



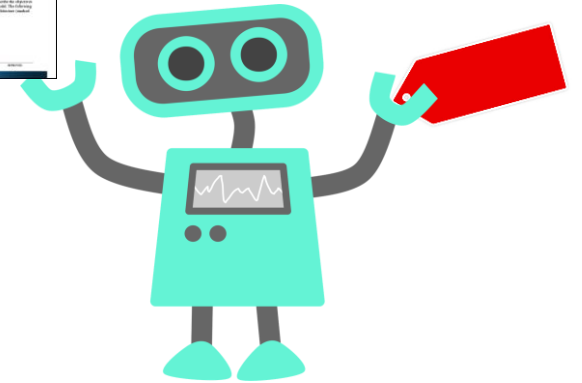
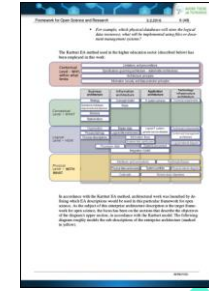
Annif & YKL

Talonmies Extra 20-10-2022

Annif ja Finto AI

annif

Tuottaa sisällönkuvailua (asiasanoitus ja luokitus) tekstin perusteella Kansalliskirjastossa kehitetty työkalu (prototyyppi 2017), avointa koodia
Soveltuu parhaiten asiateksteille, mutta testataan myös muilla aineistoilla
Perustuu **kieliteknologiaan** ja **koneoppimiseen**
Voidaan käyttää **rajapinnan kautta** tai **itsenäisenä työkaluna**



annif

kehitys alkoi 2017

Yleiskäyttöinen avoimen lähdekoodin **työkaluohjelmisto** automaattiseen sisällönkuvailuun

Tukee monia kieliä ja eri sanastoja

Verkkosivut, GitHub-koodivarasto, demolomake ja rajapintapalvelu testikäyttöä varten

Maailmanlaajuinen käyttäjäyhteisö

Tekninen dokumentaatio englanniksi, verkkokurssi, käyttäjäryhmä annif-users

annif.org

fintoai

julkaistu 5/2020

Annifiin perustuva automaattisen sisällönkuvailun **verkkopalvelu**

Tukee YSO-kuvailua kolmella kielellä (YKL tulossa)

Verkkopalvelu, jossa lomake ja rajapintapalvelu

Kotimaisen kuvailutyön tukipalvelu, Finton sisar

Käyttöönotto-ohjeistus suomeksi, ruotsiksi ja englanniksi, Finton tiedotuskanavat

ai.finto.fi

API-palvelu

Finto AI toimii myös rajapintapalveluna, joka voidaan integroida omiin järjestelmiin.

[Lisätietoja](#) | [OpenAPI-kuvaus](#)

Kuvailtava teksti

Suurteholaskennasta puhtia automaattiseen sisällönkuvailuun: High-Performance Digitisation -hankkeen yhteistyökokemuksia ✕

High Performance Digitisation -hanke on päättymässä vuoden lopussa. Järjestämme sen kunniaksi kaikille avoimen webinaarin otsikolla Suurteholaskennasta puhtia automaattiseen sisällönkuvailuun: High-Performance Digitisation -hankkeen yhteistyökokemuksia.

Webinaarissa esitellään hankkeen saavutukset ja tulevaisuuden suunnitelmat. Kerromme mm. CSC:llä hankkeen puitteissa tehdystä automaattisen kuvailun algoritmien testauksesta ja tulosten hyödyntämisestä Kansalliskirjastossa tehtävässä kehitystyössä.

Mukaan on myös lupautunut kaksi ulkomaista puhujaa: Annemieke Romein Hollannin kansalliskirjastosta kertoo automaattisen asiasanoittajan Annifin käytöstä vanhoja lakitekstejä käsitelleessä tutkimusprojektissa ja Koraljka Golub Linné-yliopistosta kertoo asiasanoituksen laadun arvioinnista.

ai.finto.fi

Sisällönkuvailu

Sanasto ja tekstin kieli

YSO suomi (2020-10-05) ▼

Ehdotusten enimmäismäärä

10 15 20

Aihe-ehdotusten kieli

sama kuin tekstin kieli ▼

Anna aihe-ehdotukset

Ehdotetut aiheet

- [sisällönkuvailu](#)
- [projektit](#)
- [kansalliskirjastot](#)
- [automaattinen sisällönkuvailu](#)
- [arviointi](#)
- [asiasanoitus](#)
- [automaatio](#)
- [tutkimusprojektit](#)
- [indeksointi](#)

Annifin / Finto AI:n käyttöönottoja

Finto AI-palvelun käyttö

Omatoiminen käyttö verkkoselaimella

API-integraation kautta:

- **Julkaisuarkistot:** JYX (JYU), Osuva (Vaasan yo), Theseus (amk:t), Trepo (Tampereen yo), Taju (Taideyliopisto), Lauda (Lapin yo) ...
- 40 **museon** Collecte-kokoelmanhallinta
- **Kirjavälitys Oy**, ennakkotiedot uutuuksikirjoista
- KK:n **E-vapaakappaleiden** vastaanotto

Itsenäinen Annif-asennus

- **Yle** / verkkouutisten asiasanoitus
- Saksan kansalliskirjasto **DNB**
- Saksan taloustieteen instituutti **ZBW**
- Alankomaiden kansalliskirjasto **KB.nl**
- Alankomaiden tilastokeskus **CBS**
- Puolan kansalliskirjasto
- Dissem.in -palvelu

... ja muita, joista emme välttämättä edes tiedä!

Summa summarum: Miten Annifin / Finto AI:n saa käyttöön

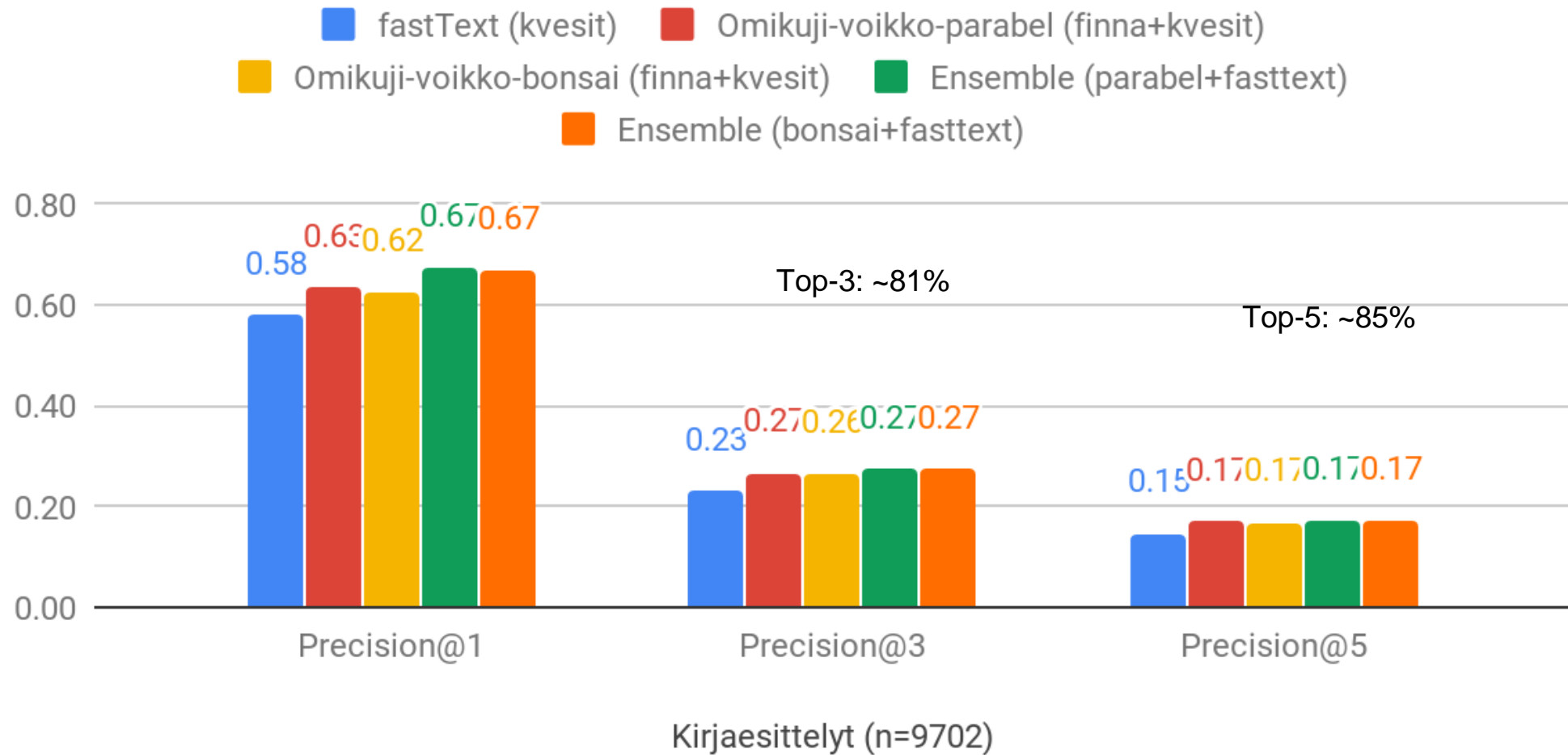
- Lomakkeet osoitteessa ai.finto.fi ja annif.org
- Finto AI on helppo saada käyttöön rajapinnan kautta
- Annifin voi asentaa itse

Automaattinen kuvailu & YKL

YKL-aineistoja koulutukseen ja testaukseen

1. Finnan tietueet, joissa YKL-luokkia
 - a. **suomenkielisiä 6,4M tietuetta**
 - b. ruotsinkielisiä 450k tietuetta
 - c. englanninkielisiä 1,1M tietuetta
2. Kirjavälitykseltä saadut kirjaesittelyt (113k tietuetta)
3. [Makupalat](#)-linkkikokoelman linkit (16k tietuetta)

YKL-algoritmien suorituskyky eri mittareilla (testiaineistona Kirjavälityksen kirjaesittelyt, n=9702)



Tämänhetkistä kehitystyötä

- Luokitukset: osoitteessa <https://dev.annif.org/> on testattavissa YKL- ja Thema-luokitusta Annifilla
- Kommentteja ja toiveita saa antaa joko [yleiseen palautteen kautta](#) tai suoraan allekirjoittaneelle (Mona Lehtinen, etunimi.sukunimi@helsinki.fi)

Tämänhetkistä kehitystyötä

- Sisällönkuvailun asiantuntijaryhmä on perustanut automaattisen kuvailun työryhmän
- Tällä ryhmällä on tarkoitus v. 2023 testata YKL-kuvailua

Automaattisen kuvailun projektisuunnitelma 2022

- Parannetaan automaattisen kuvailun laatua mm. hyödyntämällä koneoppimista ja kieliteknologiaa sekä tekemällä yhteistyötä
- Laajennetaan automaattista sisällönkuvailua luokituksiin (YKL, Thema) ja kaunokirjallisuuteen (KAUNO)
- Kehitetään Finto AI -palvelun toiminnallisuuksia, käytettävyyttä ja saavutettavuutta
- Edistetään automaattisen kuvailun käyttöönottoja Suomessa ja maailmalla

Automaattisen bibliografisen kuvailun kehitys

Hyödyllistä mm. e-vapaakappaleiden vastaanotossa ja julkaisuarkistoissa

- **PDF-tiedosto** → **perusmetatiedot** (tekijä, julkaisuaika, nimeke, sarja....)

Aiemmin testattu **Grobid**-ohjelmaa, nyt käynnissä alustava pilotti **Lingsoft Oy:n** kanssa.

Kiitos!



Juho Inkinen



Mona Lehtinen



Osma Suominen