerättyä informaatiota suositusjärjestelmien algoritmien parantamiseen (Barragáns-Martínez et al. 2010, s. 24). **Hybridisuositusjärjestelmät** perustuvat useista eri lähteistä kerättyyn dataan, esimerkiksi kyseinen järjestelmä voi perustaa suosituksensa tageista, folksonomiasta ja arvioista (Barragáns-Martínez et al. 2010, s. 24-25). **Populaarisuuteen perustuva suositusjärjestelmä** perustaa suosituksensa järjestelmän käytetyimmistä tageista ja on siten luonteltaan yksinkertaisin suositusjärjestelmä (Gemmell et al. 2010, s. 831). Suositusjärjestelmät voivat myös hyödyntää **tensoriteknikoita** suosituksissaan (Gemmell et al. 2010, s. 830) (Lu et al. 2011, s. 562-563). **Personoidussa suositusjärjestelmässä** käyttäjän profiili ja siihen liitetyt tagiklusterit otetaan huomioon suosituksessa (Shepitsen et al. 2008, s. 261-262). **LDA-pohjainen suositusjärjestelmä**, jossa dokumentin sisältämät sanat esitetään sanojen sijaan tageina, mallintaa dokumentit ja käyttäjät epävakaiden aiheiden kanssa, jotka mallinnetaan tagijakaumiksi (Lu et al. 2011, s. 563). **Käyttäjöpohjaisissa suositusjärjestelmissä** käyttäjä voi itse hallita tarjottuja suosituksia (Vig et al. 2010, s. 325).

Käyttäjälle tarjottuja tagisuosituksia hyödynnetään kaikissa kolmessa tutkimukseen valituissa sivustoissa. Delicious hyödyntää tagisuosituksia, kun käyttäjä tallentaa objektia (tämän sivuston tapauksessa objekti on kirjanmerkki, joka sisältää sivuston linkin ja siihen liitettyä metatietoa, kuten tageja) sivustoon. Tällöin käyttäjälle tarjotaan tagisuosituksia objektiin liittyen. Objektista tarjottuja tagisuosituksia oletettaneen perustuvan itse lisättävään sivustoon tai muiden mahdollisten käyttäjien lisäämiin tageihin samasta sivustosta.

Bibsonomy-sivusto hyödyntää myös tagisuosituksia. Kun käyttäjä lisää uutta kirjanmerkkiobjektia, hänelle tarjotaan tagisuosituksia, jotka perustuvat käyttäjän aikaisemmin lisäämiin tageihin sekä informaatiosta, jota kirjanmerkkiobjektia luodessaan tarvitsee objektiin lisätä. Jos käyttäjä lisää sivustoa, joka on jo aikaisemmin lisätty sivustoon, suosittelee sivusto tageja sivustolle aikaisemmin lisättyjen objektien tageista. Bibsonomyssa on myös mahdollisuus luoda ”käyttäjäverkkoja” itsensä ympärille (engl. *spheres*), joihin voi lisätä sivuston muita käyttäjiä, ryhmittää heitä ja tarkastella heidän lisäämiään kirjanmerkkejä tai julkaisuja. Tagisuosituksia ei kuitenkaan havaittu tarjottavan muiden käyttäjien lisäämistä tageista käyttäjäverkon kautta.

Flickr hyödyntää myös tagisuosituksia, mutta suositukset eroavat muista sivustoista sivus-



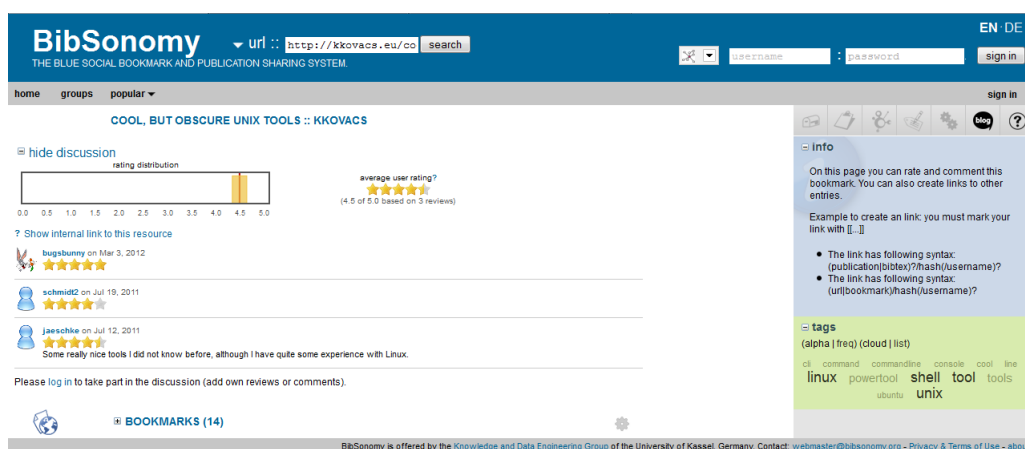
Kuvio 5: Esimerkki käyttäjille tarjottavista tagisuosituksista Delicious-sivustossa.

toon lisättävien objektien osalta (tässä tapauksessa kuvia, jotka sisältävät erilaista metatietoa). Kuvaa lisättäessä sivustoon käyttäjälle tarjotaan tagisuosituksia hänen omasta tagivastostaan, kun käyttäjä ei vielä ole luonut kontakteja muiden käyttäjien kanssa. Muutoin sivusto voi tai voisi tarjota tagisuosituksia laajemmalla tavalla käyttäjän kontaktien tai ryhmien kautta (kuten Sigurbjörnsson ja van Zwol tutkimuksissaan osoittavat (Sigurbjörnsson ja Zwol 2008) (Rae, Sigurbjörnsson ja Zwol 2010)), mutta tämän havaitsemiseen tässä tutkimuksessa olisi vaadittu laajempi katselmus sivuston tagisuosituksen toiminnasta. Tämän lisäksi on huomioitava, että käyttäjäpohjainen tagien suosittelu on ymmärrettävää, sillä tagisuositusten antaminen kuva-analyysin avulla voisi olla hankalaa, sillä kuvan analysoiminen ja kuvassa olevan tiedon ymmärtäminen koneellisesti on haastavaa ja voisi antaa helposti epäolennaisia suosituksia.

5.3 Kohteiden arviointi tagien avulla

Kohteiden arviointi tai **rankkaus** on keino, jolla voidaan arvioida kohteita ja esittää arvioinnit käyttäjille verkkopalvelun palvelun parantamiseksi. Esimerkiksi sosiaalisissa musiikkikatalogipalveluissa käyttäjä voi rankata musiikkikappaleita ja -albumeita eri asteikoilla (*RateYourMusic.com* 2013). Arviointi voidaan myös toteuttaa käyttämällä arvoasteikon sijaan tageja arvon ilmaisemiseen. Kyseisessä menetelmässä käyttäjän valittuun kohteeseen liittämä tagi sisältää metadatan arvon tietyllä asteikolla, esimerkiksi asteikolla 1-10 tai hyvä/huono (esim. tagi[toiminta]-arvio[hyvä]). (Vig et al. 2010, s. 324)

Kohteiden arviointia hyödynnettiin vain Bibsonomyssa. Bibsonomyssa käyttäjät voivat arvioida objekteja numeroasteikoilla 1-5 ja arvioiduista objekteista näytetään sille annettu keskiarvo. Tagiarviointia ei hyödynnetty yhdessäkään tutkitussa sivustossa, mutta Delicious-sivuston objekteissa esiintyi myös subjektiivisia tageja tai niitä pystyi lisäämään omiin objekteihinsa, kuten esimerkiksi sanoja "*great*", "*best*" ja "*magnificent*". Mahdollisuus kuvailevien tagien lisäämiseen pätee tosin jokaiseen tutkittavaan sivustoon.



The screenshot shows the BibSonomy interface for a bookmark titled "COOL, BUT OBSCURE UNIX TOOLS :: KKOVCAS". The page features a navigation bar at the top with "home", "groups", and "popular" options. Below the title, there is a "hide discussion" button and a "rating distribution" bar chart. The average user rating is displayed as 4.0 out of 5.0, based on 3 reviews. A list of reviews follows, with the most recent one from "jaeschke" on Jul 12, 2011, stating "Some really nice tools I did not know before, although I have quite some experience with Linux." The right sidebar contains an "Info" section with instructions on how to rate and comment, and a "tags" section with a list of tags including "linux", "powertool", "shell", "tool", "tools", "ubuntu", and "unix".

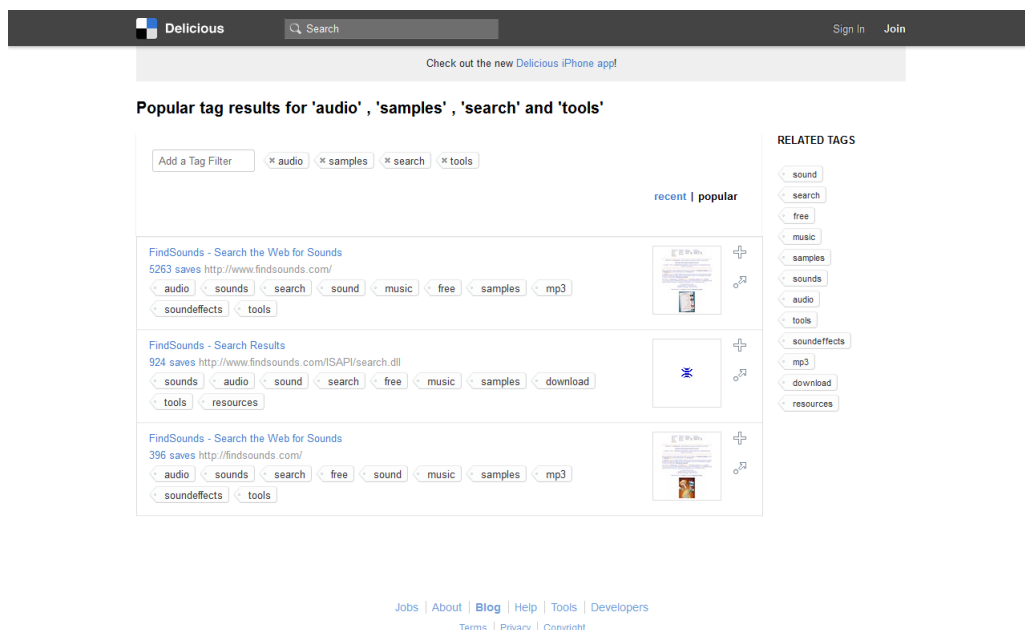
Kuvio 6: Esimerkki kohteiden rankkauksesta Bibsonomy-sivustossa.

Flickr-sivustolla ei ole mitään selkeää kohteiden arviointia, mutta sivuston objekteilla on näkyvissä käyttäjien suosikkilisäykset objektista. Tällä tarkoitetaan sitä, että käyttäjän on mahdollista ”tykätä” tai lisätä suosikkeihinsa kuvia, jolloin ne jäävät käyttäjän omaan käyttäjätiliin muistiin. Suosikkilisäyksien määrän voivat nähdä muut käyttäjät, joten suurempi luku kuvastaa kuvan suosittavuutta käyttäjien joukossa. Luku ei sinänsä ole arvio kuvan laadusta tai paremmuudesta, mutta sitä voidaan pitää eräänlaisena arviona, jonka käyttäjät ovat kuvasta antaneet. Tosin Flickr-sivuston kuvien subjektiivisen luonteen takia on ymmärrettävää, että kuvia ei pysty arvioimaan arvoasteikoilla, sillä mahdollisuus kuvien arvioimiseen objektiivisesti voisi aiheuttaa suuria ongelmia sivuston käyttäjien välille.

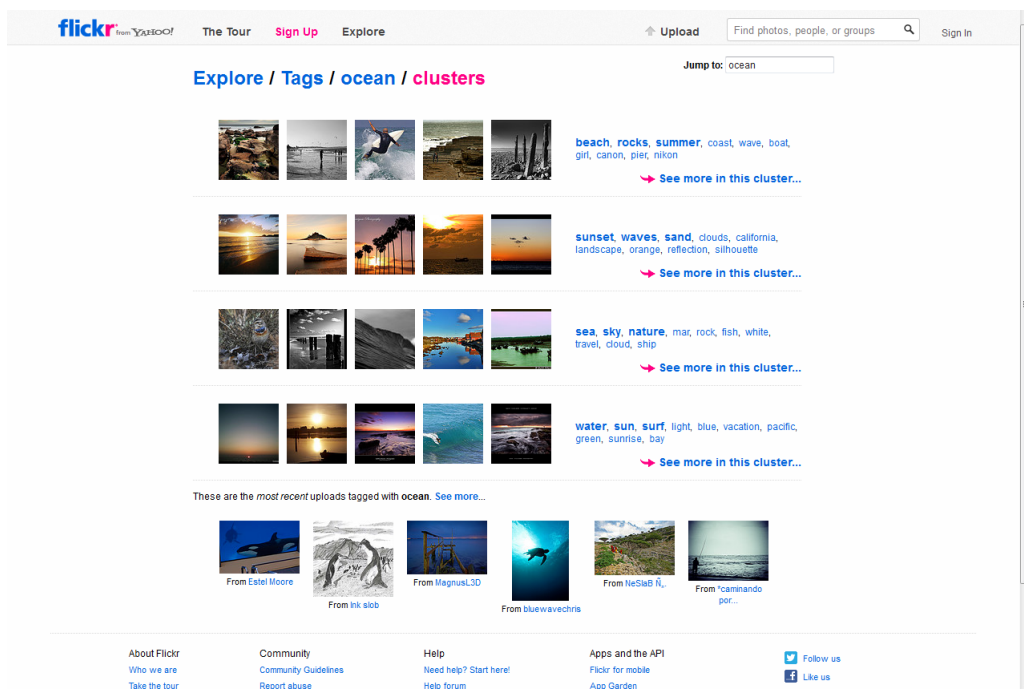
5.4 Tagien avulla hakeminen

Erilaisia **hakuja** voidaan suorittaa tagien kautta tai niiden avulla. Kun verkkopalvelu hyödyntää resursseja ja niiden välisiä huomautuksia, voidaan muodostunutta hierarkiaa käyttää hakemiseen, jolloin käyttäjä voi navigoida hierarkian sisällä kohteesta toiseen (Yang ja Shi 2009, s. 182). Käyttäjä voi myös hakea kohteita tagipilven kautta, jolloin käyttäjän valitseman tagin kautta palautuu hakutulos, joka sisältää painettua tagia vastaavia kohteita (Yang ja Shi 2009, s. 182). Vaikka tagihaku ja avainsanahaku ovat luonteeltaan samanlaisia, on tagihaussa erona itse palvelu, jossa haku suoritetaan (Papakonstantinou, Panagiotou ja Verbeek 2009, s. 356).

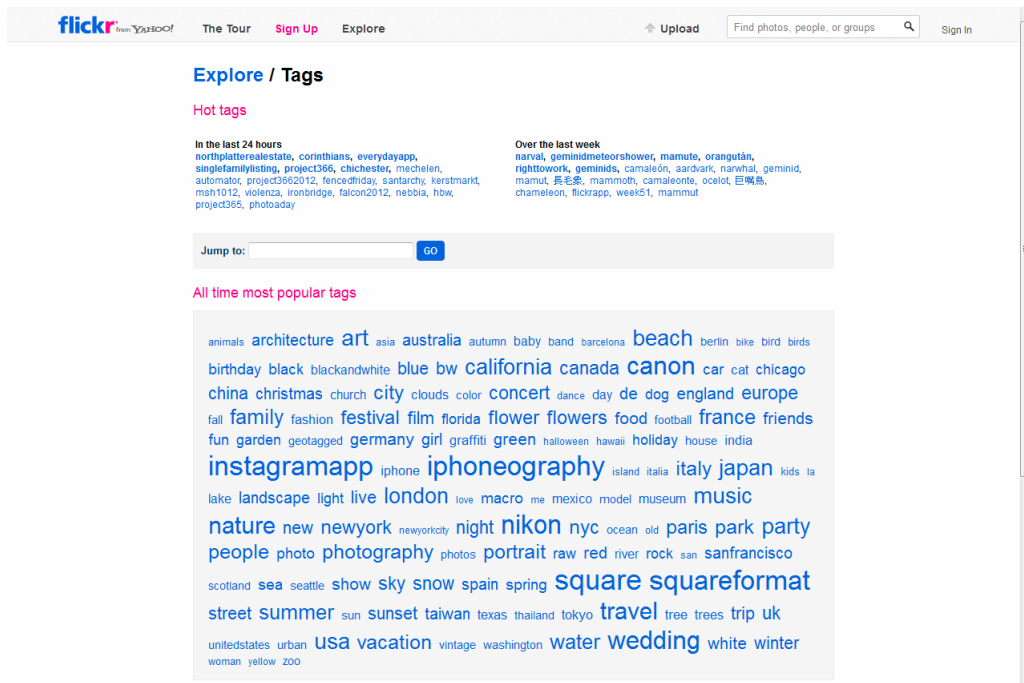
Jokainen tutkimuksen sivusto hyödyntää tiedonhakemista tagien avulla. Delicious-sivustossa tagihaku tapahtuu niin, että objektin tagia painamalla avautuu tagihaku kyseisellä tagilla ja hakutuloksessa näkymän 'Related Tags'-kohdan tageja painamalla avautuu navigoiminen tagirakenteen kautta, josta voi edetä hakua syvemmälle tagirakenteeseen niin kauan kuin käyttäjä itse haluaa. Bibsonomyssa hakemisen tagien avulla pystyy toteuttamaan monella eri tavalla joko tagipilven, hakukentän tai objektiin liitettyjen tagien avulla. Tageja linkittämällä yhteen voi käyttäjä hakea uutta tietoa tagirakenteesta ja aina lisättäessä yhtä uutta tagia toisen jälkeen sivusto ilmoittaa ennakkoon, kuinka monta hakutulosta halutulla tagihaulla tulisi. Flickr-sivustossa pystyy myös tekemään hakuja tagien avulla sekä hakemaan kuvia tagiklusterin kautta. Tagirakenteessa ei tosin pystynyt yhtä syvällisesti hakemaan tietoa kuin tutkimuksen kahdessa muussa sivustossa, käyttäjän haun mahdollisuuksien rajoituessa joko tagipilven tagiin tai tagia vastaavaan tagiklusteriin.



Kuvio 7: Esimerkki tagihausta Delicious-sivustossa.



Kuvio 8: Esimerkki tagiklusterihausta Flickr-sivustossa.



Kuvio 9: Esimerkki tagipilvestä Flickr-sivustossa.

5.5 Käyttäjän profilointi tagien avulla

Kun käyttäjä käyttää sosiaalista verkkopalvelua ja lataa tai tagaa verkkopalvelun kohteita, luo hän verkkopalveluun dataa, joka voidaan linkittää käyttäjään. Tätä muodostunutta dataa voidaan käyttää käyttäjän mieltymysten tunnistamiseen sekä käyttäjälle tarjottujen palveluiden, kuten suositusten, arvioiden tai hakujen parantamiseen, samalla tehden hänelle niistä hyvin henkilökohtaisia toimintoja.

Käyttäjän käyttäjäprofiili ilmaisee yleensä hänen mielenkiintonsa kohteita ja suosituksiaan sekä profilointiin voidaan liittää myös hänen kohde- ja tagisuosituksensa (Liang et al. 2010, s. 56). Käyttäjän käytöksen mallintamiseen voidaan käyttää luvussa 2.2 mainittuja asioita. Käyttäjä voidaan identifioida kuvailevaksi tai kategorisoivaksi tagaajaksi käyttämällä suhdelukuja, kuten sanastokoko, tagiresurssisuhde, tagien keskiarvo kohdetta kohden ja orposuhde. Käyttäjän muihin identifioitaviin ominaisuuksiin kuuluu thumbnail, tarkempi profiilitieto, kaverilista, käyttäjän luoma/lataama sisältö sekä kommentit, joihin kuuluu käyttäjän tilapäivitykset, lausunnot, suositukset ja tagit käyttäjästä tai käyttäjän kohteista (Krishnamurthy ja Wills 2008, s. 38).

Käyttäjän lisäämät tagit ovat tärkeä osa käyttäjän identiteettiä jokaisessa tutkimuksen sivustossa. Pääasiallisesti käyttäjän kaikki sivustoon lisäämät tagit tallentuvat käyttäjän tilille ja ovat hänelle näkyvissä käyttäjän profiilisivulla. Delicious-sivustossa käyttäjän lisäämien tagien lisäksi tageista näytetään myös tagien määrä käyttäjän lisäämissä objekteissa. Käyttäjä voi myös luoda tagiryppäitä, jotka voivat sisältää joko uusia tai käyttäjän aikaisemmin lisäämiä tageja. Bibsonomyssa käyttäjien lisäämät tagit tallentuvat käyttäjän tilille ja niitä käytetään tagisuositukseen ja tagien avulla hakemiseen. Kun käyttäjä hakee sivustossa olevia tekijöitä (engl. *authors*), niin heidän lisäämänsä objektit ja tagit näkyvät käyttäjälle sivulla. Bibsonomyssa käyttäjä voi myös luoda konsepteja, jotka koostuvat pääluokkatagista (engl. *supertag*) ja siihen liitetystä alatageista. Flickr-sivustossa käyttäjän tagit tallentuvat myös käyttäjän tilille ja käyttäjän tageja voivat myös nähdä sivuston muut käyttäjät. Muut käyttäjät voivat myös toteuttaa tagihakuja jonkun muun käyttäjän lisäämien tagien kautta.

5.6 Tutkimuksen tulokset

Tutkimukseen valitut sivustot voidaan jakaa kahteen ryhmään. Delicious ja Flickr kuuluvat enemmän sosiaalisiin verkkopalveluihin, joissa yhteisöllisyys sekä verkostoituminen muiden käyttäjien kanssa on asetettu sisällöntuottoa arvokkaammaksi asiaksi, kun taas Bibsonomy tarjoaa enemmän asialliseen ja laadukkaaseen sisällöntuottoon vaadittavia työkaluja. Nämä erot voidaan selvimmän nähdä jo sivustojen ulkoasuissa sekä erilaisissa sivuston tarjoamis- ja toiminnoissa. Sivustot jakaantuvat myös kahteen leiriin sivustoon lisättävien objektien osalta, missä Flickr eroaa kahdesta muusta siten, että sivustoon lisättävät objektit koostuvat kuvista ja muiden tekstipohjaisista objekteista.

Taulukko 1: Tutkittavien verkkosivustojen perustietoja.

	Delicious	BibSonomy	Flickr
Sivustoon lisättävä objekti:	kirjanmerkki	kirjanmerkki, julkaisu	kuva
Folksonomiarakenne:	leveä	leveä	kapea
Sivuston tyyli:	casual	professional	casual

Taulukko 2: Tutkittavissa verkkosivustoissa tutkittuja parametreja.

	Delicious	BibSonomy	Flickr
Tagisuositusten hyödyntäminen:	sisältöpohjainen, tagipohjainen	sisältöpohjainen, populaarisuus	personoitu
Kohteiden rankkaus/arviointi:	ei	arvoasteikko	ei
Tageilla rankkaus/arviointi:	ei	ei	ei
Tiedonhaku tagien avulla:	tagihierarkia	tagipilvi, tagihierarkia	tagipilvi, tagiklusteri
Tagit osana käyttäjän identiteettiä:	kyllä	kyllä	kyllä

Tutkimuksen sivustot hyödyntävät tageja monipuolisesti. Tagisuositusten hyödyntäminen, tiedonhaku tagien avulla sekä tagien käyttäminen osana käyttäjän identiteettiä löytyi kaikista tutkimuksen sivustoista, tosin niiden toteutustavoista löytyi pieniä eroavaisuuksia. Esimerkiksi Flickr-sivuston antamien tagisuositusten erilainen toteutustapa voidaan selittää sivustoon lisättävien objektien, eli kuvien avulla. Tutkimuksen sivustoissa hyödynnettiin vähiten

kohteiden arviointia: vain Bibsonomyssa oli mahdollista arvioida sivuston objekteja. Tagiarviointia ei esiintynyt sivustoissa, mutta Delicious-sivuston tageista löytyi eniten subjektiivisia tageja. Täten sivustojen käyttäjien olisi teoriassa mahdollista lisätä objekteihinsa tageja, jotka ovat arvioivia tai kuvailevia.

Tutkimuksen tuloksia tulee arvioida varovasti, sillä tutkimukseen valittiin sivustoja, joissa tageja hyödynnettäisiin hyvin suurella todennäköisyydellä. Myös tutkimukseen valittujen sivustojen määrä on pieni. Täten tutkimuksen tuloksista ei voi päätellä, miten tageja hyödynnetään kokonaisuudessaan sosiaalisissa verkkopalveluissa. Tutkimus antaa kuitenkin suuntaa sille, miten tageja oletetaan hyödynnettävän sellaisissa sosiaalisissa verkkopalveluissa, joissa hyödynnetään tageja. Tutkimuksesta voi myös olettaa, että sivuston keskittyessä enemmän yhteisöllisyyteen ja verkostoitumiseen ja sivuston objektien ollessa subjektiivisesti arvioitavia ja käyttäjän itse tuottamia, on kohteiden arvioiminen poissuljettu ominaisuus siitä mahdollisesti aiheutuvien käyttäjien ristiriitatilanteiden vähentämiseksi.

6 Yhteenveto

Tagit ovat avainsanoja tai termejä, joita voidaan liittää erilaisiin asioihin kuvaamaan niitä. Tägeja hyödynnetään erityisesti informaatioteknologian alalla erilaisin tavoin, joista keskeisimpiä ovat tiedon kuvaileminen metadatan avulla, tietoalueiden yhdistäminen tietorakenteiksi sekä tiedon hakeminen tietorakenteista. Tagit ovat yksinkertaisia määritellä ja niitä voi luoda kukin millä perusteella tahansa, mikä tekee niiden käyttämisestä ja tutkimisesta mielenkiintoista ja monipuolista.

Tämä tutkimus keskittyi tutkimaan tagien käyttöä sosiaalisissa verkkopalveluissa. Tägeja pystytään hyödyntämään verkkopalveluissa monella eri tavalla, jotka tarjoavat sivuston käyttäjille lisäarvoa ja mielenkiintoisia tapoja käyttää sivustoja. Käyttäjä voi tageja hyödyntävissä sivustoissa luoda tageja, kuvailla kohteita tageilla sekä hakea uutta informaatiota tagien kautta. Tägeina voi myös käyttää muiden sivuston käyttäjien tageja, mikä voi tehdä tagaamisesta myös sosiaalista toimintaa. Merkittävimmät sovelluskohteet tageja hyödyntävissä sosiaalisissa verkkopalveluissa ovat tagisuositukset, tiedonhaku sekä kohteiden ja käyttäjien kuvaaminen tagien avulla.

Tiedon käsittelemiseen, luomiseen ja hakemiseen kuluu paljon aikaa. Metadatan tutkiminen ja tutkimisen kautta löytyvien teorioiden ja sovellusten löytäminen edesauttanees yhteiskuntaa tekemään parempia sovelluksia löytämään ja käsittelemään tietoa. Tulevaisuudessa on mahdollista käyttämällä metadatojen sovelluksia (jotka pohjautuvat tageihin ja/tai ontologioihin) luoda järjestelmiä, joissa tietokoneet luovat itse uutta tietoa olemassa olevasta tiedosta. Tällaisesta käytetään nimeä Semantic Web, jossa on informaatiota, jota tietokoneetkin ymmärtävät.

Lähteet

Barragáns-Martínez, Ana Belén, Marta Rey-López, Enrique Costa-Montenegro, Fernando A. Mikic-Fonte, Juan C. Burguillo-Rial ja Ana Peleteiro-Ramallo. 2010. “Exploiting Social Tagging in a Web 2.0 Recommender System.” *IEEE Internet Computing* 14 (6): 23–30. <http://dblp.uni-trier.de/db/journals/internet/internet14.html#Barragans-MartinezRCMBP10>.

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. ACM ID: 1557969. Torino, Italy: ACM. ISBN: 978-1-60558-486-7. doi:10.1145/1557914.1557969.

Benz, Dominik, Andreas Hotho, Robert Jäschke, Beate Krause, Folke Mitzlaff, Christoph Schmitz ja Gerd Stumme. 2010. “The Social Bookmark and Publication Management System BibSonomy”. *The VLDB Journal* (Berlin / Heidelberg) 19, numero 6 (): 849–875. ISSN: 1066-8888. doi:10.1007/s00778-010-0208-4. <http://www.kde.cs.uni-kassel.de/pub/pdf/benz2010social.pdf>.

Candan, K. Selçuk, Luigi Di Caro ja Maria Luisa Sapino. 2008. “Creating tag hierarchies for effective navigation in social media.” Teoksessa *SSM*, toimittanut Ian Soboroff, Eugene Agichtein ja Ravi Kumar, 75–82. ACM. ISBN: 978-1-60558-258-0. <http://dblp.uni-trier.de/db/conf/cikm/ssm2008.html#CandanCS08>.

Caro, Luigi Di, K. Selçuk Candan ja Maria Luisa Sapino. 2008. “Using tagFlake for condensing navigable tag hierarchies from tag clouds.” Teoksessa *KDD*, toimittanut Ying Li, Bing Liu ja Sunita Sarawagi, 1069–1072. ACM. ISBN: 978-1-60558-193-4. <http://dblp.uni-trier.de/db/conf/kdd/kdd2008.html#CaroCS08>.

Christie, Timothy, Christopher Lueg ja Nilufar Baghaei. 2010. “Tag clouds as social signalers.” Teoksessa *OZCHI*, toimittanut Margot Brereton, Stephen Viller ja Ben Kraal, 360–363. ACM. ISBN: 978-1-4503-0502-0. <http://dblp.uni-trier.de/db/conf/ozchi/ozchi2010.html#ChristieLB10>.

- 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. <http://dblp.uni-trier.de/db/journals/tois/tois28.html#ClementsVR10>.
- Gemmell, Jonathan, Thomas Schimoler, Bamshad Mobasher ja Robin D. Burke. 2010. “Hybrid tag recommendation for social annotation systems.” Teoksessa *CIKM*, toimittanut Jimmy Huang, Nick Koudas, Gareth J. F. Jones, Xindong Wu, Kevyn Collins-Thompson ja Aijun An, 829–838. ACM. ISBN: 978-1-4503-0099-5. <http://dblp.uni-trier.de/db/conf/cikm/cikm2010.html#GemmellSMB10>.
- Gupta, Manish, Rui Li, Zhijun Yin ja Jiawei Han. 2010. “Survey on Social Tagging Techniques”. *ACM SIGKDD Explorations Newsletter* (New York, NY, USA) 12, numero 1 (): 58–72. ISSN: 1931-0145. doi:10.1145/1882471.1882480.
- Hearst, M., ja D. Rosner. 2008. “Tag Clouds: Data Analysis Tool or Social Signaller?” Teoksessa *Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS)*. Hawaii, USA.
- Heymann, Paul, ja Hector Garcia-Molina. 2006. *Collaborative Creation of Communal Hierarchical Taxonomies in Social Tagging Systems*. Tekninen raportti. Computer Science Department, Stanford University. </brokenurl#dbpubs.stanford.edu:8090/pub/2006-10>.
- Iturrioz, Jon, Oscar Díaz ja Iker Azpeitia. 2011. “Reactive tags: associating behaviour to prescriptive tags.” Teoksessa *HT*, toimittanut Paul De Bra ja Kaj Grønbaek, 191–200. ACM. ISBN: 978-1-4503-0256-2. <http://dblp.uni-trier.de/db/conf/ht/ht2011.html#IturriozDA11>.
- Kim, Hak L., Suk H. Hwang ja Hong G. Kim. 2007. “FCA-based approach for mining contextualized folksonomy”. Teoksessa *SAC '07: Proceedings of the 2007 ACM symposium on Applied computing*, 1340–1345. New York, NY, USA: ACM Press. ISBN: 1595934804. doi:10.1145/1244002.1244292. <http://dx.doi.org/10.1145/1244002.1244292>.
- Koivisto, Antti. 2010. “Kontekstitietoon perustuva automaattinen tagien luonti mobiililaitteilla”. Thesis Required for a Diploma, Tampere University of Technology.

- Körner, Christian, Dominik Benz, Andreas Hotho, Markus Strohmaier ja Gerd Stumme. 2010. “Stop Thinking, start Tagging - Tag Semantics emerge from Collaborative Verbosity”. Teoksessa *Proc. of 19th International World Wide Web Conference (WWW), April 26-30 2010, Raleigh, NC, USA*.
- Koskinen, Riikka. 2007. “Folksonomioiden semantiikkaa”. Bachelor Thesis, University of Helsinki.
- Krishnamurthy, Balachander, ja Craig E. Wills. 2008. “Characterizing privacy in online social networks”. Teoksessa *Proceedings of the first workshop on Online social networks*, 37–42. WOSP '08. Seattle, WA, USA: ACM. ISBN: 978-1-60558-182-8. doi:<http://doi.acm.org/10.1145/1397735.1397744>.
- Liang, Huizhi, Yue Xu, Yuefeng Li, Richi Nayak ja Xiaohui Tao. 2010. “Connecting users and items with weighted tags for personalized item recommendations.” Teoksessa *HT*, toimittanut Mark H. Chignell ja Elaine Toms, 51–60. ACM. ISBN: 978-1-4503-0041-4. <http://dblp.uni-trier.de/db/conf/ht/ht2010.html#LiangXLNT10>.
- Lu, Caimei, Xiaohua Hu, Jung ran Park ja Jia Huang. 2011. “Post-based collaborative filtering for personalized tag recommendation.” Teoksessa *iConference*, 561–568. ACM. ISBN: 978-1-4503-0121-3. <http://dblp.uni-trier.de/db/conf/iconference/iconference2011.html#LuHPH11>.
- Nurminen, Miika. 2005. “Tiedonlouhinta rakenteisista dokumenteista”. Tutkielma, University of Jyväskylä. <http://urn.fi/URN:NBN:fi:jyu-200594>.
- Papakonstantinou, Christina, Ioannis Panagiotou ja Fons J. Verbeek. 2009. “The TicTag application: towards tag-based meta-search for browsing the web.” Teoksessa *BCS HCI*, 354–361. ACM. <http://dblp.uni-trier.de/db/conf/bcshci/bcshci2009.html#PapakonstantinouPV09>.
- Plangprasopchok, Anon, Kristina Lerman ja Lise Getoor. 2010. “A Probabilistic Approach for Learning Folksonomies from Structured Data”. *CoRR* abs/1011.3557. <http://dblp.uni-trier.de/db/journals/corr/corr1011.html#abs-1011-3557>.

Porter, Joshua. 2006. *The Del.icio.us Lesson*. <http://bokardo.com/archives/the-delicious-lesson/>. [Online; accessed 28-January-2013].

Rader, Emilee, ja Rick Wash. 2008. "Influences on tag choices in del.icio.us". Teoksessa *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, 239–248. CSCW '08. San Diego, CA, USA: ACM. ISBN: 978-1-60558-007-4. doi:10.1145/1460563.1460601. <http://doi.acm.org/10.1145/1460563.1460601>.

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>.

RateYourMusic.com. 2013. <http://rateyourmusic.com>. [Online; accessed 28-January-2013].

Shepitsen, Andriy, Jonathan Gemmell, Bamshad Mobasher ja Robin Burke. 2008. "Personalized recommendation in social tagging systems using hierarchical clustering". Teoksessa *Proceedings of the 2008 ACM conference on Recommender systems*, 259–266. RecSys '08. Lausanne, Switzerland: ACM. ISBN: 978-1-60558-093-7. doi:10.1145/1454008.1454048. <http://doi.acm.org/10.1145/1454008.1454048>.

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. New York, NY, USA: ACM. ISBN: 978-1-60558-085-2. doi:10.1145/1367497.1367542.

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. WOSN '08. Seattle, WA, USA: ACM. ISBN: 978-1-60558-182-8. doi:10.1145/1397735.1397737. <http://doi.acm.org/10.1145/1397735.1397737>.

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. Washington, DC, USA: IEEE Computer Society. ISBN: 978-1-4244-3453-4. doi:<http://dx.doi.org/10.1109/ICSE.2009.5070504>. <http://portal.acm.org/citation.cfm?id=1555001.1555018>.

Vig, Jesse, Matthew Soukup, Shilad Sen ja John Riedl. 2010. "Tag expression: tagging with feeling." Teoksessa *UIST*, toimittanut Ken Perlin, Mary Czerwinski ja Rob Miller, 323–332. ACM. ISBN: 978-1-4503-0271-5. <http://dblp.uni-trier.de/db/conf/uist/uist2010.html#VigSSR10>.

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. WSDM '10. New York, New York, USA: ACM. ISBN: 978-1-60558-889-6. doi:10.1145/1718487.1718497. <http://doi.acm.org/10.1145/1718487.1718497>.

Yang, Kun, ja Zhongzhi Shi. 2009. "Exploring Web Resources in Social Annotation Services in Hierarchy." Teoksessa *GCC*, 182–187. IEEE Computer Society. ISBN: 978-0-7695-3766-5. <http://dblp.uni-trier.de/db/conf/gcc/gcc2009.html#YangS09>.

Zhang, Ning, Yuan Zhang ja Jie Tang. 2009. "A tag recommendation system for folksonomy." Teoksessa *CIKM-SWSM*, toimittanut Irwin King, Juan-Zi Li, Gui-Rong Xue ja Jie Tang, 9–16. ACM. ISBN: 978-1-60558-806-3. <http://dblp.uni-trier.de/db/conf/cikm/swsm2009.html#ZhangZT09>.

Zhou, Tom Chao, ja Irwin King. 2009. "Automobile, car and BMW: horizontal and hierarchical approach in social tagging systems." Teoksessa *CIKM-SWSM*, toimittanut Irwin King, Juan-Zi Li, Gui-Rong Xue ja Jie Tang, 25–32. ACM. ISBN: 978-1-60558-806-3. <http://dblp.uni-trier.de/db/conf/cikm/swsm2009.html#ZhouK09>.

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. Eindhoven, The Netherlands: ACM. ISBN: 978-1-4503-0256-2. doi:10.1145/1995966.1995981. <http://doi.acm.org/10.1145/1995966.1995981>.