

YSAsta YSOon: kesäkuulumiset ja syysuunnitelmat

Jarmo Saarikko, Kansalliskirjasto
Missä mennään -webinaari 20.9.2019

Kesäkuulumiset

- **Melindan yso-konversio** toteutettiin heinäkuussa
- Kaikki **ysa, allars, musa** ja **cilla** -merkityt termit konvertoitiin
- Konversio käsitteli em. termit kentissä 385, 567, 648, 650, 651 ja 655
- Konversiossa purettiin ketjut em. kentissä
- Termejä siirrettiin myös kenttiin: 257, 370, 382, 386, 388 ja 653
- Käytettiin sanastoja: **yso, yso-paikat, slm, seko**
 - Näihin sanastoihin kuulumattomat termit siirrettiin kenttään 653 (+ 2. indikaattori)

Konversion toteutus oli siis huomattavasti laajempi ja monimutkaisempi kuin alun perin kaavailtu “pelkkä” asiasanakenttien YSA:sta YSOon konversio. Konvertoitiin kaikki aineistotyypit, joissa oli käytetty em. sanastoja kuvailussa.

Melindan konversion tuloksia

- Noin 10 miljoonaa tietuetta, joita ei konvertoitu mitenkään
 - Noin 4.9 miljoonaa tietuetta konvertoitiin
 - Poistettiin noin 23 miljoonaa kenttää
 - Tuotettiin 45 miljoonaa **YSO** ja **SLM**-termiä (n. 22 milj. suomeksi ja ruotsiksi)
 - Tuotettiin hieman vajaa 0.5 miljoonaa **seko**-termiä
-
- Fennica ja Viola konvertoitiin ennen Fikkaan viemistä
 - <https://kansalliskirjasto.finna.fi/>

Musiikki- ja elokuva-aineistot, pelit

- Nämä konvertoitiin eri tavalla kuin tekstiaineistot
- Vastaavat termit valittiin ensin SLM sanastosta ja vietiin 655 kenttään
- “aiheet” -termin jälkeiset termit valittiin aina YSO:sta
- “aiheet” ja “musiikki” -termit poistettiin
- Samaan ketjuun kuuluneiden termeille lisättiin ko. termin uuden kentän $\ddagger 8$ -osakenttään ketjun järjestysnumero muodossa $\ddagger 81 \setminus u$, $\ddagger 82 \setminus u$, jne.

Melindan konversiossa havaittuja ongelmia

Kaikki havaitut ongelmat on pyritty korjaamaan konversio-ohjelmaan tai sanastoihin, jotteivät ne toistuisi seuraavissa konversioissa.

- Joissakin tapauksissa sanastoista löytynyt “oikea” termi ei viittaakaan samaan käsitteeseen ja sanastossa on vain tuo yksi (väärä) vaihtoehto
 - Esimerkiksi *Viborg*, jolla tarkoitettiin tanskalaista kaupunkia, konvertoitui Suomen Viipuriksi.
 - Muita hankalia tapauksia mm. *Kemi*, *Georgia*, *Lappi*
- Ruotsinkielinen soitin on korvautunut suomenkielisellä seko-termillä.
 - Koska seko on yksikielinen, ruotsinkielinen termi poistui kokonaan.
- Joissakin tapauksissa aihe-termi oli korvattu virheellisesti muototermillä tai päinvastoin
 - Tämä johtui eri aineistotyyppien erilaisesta käsittelyjärjestyksestä vastintermien haussa

Konversio-ohjelma päivitetty

- Kaikki tähän mennessä havaitut tulkintaongelmat konversion tuloksissa ja ohjelman tuottamilla tarkistuslistoilla on korjattu ohjelmaan. Mm. aineistotyyppien tunnistamista ja termien käsittelyjärjestystä on parannettu
- Konversio-ohjelmaan tehdyt muutokset ja versiopäivitykset on dokumentoitu githubissa.
- Konversiota harkitessanne poimikaa aina uusimmat versiot ohjelmasta ja sanastoista
- Korjauksia tehdään edelleen, jos konversioissa havaitaan ongelmia.

Syysuunnitelmat

- Loppukesälle suunniteltu Voyageria käyttävien ja Voyagerista Kohaan siirtyneiden korkeakoulukirjastojen paikalliskantojen konversio on siirtynyt syyskuun lopulle
- Uudet termit näkyvät **Finnassa** vasta paikalliskannan konvertoinnin jälkeen
 - Fennica ja Viola on konvertoitu. Konvertoidut yso- ja slm-termit näkyvät **kansalliskirjasto.finna.fi** -palvelussa haettaessa *Kansalliskirjaston kokoelmat*
- Konversio-ohjelman tuottamat **tarkistuslistat** käydään läpi ja tehdään niiden mukaisesti päivityksiä Melindan MARC-tietueisiin

Konversio-ohjelman tuottamat tarkistuslistat

Konversio-ohjelma tuottaa tarkistuslistalle kaikki ne kentät, joita ei voida konvertoida suoraan. Runsaimpia näistä voi konvertoidaan joko erillisajoina, mutta suurin osa käsiteltävä yksitellen

- Termit, joita ei löytynyt sanastoista
 - Vapaan indeksoinnin termit, kirjoitusvirheet, muunkieliset termit
- Käsitteet, joille ei löytynyt yksiselitteistä vastinetta
 - Useimmissa tapauksissa käsitteelle löytyi useampia vastineita
 - Esim. termillä on sulkutarkenteellinen muoto kahdessa tai useammassa käsitteessä
 - Osa yso-termejä, kun olisi pitänyt olla slm-termi tai päinvastoin
 - Kaikki 653 kenttään siirretyt termit ovat mukana tarkistuslistalla
- Joitakin erikoistapauksia (‡g-, ‡3-, ‡6-osakenttä, ym.)

Melinda-konversion tarkistuslista

Noin **800.000** kenttää (3,5% poistetuista kentistä)

- Termille löytyi useampi mahdollinen vaihtoehto
 - **160.000** kenttää, 1.400 eri termiä, joista 300 esiintyi vain 1-2 kertaa
- Jokin kentän termi ei löytynyt käytetyistä sanastoista
 - **600.000** kenttää, 170.000 eri termiä, joista 300 esiintyi >100 ja 135.000 1-2 kertaa
- Poistetut termit
 - **21.000** kenttää (musiikki, musik, aiheet, fiktio, fiktion)
- Kentässä ±6 osakenttä - kyrilliset kentät käsiteltävä manuaalisesti
 - **3.400** kenttää
- Melindan muissa kuin konvertoiduissa kentissä olleet ysa- ja allars-termit on myös kartoitettu ja suurelta osin käsitellään manuaalisesti (**3.400** kenttää)

Käsitteitä, joita ei voi konvertoida suoraan

- ❖ Syynä eroihin mm. yksikielisestä monikieliseen sanastoon siirtyminen sekä muut sanastoissa tapahtuneet muutokset
 - mallit
 - mallit (mallintaminen)
 - mallit (ammatit)
 - käsityömallit
 - roolimallit
 - käsittemallit
 - valumallit
 - messut
 - messut (tapahtumat)
 - messut (sävellykset)
 - messut (jumalanpalvelus)
 - opetusohjelmat
 - opetusohjelmat (oppilaitokset)
 - opetusohjelmat (tietokoneohjelmat)
 - opetusohjelmat (media)
 - ohjelmat
 - ohjelmat (suunnitelmat)
 - ohjelmat (media)
 - tietokoneohjelmat
 - ohjaus
 - ohjaus (taiteet ja media)
 - ohjaus (neuvonta ja opastus)
 - ohjaus (hallinta)
- ❖ Tässä vain muutamia esimerkkejä. Oikea vaihtoehto on valittava tutkimalla tietuetta.

Lisätietoja

- ❖ Konversiosäännöt sekä konversion valmistelu ja toteutus
 - <https://www.kiwi.fi/display/ysall2yso>
- ❖ Konversio-ohjelma
 - <https://github.com/NatLibFi/yso-marcbib>
 - Python-kielinen avoimen lähdekoodin ohjelmisto

Kiitos mielenkiinnosta!

jarmo.saarikko@helsinki.fi

<https://orcid.org/0000-0002-6801-6151>

Runsaimpien 653 kenttään siirrettyjen termien tyypit

- Yht. 293 termiä, joita yli 100 kpl Melindan konversion tarkistuslistalla

▪ aihe	81
▪ muoto	75
▪ paikka	75
▪ yhteisö	21
▪ teos	11
▪ henkilö	8
▪ seko	7
▪ tuote	3
▪ virhe	3
▪ English	6
▪ Deutsch	3

Syitä mm.

- Kirjoitusvirhe
- Tieto oli väärässä kentässä
- Aihe-, paikka tai muototermi puuttuu sanastoista