

Jarmo Saarti

Folksonomiat ja kontrolloidut sanastot kaunokirjallisuuden kuvailussa

Jarmo Saarti: Folksonomiat ja kontrolloidut sanastot kaunokirjallisuuden kuvailussa. [Folksonomias and controlled vocabularies as tools for the content representation of fiction]. Informaatiotutkimus 28(1), 2009.

This paper discusses differences and possibilities between two approaches of the fiction knowledge organization: folksonomias and controlled vocabularies. Due to the nature of fiction communication it is suggested that these both tools should be used in order to achieve both structured and managed tools as well as in order to give the users as flexible means for describing their experiences as possible.

Address: Jarmo Saarti, Kuopio University Library, P.O.Box 1627, FIN-70211 Kuopio, Finland. e-mail: jarmo.saarti@uku.fi

Johdanto

Fiktiivisten aineistojen, erityisesti romaanien, sisällönkuvailussa on hyödynnetty perinteisiä kirjastojen työkaluja, kuten luokitusta, asiasanoitusta ja tiivistelmiä (ks. tästä yleisesitys Saarti 1999). Painetun kirjan aikakaudella kaunokirjallisuuden tiedonjärjestäminen liittyi pitkälti hyllyluokitusten kehittämiseen ja niiden hyödyntämiseen avokokoelmien löytyvyyden parantamisessa ja selailuhaun tukemisessa. Tämän vaiheen luokitusten perusteina olivat pitkälti kaunokirjalliset lajityypit eli genret ja puhuttiin paljon genreluokituksista.

Tieto- ja viestintätekniikan kehittymisen myötä tarve kehittää kaunokirjallisuuden aihehakua ja teosten indeksointia nousi ajankohtaisemmaksi, koska uusilla tietoteknisillä välineillä pystyttiin paremmin vastaamaan siihen haasteeseen, jonka kaunokirjallisen teoksen sisällön määrittelyn moniulotteisuus tai alan terminologiaa käyttäen, monifasettisuus, asettaa. Tiedonjärjestämistä oli jo tätä ennen pyritty tekemään mm. erilaisilla hakemistoilla, jotka liittyivät joko yksittäisiin teoksiin tai teosjoukkoihin (vrt. teosten suositteluoppaat, valikoimaluettelot erilaisista teoksista ja niiden esittelyistä). Korttiluetteloiden aikakaudella alkoivat myös ensimmäiset

kaunokirjallisten teosten asiasanoituskokeilut. Systematiikkaan asiasanoituksessa päästiin varsin myöhään, vasta oikeastaan 1990-luvulla alkoi ilmestyä laajempia kaunokirjallisuuden tesauroksia ja asiasanastoja.

Kaunokirjallisten teosten sisällönkuvailun suurin haaste on sen kohteen olemus: teoksen sisällön lukija tulkitsee aina omista lähtökohdistaan. Koska tulkinta viime kädessä on aina subjektiivinen, on ongelmana löytää kaunokirjallisuudesta ne aspektit, joiden voidaan katsoa olevan yksittäisestä kokemuksesta riippumattomia, naivistisesti sanottuna objektiivisia elementtejä kaunokirjallisessa teoksessa (ks. Saarti 1999).

Näiden elementtien listauksia on tehty useita alamme tutkimuksessa (hyvä yleisesitys näistä löytyy teoksesta Saarti 1999). Sisällönkuvailun perusfasetteina on pidetty mm. teosten lajityyppiä (genreä), teosten tapahtuma-aikaa ja paikkaa. Selkeimpiä ja aidoimmin objektiivisia elementtejä ovat esimerkiksi teoksen tapahtuma-ajan ja paikan lisäksi tyyliin, kieleen ja teoksessa käsiteltyihin faktisiin objekteihin, kuten esimerkiksi henkilöihin, liittyvät asiat. Teoksen kieli lienee näistä selkeimmin määriteltävissä: teos on kirjoitettu englanniksi, suomeksi, puhekielellä jne. Mutta jo teoksen objektit ovat hankalammin määriteltävissä:

onko teoksessa a esiintyvä Helsinki sama kuin teoksessa b? Mitä tekemistä tieteiskirjallisuuden Helsinki nimisellä paikalla on sen paikkakunnan kanssa, johon me viittaamme, kun puhumme Helsingistä?

Tulkinnan käsitteen luonteesta fiktiivisten aineistojen kohdalla seuraa se, ettemme voi sanoa, että kustakin teoksesta on olemassa vain yksi oikea tulkinta tai että jonkin lukijan tulkinta on väärä (vrt. Lehtonen 2000, 112 – 115 ja Gadamer 2005, 10 – 11). Teoksen tulkinta muuttuu myös sen ja sen lukijoiden historian myötä. Niinpä kussakin fiktiivisten aineistojen tietojärjestelmässä tulee määritellä ne elementit, joista voidaan päästä tarpeeksi suureen yhdenmukaisen mielipiteen muodostukseen, jotta voidaan määritellä esimerkiksi luokituksen tai asiansanoituksen tyyppisiä välineitä sisältöjen hallitsemiseksi. Lisäksi tulee määritellä, mitkä elementeistä tulee jättää avoimen keskustelun ja jopa yksittäisen ihmisen tulkinnan alueeksi ja miten hallita tuo tulkinnan muutos historian kuluessa.

Seuraavassa käsitellään folksonomioiden ja kontrolloitujen sanastojen roolia ja ominaispiirteitä kaunokirjallisuuden kuvailussa. Esitys on luonteeltaan käsitteellinen analyysi ja siinä käytetyt esimerkit siten yksittäisiä, niiden systemaattinen analyysi jää tutkimuksen seuraavaan vaiheeseen. Artikkelin perustuu Informaatiotutkimuksen yhdistys ITY:n syysseminaarissa Tampereella 15.11.2008 pidettyyn esitelmään.

Käyttäjä- ja teoslähtöisyys fiktiivisten aineistojen sisällönkuvailussa

Teosten ja niiden käyttäjien suhdetta kuvataan informaatiotutkimuksessa relevanssin käsitteellä. Relevanssin käsite on sumea ja hankala määritellä, mutta karkeasti sen voi sanoa tarkoittavan löydetyn dokumentin ja tiedonhakijan tarpeen välistä positiivista kohtaamista. Käsitteen sumeus tulee siitä, että teoksen ja kokijan positiivinen kohtaaminen tai tiedonhaussa tiedontarpeen ja teoksen kuvailun täsmäytys voi olla hyvin monimuotoista. Kohtaamisen seurauksena voi olla jopa se, että vaikka löydetty dokumentti ei vastaa hakijan tiedontarpeeseen, se auttaa tarpeen muotoilussa sellaiseksi, että tiedonhaku voi jatkua.

Jo Saracevic (1975) erotti relevanssin käsitteessä eri tasoja, joita hänen mukaansa olivat aiheenmukaisuus eli topikaalisuus (liittyy teokseen

ja sen sisältöön), motivaationmukaisuus (liittyy tiedonhakijan tarpeeseen) sekä tulkinnallisuus (liittyy teoksen tulkintaan). Fiktiivisissä aineistoissa kukin näistä tasoista asettaa spesifit haasteensa tiedonjärjestämisessä käytetyille menetelmille; olivat ne sitten inhimilliseen toimintaan, tietojärjestelmiin perustuvia tai näiden hybridejä.

Kuten jo edellä todettiin, fiktiivisessä aineistossa aiheenmukaisuuden käsite on hankala. Toisessa ääripäässä ovat teokset, jotka selkeästi haluavat sanoa nk. reaalityodellisuudesta jotakin. Hyvänä esimerkkinä ovat mm. elämäkerrat ja realistiset historialliset romaanit. Fiktiivinen elämäkerta kertoo jonkin historiallisen ihmisen elämästä, mutta se, että teos on julkaistu fiktiivisenä eikä historiallisena elämäkertana kertoo sen, että tekijän intentiossa on ollut jotakin, joka siirtää tuon historiallisen henkilön fiktiivisen maailman puolelle ja sen jälkeen kytkös historialliseen henkilöön muuttuu toisenlaiseksi – se on hänen historiallisen elämänsä tulkinta ja uudelleenluominen. Tämän vuoksi vähimmäisvaatimuksena tietojärjestelmille on, että fiktiiviset ja faktuaaliset aineistot pystytään selkeästi erottamaan. Kuten tutkimuksissa on jo pitkään todettu, fiktiivisten aineistojen tietojärjestelmien tulee olla monifasettiseen sisällön analyysiin ja tiedonhakuun vastaamaan pystyviä, erityisesti, jos samassa järjestelmässä pyritään järjestämään sekä fiktiivistä että fakta-aineistoa.

Fiktiivisen aineiston relevanssin motivaationmukaisuus on keskeinen haaste fiktiivisen aineiston tiedonhaussa¹ ja siihen liittyvien tietojärjestelmien laatimisessa. Fiktiivisten aineistojen tiedonhaun voi jakaa karkeasti kahteen keskeiseen motivaatioon:

1. halutaan löytää jonkin aiheen mukaisia teoksia tai
2. halutaan löytää hyviä teoksia – hyviä elokuvia katsottaviksi, hyviä romaaneja luettavaksi, hyvää musiikkia kuunneltavaksi jne.

Ensimmäinen motivaatio liittyy myös aiheenmukaisuuden haasteeseen fiktiivisissä aineistoissa. Tietojärjestelmien tulisikin kyetä vastaamaan esimerkiksi siihen kysymykseen, miten hyvin tai huonosti reaali maailman aiheenmukaisuuden mukaisesti teoksessa A käsitellään historiallista tapahtumaa X, tai jotakin faktamaailman asiaa, esimerkiksi sairastumista, identiteetin rakentumista? Topikaalisesti kirjallisuutta hakevien motivaatio teosten löytymiseen on usein myös ei-esteettinen lukemisen tapa. Haetaan

teoksia esimerkiksi opinnäytetyötä tai historiallista tutkimusta varten tai matkaoppaaksi.

Toinen motivaatio on hankalammin hallittavissa perinteisen sisällönkuvailun menetelmillä. Hyvän teoksen etsimisessä on erityisesti fiktiivisissä aineistoissa kyse erilaisuuden ja samankaltaisuuden (vrt. reema-teema käsitepari) optimoinnista kutakin henkilöä varten (vrt. Richin tutkimukset oppivasta kaunokirjallisuuden suosittelijätietojärjestelmästä 1979 ja 1983). Eli etsitään hyvää lukemista sen pohjalta, mitä aikaisemmin on pidetty hyvänä, mutta halutaan samalla jotakin uutta kokemuksellisuutta. Toisaalta fiktiivisissä aineistoissa usein suurin löytämisen riemu koetaan löydettyessä teos, joka avaa uuden fiktiivisen maailman eli joka ei vastaa mitään aikaisempaa kokemusta.

Tietojärjestelmien laatiminen tälle jälkimäiselle motivaation kokonaisuudelle on haastava. Ensinnäkin siksi, että käsitteellisesti ”hyvä teos” on täysin sumea ja riippuvainen lisäksi yksilön omasta mielipiteestä, joka muuttuu historiallisesti ajan kuluessa. Teos, jonka koen hetkellä A merkitykselliseksi saattaa hetkellä B olla merkityksetön ja palata hetkellä C taas merkitykselliseksi omassa lukemishistoriassani. Hyvä esimerkki tästä on lapsuuden kirjojen tai äänitteiden uudelleen löytäminen vanhuuden aikana tai jonkin teoksen merkityksellisyys oman elämän tai kansakunnan murroskohdassa.

Toisaalta hyvä fiktiivinen teos määrittäytyy myös aina fiktiivisen kokemisen historiallissosiaalisesta kontekstista. Teosten tulkinta on aina sosiaalinen tapahtuma koska kieli on sosiaalinen ja historiallinen konstruktio (dadaistit yrittivät toki kyllä luoda täysin vapaata kieltä). Tämän aspektin hallinta on näyttäytynyt jo pitkään esimerkiksi erilaisten kirjallisten kaanoneiden luomisena, joko yläkulttuurisena toimintana (”maailman merkittävimmät elokuvateokset”, ”Suomen kansalliskirjailijat”) tai osakulttuurisena toimintana (”sata jatsilevyä, jotka sinun on kuunneltava”, ”Death metallin viime vuoden parhaat”). Tällöin teosten käyttämiseen ja niihin liittyvään tiedonhakuun tulee mukaan myös sosiaalinen toiminta ja sosiaalisessa yhteisössä mukana olemisen ja yhteisöön kuulumisen tarve.

Fiktiivisissä aineistoissa voidaan erottaa myös neljäs relevanssin piirre, joka todennäköisesti näyttäytyy myös fakta-aineistojen kohdalla, mutta ei niin selkeänä. Nimitän sitä toiminnallisuuden relevanssiksi. Tämä toteutuu ensinnäkin teosten

tulkinnoissa, jotka näyttäytyvät kirjallisuuden arvosteluina tai verkkomaailmassa omien esteettisten kokemusten jakamisena erilaisissa verkkoyhteisöissä. Toisaalta tämä toteutuu uusien fiktiivisten teosten tekemisenä, jotka liittyvät johonkin olemassa olevaan teokseen, lajiytyppiin tai fiktiiviseen maailmaan (vrt. esimerkiksi roolipelien tarinat ja niiden luominen). Erityisesti tämän näyttäytyy populaarikulttuurigenreissä ja verkkomaailma on antanut tälle toimintatavalle työkalut ja synnyttänyt oman toimintaympäristönsä esimerkiksi nk. fanifiktiosivustoissa.

Lukijat teosten tulkitsijoina ja fiktiivisen merkitysmaailman rakentajina

Kirjastojen ja muiden tiedontallennusta tekevien organisaatioiden tekemä asiansanoitus voidaan jakaa kahteen pääalueeseen, jotka ovat ihmisten tekemä tulkinnallinen ja liittävä sanoitus ja automaattinen, algoritmeihin perustuva tekstidatasta luotu avainsanahakemistojen laatiminen. Näistä edellisessä työkaluina ovat erityyppiset sanastot ja jälkimmäisessä algoritmit, jotka pyrkivät mallintamaan mm. tilastollisesti ja kieliteknologisesti järkeviä tapoja laatia hakemistoja, joilla pyritään sisältöanalyttisesti kuvailemaan teosten sisältöä.

Tietoverkon sosiaalisten teknologioiden kehittyessä käyttäjien mahdollisuus tehdä tulkinnallista ja liittävää sanoitusta on lisääntynyt. Tässä yhteydessä on käytetty termejä folkosomiat ja tagittaminen. Folkosomia voidaan määritellä käyttäjien luomaksi taksonomiaksi erityisesti verkkoyhteisössä tapahtuvaa dokumenttien kuvailua avainsanojen - tai tagien (tags) kuten verkossa näitä nimitetään – avulla (Folksonomy 2009, Folksonomia 2009, Noruzi 2007).

Folkosomioiden rakentumisen syynä on ollut sama kuin muidenkin dokumenttien: teosmassa tarvitsee liittävää asiansanoitusta, jotta sisällöllinen tiedonhaku olisi edes jotenkin mahdollista. Sosiaalisten verkkoteknologioiden kehittyminen on puolestaan antanut mahdollisuuden niiden käyttöön: verkossa olevissa palveluissa on mahdollista antaa käyttäjille vapaa oikeus ja mahdollisuus indeksoida teoksia.

Sanastojen rakentuminen on alkanut vilpisti, mutta varsin pian asiansanoittamiselle (verkkokielessä: tagittamiselle) on alettu vaatia rakenteita: sanastoja ja jopa tesaurusrakenteita. Noruzi (2007) esittää seuraavat perinteiset

perusteet tesaurusmaisen rakenteellisuuden tarpeelle: termistön ja niiden muodon ja merkityksen kontrollointi, hierarkkiset linkitykset termien välillä, oikeinkirjoituksen hallinta ja sanastorakenteen luominen tiedonhakijoiden tueksi. Spiteri (2007) puolestaan näkee folksonomian mahdollisuutena laajentaa yleisen kirjaston palveluita, mutta hän pitää tärkeänä vastaavien rakenteellisten ratkaisujen tekemistä, jos kirjastot avaavat käyttäjille mahdollisuuden indeksoida teoksia omaan aineistoluetteloonsa. Verkkoympäristö antaa tesaurusten kehittämiselle myös uusia mahdollisuuksia, esimerkiksi niiden linkittäminen sanakirjoihin ja hakuteoksiin merkitysten määrittämiseksi ja sanastojen keskinäisiin linkittämisiin.

Avoin asiansanoituskin näyttäisi tarvitsevan rakenteen. Lisäksi näyttää siltä, että tuo rakenne tulee suunnitella ja laatia ammattimaisesti. Suomessa meillä alkaa olla mahdollisuus integroida ontologisoidut verkkosanastot – sekä VESA että Kaunokki – tähän käyttöön. Saarti (1999) on esittänyt, että käyttäjien vapaan indeksoinnin kontrolloimiseksi kannattaa harkita tilastollisia välineitä, joilla asetetaan annetut asiansanat järjestykseen niiden yleisyyden mukaan. Toinen mahdollisuus on hyödyntää useissa verkkopalveluissa käytettyä äänestysmenetelmää, jolla tiedonhakijat arvioivat käytetyn kuvailun relevanssia. Kolmas mahdollisuus on moderointi eli ammattilaisen tekemä valinta ja toimitustyö annetuille asiansanoille, tätä hyödynsivät jo Hilderley ja Rafferty (1997) omassa, nk. demokraattisen indeksoinnin kokeilussaan.

Taulukossa 1 vertaillaan folksonomioita ja kontrolloituja sanastoja sisällönkuvailun välineinä. Peruserona voidaan pitää sitä, että folksonomia rakentuu yhteisön toiminnasta ja testautuu samanaikaisesti kun sitä luodaan. Kontrolloidun sanaston rakentuminen on yleensä tälle vastakkainen prosessi. Tämä johtuu tietenkin näiden

sanaston rakentamisen muotojen perustavasta erosta: toinen vastaa nopeasti kysymykseen mitä minä ajattelen teoksesta (käyttäjärelevanssi) toinen pyrkii yleistämään teoksen sisällön laajemmalle yhteisölle (topikaalinen relevanssi). Jatkotutkimuksen kohteeksi jää se, kuinka nämä kaksi välinettä yhdistetään eli selkeä ja kontrolloitu rakenne sanaston ja hierarkioiden joustavaan luomiseen. Samassa yhteydessä on syytä testata myös tilastollisten menetelmien ohjelmallista hyödyntämistä.

Johtopäätöksiä

Nettiyhteisöjen vahvuus ja innovatiivisuus tulee aktiivisesta sosiaalisesta toiminnasta, jota kukin osa- ja alakulttuuri harrastaa omassa toimintaympäristössään. Näyttää kuitenkin siltä, että käyttäjien ja tallennetun informaation määrän kasvaessa, sen hallintaan tarvitaan perinteisiä tiedonhallinnan ja dokumenttien tiedonjärjestämisen menetelmiä.

Mielenkiintoisen haasteen muistiorganisaatioille tarjoaa niiden kehittyneiden sisällönkuvailun välineiden integrointi sosiaalisiin verkkoyhteisöihin ja näiden tapaan tuottaa ja kuvailla tuottamaansa informaatiota. Sama haaste toisinpäin: käyttäjien päästäminen kuvailemaan dokumentteja esimerkiksi kirjastojen tietokantoihin sekä sentyyppisillä välineillä, mitä esimerkiksi verkkokirjakaupoissa käytetään (teosten arviointi, niiden käyttö, omien käyttökokemusten julkaiseminen) että kirjastojen perinteisillä välineillä (luokituskaavoilla ja asiansanoilla).

Informaatiotutkimukselle uuden ja haastavan kohteen tarjoavat sosiaaliset verkkoympäristöt, joissa käyttäjät jakelevat tuottamia aineistoja ja samalla järjestävät niitä käyttöä varten toisille käyttäjille. Tässä on aito empiirinen

Folksonomia

-vastaa avoimesti kysymykseen fiktion merkityksestä ja aiheenmukaisuudesta
-tukee myös pienyhteisöjen tulkinnallista toimintaa
-on joustavasti määritelty
-siirtymä sisällönkuvailusta elämyksellisen kokemuksen kuvailuun helppoa

Kontrolloitu sanasto

-vastaa kontrolloidusti kysymykseen fiktion merkityksestä ja aiheenmukaisuudesta
-tukee (kirjastojen) bibliografista valvontaa
-on tiukasti määritelty
-siirtymä sisällönkuvailusta elämyksellisen kokemuksen kuvailuun periaatteessa mahdotonta

Taulukko 1. Folksonomioiden ja kontrolloitujen sanastojen vertailua

laboratorio, jossa on mahdollista seurata, kuinka tiedonjärjestämisen välineet kehittyvät sosiaalisen evoluution tuloksena.

Hyväksytty julkaistavaksi 27.3.2009.

Viittaus:

¹Tiedonhaku on fiktiivisiin aineistoihin liittyen hankala käsite. Yritin joskus ehdottaa suomenkielen *sepite*-termin käyttöönottoa tässä yhteydessä ja määrittellä fiktiivisten aineistojen haun *sepite*hauksi. Termin arkaainen konnotaatio ikävä kyllä tekee tämän käytön jos nyt ei mahdottomaksi niin ainakin hankalaksi. Yksi mahdollisuus olisi käyttää käsitettä kaunohaku.

Lähteet

- Folkosonomia (2009). Teoksessa *Wikipedia*. [Luettu 11.1.2009.] Saatavissa: <http://fi.wikipedia.org/wiki/Folkosonomia>
- Folkosonomy (2009). In *Wikipedia, The Free Encyclopedia*. [Luettu 11.1.2009.] Saatavissa: <http://en.wikipedia.org/wiki/Folkosonomy>
- Gadamer, Hans-Georg (2005). Totuus hengentieteissä. Teoksessa: *Hermeneutiikka: ymmärtäminen tieteissä ja filosofiassa*, s. 3 – 11, valikoinut ja suomentanut Ismo Nikander.
- Hidderley, Rob & Rafferty, Pauline (1997). Democratic indexing: an approach to the retrieval of fiction. *Information Services & Use* 17 (2-3):101-109.

- Lehtonen, Mikko (2000). Merkityksen maailma: kulttuurisen tekstintutkimuksen lähtökohtia. 4.p. Tampere, Vastapaino.
- Noruzi, A. (2007). Editorial. *Webology*, 4(2), editorial 12. [Luettu 11.1.2009.] Saatavissa: <http://www.webology.ir/2007/v4n2/editorial12.html>
- Rich, Elaine (1979). User modeling via stereotypes. *Cognitive Science* (3):329-354.
- Rich, Elaine (1983). Users Are Individuals: Individualizing User Models. *International J. Man-Machine Studies Journal*, 18, 199-214.
- Saarti, Jarmo (1999). Kaunokirjallisuuden sisällönkuvailun aspektit: kirjastoammattilaisten ja kirjastonkäyttäjien tekemien romaanien tiivistelmien ja asiasanoitusten yhdenmukaisuus. (*Acta Universitatis Ouluesis. B, Humaniora*, 33.) Oulu : Oulun yliopisto. Saatavissa myös: <http://herkules oulu.fi/isbn9514254767>.
- Saracevic, Tefko (1975). Relevance: A Review of and a framework for the thinking on the notion in information science. *Journal of the American Society for Information Science*, 26, (6), 321-343.
- Spiteri, Louise F. (2007). Structure and form of folksonomy tags: the road to the public library catalogue. *Webology*, 4(2), Article 41. [Luettu 11.1.2009.] Saatavissa: <http://www.webology.ir/2007/v4n2/a41.html>