

Ajankohtaista Finto-palvelusta

Mikko Lappalainen

Missä mennään -webinaari 15.2.2022



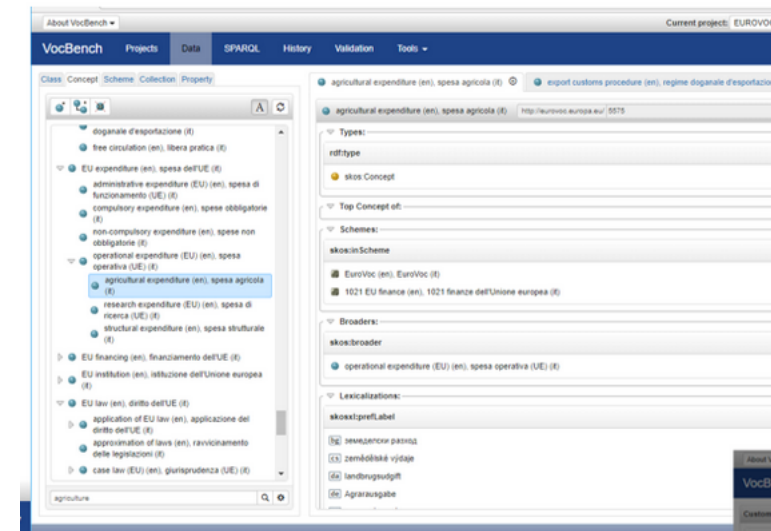
Finto-palvelun 2022 painopistealueet

- Uuden ontologiaeditorin käyttöönottoprojektin jatkaminen
- Finto.fi/Skosmos-kehitys käyttäjälähtöisesti
- Keskeisten ontologioiden (YSO, KOKO ym.) kehittäminen erityisesti linkitysten, päivitysnopeuden ja laadun näkökulmasta
- Koulutus- ja ohjeistustarjonnan kehittäminen

- Vuoden 2022 toimintasuunnitelma:
<https://www.kiwi.fi/display/Finto/Toimintasuunnitelmat>

Uuden ontologia/sanastoeditorin käyttöönotto

- Uuden verkkopohjaisen VocBench-sanastoeditorin käyttöönottoa edistetään yhteistyössä Sanastokeskuksen kanssa
- Tavoitteena sanastojen ja ontologian muokkauksen helpottaminen, sekä sanastojen päivityspotken virtaviivaistaminen
- Mahdollistaa tulevaisuudessa myös eri sanastojen entistä tiiviimmän yhteiskehittämisen

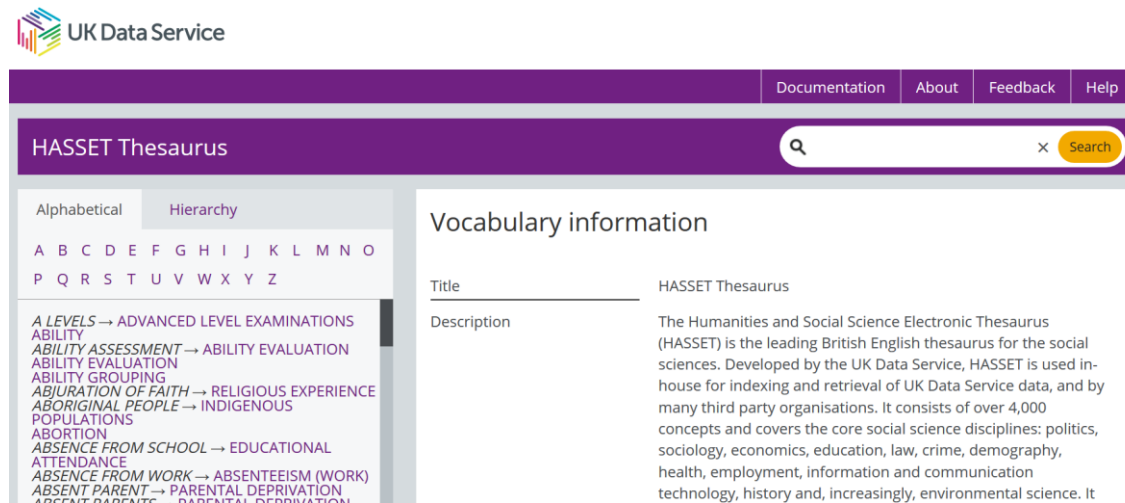


Skosmoksen/Finto.fi:n kehitys

- Skosmos on Finto.fi:n taustalla oleva, Kansalliskirjastossa kehitettävä avoimen lähdekoodin ohjelmisto
- Skosmos-kehitystä tehdään aktiivisesti teemoittaisissa kehityssprinteissä, [linkki Githubiin roadmap-sivulle](#)
- Kehitystä voi seurata [GitHubissa Skosmoksen työjonojen kautta](#)
- Kehitysehdotuksia voi tehdä suoraan Githubiin, tai laittaa sähköpostilla os. finto-posti@helsinki.fi

Skosmos-ohjelmiston käyttö maailmalla

- Uusin käyttöönotto: **UK Data Service** julkaisi [HASSET-thesauruksen Skosmos-alustalla](#) elokuussa



- Muita Skosmos-käyttäjiä mm. UNESCO, FAO, CNRS, EHESS, CESSDA, Norjan ja Latvian Kansalliskirjasto, Oslon yliopiston kirjasto jne.

Ajankohtaista sanastoista ja ontologioista (1/4)

- **Lajisto julkaistaan Finto.fi:ssä kevään 2022 aikana**
 - Lajitietokeskuksen lajiluettelo
 - Laji.fi:n tietoihin pohjautuva, Luomuksen ylläpitämä sanasto

finto Suomalainen asiasanasto- ja ontologiapalvelu

Sanastot Tietoja Palaute Ohje

LAJISTO - Lajitietokeskuksen lajiluettelo

Sisällön kieli suomi

Aakkosellinen Hierarkia

nisäkkäät
hyönteissyöjät
jyrsijät
jäniseläimet
kädelliset
lepakot
petoeläimet
hylkeet
karhut
kissaeläimet
koiraeläimet
ketut
koirat
susi
supikoirat
näätäeläimet
sorkkaeläimet
valaat
dvrstökkäät

eliökunta > aiotumaiset > eläimet > selkäjänteiset > nisäkkäät > petoeläimet > koiraeläimet > koirat > susi

KÄYTETTÄVÄ TERMI **susi**

TYYPPI Laji

YLÄKÄSITE koirat

TIETEELLINEN NIMI Canis lupus

AUKTORI Linnaeus, 1758

MUUNKIELISET TERMIT varg ruotsi

URI http://tun.fi/MX.46549

LATAA TÄMÄ KÄSITE: RDF/XML TURTLE JSON-LD

Ajankohtaista sanastoista ja ontologioista (2/4)

- “Aikaontologia” julkaistaan Finto.fi:ssä kevään aikana
 - Erityisesti kirjastojen ja KAM-sektorin kuvailutarpeisiin kehitetty, ajan ilmaisuja sisältävä sanasto

.....finto Suomalainen asiasanasto- ja ontologiapalvelu Sanastot Tietoja Palaute Ohje

KRONO Sisällön kieli suomi

Aakkosellinen Hierarkia

0000/0999 Ensimmäinen vuosituhat
1000/1999 Toinen vuosituhat
2000/2999 Kolmas vuosituhat
2000/2099 2000-luku
2000/2009 2000-luku (vuosikymmen)
2010/2019 2010-luku
2020/2029 2020-luku
2030/2039 2030-luku
2040/2049 2040-luku
2050/2059 2050-luku
2060/2069 2060-luku
2070/2079 2070-luku
2080/2089 2080-luku
2090/2099 2090-luku
2100/2199 2100-luku
2200/2299 2200-luku
2300/2399 2300-luku
2400/2499 2400-luku
2500/2599 2500-luku

Kolmas vuosituhat > 2000-luku > 2010-luku

KÄYTETTÄVÄ TERMI **2010/2019 2010-luku**

TYYPPI Vuosikymmen

OSA KÄSITETTÄ 2000/2099 2000-luku

LOPPUAIKA 12/31/19

ALKUAIKA 1/1/10

NOTAATIO 2010/2019

MUUNKIELISET TERMIT **2010-talet** ruotsi

URI <http://urn.fi/URN:NBN:fi:au:krono:2010/2019>

LATAA TÄMÄ KÄSITE: [RDF/XML](#) [TURTLE](#) [JSON-LD](#)

Ajankohtaista sanastoista ja ontologioista (3/4)

■ Yhteistyö

- Kotimaisten kielten keskuksen (Kotus) kanssa keskustellaan yhteistyöstä kielten nimien sanaston osalta
- YSON pohjoissaamentamista edistetään erillisen hankeidean kautta yhteistyössä saamelaisalueiden toimijoiden kanssa

Ajankohtaista sanastoista ja ontologioista (4/4)

- Työn alla olevia sanastoja
 - **ISO 5127:2017** - Information and documentation, suomennos
 - **Mediataiteen ontologia OMA**
 - Mediataiteen historia (MEHI) –hankkeessa kehitettävä YSO-pohjainen erikoisontologia

Julkaisuja

- *Annif and Finto AI: Developing and Implementing Automated Subject Indexing* julkaistu J LIS.it -lehdessä (Italian Journal of Library, Archives, and Information Science)
- *Automaattisen kuvailun arvoituksia Kirjastoverkkopäivillä 2021* julkaistu Tietolinjassa, 2021(2).
- *Sisällönkuvailun ammattilaisten kokemuksia Finto AI:n käytöstä*, julkaistu Doriassa
- *Automaattisen sisällönkuvailun ohjelmiston rakentaminen – case Annif* julkaistu Signumissa

Loppukevennys: *kuinka hyödynnät Finton ontologioita ja Annifia yhteiskunnallisen somekeskustelun analysoinnissa*

#minätutkin-twiittien aiheet

Oct 6, 2021 tekijä Tuija Sonkkila

Tägit: [twitter](#) [finto](#)

Kategoriat: [R](#) [asiasanoitus](#) [ontologiat](#) [visuaalisointi](#)

Kuuksi sitten, syyskuun alussa 2021, Twitterin #minätutkin oli suosittu. Itä-Suomen yliopiston akateeminen rehtori Tapio Määttä [sysäsi ajatuksen liikkeelle](#), ja pian mukaan liitettiin sopiva aiheutunniste. Tutkijat innostuivat tiivistämään työnsä twiitin merkimmäkseen, ja suuri yleisö innostui yhtä lailla lukemastaan. Vaikka Suomesta löytyy ajantasaisia sivustoja kotimaisesta tutkimuksesta, uusimpana [tutkimustietovaranto](#), niiden varsinaisen fokus ei ole tehdä tiedettä helposti lähestyttäväksi. Toki ne saattavat onnistua siinäkin, mutta vain välillisesti.

Mistä sitten twiitattiin? Mitä olivat ne tieteenalat, joiden tutkijat ehdotivat tai kiinnostuivat twiittaamaan? Kaikki eivät tietenkään edes kuulleet asiasta.

Päätin tehdä pienimuotoisen testin siitä, miten Kansalliskirjaston [Finto AI -rajapintapalvelu](#) palvelisi twiittien automaattisessa asiasanoituksessa. Heti alkuun pitää painottaa siitä, että Finto AI/Annif on toistaiseksi pelkkä suosittelija; ihmisen on tarkoitus tehdä lopulliset asiasanavalinnat. Näin toimitaan mm. Jyväskylän yliopiston JYX-julkaisuarkistossa, jossa palvelua käytetään opinnäytetöiden kuvailun apuna:

Opinnäytetyönsä järjestelmään jättävä opiskelija saa nähtäväkseen Annifin ehdottamat asiasanat, joita hän voi halutessaan muokata. Lopulliset asiasanat hyväksyy kirjastoavustaja tarkistuksen jälkeen. Myös Kirjavalitus Oy käyttää rajapintapalvelua tuottaessaan ennakkotietoja tulevista kirjajulkaisuista. Vastaavanaista käyttöä pilotoidaan tällä hetkellä mm. Vaasan yliopiston Osuva-julkaisuarkistossa.

Tässä harjoituksessa annan kuitenkin mennä täysautomaatilla, kädet poissoita ratilta.

Alla olevat koodit ovat [GitHub-repossa](#).

Twiittien haku ja siivous

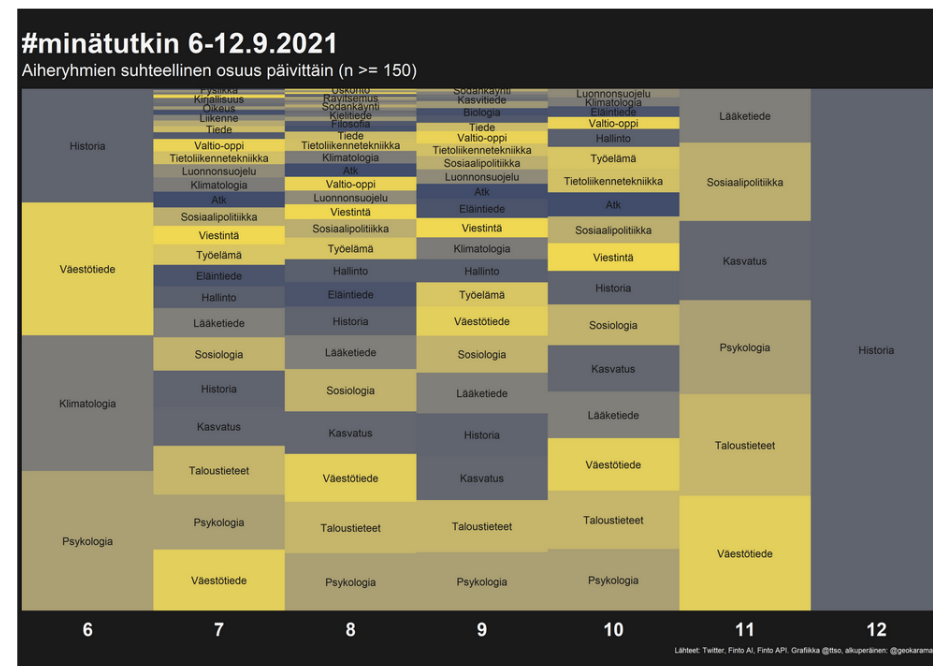
Kun haluaa kerätä twiittejä ilman erityisjärjestelyjä, on toimittava ripeästi. Vaikka #minätutkin-häntää on näkynyt aina näihin päiviin asti, hain tässä käsitellyt twiitit jo sunnuntaina 12.9. alkuiltapäivästä enkä sen jälkeen enää uudistanut hakua. Viikkaimmat päivät olivat tiistai-torstai 7-9.9.

Harto Pöngän [10.9 keräämien tilastojen mukaan](#) 5000 twiittiä vaikutti sopivalta ylärajalta. Ei uudelleentwiittauksia.

```
library(rtweet)
q <- "#minätutkin"
tweets <- search_tweets(q, n = 5000, include_rts = FALSE)
```

Siivous on enimmäkseen käsityötä. Aivan aluksi poistin selkeimpiä trollauksia, lukijoiden kiittäviä postauksia ja muita tässä yhteydessä epärelevanttejä. Myöhemmin kävin vielä twiitit kertaalleen läpi ja filteroitiin pois ne, jotka eivät olleet tutkijoilta itseltään vaan suurelta yleisöltä, tutkimusorganisaatioilta tai rahoittajalta. En edes pyrkinyt täydelliseen siivoukseen, ja lisäksi twiitteissä oli rajatapaaksia.

Karamanisilla musiikkityylit ovat nimen mukaisessa aakkosjärjestyksessä, suuruuserot käyvät ilmi kirjaimen koosta. #minätutkin-esitykseen halusin sen sijaan käsiteryhmät koon mukaisessa järjestyksessä, mielellään niin että samalla ryhmällä on aina sama väri. Ollaan faktoreiden (factor) järjestelyn alueella. Lopputuloksessa on parantamisen varaa, sillä löytämäni [keino tehdä tämä](#) rikkoo ryhmäväriyksen, kun ryhmä on enemmän kuin 35. Näin olen rajasin niiden lukumäärä niin, että jätin pois harvemmin esiintyneet.



#minätutkin-twiittien aiheet 6-12.9.2021

Lisää ajankohtaisia asioita

- [Finto-palveluiden asiakaswikin ajankohtaista-osio](#)
- Finto-palvelun twitter: @Fintopalvelu
- [Finto-tiedotus –sähköpostilista](#) asiakaswikissä
- [Kansalliskirjaston uutiskirjeet](#)
- Palvelupostiosoite: finto-posti@helsinki.fi



www.kansalliskirjasto.fi

Finto-posti@helsinki.fi

Mikko.lappalainen@helsinki.fi