

YSAsta YSOon kirjastoissa: miksi ja miten

YSAsta YSOon kirjastoissa: miksi ja miten

- YSAsta YSOon kirjastoissa: miksi ja miten
 - Johdanto
 - YSOon siirtymisen hyödyt
 - Monikielisyys ja rikastettu termistö
 - Pysyvät koneluettavat tunnisteet
 - Käsitetaso
 - Linkittyvyys
 - Siirtymä käytännössä
 - YSAlla kuvailtujen tietueiden konvertointi
 - YSON käyttöönotto kuvailuvälineissä
 - YSON käyttöönoton seuraukset

Johdanto

Yleinen suomalainen ontologia YSO on semanttisen webin ja avoimen linkitetyn datan ominaisuuksilla rikastettu koneymmärrettävä versio yleisestä suomalaisesta asiasanasto YSasta. YSAn ontologisointi aloitettiin Aalto-yliopiston ja Helsingin yliopiston yhteisessä FinnONTO-tutkimusprojektissa (2003–2012). Vuonna 2013 ontologiatyö siirtyi Kansalliskirjastoon, jossa YSOa on kehitetty edelleen erityisesti kuvailu- ja tiedonhakutarpeet huomioiden.

YSA on ollut kirjastokentän käytetyin sisällönkuvailusanasto sen kehittämisestä 1980-luvun loppupuolelta lähtien. YSAlla on kuvailtu muiden muassa Kansallisbibliografian keskeiset ei-fiktiiviset aineistot. Vaikka YSA on osoittanut käyttökelpoisuutensa vuosien varrella useassa yhteydessä, on se nykyisen verkkomaailman tarpeet huomioon ottaen kuitenkin vanhentunut.

YSO pyrkii tuomaan YSAn keskeiset ominaisuudet verkkoajan järjestelmien hyödynnettäviksi sekä rikastamaan näitä ominaisuuksia edelleen. Johtajatuksena on, että YSON myötä ontologioilla kuvailtuja sisältöjä voidaan hyödyntää tehokkaammin moderneissa tiedonhakujärjestelmissä kuten Finnassa. Lisäksi moniin muihin sanastoihin linkitetyn YSON myötä kirjastojen tietosisällöt yhdistyvät osaksi koko julkisen sektorin tietoverkosta. Tätä kautta mahdollistetaan uusia tapoja löytää ja hyödyntää erilaisia sisältöjä tiedontuottajatahosta riippumatta.

YSOon siirtymisen hyödyt

YSOa kehitettäessä on pyritty ottamaan entistä paremmin huomioon nykyaikaiseen kuvailuun, tiedonhakuun ja tietosisältöjen yhteentoimivuuteen liittyvät tarpeet. YSO tarjoaakin käyttöön seuraavat YSAan nähden uudet ominaisuudet:

Monikielisyys ja rikastettu termistö

YSON termistö on esitetty kolmella kielellä: suomeksi, ruotsiksi ja englanniksi. Monikielisyys mahdollistaa tiedonhaun yli kielirajojen: tiedonhakija voi hakea tietoa suomeksi ja saada hakutuloksiinsa myös ruotsin- ja englanninkieliset tulokset.

Kuvailussa monikielinen sanasto voi vähentää päällekkäistä työtä: kuvailu tarvitsee tehdä vain kertaalleen kullekin käsitteelle. Tällä hetkellä esimerkiksi Melindasta löytyy lukuisia tietueita, joissa on erikseen suomen- ja ruotsinkieliset asiasanat, vaikka itse käsitteet termien takana ovatkin samoja (ks. kuva 1). Näissä tapauksissa sisällönkuvailu on tehty kahteen kertaan, ensin suomenkielillä (YSA) ja sitten ruotsinkielisellä sanastolla (Allärs). Kun kuvailu tehdään monikielisellä sanastolla kertaalleen, päästään myös eroon nykyisen erikielisten termien määrän epäsuhdasta (kuva 1).

Viitenäyttö

Koko viite | Lyhyt viite | MARC-näyttö

Viite 1 (yhteensä 1)

Tietueen ID	006945110
Tekijä	■ Bowen, James, kirjoittaja.
Teos	■ The world according to Bob, suomi
Nimeke	■ Bobin maailma / James Bowen ; suomentanut Kimmo Paukku.
Muut nimekkeet	■ Alanimeke kannessa: Viisaan kissan ja katusoitajan myöhemmät seikkailut
Julkaistu	Helsinki : WSOY, © 2015 (painettu EU:ssa)
Ulkoasu	263 s. : kuv. ; 20 cm.
Aineisto	Tekstiaineisto
Kieli	fin
Alkuteoksen kieli	eng
ISBN	■ 978-951-0-40913-8 (sid.)
EAN	9789510409138
Muu(t) tekijä(t)	■ Paukku, Kimmo, käänttäjä.
Asiasanat ja luokitukset Kohdehenkilö Asiasana ■ Bowen, James. ■ löytöeläimet (ysa) ■ kissa (ysa) ■ vuorovaikutus (ysa) ■ ihminen (ysa) ■ eläimet (ysa) ■ lemmikkieläimet (ysa) ■ kodittomuus (ysa) ■ miehet (ysa) ■ muistelmat (ysa) ■ löytöeläimet -- kissa (ysa) ■ vuorovaikutus -- ihminen -- eläimet (ysa) ■ selviytyminen -- irtolaisuus (ysa) ■ upphittade djur (allars) ■ katt (allars) ■ djur (allars) ■ sällskapsdjur (allars) ■ människan (allars) ■ vaxelverkan (allars)	

Kuva 1. Nykyinen sisällönkuvailuterminien päällekkäisyys ja erikielisten termien lukumäärien epäsuhta.

YSON käsitteiden termistöä on rikastettu myös muun muassa yksikkömuotoisilla termeillä. Toisin kuin YSAssa, YSOssa kuvailija ja tiedonhakija löytävät käsitteen ”joet” myös hakusanalla ”joki”. Käyttäjän ei tarvitse siis tuntea sanastostandardien määräytyksiä termien muodoista käyttääkseen sanastoa tehokkaasti.

Pysyvät koneluettavat tunnisteet

Asiasanastoilla tehtävä kuvailu on perustunut termeihin: bibliografisiin tietueisiin on merkitty kutakin asiasanaston käsitettä vastaava merkkijono, sekä käytetyn asiasanaston tunnus. Asiasanaston käsitteen muut tiedot kuten ohjaustermit ja suhteet muihin käsitteisiin on ilmaistu erillisissä auktoriteettitietueissa.

Termipohjaisessa kuvailussa kone ei pysty ”tekemään” kuvailun termin pohjalta paljoakaan; bibliografisen tietueen termi näyttäytyy tietokoneelle mielivaltaisena merkkijonona. Termi on myös aina sidottu tiettyyn kieleen, jonka sisällä se voi viitata useisiin asioihin eli käsitteisiin. Monikielisen kuvailun ollessa tavoitteena on aina tehtävä lisää työtä ja lisättävä erikseen haluttavilla kielillä olevat termit tietueisiin. Termipohjaisen kuvailun päivittäminen on myös hankalaa.

Ontologioilla tehtävä sisällönkuvailu perustuu pysyviin URI-tunnisteisiin. Tunnisteet viittaavat asioihin; asioilla voi puolestaan olla erilaisia nimityksiä usealla kielellä, suhteita toisiin asioihin ja muita ominaisuuksia. Tunnisteen avulla saadaan kaikki yhtä asiaa merkitsevän käsitteen ominaisuudet koneiden käsiteltäviksi ja hyödynnettäviksi.

Ontologisessa kuvailussa tunniste tallentuu bibliografisiin tietueisiin. Tunnisteen avulla kone voi päätellä asioita kuvailun käsitteen pohjalta, tehdä automaattisia haunlaajennoksia alakäsitteisiin, ohjata ohjaustermeillä tehtyt haut oikeisiin käsitteisiin jne. Tunniste on myös joustavampi käsitteistön päivittyessä: jos jonkin asian nimitys muuttuu, voidaan sen tunniste kuitenkin pitää samana.

Kuvailutyötä tekevän ei tarvitse nähdä tai käsitellä tunnisteita. Tavoitteena on, että kuvailija voi tehdä kuvailun millä tahansa monikielisen ontologian kielellä ja tunniste tallentuu tietueeseen automaattisesti.

Käsitetaso

YSAssa joidenkin asiasanojen merkitys voi olla hämärä: asiasanalla ei ole suhteita toisiin sanoihin eikä sillä ole sen merkitystä selventävää huomautusta. Tämä voi tietyissä tapauksissa hidastaa kuvailua; kuvailija joutuu miettimään onko kyseessä oleva asiasana juuri se hänen tarvitsemansa. YSAssa myös yksi termi on voinut viitata useampaan käsitteeseen; tällöin esimerkiksi hakutuloksiin tulee väistämättä hälyä.

YSON kehityksen lähtökohtana ovat käsitteet eli merkityksen tai tiedon yksiköt. Käsitteet pyritään rajaamaan niin, että sekä kuvailija että tiedonhakija ymmärtää mistä asiasta kulloinkin on kyse. Käsitteisiin viittaavat termit valitaan niin, että tietty termi viittaa aina ainoastaan yhteen tarkoitteeseen.

YSON koko ontologian kattava hierarkkinen rakenne takaa sen, että jokaiselle käsitteelle on olemassa konteksti. Näin kuvailijan tai tiedonhakijan ei tarvitse arvailla termien takana mahdollisesti olevien käsitteiden merkitystä.

Yhtenäinen hierarkia mahdollistaa myös erilaisia tiedonhakusovelluksia kuten automaattisen haunlaajennoksen ja hakutulosten ryhmittelyn. Näiden pohjalta voidaan kehittää hakupalveluihin uudenlaisia tehokkaita tiedonhakuominaisuuksia.

Linkittyvyys

YSON kautta kirjastodata saadaan osaksi laajaa julkisen sektorin tietosisältöjen verkostoa. Linkitysten avulla sisältöjä voidaan yhdistää uusilla ja innostavilla tavoilla ja tiedonhakijoille voidaan tarjota uudenlaisia mahdollisuuksia navigoida verkossa järjestelmästä tai palveluntuottajasta riippumatta.

YSO on linkitetty tällä hetkellä seuraaviin sanastoihin ja ontologioihin:

KOKO-ontologia

KULO - Kulttuurien tutkimuksen ontologia

JUHO - Julkishallinnon ontologia

Terveysten ja hyvinvoinnin ontologia (Tero)

MERO - Merenkulkualan ontologia

MUSO - Musiikin ontologia

KAUNO - Kaunokki-ontologia

KITO - Kirjallisuudentutkimuksen ontologia

VALO - Valokuvausalan ontologia

LIITO - Liiketoimintaontologia

MAO/TAO - Museoalan ja taideteollisuusalan ontologia

TSR-ontologia

KTO - Kielitieteen ontologia

AFO - Luonnonvara- ja ympäristöontologia

JUPO - Julkisen hallinnon palveluontologia

PTO – Paikkatieto-ontologia

KEKO – Kestävän kehityksen kasvatuksen ontologia

LCSH - Library of Congress Subject Headings

Käytännössä tämä tarkoittaa sitä, että millä tahansa näillä ontologioilla kuvaillut tietosisällöt voidaan yhdistää kirjastojen YSOlla kuvailtuihin aineistoihin. Tiedonhakija voidaan siis ohjata esimerkiksi Melindasta Kirjasampoon (KOKOn käyttäjä) tai YLEn (KOKOn käyttäjä) verkkosivuille hakemaan lisätietoa etsimästään aiheesta. Kongressin kirjaston asiasanastolla kuvailtujen tietueiden sisällönkuvailutermit voidaan automaattisesti muuntaa YSON käsitteiksi (jos vastaavuus on olemassa).

YSO - Yleinen suomalainen ontologia

Sisällön kieli suomi x Hae

Aakkosellinen

Hierarkia

Ryhvät

- oliot
 - abstraktit objektit
 - fyysiset kokonaisuudet
 - fyysiset objektit
 - järjestelmät
 - paikka
 - alueet
 - ihmisen määrittämät paikat
 - luonnon muodostamat paikat
 - hiidenkirmut
 - kasvupaikat
 - koralliriutat**
 - kraatterit
 - muut paikat
 - ominaisuudet
 - tapahtumat ja toiminta

oliot > paikka > luonnon muodostamat paikat > koralliriutat

KÄYTETTÄVÄ TERMI

koralliriutat

YLÄKÄSITE

luonnon muodostamat paikat

ASSOSIATIIVISET KÄSITTEET

laguunit

KORVATUT TERMIT

atollit

HUOMAUTUS

Runkokunnissa elävien korallien kalkkirangoista, simpukoista ym. muodostunut särkkä.

LÄHDE

Kielitoimiston sanakirja

KUULUU RYHMÄÄN

11 Maantiede. Kartografia. Kartoitus. Geodesia. Geologia. Paleontologia
13 Hydrologia

MUUNKIELISET TERMIT

coral reefs

englanti

korallrev

ruotsi

atoller

URI

http://www.yso.fi/onto/ys/p14886

Lataa tämä käsite:

RDF/XML Turtle JSON-LD

LÄHES VASTAAVA KÄSITE

Coral reefs and islands (en)

Library of Congress Subject Headings

koralliriutat

YSA - Yleinen suomalainen asiasanasto

korallrev (sv)

Allärs - Allmän tesaurus på svenska

Kuva 2. YSON käsite "koralliriutat" Finton käsitenäkylässä.

Siirtymä käytännössä

YSON käyttöönoton vaatimien toimenpiteiden määrä riippuu kirjastojen aiemmista sisällönkuvailukäytännöistä sekä käytössä olevista järjestelmistä. Jos kuvailussa on käytetty YSAa, tulee bibliografisen datan asiasanaketät konvertoida vastaamaan YSON käsitteitä. Tämän lisäksi YSO tulee tuoda osaksi kuvailijan työpintaa sujuvan sisällönkuvailun mahdollistamiseksi.

YSAlla kuvailtujen tietueiden konvertointi

YSA on ollut käytössä pitkään kirjastojen sisällönkuvailuvälineenä. YSON käyttöönoton yhteydessä YSAlla kuvailtujen bibliografisten tietueiden sisällönkuvailukentät on hyvä muuntaa YSO-kentiksi. Tämä on tehtävissä automaattisesti YSON sisältämien YSA-linkitysten avulla (Kuva 3.).

YSA - Yleinen suomalainen asiasanasto Sisällön kieli suomi x Hae

dokumentaatiokielet > asiasanastot

KÄYTETTÄVÄ TERMI **asiasanastot**

YLÄKÄSITE [dokumentaatiokielet](#)

ASSOSIIATIIVISEET KÄSITTEET [sanastot](#)
[tesauрукset](#)

KUULUU RYHMÄÄN [81 Kirjastot. Arkistot. Tietopalvelu. Museot. Näyttelyt](#)

URI <http://www.yso.fi/onto/ysa/Y94743>

Lataa tämä käsite: [RDF/XML](#) [TURTLE](#) [JSON-LD](#) Luotu 31.8.1986, viimeksi muokattu 23.5.2016

LÄHES VASTAAVA KÄSITE [asiasanastot](#) YSO - Yleinen suomalainen ontologia

VASTAAVA KÄSITE [ämnasregister \(sv\)](#) Allars - Allmän tesaurus på svenska

Kuva 3. YSAn käsitteen linkki vastaavaan YSO-käsitteeseen.

Konvertoinnissa tulee kuitenkin määrittää se, miten tunnistepohjaiset käsitteet ilmaistaan käytössä olevan kuvailuformaatin (MARC) mukaisissa tietueissa termipohjaisten asiasanojen sijaan. Keskeisiä kysymyksiä ovat mm. tunnisteen sopivat osakentät sekä monikielisten termien tallentaminen. Lisäksi on mietittävä miten YSON ja YSAn eroavuudet otetaan huomioon muutoksen yhteydessä ja kuinka paljon tarkistustyötä vaaditaan automaattisen konvertoinnin lisäksi.

Konversioista on tekeillä oma yksityiskohtainen suunnitelmansa Finto-projektissa.

YSON käyttöönotto kuvailuvälineissä

YSON voi saada käyttöönsä usealla tavalla. Helpoin ja suositeltava tapa on kytkeä Finto-palvelun rajapinnat (api.finto.fi) omaan kuvailujärjestelmään, jolloin mitä tahansa Finton sanastoa voidaan käyttää suoraan lähteestä aina ajan tasalla olevana versiona. Näin toimivat jo nyt esimerkiksi Axiellin Aurora-järjestelmää käyttävät kirjastot.

Jos käytössä olevaan järjestelmään ei voida kytkeä Finton rajapintoja, voidaan YSO tuoda järjestelmän sisäiseksi sanastoksi. MARC-formaattia käyttävissä kirjastoissa tämä tarkoittaisi YSON käsitteiden muuntamista auktoriteettitietueiksi, joita voitaisiin sitten käyttää kuvailussa samaan tapaan kuin YSAn käsitteitä (niillä joilla YSA on ollut järjestelmässä).

YSOa voidaan käyttää myös suoraan Finton selailukäyttöliittymän kautta, jolloin käsitteiden tunnistet ja/tai termit kopioidaan Fintosta kuvailtaviin tietueisiin. Tämä ei kuitenkaan ole järkevä ratkaisu kuvailua täysipäiväisesti tekeville.

YSON käyttöönoton seuraukset

YSON käyttöönotto mahdollistaa monia kirjastokuvailun kannalta toivottavia asioita: monikielistä sisällönkuvailua voidaan helpottaa, linkityksiä hyödyntämällä voidaan vähentää päällekkäistä työtä ja kirjastoaineisto tuodaan osaksi julkista tiedon verkostoa. Ontologiat mahdollistavat myös hienostuneemmat ja älykkäämmät tiedonhauet aineistoon ontologian rakennetta hyödyntäen, joka helpottaa kirjaston käyttäjien arkea. YSOa voidaan lisäksi hyödyntää sisällönkuvailun ohella myös muussa kuvailutyössä.

Ontologioiden käyttöönoton myötä kirjastot siirtyvät kohti linkittyvää semanttista verkkoajaka, joka perustuu ihmisten ja koneiden saumattomaan yhteistyöhön. Keskiöön nousevat tiedonhakijoita kiinnostavat asiat, eivätkä painetun sanan aikakaudelta tutut, tietokoneille kasvottomat merkkijonot. "Things, not strings", kuten Google asian ilmaisee. Kirjastomaailmassa asiaa on edistetty esimerkiksi OCLC:n toimesta ja suuret kirjastoalan kehittämishankkeet kuten FRBR, RDA ja BIBFRAME ovat kaikki askeleita kohti linkitettyä, koneiden hyödynnettävissä olevaa kirjastodataa. Ontologiat ovat osa tätä kehitystä. Niiden käyttöönotolla kirjastojen rikkaat tietoaaineistot voidaan saada tehokkaasti ja uudella tavalla verkkomaailman hyödynnettäviksi.