

Kaisa Hypén

Kaunokirjallisuuden sisällöt ja kontekstit:

Kaunokirjallisuuden verkkopalvelu semanttisen webin sovelluksena

Kaisa Hypén: Kaunokirjallisuuden sisällöt ja kontekstit: Kaunokirjallisuuden verkkopalvelu semanttisen webin sovelluksena. [Contents and contexts of fiction: Finnish fiction web service as a semantic application.] Informaatiotutkimus 28(1), 2009.

Turku City Library is leading a project, which aim is to make fiction collections, even the older books, or the “long tail” of literature, more accessible for library users and librarians. The main task is to create a new web service, where fiction books can be introduced in many ways. The concept of the web service is based on a broad model for a search and retrieval system for fiction, provided by Jarmo Saarti in his dissertation in 1999. Modern web technologies are used to improve this model. Tools of the semantic web, like ontologies and mash ups, make the content description of fictional works more precise and help to show the complex cultural historical contexts of literature, also the information about the authors. The user interface will be built with Drupal open source system which gives users many possibilities to reflect their views on books. Finally, the service allows the librarians share their “tacit knowledge” for instance by showing the intertextual references between books. The most important partners in the project are the editorial staff of the libraries.fi web service and The Semantic Computing Research Group at the Helsinki University of Technology. Service will be available in the web in 2009.

*Address: Kaisa Hypén, Turku City Library, Linnankatu 2, 20100 Turku, Finland.
E-mail: kaisa.hypen@turku.fi*

Turun kaupunginkirjastossa on meneillään Kaunokirjallisuuden verkkopalvelu -projekti, jonka tavoitteena on rakentaa verkkoon yleisten kirjastojen yhteinen kaunokirjallisuuteen liittyvä palvelu. Siinä yhdistyvät kirjastojen kaunokirjallisuuskokoelmien sisällöt, kirjastoammattilaisten kaunokirjallisuuteen liittyvä tietämys ja käyttäjien panos.

Projektin keskeiset toimijat ovat palvelun sisällöllisestä kehittämisestä vastaava Turun kaupunginkirjasto ja tekniikasta vastaava Kirjastot.fi -toimitus. Hanketta rahoittaa opetusministeriö ja se on osa liikenne- ja viestintäministeriön Arjan tietoyhteiskunta -

toimenpideohjelmaa (Arjen tietoyhteiskunta). Projekti käynnistyi kesällä 2007 esiselvityksellä, palvelun pitäisi olla verkossa käytettävissä vuoden 2009 aikana.

Projekti on toiminut eräänlaisena laboratoriona ja työn edetessä on otettu kantaa uusiin verkkotekniikoihin ja työmuotoihin – semanttinen web, ontologiat, kirjastojen verkkonäkyvyys, verkostomainen työskentelytapa, long tail -ilmiö, open source, kirjasto 2.0 – ja pohdittu sitä, miten niitä voitaisiin soveltaa Kaunokirjallisuuden verkkopalvelussa¹. Tässä artikkelissa käyn soveltajan näkökulmasta läpi sitä, miten edellä mainittuja tekniikoita ja ajattelumalleja on otettu

palvelussa huomioon. Yritän samalla antaa kokonaiskuvan siitä, missä vaiheessa hanke on nyt.²

Kirjastojen on tärkeää testata uutta tietotekniikkaa. Siksi verkkopalvelun ydin, teosrekisteri, toteutetaan semanttisen webin ratkaisulla. Semanttisilla sovelluksilla toteutetuissa palveluissa tietoa voidaan prosessoida eri tavoin kuin perinteisissä kirjastojärjestelmissä. Esimerkiksi tiedon ”kontekstualisoiminen” – asettaminen eri yhteyksiin, rikastaminen semanttisten verkkopalveluiden muilla sisällöillä – on näissä sovelluksissa mahdollista.

Semanttisissa sovelluksissa sisällönkuvailuun käytetään ontologioita. Tällä tavoin syntyvä metadata on myös tietokoneiden ”ymmärtämässä” muodossa, joten erilaiset ohjelmallisesti tuotettavat semanttiset suosittelut ja automaattiset linkitykset ovat mahdollisia. (Hyvönen 2008, 12.) Webissä kulttuurisiin ilmiöihin liittyvät kokemukset syntyvät paitsi itse kohteista, myös niiden välisistä kiinnostavista yhteyksistä ja niistä syntyvistä temaattisista kokonaisuuksista.

Näistä yhteyksistä rakentuu semanttinen verkko, jossa tieto esitetään osana laajempia kokonaisuuksia. (emt, 15; Hyvönen 2007, 82.) Teknillisessä korkeakoulussa Semanttisen laskennan tutkimusyksikössä on kehitetty monialaisesti semanttisen webin tekniikoille perustuvia tietojärjestelmiä ja verkkopalveluita, joissa näitä ominaisuuksia on testattu (Museosuomi; Terve-suomi; Kulttuurisampo).

Apuneuvoja arjen sankareille

Projektin lähtökohdat ovat yleisten kirjastojen arjessa. Kaunokirjallisuus on keskeinen osa yleisten kirjastojen kokoelmia ja lainausta. Tästä kokoelmien ja niiden käytön laajuudesta huolimatta kaunokirjallisuuden kirjastonomaista käyttöä ei ole juurikaan tutkittu, eikä sen erityispiirteet huomioonottavia tiedonhaun välineitä ole kehitetty.

Kaunokirjallisten teosten elämykselliset ja esteettiset ominaisuudet pakenevat helposti eksakteja määritelmiä ja siten myös kirjastolaisten systemaattisia sisällönkuvailun ja kokoelmanhallinnan menetelmiä. Esimerkiksi kaunokirjallisten teosten kielelliset erityispiirteet vaikuttavat niiden lukutapaan ja tulkintaan sekä siihen, miten niihin liittyvää tietoa muistetaan ja etsitään ja millaista metadataa niihin liittyen pitäisi tallentaa.

Nykyisillä kirjastojärjestelmillä ja sisällönkuvailumenetelmillä voidaan vastata kaunokirjallisuuden tietopalvelu- ja neuvontatyön tarpeisiin vain osittain. Neuvontatyötä haittaa myös se, että kertomakirjallisuutta (YKL-luokka 84.2) on sisällönkuvailtu systemaattisesti vasta vuodesta 1997, runokokoelmien, sarjakuvien ja lastenkirjallisuuden osalta 2000-luvun alusta.

Asiakkaat esittävät usein kaunokirjallisuuteen liittyviä kysymyksiä, joihin ei voi vastata tyydyttävästi nykyisten tietokantojen avulla. Jos asiakas haluaa ”Daniil Harmsin tyyppistä satiiria” (perustuu todelliseen tiedonhakutilanteeseen), vaatii se vastaajalta monimutkaista prosessointia ja asioiden yhdistelyä. On tiedettävä, kuka on Daniil Harms, ja mikä hänen tavassaan kirjoittaa satiiria on erityistä.

On tutkailtava, mitkä muiden satiirikkojen teokset voisivat olla asiakkaan tarkoittamalla tavalla ”Daniil Harmsin tyyppistä”. Kirjastojärjestelmästä on toki apua – sieltä voi etsiä asiasanojen avulla satiirisia teoksia, mutta viime kädessä asiakaspalvelijan on pääteltävä, mitkä niistä voisivat olla asiakkaan tarkoittamia. Arvioinnissa tarvitaan kirjallisuudentutmuksen lisäksi myös kykyä haastatella asiakasta, kysyä relevantteja kysymyksiä ja tulkita hänen vastauksiaan.

Näitä työläitä tiedonhakuja tehdään päivittäin kirjastoissa eri puolilla Suomea, ja niiden tuloksena syntyneitä ahaa-elämyksiä ja hyviä vastauksia ei tällä hetkellä voida hyödyntää muualla. Kun Daniil Harmsiin liittyvä kysymys esitetään seuraavan kerran jossakin kirjastossa, alkaa tiedonhakuprosessi alusta.

Nämä kaunokirjallisuuden ja siihen liittyvän neuvonta- ja tietopalvelutyön erityispiirteet ovat herättäneet kysymyksen, voitaisiinko modernin tietotekniikan avulla luoda järjestelmä, joka helpottaisi kirjastoissa kaunokirjallisuuden parissa tehtävää työtä ja joka tarjoaisi mahdollisuuden tuoda entistä monipuolisemmin esiin kirjastojen kaunokirjallisuuskokoelmia ja kirjastolaisten niihin liittyvän osaamista. Tällaisen järjestelmän pitäisi

- ottaa sisällönkuvailun lähtökohdaksi fiktiivisen aineiston erityispiirteet
- mahdollistaa toisiinsa jollakin tavoin liittyvien teosten suosittelu ja linkittäminen myös muuten kuin yksittäisten asiasanojen avulla
- mahdollistaa kirjastoammattilaisten verkostomainen työskentelytapa, jolloin myös vanhempaa kaunokirjallisuutta, ns. kirjallisuuden

pitkää häntää, voidaan sisällönkuvailla takautuvasti, ja näin saada haun piiriin nyt hankalasti löydettävissä olevaa aineistoa (pitkästä hännästä ja verkon mahdollisuuksista tuoda sitä esiin ks. esim. Anderson 2006, 36)

-toimia kirjastolaisten hiljaisen tiedon tallennusalustana
 -tuoda teoksiin liittyvää verkossa olevaa lisätietoa (kritiikit, teoskuvaukset)
 -tarjota kanava myös lukijoille.

Näiden ominaisuuksien käytännön toteutuksen pohdinta on ollut Kaunokirjallisuuden verkkopalvelun lähtökohtana.

Kohti kehittyneempää kaunokirjallisuuden tietojärjestelmää

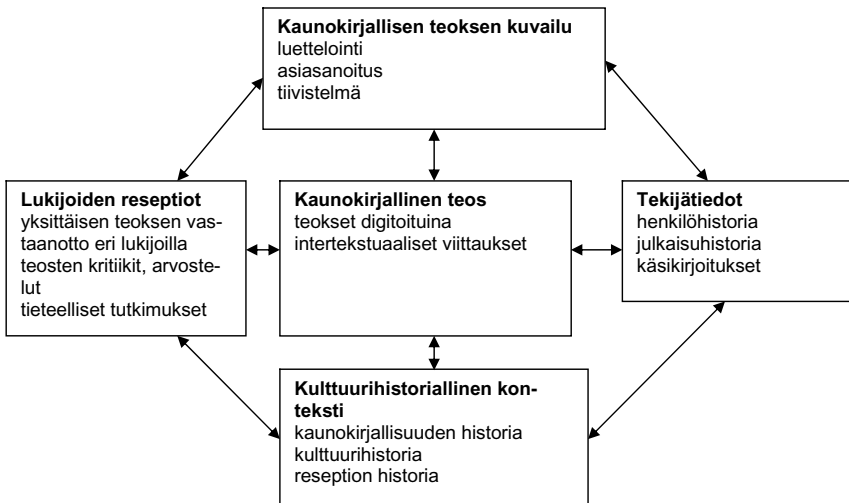
Ajatus kirjastojen yhteisestä kaunokirjallisuuteen keskittyvästä tietojärjestelmästä ei ole uusi. Onhan meillä esimerkiksi Roma-tietokanta (Roma), johon on asiasanoitettu noin 12 000 ennen vuotta 1996 ilmestynyttä romaania. Toistaiseksi systemaattisimmin kaunokirjallisuuden sisällönkuvausta, tiedonhakua ja kokoelmien hallintaa on käsitellyt nykyisin Kuopion yliopiston kirjaston johtajana toimiva Jarmo Saarti, joka on tutkinut esimerkiksi romaanien hyllyluokituksen vaikutusta niiden lainaukseen (Saarti 1996). Saarti on myös vaikuttanut merkittäväällä tavalla *Kaunokin*, fiktiivisen aineiston asiasanaston syntyymiseen

(Kaunokki 1996; Kaunokki 2000).

Kaunokirjallisuuden verkkopalvelun kehittämisen kannalta tärkein on kuitenkin Saartin väitöskirjassaan *Kaunokirjallisuuden sisällönkuvausten aspektit* esittelemä *kaunokirjallisuuden informaatiojärjestelmän ideaali- (tai laaja) malli* (Saarti 1999, 177-180). Siinä hän ottaa esille niitä kaunokirjallisuuteen liittyviä eri аспекteja, jotka ideaalisessa tietojärjestelmässä pitäisi ottaa huomioon. Kymmenen vuotta sitten Saarti oli sitä mieltä, ettei tällaisen järjestelmän rakentaminen (ainakaan silloisella tekniikalla) ole mahdollista.

Saartin kehittämä malli (kuva 1) on ollut kaunokirjallisuuden verkkopalvelun kehittämisen teoreettinen viitekehys ja työn lähtökohta. Olemme pyrkineet ottamaan kantaa mallissa esitettyihin aspekteihin ja pohtimaan, ovatko ne nyt kehittyneemmällä tekniikalla toteutettavissa ja millä tavoin.

Tämän mallin keskiössä ovat ”digitoidut kaunokirjalliset teokset” (Saarti 1999, 180), Kaunokirjallisuuden verkkopalvelun ytimessä on semanttinen teosrekisteri, johon tallennetaan ja josta ovat haettavissa kaunokirjallisten teosten eri tavoin rikastetut viitetiedot, tulevaisuudessa myös digitaalisessa muodossa olevat tekstit (suomalaisen kirjallisuuden digitoinnista ks. Nummi 2008, digitaalisista teksteistä esim. Niemi-Pynttari 2007).



Kuva 1. Kaunokirjallisuuden informaatiojärjestelmän laaja malli. (Saarti 1999, 179)

Jotakin uutta, jotakin vanhaa...

Uusi tekniikka edellyttää uusia työvälineitä. Semanttisissa järjestelmissä asiasanastojen sijaan käytetään *ontologioita*. Metatietojen tuottamista, luettelointia ja sisällönkuvailua, kutsutaan näissä järjestelmissä *annotoinniksi*. (Hyvönen 2008, 19.) Annotointiin on Teknillisessä korkeakoulussa kehitetty työkalu, selainpohjainen Saha-annotaatioeditori, jota voidaan muokata eri tiedonalojen tarpeisiin. Kullekin alakohtaiselle Sahalle laaditaan *metadataskeema*, jossa määritellään siinä tarvittavat tietokentät.

Terminologian ei pidä antaa hämätä – sisällönkuvailutyön perusteet säilyvät entisellään. Edelleen se edellyttää sisältöön tutustumista, sen analysointia ja relevanttien käsitteiden valintaa. Myös tiedontallennustyön prosessi etenee kutakuinkin entiseen malliin, lisäetuna on se, että ontologioiden avulla tuotetut sisällönkuvailut ovat paitsi ihmisten myös entistä paremmin tietokoneohjelmien hyödynnettävissä.

Ontologioista

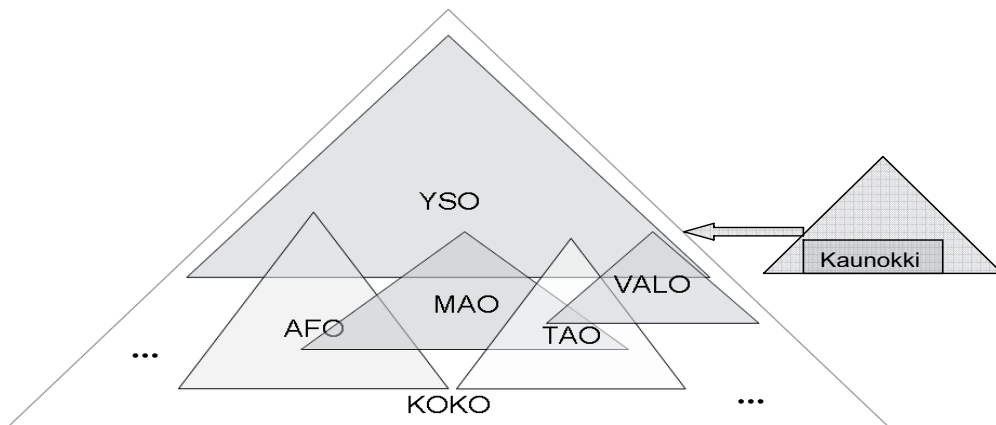
Ontologiat ovat tesauruspohjaisia, tietokoneohjelmistojen näkökulmasta kehitettyjä sisällönkuvailujärjestelmiä. Asiasanastot kehitettiin kortistoja ja inhimillistä tulkintaa varten,

ontologioiden avulla tuotettua sisällönkuvailutietoa voidaan hyödyntää verkkoympäristössä, verkon keinoin.

Ontologiat kuvaavat jonkin mikromaailman käsitteet, järjestävät käsitteiden väliset suhteet, ne ovat formaaleja ja eksplisiittisiä (Airio 2007). Ontologiat ovat myös hierarkisia, eli luokista luotavien yksilöiden tulee olla kaikkien yläluokkiensa yksilöitä (Hyvönen 2008, 21).

Ontologinen järjestelmä muodostuu luokkaontologioista ja yksilöontologioista. Luokkaontologiat on muodostettu käytössä olevista asiasanastoista. Yksilöontologioita ovat paikka- ja henkilöontologiat, kuten esimerkiksi Suomalainen paikkaontologia SUO ja Suomen historiallinen paikkaontologia SAPO. (Hyvönen 2008, 21, 25.)

Tärkein yksittäinen luokkaontologia on Yleiseen suomalaiseen asiasanastoon (YSA) perustuva Yleinen suomalainen ontologia (YSO). Se on perusontologia, joka yhdessä alakohtaisten luokkaontologioiden kanssa muodostaa kokonaisuontologia KOKOn. Siinä erikoisalojen käsitteistöt rikastavat perusontologiaa, eikä käyttäjän tarvitse miettiä, minkä alan ontologiaa hänen pitäisi käyttää, kun hän tarvitsee vaikkapa käsitteitä ”sukkulamato” tai ”elämäkertaromaani”, koska kaikki käsitteistöt löytyvät jatkossa KOKOsta. (Hyvönen 2008, 21.)



Kuva 2. ”Yhteisöllinen kokonaisuontologia KOKO koostuu yläontologiasta YSO ja sitä tarkentavista alaontologioista, joilla on yhteisiä käsitteitä YSO:n ja toisten alaontologioiden kanssa. Ontologioiden keskinäiset leikkaukset kuvassa ovat vain viitteellisiä ja tarkoitettu yleiskuvan havainnollistamiseksi” (Hyvönen 2008, 23). Kuvassa mainitut alaontologiat ovat MAO = Museoalan, AFO = Maa- ja metsätalouden, TAO = Taideteollisuusalan, VALO = Valokuvausalan ontologia. KOKOon pyrkivän Kaunokin kuvioon lisäsi tämän tekstin kirjoittaja.

The screenshot shows the ONKI web application interface. The main content area displays the concept 'lapset (ikään liittyvä rooli)' (children (age-related role)). The interface is divided into several sections:

- Käsitteet:** A list of related concepts, including 'lapset (perheenjäsenet)', 'lapset (ikään liittyvä rooli)', 'lapset (kooste)', 'lapset (sosioekonomiseen ryhmään liittyvä rooli)', 'lapset (oikeudet)', 'lapset (perheenjäsenet)', 'lapset (sosioekonomiseen ryhmään liittyvä rooli)', 'lapset - oikeudet', 'lapset_kooste', and 'lapsettomuus'.
- Konteksti:** A hierarchical tree structure showing the concept's context, including 'yso-käsitteet', 'pysyvä', 'ominaisuudet', 'ulkoiset ominaisuudet', 'rooli', 'väestöryhmään liittyvä rooli', 'ikään liittyvä rooli', and 'lapset (ikään liittyvä rooli)'. The 'lapset (ikään liittyvä rooli)' node is expanded, showing sub-roles like 'pojat (ikään liittyvä rooli)', 'tytöt', and 'vauvat'.
- Ominaisuudet:** A list of properties for the concept, including 'Asiasana (en): children', 'Asiasana (fi): lapset (ikään liittyvä rooli)', 'Korvaa asiasanan (fi): imeväisikäiset', 'Asiasana (sv): barn', 'Korvaa asiasanan (sv): spädbarn', 'Lahikasite: alaikäiset', 'esikouluikäiset', 'ikäryhmät', 'kouluikäiset', 'lapsenmurha', 'lapset (sosioekonomiseen ryhmään liittyvä rooli)', 'leikki-ikäiset', 'sukupohjat', and 'Tyyppi: YSO-käsite'.

Kuva 3. Näkymä YSO:sta, käsitteet on tehty käsitteellä ”lapset”. Käsitteeseen on YSO:ssa kolme näkökulmaa, ”lapset” perheenjäsenenä, ikäryhmän ja sosioekonomisen ryhmän edustajana. Kullakin näistä on oma yksilöllinen URI-tunnisteensa, jonka perusteella ohjelma tunnistaa ”lapset”-käsitteen eri aspektit, vaikka kirjoitusasu on sama. Ontologian käyttäjälle nämä eri roolit on kerrottu selväkielisenä (suluissa käsitteen jälkeen).

Kullakin ontologiakäsitteellä on yksilöllinen, kirjaimista ja numeroista muodostettu, kieli-riippumaton verkkotunnus URI-tunniste (Hyvönen 2008, 22, 27). URI-tunnisteiden avulla käsitteet, joilla on sama kirjoitusasu mutta eri merkitys, voidaan erottaa toisistaan (kuva 3). Tämä tekee tiedontallennuksesta ja -hausta aiempaa täsmällisempää.

Yksi alakohtaisista ontologioista on Verkko-Kaunokkiin (Verkko-Kaunokki) perustuva KAUNO. Kaunokin ontologisointi tehdään Protégé -ontologiaeditorilla osana Kaunokirjallisuuden verkkopalvelu -projektia. Ontologisoitintyössä kukin Kaunokin noin 4 500 termistä käydään läpi käsitteanalyysin keinoin, käsitteet ja niiden väliset suhteet määritellään ja ”ripustetaan” YSO:an. Samalla pyritään lisäämään

KAUNON sisäistä hierarkiaa ja määrittelemään siinä käytettäviä käsitteitä.³

Ontologiat ovat kaikkien vapaasti käytettävissä ONKI-palvelimella (ONKI-palvelin). Ontologiat on myös integroitu Sahaan, joten niiden käyttö tiedontallennuksessa on vaivatonta.

Metadataskeema

Kutakin erillistä alakohtaista sovellusta varten laaditaan tietomalli, *metadataskeema*, jossa määritellään kyseisessä järjestelmässä tarvittavat tietokentät, ja joiden mukaisesti rakennetaan alakohtaiset Saha-annotaatioeditorit. Esimerkiksi Kulttuurisampo-palvelussa on käytössä 16 skeemaa eri aineistotyypeille. Niiden pohjalta laaditaan malli järjestelmän ulkoiseksi

metatietorajapinnoiksi. (Hyvönen 2008, 19.)

Kaunokirjallisuuden verkkopalvelun teosrekisterin lähtökohtana on kaunokirjallinen teos, *work*, joka voidaan määritellä ”a distinct intellectual or artistic creation” (FRBR 1998, 16) tai ”kaikkien kielellisten versioiden kuljettama ajatussisältö” (Tillett 2004, 2). Teos voi olla muodoltaan esimerkiksi romaani, runo tai novelli. Skeeman ja Kaunokki-Sahan laatimisen apuna on käytetty funktionaalisen luetteloinnin mallia (FRBR 1998; Tillett 2004) sekä Dublin Core - metadataformaattia (Stenvall 2002; Dublin Core 2001).

Kukin teos on rekisterissä yhtenä tietueena, johon tallennetaan luettelointimallista poimittuja, teosten ”olemisentasoisiin” (entiteetteihin, *entities*) liittyviä ominaisuuksia (attribuutteja, *attributes*).⁴

Kaunokki-Sahassa ovat nämä skeeman mukaiset tietokentät:

- tekijä(t) eri funktioissa (kirjoittaja, kuvittaja, kääntäjä),
- nimeke eri muodoissaan (pää- ja alanimeke, alkuteos, millä muilla nimekkeillä teos on ilmestynyt),
- ensimmäisen painoksen tai käännöksen julkaisuvuosi,
- sivumäärä,
- kustantaja(t),
- sarja(t),
- mahdolliset sovitukset tai elokuvaversiot,
- alkuteoksen kieli (koska rekisteri on nimenomaan suomenkielisen kaunokirjallisuuden tietokanta, lähtöoletus on, että teosten kieli on suomi)
- kansi
- kirjallisuuspalkinnot
- erityiskokoelma, johon teos kuuluu (esimerkiksi maakuntakirjastojen maakuntakokoelmat)
- sisällönkuvailuun liittyvät kentät

Kaunokirjallisuuden verkkopalvelun metadata-skeemassa kiinnitettiin erityistä huomiota sisällönkuvailuun ja suositteluun. Niinpä kullakin Kaunokin kuudesta fasetista (genre, teema, toimija, aika, paikka, muut asiasanat) on oma kenttänsä Kaunokki-Sahassa. Tämä jäntevöittää tiedontallennusta ja tuo uusia mahdollisuuksia tiedonhakuun ja tiedon esittämiseen.

Ontologioiden hierarkiarakenne todennäköisesti vähentää tarvetta ontologiakäsitteiden ketjutukseen. Enää ei esimerkiksi tarvita käsitteketjua ”perheenjäsenet : lapset”, koska ontologiassa

on määritelty, että ”lapset (perheenjäsenet)” ovat ”perheenjäsenet” -käsitteen alakäsite. Kun ontologian avulla tuotetusta metadatasta etsitään käsitteellä ”perheenjäsenet”, se tuo tarvittaessa esiin myös sen alakäsitteillä kuvailtuja teoksia. Tai toisinpäin: haun voi määritellä koskemaan myös yläkäsitteitä.

Kaunokki-Sahassa on kuitenkin mahdollisuus liittää käsitteitä toisiinsa. Esimerkiksi sinne on mahdollista luoda pysyvä käsitepari ”odotus : jouluku”, jolla kuvatut teokset saadaan Sahassa listattua. Kaunokki-Sahaan voi tallentaa myös suorasanaisia kuvauksia teoksista, kuten tiivistelmiä, kritiikkejä ja esittelyjä.

Skeemaan on lisätty kenttiä, jotka eivät ole kuuluneet perinteisiin kirjastojärjestelmien metatietomalleihin. Tärkein näistä on mahdollisuus suositella ja linkittää teoksia toisiinsa. Tämä liittyy Saartin mallissa kohtaan *intertekstuaaliset viittaukset*. Hän lienee tarkoittanut sillä kirjallisuudentutkimuksen keinoin todennettavissa olevia teosten välisiä, esimerkiksi teemoihin liittyviä yhteyksiä. Termiä on määritelty myös seuraavasti: ”tekstin suhde muihin teksteihin; tekstienvälisyys. [...] Hyvin laajasti ymmärrettyinä intertekstuaalisuus on yhteisten kirjallisten koodien ja konventioiden omaksumista” (Hosiaisuus 2003, 357). Projektissa olemme ottaneet vapauden tulkita sitä käytännön kirjastotyön näkökulmasta.

Ymmärrämme sen kirjastoammattilaisten hiljaisen tiedon muotona, ”päänsisäisenä intertekstuaalisuutena”, joka perustuu laajaan kirjallisuudentuntemukseen. Se on se intuitiivinen tieto, joka ohjaa kirjastonhoitajaa ja jonka varassa hän toimii, kun hän etsii asiakkaalle ”samantyyppistä luettavaa” tai ”hyvää kirjallisuutta”. Tämä tieto liittyy usein hankalasti ilmaistaviin seikkoihin, kuten teosten tyyliin, aiheenkäsittelytapaan tai tunnelmaan, ja sitä on vaikea ilmaista yksittäisillä asiasanoilla.

Edellä kuvattu ”Daniil Harmsin tyyppinen satiiri” -tapaus sopii hyvin esimerkiksi suosittelukentän käytöstä. Kun kirjastoammattilainen on löytänyt sopivat teokset, hän voi lisätä niistä maininnan teosrekisteriin, jolloin ne ovat seuraavan ammattilaisen – ja myös palvelun käyttäjän – hyödynnettävissä. Toinen ammattilainen voi tarvittaessa merkitä lisää suositteluja, ja näin rekisteristä voi kehittyä hiljaisen tiedon tallennuspaikka, kunhan se on riittävän helppokäyttöinen niin, että tietojen täydentäminen on vaivatonta osana kirjaston arkea. Tämän tiedon

tallentaminen on erityisen tärkeää nyt, kun kaunokirjallisuuteenkin liittyvä asiantuntemus on katoamassa kirjastoista eläköitymisen myötä. Kymmeniä vuosia kirjallisuutta ja kirjallista elämää seuranneista kirjastolaisista on jäämässä eläkkeelle ja heidän mukanaan katoaa merkittävä määrä hiljaista tietoa kirjastoista (ks. esim. Kauhanen 2004).

Kaunokki-Sahan metadataskeemassa on mielletty teokseksi yksittäiset tekstit, siis myös runot ja novellit, jotka on mahdollista tallentaa rekisteriin omina tietueinaan. Runoon tai novelliin liittyvä tietue linkitetään niihin teosrekisterissä oleviin kokoelmateoksiin, joista teksti löytyy painetussa muodossa. Runot ja novellit voidaan annotoida hyvinkin tarkalla tasolla (esimerkiksi

kunkin novellin ja runon ensimmäisen version julkaisu-aika, sisällönkuvailu, mahdollisen alkuteoksen nimi), jos niin halutaan ja jos siihen on resurssseja.

Tavoitteena on, että kaikki järjestelmään tallennettu tieto on haettavissa, myös sivumäärä (käyttäjä voi hakea ”paksua” tai ”ohutta” kirjaa) ja kannen kuvailutiedot.

Kaunokki-Saha ja sisällönkuvailijoiden verkosto

Kaunokki-Sahan pohjaksi on konvertoitu HelMetin kaunokirjallisuusaineisto. Mukana ovat omina luokkinaan romaanit, novellit ja muu lyhytproosa, näytelmät, runot, kansan-

Kuva 4. Kaunokki-Sahan testiversion tiedontallennusnäky, jossa näkyy osa sisällönkuvailuun liittyvistä kentistä. Kutakin kenttää voi tarvittaessa toistaa. Sisällönkuvailukäsitteet on helppo poimia ontologioista, siinä avustaa myös ”semanttinen täydentäjä” (semantic autocompletion). Sahasta voi myös siirtyä selaamaan ontologiaa, ja tarkastella yksittäisiä käsitteitä osana kokonaisuutta.

runot, murrekirjallisuus ja sarjakuvat. Ennen Kaunokki-Sahaan siirtoa tietokannan pohja-aineistoa – ”dumpia” – käsiteltiin ohjelmallisesti Teknillisessä korkeakoulussa ja manuaalisesti Turun kaupunginkirjastossa. Ohjelmallisessa käsittelyssä yhdisteltiin samaan nimekkeeseen liittyviä tietueita, manuaalisessa läpikäyntivaiheessa täydennettiin jonkin verran asiasanoja ja luettelointitietoja muiden tietokantojen avulla (lähinnä Fennica, Piki-verkkokirjasto, Vaski ja Kainet). Kaunokki-Sahassa on aloitusvaiheessa noin 36 500 tietuetta. Mennee pitkään, ennen kuin Kaunokki-Sahassa tiedot ovat ”oikein” ja teosrekisterin funktionaalisia ominaisuuksia voidaan tiedonhaussa täysimääräisesti hyödyntää.

Uutuusaineiston luettelointi- ja sisällönkuvailutiedot ostetaan jatkossa BTJ Finland Oy:ltä. Kaunokki-Sahaan voidaan lisätä sieltä mahdollisesti puuttuvia teoksia, esimerkiksi paikalliskokoelmien aineistoa. Teosten ”manifestaatiot”, fyysiset niteet, paikannetaan kirjastoista ainakin alkuvaiheessa Frank-monihauun avulla. Ihanteellista olisi, jos verkkopalvelu voitaisiin integroida osaksi kirjastojärjestelmiä niin, että kirjaston tietokantaan tehty haku kohdistuisi samalla myös teosrekisteriin.

Kaunokki-Sahassa on siis keskeinen suomenkielinen kaunokirjallisuus tallennettuna perusluettelointitiedoin, osittain sisällönkuvailtuina. Kirjastoammattilaisten verkoston tehtäväksi jää teosten sisältöjen kuvailu suorasanaisesti ja ontologioiden avulla sekä teosten suosittelu. Urakkaa helpottaa se, että Roma-tietokannan sisällönkuvailut siirretään Kaunokki-Sahaan kevään 2009 aikana.

Verkostona tehty sisällönkuvailu vaatii yhteisiä periaatteita ja pelisääntöjä, opastusta ja koulutusta. Sisällönkuvailuyhteistyö on jo alkanut Turun ammattikorkeakoulun kirjasto- ja koulutusohjelman kanssa, kirjastojen kanssa konkreettinen sisällönkuvailutyö aloitetaan keväällä 2009, kunhan Kaunokki-Saha on teknisesti ja toiminnallisesti riittävän valmis. Yhteistyöstä kiinnostuneita kirjastoja ja kirjastolaisia on jo ilmoittautunut, ja kaikille halukkaille pyritään antamaan perehdytys tallennusalan käyttöön.

Kirjastolaisilla on jo kokemusta yhteisesti ylläpidetyistä verkkopalveluista, kuten Kysy kirjastonhoitajalta. Uskomme, että kirjastolaiset omaksuvat myös Kaunokki-Sahaan tallennuksen,

kunhan sen käyttöliittymä on helppokäyttöinen ja osallistumiskynnys sisällönkuvailuun on matala. Tavoitteena on positiivinen kehä, jossa kokemus palvelun hyödyllisyydestä innostaa rikastamaan sen sisältöjä entisestään. Toivomme, että teosrekisteristä kasvaa kirjastoille arjen työkalu, kaikille kirjallisuuteen liittyvä inspiraation, ilon ja elämysten lähde.

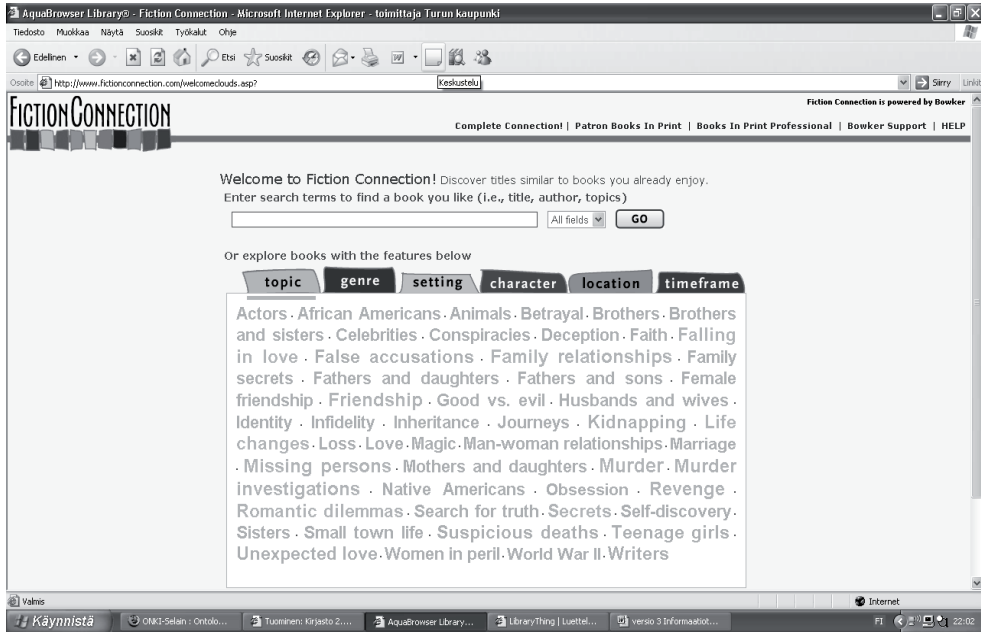
Tähän on edellytykset, ainakin jos tanskalaisten kokemuksista voidaan jotakin päätellä. Siellä on toiminut vuodesta 2002 www.litteratursiden.dk -palvelu, jota ylläpidetään ja kehitetään hajautetusti. Ylläpitäjien määrä on vuosien myötä kasvanut viidestä 78 kirjastoon ja yli sataan kirjastolaiseen. Ylläpidossa voi olla mukana pienelläkin panoksella, vaikkapa viiden teoksen kuvailulla vuodessa. Palvelu on kirjastojen tärkein tanskankielisen kaunokirjallisuuden tietopankki, ja se on merkittävä toimija laajemminkin tanskalaisessa kirjallisessa elämässä. (Vandborg 2008.)

Tekijät, lukijat ja kontekstit

Teosrekisterin rakentaminen on ollut haasteellinen tehtävä, koska sen sisällöt ja toimivuus ovat keskeisiä verkkopalvelun onnistumisen kannalta. Yhtään vähemmän haastava ei tule olemaan verkkopalvelun käyttöliittymän suunnittelu ja toteutus, jotka alkavat keväällä 2009. Käyttöliittymä rakennetaan avoimen lähdekoodin Drupal -ohjelmistolla, ja sen tulee toimia saumattomasti yhteen teosrekisterin kanssa.

Kimmo Tuominen on määritellyt ”Kirjasto 2.0 -nimellä kutsuttua uudenlaista verkkopalveluajattelua”, jolle hänen mukaansa ovat ominaisia ainakin tiedon löytämisen ja paikallistamisen vaivattomuus, verkkosisällön yksilö- ja ryhmäkohtaista räätälöitävyys sekä käyttäjien osallistuminen palvelun tuottamiseen (Tuominen 2006). Nämä tulevat olemaan myös Kaunokirjallisuuden verkkopalvelun käyttöliittymäsuunnittelua ohjaavat perusajatukset.

Palvelussa eivät kirjallisuutta esittele vain kirjastoalan ammattilaiset, vaan niin halutessaan myös palvelun käyttäjät. Samoin kuin suositussa LibraryThing -palvelussa (LibraryThing), myös verkkopalvelussa teoksiin liittyy kirjastoammattilaisen ontologioiden ja luetteloinnin keinoin tuottama *book information* ja käyttäjän vapaammin tuottama kuvailu, *social information*. Ammattilaisen ja käyttäjän



Kuva 5. Esimerkki kaunokirjallisuuden hakuun tehdystä havainnollisesta ja helppokäyttöisestä liittymästä, www.fictionconnection.com. Sovellus on kaupallinen.

tuottama informaatio ovat selkeästi omina kokonaisuuksinaan.

Käyttäjät voivat poimia kuvailukäsitteet KAUNOsta, mutta he voivat käyttää myös ns. vapaita asia- eli avainsanoja. Parhaassa tapauksessa tämä käyttäjien tekemä sisällönkuvailu antaa virikkeitä ammattilaisten sisällönkuvailulle, antaa signaaleja siitä, millaista ja miten kirjallisuutta luetaan ja millaista käyttäjälähtöinen teosten kuvailu voisi olla. Avainsanojen käytön lisäksi palveluun tulee muitakin 2.0-ominaisuuksia. Ne täsmentyvät, kun käyttöliittymäsuunnittelu pääsee vauhtiin, keskusteluissa ovat vilahdelleet mm. mahdollisuus arvioida, suositella ja esitellä teoksia, ylläpitää virtuaalisia kirjajuhllyjä ja lukupäiväkirjaa sekä keskustella teoksista.

Käyttöliittymäsuunnittelussa täsmentyy sekin, miten muista verkkopalveluista voidaan tuoda sisältöjä (kritiikkejä, teosesittelyjä, verkosta löytyviä tutkimuksia) kaunokirjallisuuden verkkopalveluun. Alustavasti on ollut puhetta yksittäisten kirjastojen, esimerkiksi Celian, ja kirjallisuusaiheisten verkkopalveluiden tuottajien kanssa (kirjavinkit.net, lukupiirit.fi). Lisää yhteistyökumppaneita etsitään, kun palvelu on

saatu valmiimmaksi. Tavoitteena on saada mukaan myös aikalaislukijoiden kuvauksia digitoimalla Arvosteleavassa kirjaluettelossa⁵ ilmestyneitä arvioita.

Kaunokirjallisuuteen liittyvä tietojärjestelmä on torso, jos siinä ei ole teosten tekijöihin liittyvää tietoa. Asiakkaat kysyvät usein tietyn maan, tietyllä kielellä kirjoitettavia tai tiettyyn lajityyppiin liittyviä kirjailijoita. Kirjailijatiedon tallennusta varten on laadittu oma metadataskeemansa, ja teosrekisterissä on kentät kirjailijaa koskeville tiedoille. Kiinteä yhteys teos- ja kirjailijatietojen välillä mahdollistaa monipuoliset haut. Palveluun on mahdollista tuoda myös kirjailijaan liittyvää lisämateriaalia, kuten haastatteluita, kuva- ja äänitallenteita, kunhan käyttö- ja tekijänoikeuksista sovitaan.

Pohja-aineisto kirjailijatietokantaan saadaan kirjastojen laatimista kirjailijaesittelyistä. Tähän liittyvää valmistelutyötä on tehty Vaasan kaupunginkirjaston hallinnoimassa maakuntakirjastojen KITKA-projektissa (KITKA). Näiden aineistojen julkaisemiseen verkossa on saatu kirjailijoiden lupa, kirjailijoiden henkilötietojen esittämisessä täytyy ottaa

huomioon lainsäädännön asettamat rajoitukset.

Kulttuurisampo (Kulttuurisampo) on muistiorganisaatioiden tuottamaan tietoon rajautuva kokonaisuus, ja se tarjoaa mielenkiintoisen kontekstin kaunokirjallisille teoksille. Kulttuurisammossa on testiaineistona HelMetistä konvertoitu, vuonna 2007 ilmestynyt kaunokirjallisuus. Jo tässä demoversiossa voidaan nähdä, kuinka teoksiin liittyvää tietoa voidaan linkittää muiden muistiorganisaatioiden Kulttuurisammossa olevan aineiston kanssa. Periaatteessa on mahdollista kytkeä esimerkiksi Turkuun liittyviä kaunokirjallisia teoksia muuhun tätä aluetta koskevaan aineistoon, kuten museoesineisiin ja maalauksiin.

Teoksia voidaan viedä niiden julkaisu- tai sisäisen tapahtuma-ajan perusteella aikajanelle tai tapahtumapaikan mukaan karttasovellukseen. Kulttuurisammon virtuaalisessa ympäristössä samoileva saa tietoa teoksista ja kirjastojen kokoelmista, vaikka hän ei hakeutuisikaan kirjastojen omiin käyttöliittymiin. Kulttuurisammon, kuten yleensäkin semanttisten palveluiden, sisältöjä voidaan myös viedä muihin yhteyksiin semanttisina ”leijukkeina” (widget). (Hyvönen 2008, 15.)

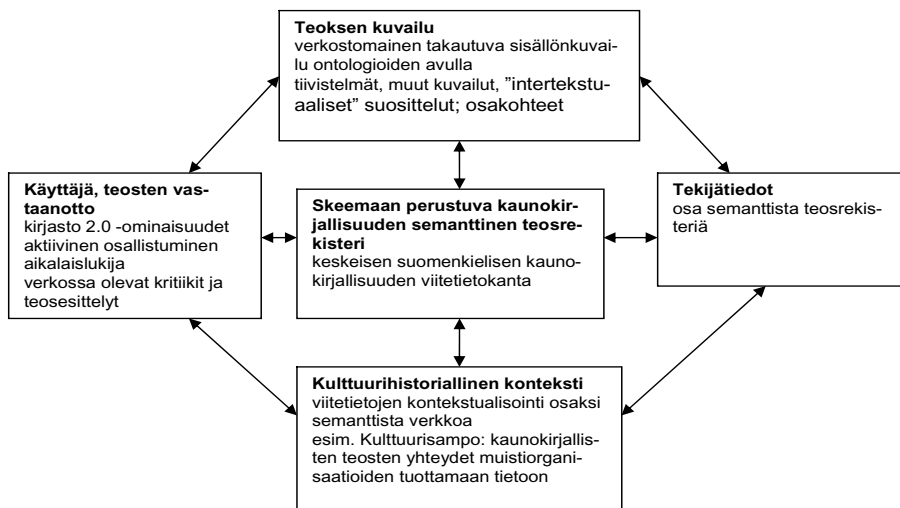
Kulttuurisampon viedään teosrekisteriinkin tuleva laaja kaunokirjallisuusaineisto. Sen myötä linkitykset ja semanttiset yhteydet

moninkertaistuvat nykyisestä. Niiden avulla voidaan havainnollistaa esimerkiksi eri taiteenlajien välisiä vaikutteita. Koskaan ei kuitenkaan voida kuvata tyhjentävästi kirjallisuuden kontekstiavaruutta, koska ”reaalimaailmassa linkitys jatkuu äärettömästi” (Saarti 1999, 180). Se ei edes ole verkkopalvelun tavoite, tärkeämpää on valikoida kokonaisuuden kannalta relevantit kontekstit ja rajata linkityksiä keskittymällä olennaiseen aineistoon.

Alla on Saartin laatima kaunokirjallisuuden tietojärjestelmän laaja malli ”semanttisesti rikastettuna” versiona, muokattuna siten kuin se on tässä kuvatussa prosessissa muotoutunut. Jokaiselle aspektille on löytynyt funktio ja toteutustapa semanttisessa mallissa, toisaalta yhtään uutta näkökulmaa ei siihen ole tarvinnut lisätä. Malli siis kuvaa yleispätevästi kaunokirjallisuuden informaatiojärjestelmää. Olisi kiinnostavaa nähdä, kuinka sitä jalostettaisiin edelleen vuonna 2019.

”Sisältö ei ole enää kuningas – konteksti on!”

Yleisten kirjastojen verkkonäkyvyys perustuu suureksi osaksi niiden tietojärjestelmiin, jotka operoivat toistaiseksi – ja ilmeisesti aika pitkälle



Kuva 6. Saartin malli ”semanttisen rikastamisen” jälkeen.

tulevaisuudessakin – viitetiedon varassa. Tietojärjestelmien kehittämisen keskeiset kysymykset liittyvät sisältöön ja kontekstiin: miten sisältöjä kuvaillaan, miten metadataa tallennetaan ja miten se esitetään. Pelkkä hyvinkään tuotettu viitetieto ei verkon visuaalisessa ympäristössä riitä, vaan sitä on rikastettava muiden palveluiden tarjoamalla sisällöillä. On tuotava esiin tietoon liittyviä relevantteja yhteyksiä ja taustoja. Sisällöt on asetettava osaksi konteksteja: ”content is no longer king – context is!” (Vandborg 2008).

On etsittävä ratkaisuja esimerkiksi näihin kysymyksiin: millaisia kirjastojen hakuliittymät ovat, onko käyttäjän hakeuduttava kirjastojen omiin liittymiin vai voidaanko mahdollisuus hakea kirjastojen kokoelmista viedä osaksi muita palveluita? Miten viitetietoja voidaan esittää osana temaattisia kokonaisuuksia, havainnollisesti ja visuaalisesti? Millainen järjestelmä mahdollistaa tiedon kontekstualisoimisen ja esittämisen eri yhteyksissä?

Toivottavasti Kaunokirjallisuuden verkkopalvelun kehittämisessä ja myöhemmin valmiin palvelun toiminnasta saatuja kokemuksia voidaan hyödyntää, kun yleiset kirjastot jatkossa pohtivat sitä, miten niiden pitäisi näkyä verkossa ja miten tietojärjestelmät voisivat tukea niiden perinteisen sisältöosaamisen esiin tuomista. Tekniikka ei ole itsetarkoitus, vaan tavoitteena on rakentaa käyttäjille parempia ja kehittyneempiä palveluita, ja synnyttää sitä kautta uusia oivalluksia ja elämyksiä.

Hyväksytty julkaistavaksi 17.3.2009.

Viitteet:

- ¹ Palvelulla ei toistaiseksi ole nimeä. Keräsimme nimiehdotuksia avoimella kyselyllä, ja saimmekin yli 160 hyvää ehdotusta. Harmi vain, että parhaiten sopivien nimien verkko-osoitteet on jo varattu. Nimi löytynee kuitenkin ennen palvelun julkistusta...
- ² Projektista on ilmestynyt artikkeleita (ks. esim. Hypén & Spåre, 2008a ja Hypén & Spåre, 2008b), ja sen etenemistä voi seurata KirjastoWikin sivulta (Kaunokirjallisuushanke). Syksyllä ja talvella 2007-08 palvelun suunnitelmia ja tavoitteita esiteltiin erilaisissa kirjastoalan koulutusilaisuuksissa. Uusi esittely- ja koulutuskierron on käynnistymässä kevättalvella 2009.

- ³ Asiasanaston muokkaamisesta ontologiaksi ks. tarkemmin Ruotsalo et al. 2008, 6-8. YSON taustalla olevasta DOLCE-järjestelmästä ks. tarkemmin esim. Hyvönen 2008, 22.
- ⁴ Näistä ominaisuuksista tarkemmin, ks. FRBR 1998, 32-48; Stenvall 2002, 8-17; Dublin Core 2001, 3-9.
- ⁵ Arvosteleva kirjaluetelo on Kouluhallituksen (nykyisin Opetushallitus) vuosina 1904-1988 ilmestynyt kirjallisuutta esittelevä ja arvioiva julkaisu. Lyhyissä esittelyteksteissä on ajoittain hyvinkin kriittinen näkökulma, sellaisena ne heijastelevat oman aikansa kirjallisuuskäsityksiä ja sitä, mitä on kulloinkin pidetty ”hyvänä kirjallisuutena”. Arvostelevaa kirjalueteloa käytettiin paljon kirjastoissakin aineistonvalinnan tukena.

Lähteet:

- Airio, Eija (2007). Ontologiat ja tiedonhaku. Esitys Semanttinen web ja funktionaalinen luettelointi -koulutuspäivässä 18.1.2007. Turun ammattikorkeakoulu.
- Anderson, Chris (2006). Pitkä häntä: Miksi tulevaisuudessa myydään vähemmän enempää. Alkuteos: The Long Tail. Why the Future of Business is Selling Less for More. Suom. Kimmo Pietiläinen. Terra Cognita.
- Arjen tietoyhteiskunta. http://www.arjentietoyhteiskunta.fi/files/47/toimintaohjelma_lopullinen_25.4.2008.pdf (18.1.09)
- Dublin Core (2001). Dublin Core – metadatatyyppien suomalainen versio. SFS 5895. Suomen Standardoimisliitto SFS.
- FRBR. (1998). Functional requirements for bibliographic records. <http://www.ifla.org/VII/s13/frbr/frbr.pdf#undefined> (16.1.09)
- Hosiaisuoma, Yrjö (2003). Kirjallisuuden sanakirja. WSOY.
- Hypén, Kaisa & Spåre, Alex (2008a). Sanoja ja semantiikkaa: Kaunokirjallisuuden verkkopalvelun ensiaskeleet. Kansalliskirjasto 2. S. 52-56.
- Hypén, Kaisa & Spåre, Alex (2008b). Kaunokirjallisuuden verkkopalvelu ja kaunokirjallisuuteen liittyvä osaaminen yleisissä kirjastoissa. Teoksessa Mahdollisuuksien ikkunoita. Turun ammattikorkeakoulun kirjasto- ja tietopalvelun koulutusohjelma 10 vuotta. Toim. Aulikki Holma ja Nina Punelpuro. Turun ammattikorkeakoulu. S. 212-224.

- Hyvönen, Eero (2007). Verkottuneen tiedon rajoja siirtämässä. Teoksessa *Rajalla: tiede rajojaan siirtämässä*. Toim. Kari Raivio, Jan Rydman ja Anssi Sinnemäki. *Gaudeamus*. S. 72-92.
- Hyvönen, Eero (2008). Kulttuurisampo – suomalainen kulttuuri semanttisessa webissä. <http://www.seco.tkk.fi/publications/2008/hyvonen-Kulttuurisampo-2008.pdf> (13.1.09)
- Kauhanen, Anna-Liina (2004). Kirjastonhoitajan hiljainen tieto uhkaa kadota. *Helsingin Sanomat* 4.12. S. 4.
- Kaunokirjallisuushanke. http://wiki.kirjastot.fi/index.php/Projekti_Kaunokirjallisuushanke (13.1.09)
- Kaunokki (1996). Kaunokirjallisuuden asiasanasto. Kokeilupainos. Vastaava toimittaja Jarmo Saarti. BTJ Kirjastopalvelu.
- Kaunokki (2000). Fiktiivisen aineiston asiasanasto. Vastaava toimittaja Jarmo Saarti. BTJ Kirjastopalvelu.
- KITKA. http://wiki.kirjastot.fi/index.php/Projekti_Kitka2 (18.1.09)
- Kulttuurisampo. <http://www.kulttuurisampo.fi> (14.3.09)
- LibraryThing. <http://fi.librarything.com/?highload=1> (17.1.09)
- Museosuomi. <http://www.museosuomi.fi/> (14.3.09)
- Niemi-Pynttari, Risto (2007). Verkkoproosa. Tutkimus dialogisesta kirjoittamisesta. *ntamo*.
- Nummi, Jyrki (2008). Kirjallisuuspankki. Suomalaisen kaunokirjallisuuden digitaalinen kirjasto (selvitystyöraportti). <http://www.kansalliskirjasto.fi/attachments/5kSvIrHoj/5xYC3fd8g/Files/CurrentFile/Kirjallisuuspankki-selvitys.pdf> (13.1.09)
- ONKI-palvelin. <http://www.yso.fi/> (18.1.09)
- Roma. <http://www.pori.fi/kirjasto/roma/> (15.1.09)
- Ruotsalo, Tuukka & Seppälä, Katri & Viljanen, Kim & Mäkelä, Eetu & Kurki, Jussi & Alm, Olli & Kauppinen, Tomi & Tuominen, Jouni & Frosterus, Matias & Sinkkilä, Reetta & Hyvönen Eero (2008). Ontology-based approach for interoperability of digital collections. *Signum* 5. S. 5-14.
- Saarti, Jarmo (1996). Romaanit hukassa? Kertomakirjallisuuden hyllyluokituksen vaikutus kirjastonkäyttöön ja kertomakirjallisuuden lainaukseen. BTJ Kirjastopalvelu.
- Saarti, Jarmo (1999). Kaunokirjallisuuden sisällönkuvailun aspektit. Kirjastoammattilaisten ja kirjastonkäyttäjien tekemien romaanien tiivistelmien ja asiasanoituksen yhdenmukaisuus. Oulun yliopisto.
- Stenvall, Jani (2002). Dublin Core -formaatin käyttöopas. Helsingin yliopiston kirjasto. http://www.kansalliskirjasto.fi/julkaisuala/dublincore/Files/liitetiedosto2/dc_opas.pdf (21.1.09)
- Terve-suomi. <http://www.terveysuomi.fi> (14.3.09)
- Tillett, Barbara (2004). Mikä on FRBR? Bibliografisen universumin käsitelmä. Suom. Päivi Pekkarinen. <http://www.loc.gov/catdir/cps0/Mika-on-FRBR.pdf> (20.1.09)
- Tuominen, Kimmo (2006). Kirjasto 2.0 ja tulevaisuuden näyttöluettelot. <http://www.lib.helsinki.fi/tietolinja/0206/kirjasto20.html> (17.1.09)
- Vandborg, Lise (2008). Litteratursiden.dk – Danish libraries website on fiction. Kaunokirjallisuuden verkkopalvelu-projektin henkilökunnalle 6.11.08 pidetty esitys. Århusin kaupunginkirjasto
- Verkko-Kaunokki. <http://kaunokki.kirjastot.fi/fi-FI/> (17.1.09)