

# Sisällönkuvailun ammattilaisten kokemuksia Finto AI:n käytöstä

*Mona Lehtinen (toim.)*

Image by DavidRockDesign from Pixabay

Toimittaja: Mona Lehtinen

Tekijät: Mirja Anttila, Maria Forsén, Taru Fröjdholm, Liisa Hilander, Outi Hintikka, Ilona Lindfors, Katriina Tiainen, Tarja Turunen

YSO-asiasanat: sisällönkuvailu, automaattinen sisällönkuvailu, metadata, tekoäly, koneoppiminen, verkkopalvelut, käyttäjäkokemus, testaus, ammattilaiset, asiasanoitus

Kansalliskirjaston raportteja ja selvityksiä 4/2021

ISBN 978-951-51-7816-9

ISSN 2242–8119



# Tiivistelmä

Tässä kokeilussa selvitimme Kansalliskirjastossa ylläpidettävän Finto AI -palvelun hyödynnettävyyttä sisällönkuvailun ammattilaiskäytössä. Testauksessa sisällönkuvailun ammattilaiset käyttivät automaattista sisällönkuvailun palvelua verkkolomakkeen kautta.

Aineistoksi valikoitui sen tyyppisiä tekstejä, joita he yleensäkin työssään käsittelevät ja jotka olivat elektronisesti saatavilla. Tulosten perusteella Finto AI on hyvä apuväline, mutta ei nopeuta sisällönkuvailun prosessia tässä käyttöyhteydessä.

## Johdanto

Erityisesti digitaalinen aineistomäärä kasvaa jatkuvasti, ja pelkästään ihmisvoimin asiansanoitusta tehtäessä työn määrä kasvaa helposti liian suureksi. Tätä kautta on syntynyt kysyntää erilaisille automatisointiratkaisuille. Finto AI on Kansalliskirjastossa kehitetty palvelu sisällönkuvailutyön tueksi.

Tässä raportissa keskitytään Finto AI:n ammatillisiin käyttökokemuksiin nimenomaan sisällönkuvailutyön näkökulmasta. Sisällönkuvailun asiantuntijaryhmä (Sisku) perusti vuoden 2021 alusta automaattisen sisällönkuvailun työryhmän. Ryhmän tavoitteena on

automaattisen sisällönkuvailun laadun varmistamiseen osallistuminen. Tähän liittyen sisällönkuvailijoista koostuva ryhmä myös testasi Finto AI -palvelua sisällönkuvailijan eli ammattilaiskäyttäjän näkökulmasta. Kyseessä oleva selvitys ei ole täysi käytettävyysselvitys eikä sen tarkoitus ole Finto AI:n toiminnan systemaattinen ja laajempi testaus. Raportin tavoite ei ole olla tieteellinen tutkimus, vaan sen on tarkoitus toimia Finto AI:n kehityksen tukena ja antaa käsitys sen toiminnasta ja soveltuvuudesta.

## Testausasetelma

### Finto AI

Finto AI:tä voidaan tällä hetkellä käyttää tekstin asiansanoituksessa suomeksi, ruotsiksi ja englanniksi perustuen kolmikielisen Yleisen suomalaisen ontologian YSO:n käsitteisiin (sis. YSO-paikat). Finto AI:lla on rajapinta, jonka kautta sen integroiminen tietojärjestelmiin on helppoa. Sitä voi käyttää myös verkkolomakkeen kautta. Verkkolomakkeelta Finto AI:n ehdottamia asiansanoja voi myös kopioida joko sanoina

tai Melinda-asiasanakenttänä. Finto AI:ta käytetään julkaisuarkistoissa, joihin korkeakoulujen opiskelijat saavat ladata opinnäytetyönsä. Finto AI on integroitu Kansalliskirjaston ylläpitämiin Lapin, Tampereen ja Vaasan yliopistojen julkaisuarkistoihin sekä ammatikorkeakoulujen julkaisuarkisto Trepoon. Julkaisuarkistojen integraatio toimii syöttölomakkeen kautta: esim. opinnäytettään arkistoon tallentava saa lomaketta käyttäessään työilleen asiansaehdotuksia, joista

hän voi valita haluamansa ja lisätä niitä tarvittaessa. Integraatio on käytössä myös Jyväskylän yliopiston JYX-julkaisuarkistossa. Automaattisen kuvailun käyttökokemuksista julkaisuarkistossa voi lukea lisää Kreodi-lehdessä julkaistusta Tiina Tolosen kirjoittamasta artikkelista.

Automaattisen asiansanoituksen palvelu Finto AI perustuu Annif-työkaluun. Annif hyödyntää useita koneoppimisen ja kieliteknologian ratkaisuja. Annifia on kehitetty vuodesta 2017 alkaen Kansalliskirjastossa seuraavien periaatteiden mukaisesti:

- Kehitettävän työkalun tulisi olla riippumaton kielestä ja sanastosta, koska kuvailussa tarvitaan useita kieliä ja sanastoja.
- Työkalun tulisi pystyä tukemaan joustavasti uusimpia ja parhaimpia automaattisen kuvailun ratkaisuja ja niitä pitäisi helposti saada integroitua työkaluun
- Työkalun tulisi olla helposti käytettävissä ja integroitavissa muihin järjestelmiin: näin ollen sitä pitää pystyä käyttämään komentoriviltä ja rajapinnan kautta
- Työkalun tulisi olla avointa lähdekoodia

Tässä selvityksessä sisällönkuvailun ammattilaiset käyttivät Finto AI -palvelua verkkolomakkeen kautta. Testi tehtiin keväällä 2021. Kuvailijat syöttivät lomakkeelle sen tyyppisiä tekstejä, joita yleensäkin työssään käsittelevät ja jotka olivat elektronisesti saatavilla. Kukin kuvaili Finto AI:n avustuksella kymmenen tekstiä. Kuvailijat täyttivät samalla lomaketta, jossa tarkennettiin kuvailtavan tekstin ominaisuuksia sekä listattiin Finto AI:n ehdottamat ja lopulliseen kuvailuun päätyneet asiansanat. Kuvailijoilta kerättiin myös perustelut sille, miksi Finto AI:n ehdottamat asiansanat oli hyväksytty tai hylätty. He antoivat myös vapaata palautetta mm. Finto AI:n toiminnasta ja soveltuvuudesta kuvailutyöhön sekä lomakkeen kautta, että erikseen sähköpostilla testauksen jälkeen (lomake liitteenä). Finto AI:n verkkokäyttöliittymä antaa mahdollisuuden valita tarjottujen asiansanojen määrä, tässä kokeessa kuvailijat rajasivat käytetyt asiansanat pääasiassa maksimissaan kymmeneen. (Huom. kuitenkin, että Finto AI ei kuitenkaan aina tarjoa maksimimäärää ehdotuksia, jos se ei löydä tarpeeksi monta luotettavana pitämäänsä asiansanaa. Mahdollisia syitä sille, ettei asiansanoja löydy, saattaa olla esimerkiksi Finto AI:n koulutuksessa käytetty aineisto.)

## Materiaali

Kuvailijat käyttivät testauksessa samantyyppistä aineistoa, jollaista työssään yleensäkin käsittelevät. Aineiston tuli olla myös elektronisesti saatavilla, mikä asetti joitain rajoituksia käytettävän materiaalin suhteen. Materiaalissa oli mukana opinnäytetöitä (pro gradu -töitä ja väitöskirjoja), tietokirjallisuutta sekä kaunokirjoja.

Materiaalia oli kolmella YSO:n tukemalla kielellä - suomeksi, ruotsiksi ja englanniksi - tosin kaikkia kieliä ei esiintynyt kaikissa materiaalityypeissä. Testausaineistona oli joko abstrakti, tiivistelmä, kirjan esittelytieto tai kokoteksti. Kopiosuojaus puolestaan vaikutti siihen, valikoituiko testattavaksi kokoteksti vai tiivistelmä: kaikesta materiaalista ei saatu lisättyä kokotekstiä Finto AI:n lomakkeelle.

## Testilomake

Käyttämämme lomake suunniteltiin yhteistyössä ryhmän kesken. Lomakkeella pyydettiin kuvailijaa kertomaan, millainen materiaali oli kyseessä, ja tarjoamaan linkki siihen. Kuvailija myös listasi hyväksymiensä Finto AI:n ehdottamien asiansanojen määrän. Kuvailija myös kertoi, jos oli joutunut lisäämään kuvailuun asiansanoja. Lisäksi lomakkeessa oli mahdollista tarkentaa, puuttuuko lisätty asiansana YSO:sta. Kuvailija myös listasi lomakkeella Finto AI:n ehdotuksista poistamansa asiansanat ja miksi ne oli poistettu. Poistamisen syyksi lomakkeella tarjottiin kolme vaihtoehtoa: väärä tarkkuus, täysi virheellisyys ja epäolennaisuus. Täysin virheellisellä asiansanalla ei siis ole mitään tekemistä tekstin aiheen kanssa. Väärällä tarkkuudella puolestaan tarkoitettiin sitä, että ehdotettu asiansana on liian laaja tai suppea kuvaamaan tekstiä kunnolla (esim. endometrioosia käsittelevän tekstin asiansanoituksessa lääketiede saattaa olla liian laaja ja epätarkka käsite). Epäolennaisuus taas viittaa siihen, että asiansana ei välttämättä ole täysin väärä, mutta ei kuvaa aineiston ydinasiaa (esimerkki asiansana perinnöllisyysneuvonta poistettiin epäolennaisena / ei ydinasiaa kuvaavana lasten geenitutkimusta käsittelevästä artikkelista).

# Tulokset

Tarkasteltaessa hyväksytyjen asiasanojen määriä havaittiin, että noin 70 % Finto AI:n ehdotuksista hyväksyttiin lopulliseen kuvailuun (taulukko 1). Tässä kuitenkin oli suuria eroja eri aineistotyyppien välillä. Esim. kaunokirjallisuuden kohdalla vain n. 24 % asia-  
sanaehdotuksista hyväksyttiin lopulliseen kuvailuun.

Kielikohtaisia tuloksia tulkittaessa täytyy myös huomioda, että englannin- ja ruotsinkielistä materiaalia oli testauksessa mukana. Puuttuvia asiasanoja myös lisättiin erittäin usein, suurimmassa osassa tapauksia joitakin keskeisiä asiasanoja oli jäänyt puuttumaan.

Aineistotyyppi	Kieli	Hyväksytyt ehdotukset	Lopullisessa kuvailussa Finto AI:n asiasanoja
Tieto / tieteellinen	Suomi	62 %	64 %
Tieto / tieteellinen	Ruotsi	73 %	56 %
Tieto / tieteellinen	Englanti	74 %	84 %
Kauno	Suomi, ruotsi, englanti	24 %	

Taulukko 1. Hyväksytyjen ehdotusten määrät aineistoittain ja kielittäin

## Huomioita

Vastauksista ilmenee, että joskus puuttumaan jääneet ehdotukset voivat olla jopa suurempi haitta kuin virheelliset ehdotukset. Hankala tilanne on myös se, jos Finto AI:n ehdotus on hyvin lähellä oikeaa, mutta vaatii kuitenkin tarkennusta. Välillä myös Finto AI:n ehdotukset ovat liian yleisellä tasolla, esimerkiksi vaikkapa syöpätaudit eikä tietty syöpätyyppi, jota artikkeli käsitteli. Vastauksissa huomautetaan myös, että Finto AI vaikuttaa suoriutuvan eri tavalla eri tieteenalojen väitöskirjojen välillä; esim. terveystieteellisen aineiston kohdalla relevanttien asiasanojen saanti oli haasteellista. Tämä saattaa johtua esim. eri alojen käyttämistä kuvailusanastoista ja -traditioista. Eräs vastaajista myös kokeili Finto AI:lla miten tulokset eroavat saman väitöskirjan erikielillä tiivistelmillä. Tässä oli havaittavissa eroavaisuuksia, ja itse asiassa käännosten testaaminen on Annifin vuoden 2021 projektisuunnitelmassa. Vastauksissa tuli esille myös se, että Finto AI toimii jossain tapauksissa jopa paremmin lyhyemmällä tekstillä (abstrakti / lähdeluettelo) kuin kokotekstillä. Fiktiivisen aineis-

ton kuvailussa Finto AI:ta ei sellaisenaan katsota hyödylliseksi.

Lisäksi vastauksissa todetaan että Finto AI kuvailun apuvälineenä ei välttämättä nopeuta sisällönkuvailun ammattilaisen työtä, koska aikaa kuluu tarkistaessa, onko aiheeseen liittyvä, mutta ei esim. tiivistelmässä mainittu termi oikea ko. aineiston kuvailussa, tai joutuu selvittämään, miksi Finto AI tarjoaa jotain tiettyä termiä, joka ei vaikuta suoraan liittyvän aiheeseen. Fintoa AI:n tekninen käyttö kuitenkin mainitaan nopeana. Finto AI:n käyttö saattaa kuitenkin auttaa kuvailun kattavuuden ja tasalaatuisuuden suhteen. Finto AI toimii hyvänä apuvälineenä sopivien asiasanojen etsimisessä varsinkin silloin, kun Finto AI:n käyttäjä ei ole varsinainen sisällönkuvailija, jolloin tutustuminen asiasanoihin ja asiansanoitukseen käy helposti.

Lisäksi kuvailijat tekivät havaintoja FintoAI:n tekemistä virheistä, esim. erisnimet saattavat sekoittaa vastaavin asiansanoihin. Mikäli Finto AI:n käyttäjä havaitsee jotain korjattavaa tai huomioitavaa, sivul-

la ai.finto.fi olevan palaute-linkin kautta voi ottaa yhteyttä kehittäjiin.

## Kehitysehdotuksia ja Finto AI:n tulevaisuuden näkymiä

Testauksen yhteydessä kuvailijoilta pyydettiin myös kehitysehdotuksia ja ideoita Finto AI:n tulevaisuuden suhteen. Finto AI:n termien relevanssijärjestys herätti vastaajien mielenkiintoa. Vastaajat esittivät myös toiveita uusien sanastojen lisäämisestä Finto AI:n valikoimaan: esim. Kanto, SLM ja tuleva Aikaontologia, joiden mukanaolo lisäisi mm. yhteisönnimien, muoto- ja lajityyppitermien, sekä ajanmääreiden saantia Finto AI:n tuomiin ehdotuksiin.

Kaunokirjallisuuden automaattista kuvailua voisi tutkia tarkemmin ja pyrkiä kehittämään esim. lisäämällä KAUNO-sanasto Finto AI:n valikoimaan tai opettamalla Finto AI tunnistamaan lajityyppejä. Kaunokirjallisuuden automaattisen kuvailun kohdalla olisi mielenkiintoista vertailla eroja käytettäessä koko-

tekstiä tai esittelytekstiä tai muuta aineistoa kuvailevaa tekstiä.

Finto AI:n tuottamien asiasanojen relevanttiutta toki lisää ajantasaisen YSO:n käyttö, joten vastaajat toivoivat myös mahdollisimman tiheitä päivityksiä.

Sanastotyön näkökulmasta olisi tärkeää, että tulevaisuudessa Finto AI:sta voisi päästä ehdottamaan lisäyksiä sanastoon samaan tapaan kuin Fintonkin kautta. Aineiston tekijällä on kuitenkin ensikäden tietoa kuvailussa tarvittavista termeistä, ja ehdotusjärjestelmän kautta on helppo tarkistaa, onko tarvittava termi jo sanastossa ja tehdä ehdotuksia ylläpidolle. Tämä prosessi voitaisiin myös automatisoida, jolloin puutteet tulisivat vaivatta ylläpidon tietoon ja YSO:n kattavuutta pystyttäisiin parantamaan eri tieteenalojen uudella termistöllä.

Käytetyn hakutekstin sanojen välisen läheisyyden huomioiminen lisäisi tarkempien ja relevantimpien asiasanojen löytymistä. Tulevaisuudessa Finto AI:n toimintaa ja käytettävyyttä voidaan tarkastella tarkemmalla tasolla.

## Kirjallisuus

Suominen, Osma. 2019. 'Annif: DIY Automated Subject Indexing Using Multiple Algorithms'. LIBER Quarterly 29 (1): 1. <https://doi.org/10.18352/lq.10285>.

Suominen, Osma; Inkinen, Juho & Lehtinen, Mona. 2021 (preprint). 'Annif and Finto AI: Developing and Implementing Automated Subject Indexing'. <https://urn.fi/URN:NBN:fi-fe2021080942632>.

Annifin verkkosivut: <https://www.annif.org> (haettu 15.11.2021)

# Liite: Testilomakkeen kysymykset

- pvm
- nimeke (+linkki)
- aineistolaji (tieteellinen, tieto, kauno)
- FintoAI:lle annettu taustateksti
- lopulliseen kuvailuun hyväksytyt/kaikki ehdotukset (lkm)
- kuvailusta pois jätetyt termit
- aihe-ehdotuksista puuttuneet YSO-termit
- YSO:sta puuttuvat tarpeelliset termit
- tekijän antamat avainsanat
- kommentit



KANSALLISKIRJASTO  
NATIONALBIBLIOTEKET

---