

Ajankohtaista Fintosta ja Annifista

Ajankohtaista

- Artikkelin *Annif and Finto AI: Developing and Implementing Automated Subject Indexing* on luettavissa preprint-versiona os. <https://urn.fi/URN:NBN:fi-fe2021080942632>. Se on hyväksytty julkaistavaksi [JLIS.it](https://www.jlis.it) -lehdessä (Italian Journal of Library, Archives, and Information Science) ja ilmestyyneen vuoden 2022 alussa. Artikkelissa esitellään Annifia ja Finto AI:ta uusine kuulumisineen, pureudutaan uusien algoritmien suorituskykyyn sekä automaattisen sisällönkuvailun käyttökohteisiin.
- Finto-palvelu on mukana ideoimassa hanketta, jonka tavoitteena on kehittää saamen kielten tukea digitaalisiin palveluihin. Finton osalta hankkeessa olisi sen toteutuessa kyse erityisesti pohjoissaamenkielisten termien lisäämisestä YSON käsitteisiin.
- Tutkimuksen datapalveluihin keskittynyt Yhdistyneiden kuningaskuntien kansallinen UK Data Service on julkaissut oman HASSET-sanastonsa uudella Skosmos-pohjaisella alustalla: <http://hasset.ukdataservice.ac.uk/> HASSET eli The Humanities and Social Science Electronic Thesaurus on brittien johtava yhteiskuntatieteellinen sanasto, jota hyödynnetään laajasti UK Data Servicen ja muiden organisaatioiden aineistojen kuvailussa ja tiedonhaussa.

Sanastot ja ontologiat

- **KAUNO**a, fiktivisen aineiston ontologiaa on rikastettu lähes 200 käsitteen osalta, joihin on nyt liitetty ryhmätieto. Poimintana voidaan mainita, että esimerkiksi käsite *ruutu aika* kuuluu ryhmään *Teema*. Lisäksi Ryhmät-välilehdellä näkyy nyt myös termiryhmä *Aika*, jota klikkaamalla voi tarkastella ryhmään kuuluvia käsitteitä.
- **KANTO** - Kansalliset toimijatiedot sisältää henkilöiden ja yhteisöjen tietoja ja on korvannut Suomalaiset yhteisönimet, joka aiemmin julkaistiin Fintossa. Kesän aikana Suomalaisen yhteisönimien yhteisöille on etsitty ajantasaiset vastineet Kantosta ja näiden välille on lisätty korvaavuussuhteet.
- Liiketoimintaontologia LIITO on yhdistetty Julkisen hallinnon palveluontologia JUPOon kesällä 2021. LIITO:n käsitteillä täydennetty JUPO on julkaistu [Finto.fi:ssä](https://finto.fi/jupo): <https://finto.fi/jupo>
- Yleisten kirjastojen luokitusjärjestelmä **YKL**:ssä on korjattu joitain hierarkian esittämiseen liittyviä ongelmia, minkä seurauksena muotoluokasta 9 eli *historia*, johdettu luokka *39.09 taloushistoria* näkyy nyt hierarkiassa oikealla tavalla ja oikeassa kohdassa.
- Finto-palvelu kehittää yhteistyössä kirjastojen kuvailuryhmien kanssa "aikaontologiaa" sisällönkuvailun tarpeisiin. Uusi ontologia tukee myös ISO 8601 -standardin mukaista aikavälien esitystapaa vuosikymmenten ja vuosisatojen tarkkuudella. Tämä vähentää monitulkintaisuutta kuvailun aineiston asiansanoituksessa, sekä helpottaa aineiston hakemista ajanjaksojen perusteella.

Finto.fi, Annif ja muu tekninen kehitys

- Finto.fi:n taustaohjelmisto Skosmokselta on julkaistu versiot 2.11 ja 2.12. Lue lisää aiheesta näistä uutisteksteistä: <https://www.kiwi.fi/x/iQBTDQ> ja <https://www.kiwi.fi/x/3oBNDQ>
- **Skosify** on työkaluohjelma SKOS-muotoisten sanastojen muodostamiseen ja laadunvarmistukseen, jota käytetään Finto.fi:n taustalla. Siitä on julkaistu **uusi versio 2.3.0**, joka mm. parantaa tulosten deterministisyyttä (sama lopputulos eri ajokerroilla) sekä yhteensopivuutta uusien Python- ja rdflib-versioiden kanssa.
- Annif-työkalusta on kesän aikana julkaistu versiot **0.53** ja **0.54**. Näissä versioissa Annifiin on mm. lisätty **SVC-algoritmi** ja tuki **YAKE-algoritmin** käytölle, monia parannuksia MLLM-algoritmin toiminnan nopeuttamiseksi ja uusi **transform-ominaisuus** helpottamaan tekstisyyteen muokkaamista ennen sen käsittelyä algoritmeissa. Versio 0.55 julkaistaan pian; se tulee sisältämään mm. transform-ominaisuutta hyödyntävän **kielensuodatustoiminnon** tekstisyyteen puhdistamiseen.
- **Finto AI** -palvelu ja annif.org-sivuston taustalla toimiva Annif-asennus ovat siirtymässä Helsingin yliopiston tietotekniikkakeskuksen OpenShift-ympäristöön nykyisestä Kansalliskirjaston omasta Docker Swarm -ympäristöstä. OpenShift tarjoaa paremmat mahdollisuudet palvelun versionhallintaan ja automatisoituihin resurssien skaalaukseen korkean kuormituksen tilanteissa.
- Olemme parhaillaan luomassa uusia, päivitettyjä koneoppimismalleja Finto AI:n käyttöön. Uudet mallit perustuvat syyskuun lopun YSOon ja YSO-paikkoihin, Annifin tulevan version 0.55 tarjoamiin algoritmeihin sekä uusimpiin koulutusaineistoihin, jotka on koottu mm. [Finna.fi](https://finna.fi)-palvelusta, Kirjavälitys Oy:n keräämistä kirjaesittelyteksteistä ja Kansalliskirjastolle luovutetuista e-vapaakappaleista. Mallien käyttöönotto Finto AI:ssa parantaa ehdotusten ajantasaisuutta uusien YSON käsitteiden osalta ja toivottavasti myös niiden osuvuutta.

Ryhmien kokoukset

- Finton ohjausryhmä kokoontui 28.9. Kokouksen aineistot ja muistio on luettavissa [ohjausryhmän sivuilta](#).
- Vuoden kolmas varsinainen YSO-kokous pidettiin 15.9.2021. Kokouksen asialistalla oli yhteensä 144 uutta käsitte- ja muutosehdotusta, joista YSOon hyväksyttiin 66 uutta käsitettä. Muutoksia tai ohjaustermejä käsiteltiin 45 käsitteessä. 21 ehdotusta ei otettu sanastoon. Käsittely ja päätökset löytyvät [GitHubista](#). Seuraava YSO-kokous pidetään 24.11.2021 ([#4/021](#)). Sitä ennen pohditaan hankalia käsitteitä YSO-ongelmakouksessa 3.11.2021 ([2021-B](#)).
- Vuoden 2021 toinen Ontologiakehittäjien yhteistyöryhmän kokous pidetään 19.11.2021 OY-kokouksen aineistot löytyvät [ryhmän sivulta](#) Finton asiakaswikissä.

Tapahtumat ja koulutukset

- Järjestimme kesän aikana kaksi virtuaalista Annif-koulutusta ([Annif tutorial](#)): kesäkuussa Uumajan yliopiston järjestämän [metatietotyöpajan](#) yhteydessä sekä elokuussa Viron kansalliskirjaston ja Tarton yliopiston järjestämän [digitaalisten ihmistieteiden kesäkoulussa](#).
- Järjestämme seuraavan Annif-koulutuksen [SWIB21](#)-virtuaalikokouksen yhteydessä 30.11. Tilaisuus on ilmainen, mutta edellyttää rekisteröintiä.
- Annif-koulutusta on luvassa myös joulukuun alussa Ranskassa (osin virtuaalisena, mm. Annif-koulutus) järjestettävässä [Fantastic Futures 21](#) -konferenssissa.
- Latvian kansalliskirjasto järjesti 8.10. verkkokokouksen nimeltään [Interoperability of Data of Cultural Heritage Institutions in the Digital Space: Terminological Aspect](#). Konferenssissa Finto-palvelun kehittämispäällikkö Mikko Lappalainen piti esityksen Finton ontologioista.
- [Kirjastoverkkopäivät](#) järjestetään Ke 27.10. - To 28.10.2021. Ohjelmassa on mm.

- 27.10. klo 12:30-13:40 Sessio [K2S3 Finto-palvelu - osallistamista ja yhteistyötä Suomessa ja maailmalla](#). Kerromme miten yleistä suomalaista ontologiaa (YSO) kehitetään ja mitä sen käsite-ehdotuksille tapahtuu esityksessä *Kaikki mukaan YSOa kehittämään!* Esityksessä *Avoimella koodilla kansainväliseen yhteistyöhön - Skosmos-ohjelmiston käyttö maailmalla* kerromme minkälaisen tiedon esittämiseen Fintossa kehitettyä Skosmos-selainta on käytetty, minkälaiseen käyttöön Skosmosta on suositeltu, ja ketkä sitä käyttävät.
 - 28.10. klo 9-12 työpaja nimellä *Kiinalaisen huoneen arvoitus: automaattisen sisällönkuvailun arviointi jatkuu*. Työpajassa keskitytään (automaattisen) kuvailun laadun arviointiin käytännönläheistä tehtävää unohtamatta. Tehtävässä osallistujat arvioivat kuvailun laatua itse. Työpaja on jatkoa aikaisemmille, vuosien 2017 ja 2019 Kirjastoverkkopäivien Annif-aiheisille työpajoille. Lisätietoa ja linkki ilmoittautumiseen löytyy Kirjastoverkkopäivien sivuilta os. <https://www.kansalliskirjasto.fi/en/node/1881>, lämpimästi tervetuloa! Muistathan ilmoittautua 25.10.2021 mennessä.
- [DCMI Virtual](#) -konferenssissa järjestetään 13.10. tekoälyyn ja automaattiseen sisällönkuvailuun pureutuva paneelikeskustelu Panel: [Why AI Automated Indexing: What Is and Is Not Possible](#). Osma Suominen edustaa keskustelussa Annifin ja Finto AI:n kehityksen näkökulmaa. Tallenteet tästä ja muista DCMI Virtual -konferenssin ohjelmanumeroista julkaistaneen konferenssin jälkeen [Dublin Core -yhteisön YouTube-kanavalla](#).
- SWIB21-konferenssissa on koulutuksen lisäksi myös muuta Annifiin liittyvää ohjelmaa: Saksan kansalliskirjaston [esitys Annif-käyttöönottoprojektista](#), Pennsylvanian yliopiston [esitys Annifin yhdistämisestä BIBFRAME-kuvailukäyttöliittymään](#) sekä myös Open Day-päivänä (torstai 2.12.) Annif-aiheinen hackathon, jossa osallistujat muokkaavat omia data-aineistojaan Annifille sopiviksi korpuksiksi.
- Osma Suominen pitää esityksen "Automated subject indexing with Annif and Finto AI" 19.11. Saksan kansalliskirjaston järjestämässä kaksipäiväisessä virtuaalikonferenssissa, jonka englanninkielinen nimi on "Network of machine processes in cataloguing". Konferenssin ohjelma julkaistaan myöhemmin.