

Monikulttuurisuus, monimerkityksisyys ja teknologioiden kehitys sisällönkuvailun tulevaisuuden haasteina

Jarmo Saarti & Susanna Nykyri & Kaisa Hypén & Kimmo Tuominen & Seija Kulmala

Julkaisujen ja sisältöjen kuvailu on monimutkaistumassa ja sisällönkuvailun rooli on muuttumassa. Sisältöjen kuvailun näkökulmia olisi tarvetta laajentaa ja huomioida käyttäjien tarpeita ja käyttöyhteyksiä. Monikulttuuristen ja monikielisten sanastojen kysyntä kasvaa. Elektroniset aineistot, datan yhdistäminen sekä automaattisen indeksoinnin ja sisällön analyysin uudet mahdollisuudet muuttavat toimintaympäristöä. Pohditaan myös, ketkä sisällönkuvailutyötä tulevaisuudessa tekevät - kirjastot, julkaisijat vai käyttäjät? Mihin suuntaan otetaan seuraavat askeleet?

Oulun yliopiston kirjaston tietopalvelutiimi järjesti 25.8.2010 eloseminaarin sisällönkuvailusta. Seminaarin teemana olivat sisällönkuvailun mahdollisuudet ja haasteet e-kirjan aikakaudella. Päivän aikana pohdittiin sekä sisällönkuvailun nykytilaa että muotoja ja roolia tulevaisuudessa.

Seminaari kokosi lähes 90 osallistujaa, joista suurin osa oli Oulun seudun kirjastoista. Lisäksi paikalla oli runsaasti kirjastolaisia muualta Pohjanmaalta, Kainuusta ja Lapista, yksittäisiä kauempaakin. Suuri osallistujamääräkin kertonee aiheen ajankohtaisuudesta ja tarpeesta keskustella tiedon järjestämisen tulevaisuudesta kirjastoissa.

Ylikirjastonhoitaja Päivi Kytömäen avaaman seminaarin ohjelmassa oli mielenkiintoisia esityksiä, jotka herättivät kuulijoissa runsaasti ajatuksia, kysymyksiä ja keskustelua. Aiheissa liikuttiin perinteisemmän sisällönkuvailun luonteen pohdintoista semanttisen webin tekniikoiden ja automatiikan tarjoamien mahdollisuuksien tarkasteluun, sisällönanalyysistä tiedonlouhintaan.

Sisällönkuvailun kysymykset ovat moninaisia. Millaisesta sisällönkuvailusta asiakas hyötyy eniten? Tunnetummeko toimintaympäristömme ja käyttäjäkuntamme tarpeeksi hyvin? Miten sisällönkuvailu pysyy mukana, kun asiakaskunta kansainvälistyy? Miten käyttäjät (tutkijat, opiskelijat,

lukijat) saadaan osallistumaan kuvailukeskusteluun? Millaiseksi kirjastojen rooli tiedonhallinnassa muotoutuu tulevaisuudessa?

Sisällönkuvailu ohjaa tiedonhakijaa

Sisällönkuvailu muodostaa keskeisen osatehtävän kirjastojen tekemässä luettelointityössä. Tiedonjärjestäminen jaetaan periaatteessa kahteen pääosa-alueeseen, luettelointiin ja sisällönkuvailuun. Edellisessä kuvaillaan pääasiassa dokumentin julkaisualustaan liittyviä аспектеja ja jälkimmäisessä dokumentin välittämään sisältöön liittyviä.

Keskeisimmät sisällönkuvailun välineet ovat olleet luokitus ja indeksointi. Jälkimmäinen jaetaan nykyisin kahtia: ihmisten tekemään ja automaattiseen. Edellisestä on käytetty myös käsitettä intellektuaalinen indeksointi.

Sisällönkuvailun tarkoitus on auttaa tiedonhakijoita löytämään heille tarpeelliset teokset. Yliopistoympäristössä nämä tarpeet liittyvät pääasiassa joko tutkimukseen tai opiskeluun. Erityisesti tutkimukseen liittyvät tarpeet ovat haasteellisia, koska sillä alueella kysymykset voivat olla täysin avoimia ja tiedonhakijan tiedontarpeena voi olla jopa varmistaa se, ettei jostakin aiheesta ole laadittu tutkimuksia.

Tieteen ja teknologian kehitys haasteina

Tiedemaailman nopea kehitys asettaa omat haasteensa erityisesti käytetyille kuvailukielille. Tieteen kehitys on usein niin nopeaa, että asiansanat ja luokituskaavat eivät pysy esimerkiksi käytetyn terminologian kanssa ajan tasalla. Tämän vuoksi viime aikoina automaattinen indeksointi on tullut entistä tärkeämmäksi välineeksi tiedonhaussa.

Automaattinen indeksointi on myös ollut vastaus suurten tekstimassojen hallintaan. Jos suomalaisissa kirjastoissa pitäisi esimerkiksi asiansanoittaa ja luokitaa kaikki maahamme hankittavat tieteelliset sarjajulkaisut ja niiden yksittäiset artikkelit, eivät käytettävissä olevat henkilöstöresurssit riittäisi siihen.

Julkaisujen määrän kasvamisesta ja tietotekniikan kehityksestä on ollut seurauksena myös se, että kirjastot ovat joutuneet pohtimaan sisällönkuvailun ja luetteloinnin roolia kirjastotyössä. Keskeistä tässä harkinnassa on kussakin kirjastossa ja kansallisissa palveluissa miettiä, mitä varten tätä työtä tehdään ja mikä työstä on suomalaisten kirjastojen vastuulla, mitkä palvelut ulkoistamme muiden tekemäksi.

Analyysin ja kustannus-hyötyanalyysin tarve

Näyttäisi siltä, että kirjastot eivät ole aina olleet kovin analyttisiä sisällönkuvailujärjestelmien kehittämisessä. Sisällönkuvailujärjestelmän kehittämisessä ensimmäinen lähtökohta on käyttäjien tarpeiden ja kuvailtavien dokumenttien rakenteiden ja sisältöjen analysointi. Tiedonhaussa voidaan erotella ainakin selailu-, analyttinen haku ja kysymyksiin vastaaminen tiedonhakijoiden käyttäminä perustekniikkoina.

Näihin kuhunkin tulisi tietojärjestelmien pystyä antamaan erilaisia vastauksia ja siten tallennetun metadatan tulisi olla riittävän rikasta, jotta näihin erityyppisiin tiedonhakutilanteisiin ja niissä esitettyihin kysymyksiin löytyisi riittävän hyvät vastaukset. Nykyisten massiivisten viitetie-

tokantojen aikakaudella se, kuinka hyvin tietojärjestelmä ja siihen tehty metadata pystyy suodatamaan jopa kymmenistä miljoonista viitteistä ne kymmenisen relevanttia viitettä kullekin hakijalle, on lisäarvo, jota asiakkaat etsivät.

Toinen keskeinen analyysi on kustannus-hyötyanalyysi. Tämä sen vuoksi, että erilaisia tiedonhakujärjestelmiä on saatavilla runsaasti. Yksi mahdollisuus Suomessa olisi keskittyä kirjastojen väliseen työnjakoon Fennica-aineistoissa ja tehdä suomalaisten julkaisujen sisällönkuvailu mahdollisimman hyvin ja luottaa siihen, että muut maat ja kaupalliset toimijat tekevät oman aineistonsa kuvailun riittävän laadukkaasti.

Elektroniset aineistot ja sovellusten toimintatavat haasteina

Suomalaisten kirjastojen viitetietokannoille näyttää myös olevan ominaista, että niihin ei juuri ole tallennettu kokotekstistä aineistoa, esimerkiksi tiivistelmiä ja teosten sisällysluetteloita. Nykyiset tietojärjestelmät pystyvät käsittelemään tehokkaasti tämäntyyppistä dataa, joka antaa puolestaan uusia mahdollisuuksia tiedonhakuun ja erityisesti ne tukevat selailuhakua.

Samalla näyttää siltä, että suomalaisessa sisällönkuvailuajattelussa fasettisuus tai moniaspektisuus on ollut taka-alalla. Käytännössä tämä näkyy siinä, että dokumenteista on pyritty kuvailemaan vain ja ainoastaan keskeinen asiasisältö – esimerkiksi tutkimuksessa käytetty metodiikka, teoksen lajityyppi jne. on jätetty kuvailematta.

Nykyisessä tietoteknisessä ympäristössä sisällönkuvailija tarvitsee myös tietoteknistä osaamista ja sen hahmottamista, miten tuotettua luettelointidataa kussakin tietojärjestelmässä käsitellään. Syvälinen ja nyansoitu asiansanoitus menee hukkaan, jos tietojärjestelmä indeksoidessaan hävittää tämän datan tiedonhakijan ulottumattomiin.

Hyvänä esimerkkinä on asiansanaketjujen käsittely eri tietokannoissa: lopputulemana näyttää olevan käsien ilmaan nostaminen ja selailuhakua hyvin tukevasta tekniikasta luopuminen vain

sen vuoksi, että nykyiset kirjastojärjestelmät eivät oikein hyvin hallitse asiasanaketjujen käsittelyä.

Monikielinen ja –kulttuurinen ympäristö tuo uusia vaatimuksia

Perinteisesti tesaurus on dokumentaatiokieltä ja se laaditaan tiedon haun ja tallennuksen apuvälineeksi. Tesauruksessa esiintyvillä termeillä löytyy luonnollisesta kielestä useita vastineita. Myös sanaston esitysmuoto on keinotekoinen - termien kontekstina ovat hierarkia- ja assosiaatiosuhteet.

Tesaurusmuodon avulla pyritään auttamaan tiedonhakijaa laajentamaan tai tarkentamaan hakuaan. Jokainen tiedonhaku tietokannassa käsittelee ainakin viisi erilaista kieltä: kirjoittajan, sisällönkuvailijan, synteettisen rakenteen, tiedonhakijan ja tiedonstrategian (Buckland 1999). Nämä kaikki edustavat toisistaan poikkeavaa diskurssia.

Monikielinen yhteiskuntatieteellinen tesaurus on usein vain yhtä tai harvaa yhteiskuntaa heijastava, eikä siten palvele tarkoituseriään monikulttuurisessa ympäristössä (Hudón 1997). Nykypäivän globaalistuvassa informaatioyhteiskunnassa englannin kielen asema on vahva.

Mihin käännöstä tarvitaan?

Kääntämisen merkitys on korostunut valtavasti esimerkiksi Internetin yleistymisen ja EU:n myötä, jotka osaltaan luovat tarvetta kehittää toimivampia tiedonhallinnan apuvälineitä. Ei ole kuitenkaan aina selvää, että mitä käännöksellä tavoitellaan – onko tavoite kotouttaa vastine kohdekulttuuriin vai heijastella lähtökulttuuria vai kenties pyrkiä kansainväliseen, mahdollisimman neutraaliin sisältöön.

Euroopassa on laadittu vuosituhannen vaihteessa lukuisia monikielisiä tesauksia niin kansalliseen kuin kansainväliseen käyttöön. Niiden laadinnan perusteita ja periaatteita ei ole kuitenkaan systemaattisesti tutkittu.

Tesauruksille asetetut odotukset ja käytännöt ovat varsin kirjavat, samaten yhteiskuntatieteilijöiden itse suorittaman dokumenttiensa asiasanoituksen motiivit. Tutkijat pyrkivät jopa kek-

simään mahdollisimman uniikkeja asiasanoja ja peräti uudissanoja – jotta heidän tutkimuksensa profiloituisi ainutlaatuisena. Silloinhan ei ole ajateltu tulevaa tiedonhakutilannetta, ja näkökulma on kaukana indeksoijan näkökulmasta. (Nykyri 2010)

Standardit tukevat käännöstyötä

Erittäin yleinen tapa laatia monikielinen tesaurus on kääntää jo olemassa oleva yksikielinen tesaurus. Toinen tapa on yhdistää jo olemassa olevia tesauksia, ja kolmas, harvinaisin, on laatia uusi tesaurus ilman aiempia esikuvia. Tesaurusten laadintaa ohjaavat kansalliset ja kansainväliset standardit (esim. SFS 5471 1988 ja ISO 5964-1985), joita par'aikaa päivitetään.

Ohjeissa suositellaan kielten tasa-arvoa, mutta käytännössä englanti on yleinen valtakieli – käännöskorpuksen kielenä, ja /tai yhteisenä työskentelykielenä. Käytännössä suomalaiset osapuolet joutuvat usein mukautumaan jo laadittuun korpuksen, ja tekemään parhaansa – kääntämällä.

Tesaurusten kääntämisessä ongelmia aiheuttavat niin kieli, kulttuuri, kuin tesaurusten erityinen diskurssi. Tesauksukset ovat kulttuurinen tuote, ja heijastavat syntyhistoriaansa ja –paikkaansa. Jos ajatellaan vaikka termiä nationalismi, on sen konteksti eri tesauksissa varsin erilainen.

Suomessa tarvitaan rinnakkaisina termeinä mm. panslavismia, slavofileja ja fennomaniaa, mutta brittiläinen indeksoija kaipaa muunlaisia termejä. Kansainvälistyvässä sisällönkuvailun maailmassa tesaurusten laatijat ja päivittäjät törmäävät lukuisiin haastaviin kysymyksiin. - Mikä on selkeästi kulttuurisidonnaisten termien asema monikielisessä tesauruksessa? Pitäisikö ne säilyttää? Entä miten kääntää ne?

Tesaurusten kääntämisestä puuttuu strategia

Tesaurusten laadintaa käsittelevissä oppaissa ja standardeissa suhdetta monikulttuurisuuteen ei käsitellä; ja kääntäminen on esitetty varsin mekaanisena ja lingvistisenä prosessina. Siinä on kui-

tenkin etenkin yhteiskuntatieteellisen asiasanaston kohdalla kyse myös laajemmista asioista – esim. miten ymmärretään kääntämisen ja käännöksen tavoite ja sanojen kantamat ideologiat.

Onko asiasanasto oma kielenkäyttönsä saareke, jolle voidaan laatia omat säännöt ja jossa sanat voidaan riisua sivumerkityksistään? Entä haluammeko monikielisen asiasanaston eri kieliverisoiden olevan identtisiä, vai voisivatko erikieliset versiot täydentää toisiaan?

Informaatiotutkimuksella ei ole tarjota riittävästi käsitteellisiä työkaluja monikielisten ja -kulttuuristen tesaurosten laadintaan. Monikielisen tesauroksen laatija hyötyisi käännöstieteen käsitteistä, etenkin dynaamisen ekvivalenssin ja kommunikatiivisten käännösteorioiden saralta. Niiden ottaminen mukaan analyysiin auttaisi selvittämään tesauroksen laadinnan tavoitteita ja strategioita.

Tekstin aikaansaama reaktio käännöksen perustana

Eugene A. Nida kehitti 1960-luvulla teorian dynaamisesta ekvivalenssista, joka on ensimmäinen kommunikatiivinen käännösteoria. 'Dynaaminen' viittaa käsitykseen, että käännöksen tulisi saada aikaan sama reaktio kohdeyleisössä kuin lähtöteksti tuottaa alkuperäisessä kontekstissään. Nida (1989, 95) on myöhemmin määritellyt kääntämisen "reproducing in the receptor language the closest natural equivalent of the message of the source language, first in terms of meaning and second in terms of style."

Funktionaalisessa ekvivalenssiteoriassa puolestaan perustana on ajatus siitä, että lähtötekstin funktion ajatellaan olevan sama tai samankaltainen kohdetekstin (eli käännöksen) kanssa. Käännöksen funktio mukautuu siis lähtötekstin funktioon. Tällöin ei painoteta vastaanottajareaktiota, vaan tekstien funktiota.

Funktionaaliin käännösteorioihin lukeutuu skopos-teoria. Se – kuten muut funktionaaliset ekvivalenssiteoriat – ovatkin hyvin lähellä Nidan dynaamista ekvivalenssiteoriaa ja nimityk-

siä käytetäänkin joskus synonyymisina. Skopos-teoria korostaa käännöksen tarkoitusta, funktiota, joka on tärkeintä ottaa käännöksessä huomioon. Teorian mukaan onkin tärkeämpää täyttää käännöksen funktio kuin kääntää tietyllä tyylillä.

Kotouttaminen vai vieraannuttaminen strategiana?

Tesauroksen laadintaa ohjaavat standardit ja metodiset oppaat keskittyvät kuitenkin varsin korostuneesti hyvää asiasanaa määritellössään sen kielellisiin piirteisiin ja kulttuurisia tekijöitä ei juuri käsitellä. Monikielistä tesaurosta laadittaessa ja käytettäessä toimitaan kuitenkin yleensä nimenomaan monikulttuurisessa ympäristössä.

Käännöstieteessä tunnustetaan perinteisesti kaksi vastakkaista käännösstrategiaa, vieraannuttaminen (*foreignisation*) ja kotouttaminen (*domestication*). Kotouttamalla lähtökielen ja –kulttuurin käsite tuodaan lähemmäksi kohdekielen ja kulttuurin puhujaa. (Jos olisi kyse vaikka machoista kansanhuveista, niin härkätaistelu voisi silloin saada suomenkielisenä vastineena eukonkannon.)

Vieraannuttaminen on strategiana päinvastainen, ja siinä tulokielen ja kulttuurin edustaja pistetään kokemaan alkukielen ja kulttuurin eksotiikka. Niiden lisäksi tesauroksiin liittyen on kolmas selkeä päästrategia, kansainvälistäminen, jossa kulttuuriset erikoispiirteet häivytetään. EU-yhteyksissä on puolestaan löydetty existential equivalence –strategia, eli käännösten ei edes odoteta olevan kommunikatiivisia, riittää että ne vaan ovat.

Monimerkisyys otettava huomioon

Sanatasolla on tärkeää myös pohtia ongelmia niin semantiikan kuin pragmatiikan kannalta. Sanoilla on paitsi ilmimerkitys, denotaatio, niin myös sivumerkityksensä, konnotaatio. Tutkimuksen (Nykyri 2010) mukaan sanojen sivumerkitys voi olla yhteiskuntatieteilijöille ratkaisevampi kuin ilmimerkitys, ja silloin yksinomaan

sanojen denotaatiosalla operoiva tesaurus on tiedonhakijalle käyttökelpoton.

Tärkeää on siis pohtia, että mitä monikielisel-
lä tesauruksella tarkoitetaan ja tavoitellaan, sa-
moin kuin monikielisel-
lä sisällönkuvailulla. On-
ko tesaurus monikielinen, monikulttuurinen ja/
tai kansainvälinen, ja onko monikielinen sisäl-
lönkuvailu eri kieliversioissaan identtistä vai en-
nemminkin toisiaan täydentävää.

Lisäksi on syytä pohtia asiasanoituksen diskursii-
visia piirteitä eli sitä, että miten se poikkeaa muusta
käyttäjien kielenkäytöstä, ja millainen poikkeaa-
ma on vielä sallittua sanaston käytettävyyden kär-
simättä. Tesauukset ovat pääfunktioltaan norma-
tiivisia, ohjaavia, mutta niiden deskriptiivisyyden
lisääminen toisi ne lähemmäksi niiden käyttäjiä.

Ihmiskunnan muisti?

Yleisen kirjaston kirjastonhoitajan oletetaan tun-
tevan kirjallisuutta ja muuta aineistoa, jota kirjas-
ton hyllyiltä löytyy. Hänen oletetaan myös osaa-
van esitellä ja suositella erilaisiin tilanteisiin so-
pivaa kirjallisuutta.

Kun katsoo oheista, Turun kaupunginkirjas-
ton kaunokirjallisuusosaston kokoelmasta otet-
tua kuvaa, tehtävän haasteellisuudesta saadaan
jonkinlainen käsitys. Miten ottaa haltuun laajat
kokonaisuudet, poimia sieltä kulloinkin kiinnostavat
sisällöt ja kuinka siitä selviytyy vastavalmis-

tunut kirjastoalan ammattilainen? Tukeeko kir-
jaston tietokanta ja millä tavalla kirjastoamat-
tilaista tässä tehtävässä?

Usein mainitaan, että kirjastojärjestelmät ovat
etenkin vanhemman aineiston osalta lähinnä va-
rastoluetteloita. Tieto kirjoista ja aineistoista löy-
tyy verkon muista palveluista, ja kirjaston tieto-
kantaan tullaan tarkistamaan aineiston saatavuus.
Systemaattista indeksointia on tehty 1990-luvun
puolivälistä lähtien, sitä vanhempaa aineistoa on
asiasanoitettu satunnaisesti. Vain uudempi ai-
neisto siis tarttuu asiasanoilla tehtäviin hakuihin.

Onko sisällönkuvailun rooli kirjastojärjestelmissä hukassa?

Kirjastojen tietojärjestelmien mahdollisuuksis-
ta ja siitä, millaisia niiden pitäisi tulevaisuudessa
olla, ei juuri käydä julkista keskustelua. Ei poh-
dita sitä, mitkä kirjastojärjestelmän roolit ja teh-
tävät jatkossa ovat.

Onko kirjastoilla velvollisuus toimia ”ihmis-
kunnan muistina” myös verkossa, onko kirjastoil-
la toisaalta varaa pitää hyllyssään aineistoa, jon-
ka sisältöä ei tunneta ja kuvailla? Riittääkö, että
käyttäjä voi jatkossakin näiden järjestelmien avul-
la paikantaa aineiston, vai ovatko kirjastot aktiivi-
sia toimijoita, jotka tuovat oman aineistoihin liit-
tyvän asiantuntemuksensa esiin esimerkiksi laa-
jempina kuvailuina ja suositteiluina?

E-kirjan ja muiden sähköisessä muodossa olevien aineistojen tulo ei millään tavalla vähennä laadukkaana metatiedon tarvetta. Päinvastoin, helpompi pääsy sähköisiin aineistoihin lisää entisestään tarvetta tietää, mitä sisältöjä missäkin resurssissa on.

Esineen kuvailusta sisältöjen kuvailuun

Sama sisältö ilmestyy yhä useammin eri muodoissa: kovakantisena kirjana, pokkarina, isotekstisenä laitoksena, äänikirjana, e-kirjana. Kiinnostavaa on myös se, millaisia suhteita ja yhteyksiä sisältöjen välillä on, mitkä teokset liittyvät toisiinsa ja millaisen verkoston ne muodostavat.

Näiden aspektien esittäminen edellyttää painopisteen siirtämistä esineen kuvailusta sisältöjen kuvailuun. Välineitä tähän on luotu jo pitkään funktionaalisen luetteloinnin (FRBR) ja sille pohjautuvan RDA-mallin kehittämisen myötä. Niissä tunnistetaan teoksen eri tasot (teos - ekspressio - manifestaatio - kappale), joita kuvaillaan erikseen.

Kun tätä mallia sovelletaan täysimääräisesti, voidaan luettelointityötä tehostaa. Kunkin teoksen sisältö kuvaillaan vain kerran ja sen eri ilmenemismuodot (käännökset, painokset...) liitetään tähän sisällönkuvailuun. Näin voidaan kertoa esimerkiksi, että *Mitta mitasta* ja *Verta verrasta* ovat molemmat Shakespearen *Measure for measure* -näytelmän suomennoksia.

Kirjasampo kokeilee funktionaalista luettelointia

Funktionaalisen luetteloinnin käytännön sovelluksia ei vielä kovin paljon ole. Yksi niistä on toteutettu Kirjasampo-hankkeessa, jossa rakennetaan kaunokirjallisuuteen keskittyvää verkkopalvelua semanttisen webin välineillä.

Siinä käytössä olevaan tietomalliin sisältyvät ns. abstrakti teos, jossa kuvaillaan teoksen sen kaikille manifestaatioille yhteistä sisältöä ja ns. fyysinen teos, jossa kuvataan abstraktin eri ilmentymiä, kuten käännöksiä. Abstraktia teosta voi ku-

vailla monin tavoin, kuten ontologiakäsittein ja tiivistelmin, lisäksi voidaan linkittää verkossa olevaa teokseen liittyvää aineistoa (kriitikkejä, esitelyjä, elokuvaversioita).

Kirjasammon pohjaksi on konvertoitu Helmetin aikuisten kaunokirjallisuus marc-muodosta semanttisessa webissä käytössä olevaan RDF-muotoon. Ohjelmallisen konversion jälkeen tätä dataa on muokattu manuaalisesti monin tavoin ja samalla on testattu sitä, miten kirjastojen tuottamaa peruskuvailua voidaan rikastaa.

Tietokantaan on mm. tallennettu kaikki Aleksis Kiven runot kokoteksteinä ja linkitetty ne niihin kokoelmiin, mistä ne löytyvät painettuina, syötetty yli 25 000 novellin tiedot ja linkitetty novellit kokoelmiin ja käyty systemaattisesti läpi 1970-luvulla ilmestynyttä kaunokirjallisuutta. Näin on voitu testata sitä, miten uusilla tekniikoilla tehty takautuva sisällönkuvailu onnistuu ja millaisia resurssia se vaatii.

Osaksi merkitysten verkkoa

Semanttisen webin tavoitteena on rakentaa verkkoon metadatarakennetta, jonka avulla verkon tietojen tehokkaampi ja monipuolisempi koneellinen käsittely on mahdollista. Tässä ontologioilla on keskeinen rooli.

Ontologiat ovat verkkoon syntyneitä ja sen ehdoilla kehitettyjä, asiansastoille pohjautuvia koneluettaavia sisällönkuvailun välineitä. Ontologioiden avulla verkon sisältöjen linkittäminen ja automaattiset suosittelut ovat mahdollisia.

Kirjasammossa sisällönkuvailuun käytetään useita eri ontologioita, jotka on integroitu semanttiseen tiedontalennusalueeseen (SAHA). Yksi Kirjasammon tavoite on, että koneellinen ”äly” kytkee teoksia yhteen tavalla, johon ei kenenkään yksittäisen lukijan lukeneisuus tai muisti riitä.

Koneäly ei tuota kuvailuja

Erilaisten päättelymekanismien avulla voidaan annettujen ontologiakäsitteiden pohjalta automaattisesti nostaa esiin teoksia, jotka ovat jollakin tavoin toistensa kaltaisia. Mitä useampia teoksia

on kuvailtu ja mitä enemmän niille on annettu kuvailevia käsitteitä, sitä rikkaampia ja monipuolisempia suosituksia voidaan näiden päättelysääntöjen avulla tuottaa.

”Koneäly” ei itsessään tuota kuvailuja, vaan tarvitaan lukijoiden massaa kuvailemaan teoksia monipuolisesti. Tähän tarvitaan kirjastoalan ammattilaisia ja lukijoita.

Toinen keskeinen tavoite on, että sisällönkuvailuja voidaan linkittää muihin palveluihin ja toisaalta tuoda muista verkkopalveluista, esimerkiksi Wikipediasta, sisältöjä Kirjasampoon. Tätä on testattu Kulttuurisammossa (www.kulttuurisampo.fi), jossa ”semanttisen” Kalevalan yksittäisten runojen yhteyteen tuodaan niiden teemaan liittyviä kaunokirjallisia teoksia, valokuvia ja maalauksia eri muistiorganisaatioiden tietojärjestelmistä.

Open Linked Data - datan yhdistely muuttaa näkökulmaa

Koko verkon tasolla datan monipuolisempaa yhdistelyä ja käyttöä tutkitaan erilaisissa Open Linked Data -hankkeissa (<http://linkeddata.org/>). Kirjastomaailmasta tässä on mukana ruotsalaisten kansallisbibliografia Libris (<http://libris.kb.se/>). Kirjasammon osalta tavoitteiden toteutumista päästään arvioimaan syksyn ja talven 2010-11 aikana, kun sen käyttöliittymää koodataan, hakuja ja muuta toiminnallisuutta testataan.

Aineistoja esitellään etenkin yleisten kirjastojen kotisivuilla ja blogeissa, näyttelyissä ja valikoi- maluetteloissa, miksi siis ei myös kirjastojärjestelmissä. Bibliografisen datan lisäksi pitäisi tallentaa myös arvioita, kuvailuja ja suositteluja.

Tarvittaisiin entistä monipuolisempaa metadattaa, myös vanhemmasta aineistosta sekä muiden verkkopalveluiden sisältöjen hyödyntämistä. Se edellyttäisi kirjastojen tietojärjestelmien idean ja funktioiden uudelleen määrittelyä, uudenlaista tavoitteenasettelua, työn organisointia sekä uudenlaisten tekniikoiden soveltamista.

Näistä elementeistä voisi vähitellen verkkoon rakentua universaali ”ihmiskunnan muisti”, joka heijastelee kirjastojen kokoelmien monipuoli-

sia sisältöjä ja joka toimii vuorovaikutuksessa verkon muiden sisältöjen kanssa.

Mikä on manuaalisen indeksoinnin rooli?

Sisällönkuvailun, luetteloinnin ja kirjastoluetteloiden tulevaisuutta on viime vuosina pohdittu mm. *Kongressin kirjaston*, *OCLC:n*, *Australian kansalliskirjaston* ja brittiläisen *RIN:n* piirissä. Keskustelu on ollut kriittistä: on mm. kysytty, muodostuvatko nykyiset luetteloinnin ja sisällönkuvailun käytännöt liian kalliiksi hyötyynsä nähden. Kirjastoluetteloiden vertaaminen vaikkapa Amazonin tai Googlen palveluihin ei ole tuottanut kirjastoammattilaisten kannalta järin mi- rittelleviä tuloksia.

Kun etenkin tieteellinen kirjallisuus siirtyy verkkoon ja kokotekstihaun piiriin, on syytä kysyä, mikä on manuaalisen indeksoinnin rooli. Miten se muuttuu sellaisessa mahdollisessa tulevaisuudessa, jossa kirjaston näyttöluettelokin toimisi googlemaisesti ja sosiaalisen median mahdollisuuksia hyödyntäen?

Käykö tulevaisuudessa niin, että näyttöluettelot lakkaavat olemasta viitetietokantoja? Entä jos meidän lintojemme ja artojemme hakuominaisuudet rakentuisivatkin vastaisuudessa automaattisen kokoteksti-indeksoinnin luomien mahdollisuuksien varaan? Vaikka automaattisen ja manuaalisen indeksoinnin kannattavuus tiedonhaun tehokkuuden kannalta on kestopuheenaihe informaatiotutkimuksessa, voitaneen kiistatta todeta, että automaattinen indeksointi on huomattavasti halvempaa ja nopeampaa kuin ihmisen tekemä käsi- ja aivotyö.

Automaattisen indeksoinnin ja sisällön analysoinnin tulevaisuudenkuva

Jos automaattinen indeksointi on mahdollista, sitä käytetään myös kirjamuotoisen aineiston kohdalla vähintäänkin manuaalista indeksointia täydentämässä.

Kokotekstin indeksoinnin ja haun menetelmät

sekä muut kokotekstiä käsittelevät teknologiat ovat jatkuvasti kiihkeän kehitystyön kohteina.

Kieliteknologiset ja tiedon louhintaan perustuvat teknologiat mahdollistavat vastaisuudessa sen, että automaattiset sovellukset käyttävät tieteellistä kirjallisuutta tutkimusaineistonaan. Koneilla on kapasiteettia löytää tieteellisestä kirjallisuudesta sellaisiakin yhteyksiä, joita ihmisäivot eivät kykene havaitsemaan.

Kirjallisuuteen perustuvien löydösten generoinnin menetelmistä (*literature-based discovery, LBD*) saattaa olla vastaisuudessa arvaamatonta hyötyä esimerkiksi uusien syöpälääkkeiden kehittäessä. Automatiikan rooli tulee kehittymään ja kasvamaan tulevaisuudessa, jos sovellukset pystyvät automaattisen indeksoinnin lisäksi myös sisällön entistä kehittyneempään analysointiin.

Informaation siirrosta keskusteluihin

Sanana *sisällönkuvailu* kantaa mukanaan tietynlaisia oletuksia siitä, minkälaisesta toiminnasta on kyse. Taustalla on ajatus siitä, että olisi olemassa tallennusmediasta riippumaton informaation sisältö, joka voidaan siirtää erilaisten teknisten ja sosiaalisten mekanismien välityksellä informaation tarvitsijoille. Tätä oletusta, informaation siirtomallia, on takavuosina kriittisesti käsitelty mm. Phil Agren, Ronald E. Dayn ja Brenda Dervinin tutkimuksissa.

Vaihtoehtoisen näkökulman mukaan sisällönkuvailussa on tai tulisi olla kyse pikemminkin keskustelusta ja keskustelullisten tai argumentatiivisten näkökulmien mallintamisesta tiedonhakijan kannalta havainnollisesti. Nykyään harjoitettua sisällönkuvailua ei tulisi hylätä, mutta sitä tulisi täydentää tieteellisten keskustelukontekstien kuvailemisella. Tämä täydentävä kuvailutyö voidaan sosiaalisen median periaatteiden mukaisesti säilyttää pitkälti tiedeyhteisön edustajien (tutkijoiden ja opiskelijoiden) harteille.

Informaation siirtomalli on johtanut substansiivisiin perustuviin sisällönkuvailun käytäntöihin. Dokumentaatiokiellemme ovat substansiivi-

rakenteita, joiden avulla ajattelemme voivamme vangita kuvailtavien tietoaineistojen käsitteellisen ydinsisällön.

Tieteellisten keskustelukontekstien mallit

Tieteellisten keskustelun mallintamiseen perustuvissa järjestelmissä, jotka tällä hetkellä ovat lähinnä luonnosasteella, substansiivikuvailua on haluttu täydentää tieteellisen keskustelun kulkua mallintavilla verbeillä. Keskusteluhan on pitkälti sanoilla tekemistä, kuten kielifilosofiaa tuntevat tietävät. Niinpä tieteellisen keskustelun visuaalisen mallintamiseen ja kuvailuun sopivat nimienomaan verbit.

Miksi sisällönkuvailua tulisi täydentää tieteellisen keskustelun kulkua mallintavilla verbinäkyillä? Informaation siirtomallin kriitikoiden mukaan tiedeyhteisön edustajat (tutkijat ja pidemmälle ehtineet opiskelijat) eivät hahmota tiedollista ympäristöään informaation siirtomallista käsin, vaan he ovat kiinnostuneita siitä, minkä siirtomalli hygieenisesti kätkee.

Tutkijaa kiehtovat tiedon muodostukseen liittyvät kiistat ja neuvottelut, teoriat ja paradigmat, erilaisten koulukuntien ja keskusteluyhteisöjen väliset tulkintayksimielisyydet ja -ristiriidat.

Sisältöjen monimutkaiset yhteydet

Tutkijan olennainen työympäristö on tekstien välinen monimutkainen keskustelullinen maasto, jonka kartoittaminen on olennaista hänen työnsä kannalta. Tekstien ilmissisällön lisäksi tutkijaa siis kiinnostavat tekstien julkilausutut ja piilevät suhteet muihin julkaisuihin, käsitteisiin ja tutkimusaineistoihin. Sisältöä jopa olennaisempaa on tutkijan kannalta tekstien keskustelullinen konteksti.

Siirtomalli ei näe sitä, minkä tutkija tietää: yksittäiset dokumentit ovat aina osallisia jossakin suuremmissa keskustelussa, kannanottoja argumentaation kohteena olevaan asiaan. Tiedejulkaisut sijoittuvat joko jonkin keskustelutradition tai paradigman sisälle tai eri traditioiden väli-

siin solmukohtiin, rajapintoihin, joissa innovaatiot yleensä tapahtuvat.

Traditionaaliset kirjastoluettelot kätkevät tahattomasti tekstien kontekstit ja monitulkintaisuudet, minkä vuoksi tekstien välinen maailma ei hahmotu niiden avulla. Kirjastoluettelot voivat auttaa tiedon paikallistamisessa, mutta eivät tiedon muodostuksellisten prosessien kartoittamisessa.

Käyttäjät sisällönkuvailijoina

Vastaisuudessa automaattisen indeksoinnin ja sisällön analyysin menetelmät tulevat kehittymään, minkä vuoksi niitä käytetään entistä enemmän traditionaalisen sisällönkuvailun korvikkeina tai jopa syrjäyttäjinä. Manuaalista sisällönkuvailua tulevat tekemään paitsi kirjastoammattilaiset, myös tieteellisten dokumenttien alkuperäiset tuottajat.

Uudenlaiset kirjastoluettelot voivat tarjota tekstien kirjoittajille ja käyttäjille helppokäyttöisen metadatan tallennusalueen, jonka välityksellä he voivat määritellä tekstinsä tieteellisen kontribuution, suhteet järjestelmän jo aiemmin sisältämiin dokumentteihin sekä alansa keskeiseen käsitteistöön.

Tallennusalueen vastapainona järjestelmän käyttäjät voivat hyödyntää monipuolisia tiedonhaku- ja visualisointimahdollisuuksia. Tieteellisen keskustelun substantiivi- ja verbipohjainen mallintaminen auttaa paitsi tiedonhaussa, myös pyrittäessä ymmärtämään tietyn dokumentin tai dokumenttiryppään laajempaa intellektuaalista kontekstia.

Tavoitteena ei ole pelkkä tiedonhaku, vaan hakukäyttöliittymä voi toimia myös tekstien välisen maailman hahmottamisen välineenä. Käsitteellisen tällaisen käyttöliittymän toiminnasta voi saada esimerkiksi tutkailemalla brittiläisen Open Universityn piirissä kehitetyn *Cohere*-järjestelmän prototyyppiä, joka on avoimesti käytettävissä verkossa.

Monitulkintaisuus on edelleen ongelma

Vaikka käytössämme olisi automaattisia luokitus-

menetelmiä sekä muita teknisesti kehittyneitä indeksoinnin ja tiedon esittämisen työkaluja, tieteellisiä keskusteluja luonnehtivan monitulkintaisuuden hahmottaminen edellyttää inhimillistä arviointia ja harkintaa.


Tämä johtuu etenkin siitä, että keskustelua luonnehtivat jännitteet voivat olla piileviä. Tietoväite on usein retorisesti konstruoitu mitätöimään tai haastamaan sen esittämistilanteessa julkilausumatonta vaihtoehtoista tulkintaa, kuten brittiläiset sosiaalipsykologit Michael Billig, Jonathan Potter ja John Shotter ovat havainneet.

Koska argumentaatiokonteksti ei käy ilmi suoraan tekstistä, tarvitaan ihmisälyä sen aistimiseen: automatiikka ei todennäköisesti vielä pitkään aikaan kykene tällaista keskustelukontekstissa piilevää tulkinnallista tietoa tavoittamaan.

Argumentaatiokontekstin automaattisen hahmottamisen vaikeus saattaa vastaisuudessa olla tärkein perustelu sille, että manuaalista indeksointia edelleen tarvitaan kokotekstin käsittelyyn perustuvan automatiikan rinnalla. Kone- ja ihmisälyn vastakkainasettelun sijaan automaattinen ja manuaalinen sisällönkuvailu on vastaisuudessaakin syytä nähdä toisiaan täydentävinä tietoresurssien kuvaamisen keinoina.

Lopuksi

Tiedon organisointi on murroksessa. Automaattiset sekä sosiaaliset tavat järjestää ja kuvailla dokumentteja ovat haastamassa perinteiset, käsityölähtöiset luetteloinnin ja sisällönkuvailun menetelmät. Keskeisin haaste kirjastoille on järkevän yhteistyön ja siihen liittyvän työnjaon sopiminen.

Kirjastojen tietojärjestelmien tekniikoita ja toiminnallisuuksia pitää kehittää käyttäjälähtöisemmiksi nopealla aikataululla. Meillä tulee olla rohkeutta myös luopua siitä työstä, jonka jo joku muu tietoverkkojen yhteisessä maailmassa tekee. Vain tällä tavalla voimme vapauttaa resursseja kirjastojen sisällä keskeisimpään tehtävään: asiakkaiden palveluun. 

Lähteitä:

Buckland, Michael (1999). Vocabulary As A Central Concept In Library And Information Science. Preprint of paper published as "Vocabulary as a Central Concept in Library and Information Science" I: Digital Libraries: Interdisciplinary Concepts, Challenges, and Opportunities. Proceedings of the Third International Conference on Conceptions of Library and Information Science (COLIS3, Dubrovnik, Croatia, 23-26 May 1999. Ed. by T. Arpanac et al. Zagreb: Lokve, s. 3-12

Hudón, Michele 1997. Multilingual thesaurus construction: integrating the views of different cultures in one gateway to knowledge and concepts. Information Services & Use, Vol. 17 Issue 2/3, 1997, p. 111, 13 pp.

ISO 5964-1985 = International Standards Organization (ISO). Guidelines for the Establishment and Development of Multilingual Thesauri. 1st ed. Geneva: ISO, 1985 (ISO 5964-1985-02-15).]

Nida, Eugene A. 1989. Science of translation. Teoksessa: Chesterman, Andrew (toim.) 1989: Readings in translation Theory, s. 80-98. Helsinki: Finn Lectura.

Nykyri, Susanna 2010. Equivalence and Translation Strategies in Multilingual Thesaurus Construction. Doctoral thesis. Åbo, Åbo Akademi University Press.
SFS 5471 1988-10-17, 1 p., 10 s. [5] Guidelines for the establishment and maintenance of Finnish language thesauri. Suomenkielisen tesauruksen laatimis- ja ylläpito-ohjeet. Suomen Standardisoimisliitto [The Finnish Standards Association], SFS.

Lisätietoja:

Seminaarin ohjelma ja esitykset verkossa osoitteessa: <http://www.kirjasto.oulu.fi/index.php?id=1384>

Tietoa kirjoittajista:

*Jarmo Saarti, kirjastonjohtaja
Itä-Suomen yliopiston kirjasto
e-mail: jarmo.saarti@uef.fi*

*Susanna Nykyri, tutkija
Åbo Akademi, Ämnet informationsförvaltning
e-mail: susanna.nykyri@abo.fi*

*Kaisa Hypén, va. palvelupäällikkö
Turun kaupunginkirjasto
e-mail: kaisa.hypen@turku.fi*

*Kimmo Tuominen, kirjastonjohtaja
Jyväskylän yliopiston kirjasto
e-mail: kimmo.k.tuominen@jyu.fi*

*Seija Kulmala, informaattikko
Oulun yliopiston kirjasto
email: seija.kulmala@oulu.fi*