

# Puoliautomaattinen sisällönkuvailu FintoAI:lla

- [Johdanto](#)
  - [Puoliautomaattinen ja automaattinen sisällönkuvailu](#)
  - [Finto AI](#)
- [Finto AI:n käyttö verkkolomakkeella](#)
- [Tarkistaminen ehdotusten hakemisen jälkeen](#)
- [Palaute](#)

## Johdanto

Tämän tekstin tarkoitus on esitellä lyhyesti puoliautomaattista (ihmisen tarkastamaa) ja automaattista (täysin koneellisesti tuotettua) sisällönkuvailua, sekä tarjota ohjeita puoliautomaattiseen sisällönkuvailuun Finto AI -työkalun avulla. Ohjeen kohderyhmänä ovat erityisesti sisällönkuvailua sivutoimisesti tai kertaluontoisesti tekevät henkilöt, jotka käyttävät Finto AI -palvelua, sekä sisällönkuvailun ammattilaiset. Ohjeen on laatinut Sisällönkuvailun asiantuntijaryhmä yhdessä Finto AI:n kehittäjien kanssa.

## Puoliautomaattinen ja automaattinen sisällönkuvailu

Sisällönkuvailun tarkoitus on ilmaista aineiston (esim. kirja, artikkeli, verkkosivu, esine jne.) aihe aineiston löytämisen helpottamiseksi. Sisällönkuvailussa voidaan hyödyntää esim. luokitusjärjestelmiä, asiasanastoja tai vapaita avainsanoja. Se on perinteisesti asiantuntijatyötä, jota tekevät sisällönkuvailun ammattilaiset kirjastoissa, arkistoissa ja vastaavissa organisaatioissa. Myös aineiston tekijät kuvailevat usein omia töitään.

Aineiston, erityisesti digitaalisen aineiston, määrän kasvaessa sen sisällönkuvailu ihmisvoimin hankaloituu. Suuren aineistomäärän sisällönkuvailun helpottamiseksi katsetta on käännetty erilaisiin koneoppimis- ja tekoälyratkaisuihin. Puoliautomaattisessa sisällönkuvailussa jokin järjestelmä antaa halutulle tekstille asiasanaehdotuksia, jotka ihminen tarkistaa. Ihmiselle jää siis päävastuu kuvailun laadusta, järjestelmän antamien alustavien ehdotusten tarkoituksena on sujuvoittaa kuvailutyötä. Täysin automaattisessa kuvailussa järjestelmän algoritmien tuottama kuvailu etenee sellaisenaan ilman ihmisen tekemää tarkistusta esimerkiksi tietokantaan. Näin pyritään siihen, että aineisto saataisiin edes jollakin tavalla kuvailtua, mutta menetelmän ongelmana ovat ennen kaikkea virheelliset asiasanat, jotka tuottavat hakutuloksiin haun kannalta vääriä aineistoja.

Automaattisen kuvailun puutteiden taustalla voi olla esim. se, ettei järjestelmä välttämättä pysy sanastokehityksen vauhdissa, sanaston ja järjestelmän päivitykset saattavat kulkea eri tahdissa. Lisäksi järjestelmä yleensä tarvitsee valmiiksi kuvailtua opetusdataa, jossa uusi sanastoon lisätty termi esiintyy tarpeeksi monta kertaa, jotta se oppisi itsekin käyttämään tätä uutta termiä oikein. Koulutusaineisto on kaiken koneoppimisen ja tekoälyn kompastuskivi: mikäli opetusaineisto on puutteellista tai jollain tavalla vääristynyttä, eivät algoritmit suoriudu oletetulla tavalla. Tekstissä esiintyvät monimerkityksiset sanat saattavat tuottaa virheellisiä asiasanaehdotuksia, jos algoritmi tekee väärän tulkinnan tekstin sanaa vastaavasta sanaston asiasanasta. Tietyn alan tekstien kuvailussa usein esiintyvät asiasanat saattavat olla ylliedustettuja algoritmien ehdotuksissa, eli kone voi ehdottaa alalla usein esiintyvää asiasanaa myös silloin, kun kyseinen aineisto ei käsittele kyseistä aihetta.

## Finto AI

Finto AI on työkalu tekstin asiasanoitukseen. Sitä voidaan tällä hetkellä käyttää asiasanoituksessa suomeksi, ruotsiksi ja englanniksi. Finto AI perustuu Kansalliskirjastossa kehitettävään Annif-työkaluun. Annif on avoimen lähdekoodin työkalu, joka hyödyntää kieliteknologiaa ja koneoppimista. Annifista löytyy lisätietoa sen verkkosivulta os. <https://annif.org> tai [asiakaswikistä](#). Finto AI:ta voidaan käyttää osoitteessa [ai.finto.fi](https://ai.finto.fi) olevan lomakkeen kautta. Lomakkeen käyttöä kuvataan tarkemmin seuraavassa luvussa.

Finto AI:lla on myös rajapinta. Sen kautta Finto AI voi keskustella muiden järjestelmien kanssa, jolloin se voidaan yhdistää osaksi laajempia kokonaisuuksia, esimerkiksi kirjastojen tietojärjestelmiä. Lisätietoja rajapinnan käytöstä löytyy Finto AI:n [asiakaswikistä](https://ai.finto.fi/v1/ui/) ja os. <https://ai.finto.fi/v1/ui/>. Demonstraatio Finto AI:n toiminnasta julkaisuarkiston syöttölomakkeella on katsottavissa os. <https://urn.fi/URN:NBN:fi-fe2020042219855>. Rajapinta on käytössä esimerkiksi usean yliopiston julkaisuarkistossa sekä Kansalliskirjastossa e-vapaakappaleiden vastaanotossa. Ajantasainen tieto käyttäjistä löytyy [asiakaswikistä](#).

## Finto AI:n käyttö verkkolomakkeella

Finto AI:n lomakekäyttö aloitetaan syöttämällä kuvailtava teksti lomakkeelle. Lomaketta voi käyttää kirjoittamalla tai kopioimalla ja liittämällä tekstiä tekstikenttään. Tekstin voi myös ladata lomakkeelle tiedostosta tai hakemalla se URL:n kautta. Eri syöttötavoille on lomakkeella omat välilehdet tekstikentän yläpuolella. Mikäli haluat ladata lomakkeelle tiedoston omalta koneeltasi, valitse välilehti "Syötä tiedosto". Pääset etsimään tiedoston selaa-painikkeella. Mikäli haluat teksti on verkossa, valitse "Syötä URL" -välilehti ja kopioi URL lomakkeen sille tarjotun kenttään ja paina "Hae teksti". Lomakkeen tukemat tiedostomuodot ovat: .txt, .pdf, .doc, .docx, .odt, .rtf, .pptx, .epub, .html sekä .htm. Tuetut tiedostomuodot ilmoitetaan myös lomakkeella ko. välilehdillä.

## Kuvailtava teksti

Syötä teksti

Syötä tiedosto

Syötä URL

Kopioi tai kirjoita tähän tekstiä ja paina "Anna aihe-ehdotukset"-painiketta



Voit myös pudottaa tiedoston tai URL:n tähän

Käytössä [Annif](#) v1.0.0

## Kuvailtava teksti

Syötä teksti

Syötä tiedosto

Syötä URL

Selaa

Valitse tiedosto

Tuetut tiedostomuodot: .txt, .pdf, .doc, .docx, .odt, .rtf, .pptx, .epub, .html

Teksti tiedostosta haetaan tähän. Paina sitten "Anna aihe-ehdotukset"-painiketta

Voit myös pudottaa tiedoston tai URL:n tähän

Käytössä [Annif](#) v1.0.0

## Kuvailtava teksti

Syötä teksti

Syötä tiedosto

Syötä URL

Syötä URL

Hae teksti

Tuetut tiedostomuodot: .txt, .pdf, .doc, .docx, .odt, .rtf, .pptx, .epub, .html

Teksti URL-osoitteesta haetaan tähän. Paina sitten "Anna aihe-ehdotukset"-painiketta

Voit myös pudottaa tiedoston tai URL:n tähän

Käytössä [Annif](#) v1.0.0

Kuva 1. Ruutukaappauksia Finto AI:n verkkolomakkeelta eri välilehdistä

Avattavasta valikosta voit valita tekstin kielen ja haluamasi kuvailusanaston. Voit myös säätää asiasanaehdotusten määrää. Asiasanojen kielen voi valita riippumatta tekstin kielestä, eli esimerkiksi englanninkielisille dokumenteille voi hakea suomenkieliset asiasanat. Asiasanat haetaan napsauttamalla "Anna aihe-ehdotukset" -painiketta. Aihe-ehdotusten määräksi voi säätää kymmenen, viisitoista tai kaksikymmentä. Sanasto ja tekstin kieli -pudotusvalikosta valitaan tekstin mukainen kieli ja sanasto (huom. valintojen tulee olla oikein, muuten asiasanoitus ei onnistu). Tällä hetkellä Finto AI tukee sanastoista Yleistä suomalaista ontologiaa YSO:a (sisältäen YSO-paikat) kolmella kielellä (suomi, ruotsi ja englanti), Yleisten kirjastojen luokitusjärjestelmää YKL:ää (myös kolmella kielellä), sekä kaunokirjallisuuden kuvailussa käytettävää KAUNO-ontologiaa (suomeksi). Huomaa, että lomakkeen valikossa ilmoitetaan myös kulloinkin käytetty YSO:n versio suluissa. Aihe-ehdotusten kielen voi valita myös niin, että aihe-ehdotusten kieli voi olla eri kuin tekstin, esim. suomenkieliselle tekstille voi valita ruotsinkieliset asiasanat. On kuitenkin hyvä huomioda, että englanninkieliset termimuodot eivät kaikissa tapauksissa ole täsmällisiä vastineita suomen- ja ruotsinkielisille, ja niiden oikeellisuus on syytä varmistaa. Tätä valintaa säädetään aihe-ehdotusten kieli -pudotusvalikosta. Finto AI:n käyttöliittymässä on mahdollisuus kopioida ehdotettu asiasana tai sen URI nappia painamalla leikepöydälle. Kopioida voi myös asiasanan Marc-kentän (termi itse, siihen liittyvä sanastokoodi kielitarkenteineen sekä URI ) esim. Aleph-luettelointiohjelmasta varten. Ohjeita tämän ominaisuuden käytöstä löytyy Melindan Talonmiehen tuokiosta 22.4.2021 kuvatulta videolta os. <https://youtu.be/vNtTVxqc0Zc?t=1389>

Kun teksti on saatu lomakkeelle, asiasanojen hakeminen etenee askel askeleelta näin:

1. Valitse oikean yläkulman valikosta tekstin kieltä vastaava kieli ja sanasto
2. Valitse haluamasi aihe-ehdotusten määrä (10-20)
3. Valitse, haluatko saada aihe-ehdotukset samalla kielellä kuin teksti, vai haluatko niille käännökset
4. Hae aihe-ehdotukset painamalla nappulaa
  - a. Huom. ehdotukset ovat koneen ehdottamassa "paremmuusjärjestyksessä"
5. Aihe-ehdotuksia voi kopioida asiasanan vieressä olevalla painikkeella

Kuva 2. Finto AI:n lomake, jolle on kopioitu haluttu teksti ja haettu asiasanat YSO:sta

## Tarkistaminen ehdotusten hakemisen jälkeen

Finto AI:n antamia ehdotuksia tarkistettaessa on tärkeää huomioida:

- Jäikö jokin tärkeä aihe puuttumaan?
  - Kone ei välttämättä huomaa ehdottaa esim. kaikkia olennaisia asiasanoja, erityisesti jos ne eivät esiinny tekstissä tai eivät ole kyseisen aihealueen aineistoissa yleisiä.
  - Lisäksi kannattaa tarkistaa, ovatko tarpeelliset paikkakäsitteet mukana, ja tarvittaessa lisätä puuttuvat kuvailuun.
  - Huom. uusimpia YSO:n asiasanoja ei välttämättä löydy heti Finto AI:sta.
  - Kuvailussa käytettäviä käsitteitä voi etsiä esim. YSO-, YSO-paikat ja YSO-aika -sanastoista, ks. <https://finto.fi>
- Onko mukana virheellisiä ehdotuksia?
  - Väärän aihealueen termit on usein helppo huomata, mutta mukana voi olla myös oikealta näyttäviä asiasanoja, jotka liittyvät esim. kyseisellä alalla tavalliseen tutkimuskohteeseen, vaikka kuvailtavassa aineistossa on käsitelty eri asiaa.
- Onko mukana ehdotuksia, jotka ovat aineiston kannalta liian yleisellä tasolla?
  - Esim. termi *sairaudet* tiettyä tautia tutkivassa aineistossa tai *eläimet* vain koiraa käsittelevässä tekstissä.
- Onko mukana ehdotuksia, jotka kuvaavat vain pientä osaa aineistosta ja ovat kokonaisuuden kannalta epäolennaisia tai liian suppeita?
  - Esim. *renkaat* kun puhutaan autoista.

Finto AI ei toistaiseksi osaa ehdottaa aiheeksi henkilöitä, yhteisöjä tai teoksia. Nämä pitää siis muistaa lisätä kuvailuun. Jos tekstin lajityyppi (esim. opinnäytteet, oppikirjat jne.) on tarpeellista ilmaista, myös se on lisättävä erikseen. Kirjastokuvailussa tähän käytetään SLM-sanastoa, ks. sisällönkuvailuopas (<https://wiki.helsinki.fi/x/uaLkFw>). Muissa tapauksissa noudatetaan oman organisaation kuvailuohjeistusta lajityypin merkinnästä. Lajityyppiä (esim. väitöskirja, pro gradu) ei merkitä opinnäytteessä aiheiden yhteyteen.

Ehdotuksista puuttuvia asiasanoja voi etsiä YSO:sta tai erikoisalojen sanastoista, esim. MeSH. Jos tarvittava asiasana puuttuu YSO:sta, se kannattaa kuitenkin lisätä kuvailuun vapaana avainsanana, ja siitä on hyvä tehdä ehdotus sanastoon (<https://finto.fi/ysa/fi/#suggestion>). Organisaation ohjeistuksen mukaisesti voidaan käyttää myös muiden sanastojen asiasanoja.

# Palaute

Palautetta tästä ohjeesta ja kehitysehdotuksia voi lähettää esimerkiksi Finton palautelomakkeen kautta ([finto.fi](https://finto.fi) -pääsivun yläreunasta kohta "palaute" tai "feedback").