

# HPD:n vaikutus Annifin ja Finto AI:n kehitykseen

HPD-webinaari 19.1.2021

Osma Suominen

# 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](http://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](http://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.

#### 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](#)

## HPD-hankkeen työkohteet

## Hyödyntäminen Annifissa ja Finto AI:ssa

CSC:n deep learning -kurssi

Neuroverkko-ensemble

Annif 0.43 - 10/2019  
api.annif.org - 3/2020  
Finto AI - 5/2020

Perinteisten koneoppimisalgoritmien testaus

Omikuji-algoritmi

Annif 0.45 - 12/2019  
api.annif.org - 1/2020

fastText-algoritmin hyperparametrien haku

fastText-päivitys

Annif 0.48 - 6/2020

optimoidut fastText-hyperparametrit

Finto AI - 12/2020

Neuroverkkopohjaisten algoritmien testaus

tekstin typistämisen idea

Annif 0.50 - 12/2020  
Finto AI - 12/2020

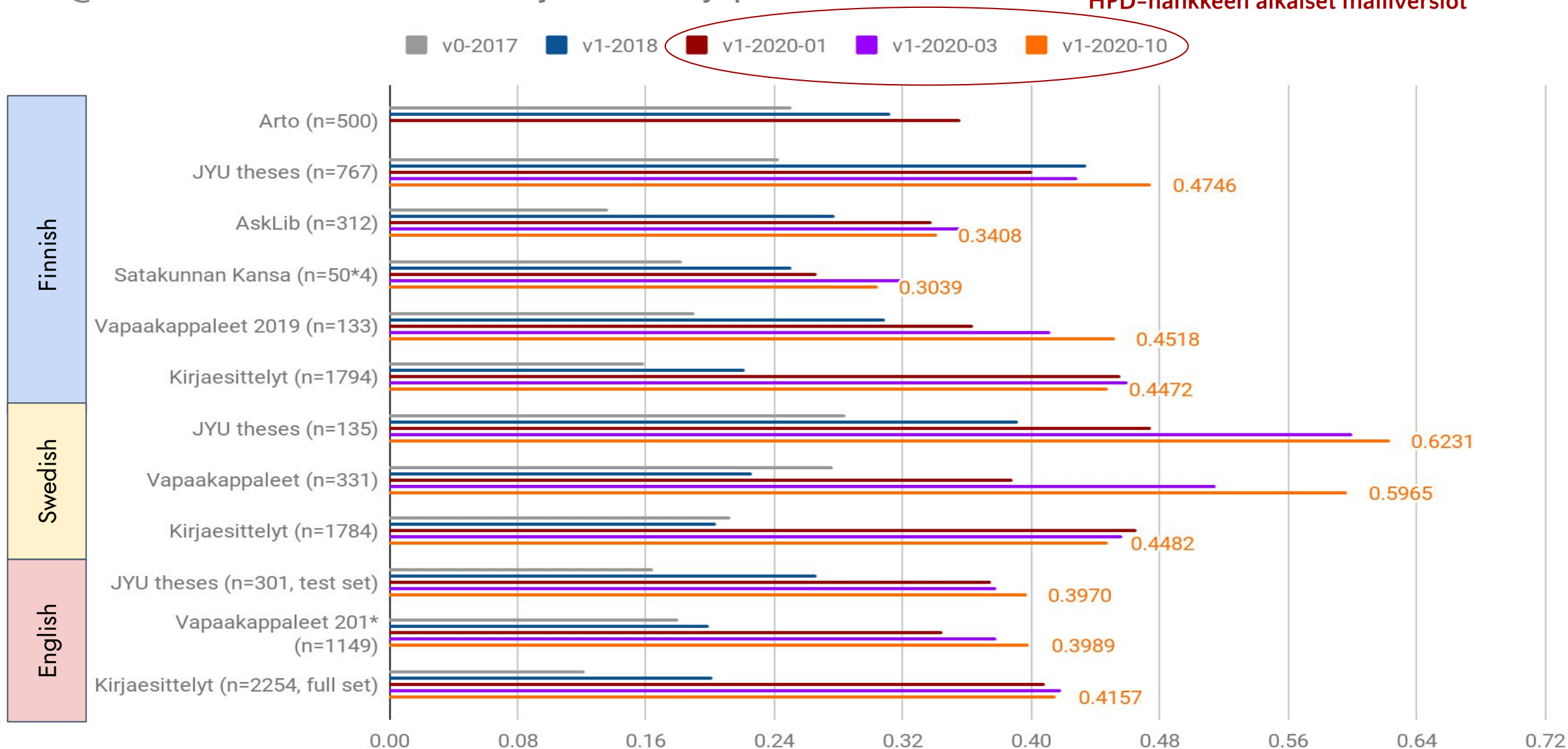
X-BERT / X-Transformer -algoritmi

Vuonna 2021

# Automaattisen sisällönkuvailun laatu on parantunut

F1@5 laatuarviot eri aineistoille ja Annif-rajapintaversioille

HPD-hankkeen aikaiset malliversiot





Kiitos!