

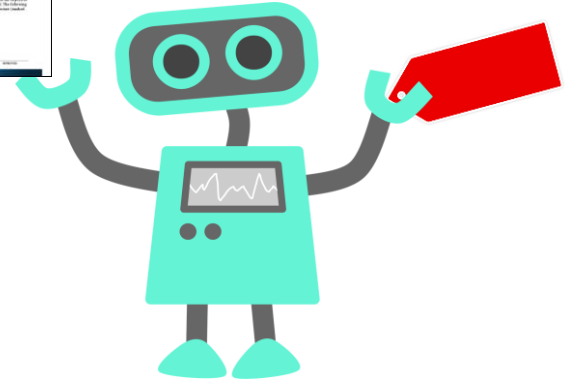
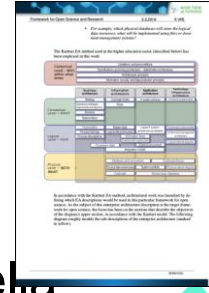
Annifin ja Finto AI:n esittelyt ja kuulumisia

Talonmiehen tuokio 22.4.2021

Mikä on Annif?

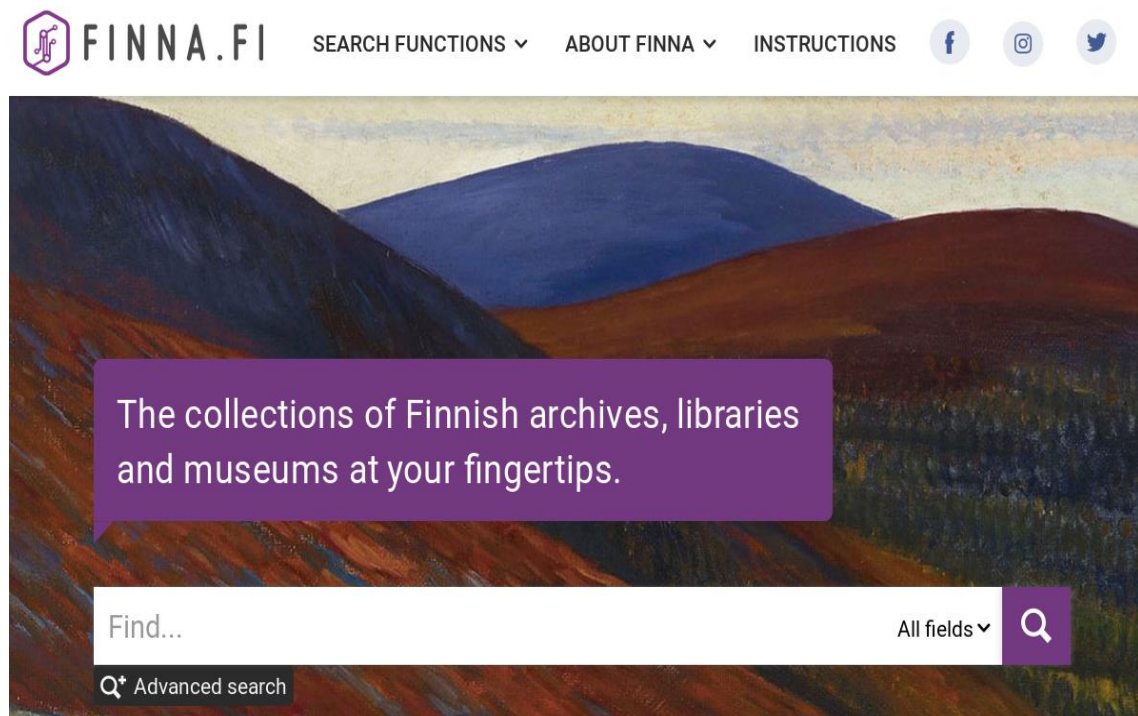
Annif

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

Annif oppii aineistosta, miten tunnistaa aiheita



annif

Annif...

...tukee useita eri algoritmeja automaattiseen sisällönkuvailuun
Algoritmeja voi käyttää **yksitellen** tai **yhdistelminä** (ensemble)



Annif GitHubissa

NatLibFi / Annif

Unwatch 7 Unstar 23 Fork 3

Code Issues 20 Pull requests 0 Projects 0 Wiki Insights Settings

Annif is a multi-algorithm automated classification and subject indexing tool for libraries, archives and museums. This repository is used for developing a production version of the system, based on ideas from the initial prototype. <http://annif.org>

subject-indexing python machine-learning code4lib classification rest-api flask-application connexion Manage topics

766 commits 7 branches 48 releases 5 contributors View license

Branch: master New pull request Create new file Upload files Find file Clone or download

osma add Zenodo DOI badge Latest commit d832514 2 days ago

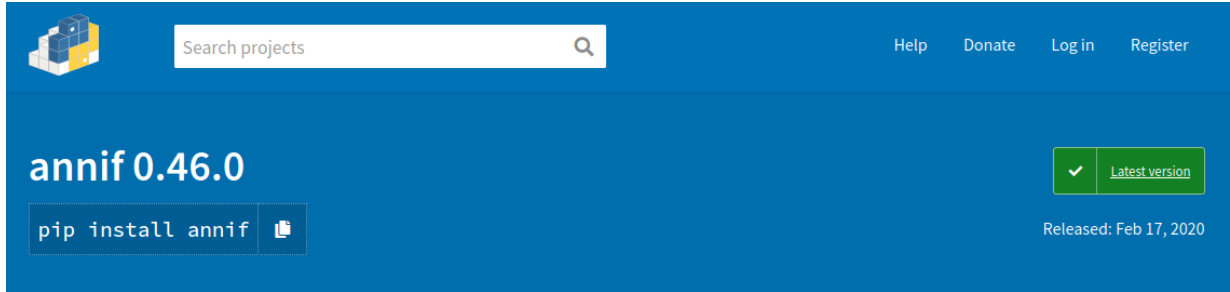
| | | |
|------------------|--|---------------|
| annif | refactor: split off JSON input to document corpus conversion in rest ... | 2 days ago |
| tests | CLI unit test for trying to learn when backend doesn't support it | 2 days ago |
| .codeclimate.yml | more comprehensive Code Climate configuration | a year ago |
| .codecov.yml | Codecov should ignore setup.py | 10 months ago |
| .coveragerc | Generate Codecov reports | 2 years ago |
| .gitignore | Add virtualenv (default? de-facto?) folder to gitignore | 15 days ago |
| .lgtm.yml | Add LGTM configuration excluding fasttext | 5 months ago |
| .scrutinizer.yml | Try to fix pipenv/pip compatibility issue pypa/pipenv#2924 within Scr... | 5 months ago |
| .travis.yml | install deb packages using apt addon (even though they're unnecessary... | a month ago |

Python 3.6+
Apache License 2.0

Kattavat yksikkötestit (99%)
PEP8 tyylisäännöt
[Tekninen dokumentaatio](#) wikissä

<https://github.com/NatLibFi/Annif>

Python CLI tool

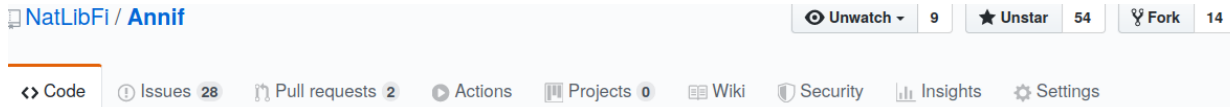


annif 0.46.0

pip install annif

Released: Feb 17, 2020

Latest version



NatLibFi / Annif

Code Issues 28 Pull requests 2 Actions Projects 0 Wiki Security Insights Settings

Annif is a multi-algorithm automated subject indexing tool for libraries, archives and museums. This repository is used for developing a production version of the system, based on ideas from the initial prototype. <http://annif.org>

subject-indexing python machine-learning code4lib classification rest-api flask-application connexion Manage topics

1,423 commits 20 branches 0 packages 61 releases 6 contributors View license

Branch: master - New pull request Create new file Upload files Find file Clone or download -

| Author | Commit Message | Time |
|------------------|--|---------------------------------|
| juhoinkinen | Bump version: 0.46.0 → 0.47.0-dev | Latest commit 2fd7db9 yesterday |
| annif | More readable way to prevent None -> str conversion | yesterday |
| docs | fix-RTD-build-error | 7 months ago |
| tests | Merge branch 'master' into issue157-support-for-notation-codes; manua... | 4 days ago |
| .codeclimate.yml | more comprehensive Code Climate configuration | 2 years ago |
| .codecov.yml | Codecov should ignore setup.py | 2 years ago |
| .codeclimate.yml | Generate Codecov reports | 2 years ago |

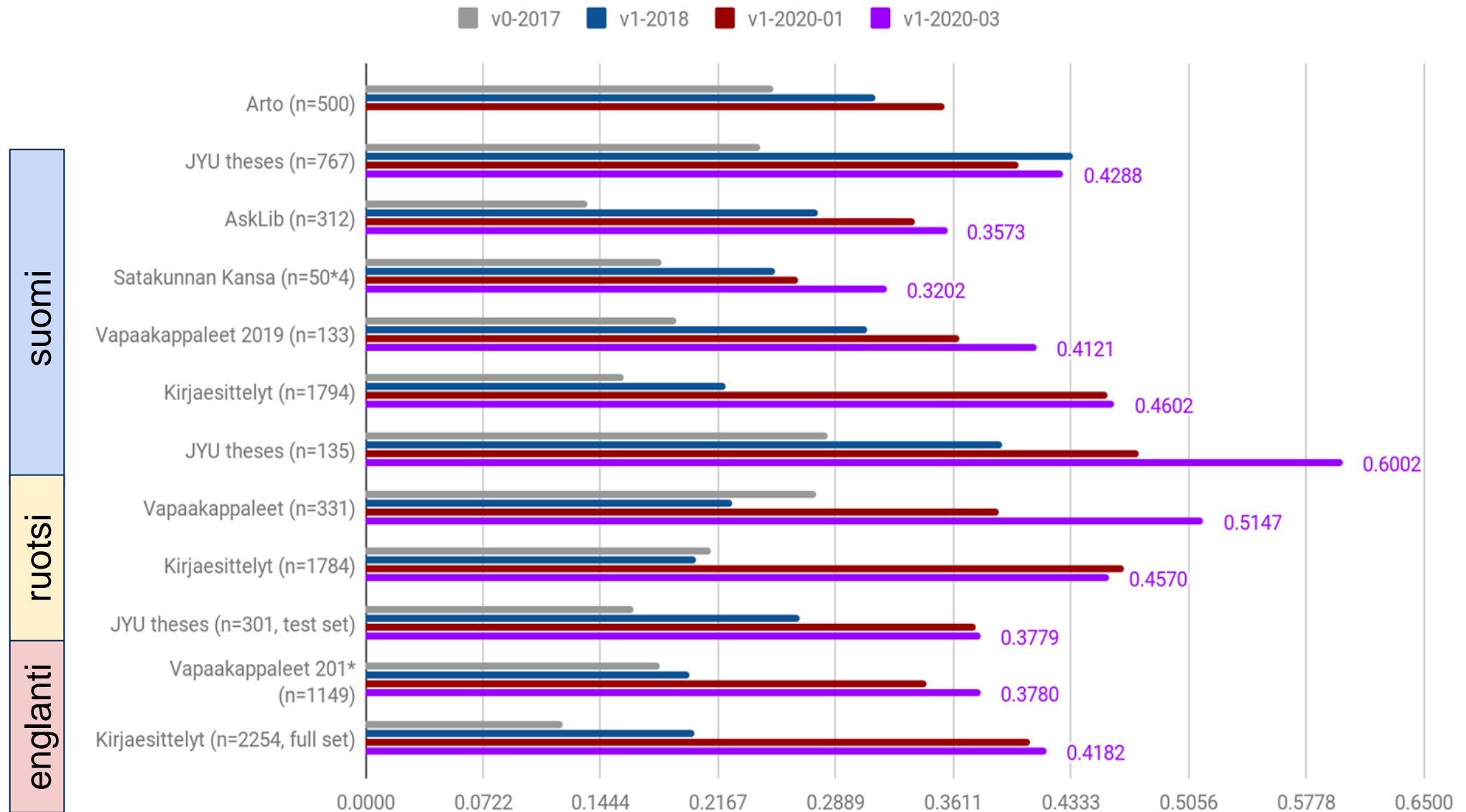


```
jmminkin@lx8-9811-008: /home/local/jmminkin/git/Annif
jmminkin@lx8-9811-008: /home/local/jmminkin/git/Annif 99x35
(Annif) jmminkin@lx8-9811-008: /home/local/jmminkin/git/Annif$ annif
Usage: annif [OPTIONS] COMMAND [ARGS]...

Options:
  --version  Show the flask version
  --help    Show this message and exit.

Commands:
  clear          Initialize the project to its original, untrained state.
  eval          Analyze documents and evaluate the result.
  index         Index a directory with documents, suggesting subjects for...
  learn         Further train an existing project on a collection of...
  list-projects List available projects.
  loadvoc       Load a vocabulary for a project.
  optimize      Analyze documents, testing multiple limits and thresholds.
  routes       Show the routes for the app.
  run          Run a development server.
  shell        Run a shell in the app context.
  show-project  Show information about a project.
  suggest      Suggest subjects for a single document from standard input.
  train        Train a project on a collection of documents.
(Annif) jmminkin@lx8-9811-008: /home/local/jmminkin/git/Annif$
```

F1@5 laatuarviot eri aineistoille ja Annif-rajapintaversioille



Annifin laatu ihmisten arvioitavana



Kuva: Mikko Lappalainen.

Finto AI julkaistiin 28.5.2020

Automaattinen sisällönkuvailu osaksi Finton palveluja

1.6.2020 13:02:40 EEST | [Helsingin yliopisto](#)



Finto laajenee automaattisen sisällönkuvailun palveluihin Finto AI -työkalun ja sen rajapintapalvelun myötä. Palvelu tukee suomen-, ruotsin- ja englanninkielisen tekstin automaattista sisällönkuvailua.

Finto AI on saatavilla osoitteessa ai.finto.fi, josta löytyvän lomakkeen avulla palvelua voi kokeilla ja käyttää. Automaattinen sisällönkuvailu toimii parhaiten opinnäytetyö- ja tietokirja-

API-palvelu

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

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

Kuvailtava teksti

Kopioi tähän tekstiä ja paina "Anna aihe-ehdotukset"-nappia 

ai.finto.fi

Sisällönkuvailu

Sanasto ja tekstin kieli

YSO suomi 

Ehdotusten enimmäismäärä **10** 15 20

Anna aihe-ehdotukset

annif

Yleiskäyttöinen **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

fintoai

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

Jyväskylän yliopiston JYX-julkaisuarkisto

Opiskelijat lataavat omat gradunsa, Annif / Finto AI ehdottaa niille aiheita (toteutus Ari Häyrinen)

Keywords

Keyword suggestions

Choose valid keywords by clicking

- information management systems [YSO]
- metadata [YSO]
- connections (technical systems) [YSO]
- content management [YSO]
- multimedia (information technology) [YSO]
- digital libraries [YSO]
- XML [YSO]
- semantic web [YSO]
- open source code [YSO]
- open data [YSO]
- user-centeredness [YSO]
- archives (memory organisations) [YSO]
- seeking [YSO]
- Works [YSO]
- cloud services [YSO]
- electronic publications [YSO]

Your own keywords

Comma separated list

keyword 1, keyword 2



Osuva-julkaisuarkisto, Vaasan yliopisto & Theseus, Ammattikorkeakoulut

Sama ajatus kuin JYXissä: Finto AI ehdottaa aiheita uusille dokumenteille

Asiasanat:

Annif-ehdotukset

- working abroad
- career development
- career
- adaptation (change)
- expatriates
- families (groups)
- managers and executives
- human resources
- work
- returnees (immigrants)

Lisää

Lisää

Syötä asiasanat, jokainen asiasana omaan kenttäänsä. Paina siis jokaisen asiasanan jälkeen Lisää-nappia. Kirjoita tarvittava määrä asiasanan alkua, jolloin ennakoiva tekstinsyöttö ehdottaa asiasanoja. Muista myös valita yllä olevasta laatikosta Annif-ehdotukset, jotka perustuvat edellisessä vaiheessa syöttämäsi kokotekstin sisältöön.

Pilottikäytössä
2.3.2020 alkaen,
toteutus
Anis Moubarik

Käyttö Kansalliskirjaston järjestelmissä

The screenshot displays the LUT University website with three search interfaces. At the top, the LUT University logo and navigation icons are visible. The Doria interface includes a search bar and a description: "Doria on Kansalliskirjaston ylläpitämä julkaisuarkisto, josta löydät tietoa Doriasta". The LUTPub interface features a search bar and a description: "LUTPub on LUT-yliopiston avoin julkaisuarkisto. LUTPub sisältää LUT-yliopiston kandidaatintöitä ja -tutkimuksia, diplomitöitä, pro gradu -tutkimuksia, lisensiaattintöitä, väitöskirjoja sekä näiden tiivistelmiä. Lisäksi LUTPubiin on tallennettu artikkeleita, tutkimusaineistoja ja esitysmateriaaleja." The Julkari interface includes a search bar and a description: "Julkari on Sosiaali- ja terveysministeriön julkaisuista saatavilla sähköinen versio laitoksen osalta myös julkaisurekisterinä. Julkariin julkaisuja voi hakea mm. julkaisujärjestetty kokoelmittain, joten sanahaun avulla voit löytää julkaisuja. Julkari on vapaasti käytettävissä. Julkaisut löydät myös julkaisurekisteristä." The Theseus interface includes a search bar and a description: "Theseus - ammattikorkeakoulujen opinnäytetyöt ja julkaisut verkossa". The Lauda interface includes a search bar and a description: "Lauda on Lapin yliopiston sähköinen julkaisuarkisto. Lauda sisältää Lapin yliopistossa tuotettuja aineistoja kuten kirjoja ja opinnäytteitä. Lisäksi Laudassa on joitakin Barentsin alueella julkaistavia kausijulkaisuja artikkeleina. Lisätietoa Laudasta".

The screenshot displays the National Library of Finland website with the digital archive interface. The header includes the logo "1640 KANSALLIS KIRJASTO" and the title "Digitaalisten julkaisujen arkistointi". Below the header, there are search options for "Pakollinen tieto" and "Vähintään toinen (sähköpostiosoite tai puhelin) on pakollinen tieto." The "Luovuttajan tiedot" section includes fields for "Yhteystietokäyttö*", "Sähköpostiosoite*", "Puhelinnumero*", and "Organisaatio". The "Julkaisun tiedot" section includes fields for "Julkaisujen lukumäärä" and "Julkaisun tyyppi". The "Perustiedot" section includes the "ISBN (viivoilla)" field.

E-vapaakappaleiden vastaanotto ja käsittely

Annifin käyttöönottoja ja testejä muualla

- Kirjavälitys oy
- YLE^{1,2}
- Kavi³
- Saksalaisia kirjastoja (kansalliskirjasto, taloustieteen kansalliskirjasto)
- Alankomaiden kansalliskirjasto⁴ ja tilastokeskus
- LCSH-testejä USA:ssa

Lisätiedot

- ¹ Suominen, Osma & Virtanen, Pia. 2020. 'Yle Meets ANNIF – an Open Source Tool for Automated Subject Indexing'. Presented at the EBU MDN Workshop 2020, June 10.
<https://tech.ebu.ch/contents/publications/events/presentations/mdn2020/yle-meets-annif--an-open-source-tool-for-automated-subject-indexing>.
- ² Nikkarinen, Irene. 2021. 'Annif <3 Yle 2.0: Annifin osittainen käyttöönotto artikkeleiden koneavusteisessa asiasanoituksessa'. Esitelmän [kalvot](#)
- ³ Lehtonen, Tommi, and Juha Piukkula. 2020. 'Automaattinen asiasanoitus Radio- ja televisio-ohjelmatietokanta Ritvassa'. Informaatiotutkimus 39 (1): 27–45–27–45.
<https://doi.org/10.23978/inf.88107>.
- ⁴ Haighton, Thomas, and Sara Veldhoen. 2020. 'Assisted Keyword Assignment Using Annif. KB Lab: The Hague.' 2020. <http://kbresearch.nl/annif/>.

Automaattisen kuvailun projektisuunnitelma

- Julkaistu os. <https://www.kiwi.fi/x/8YBND>.
- Keskeisiä tavoitteita ovat [Annif](#)-työkalun jatkokehitys sekä [Finto AI](#) -palvelun parantaminen mm. laajentamalla tarjolla olevien sanastojen valikoimaa.

Automaattisen kuvailun projektisuunnitelma

- Pyrkimyksenä on myös edistää sisällönkuvailun ja bibliografisen kuvailun automatisointia Kansalliskirjaston omissa järjestelmissä sekä rakentaa yhteistyötä automaattisen kuvailun tiimoilta kotimaassa ja kansainvälisesti.

Annifin tekninen kehitystyö

- <https://github.com/NatLibFi/Annif/releases/tag/v0.52.0>
- Finton [wiki](#)

Verkostot ja yhteistyö: HPD-projektista

- Kansalliskirjaston, CSC:n ja Kansallisarkiston yhteinen High Performance Digitisation -hanke on taannoin päättynyt.
- Hanke pureutui kasvavan digitaalisen aineistomäärän käsittelyn haasteeseen.
- Tarkoituksena oli helpottaa muistiorganisaatioiden aineiston käsittelyä / annotointia älykkäillä puoliautomaattisilla ratkaisuilla. Hankkeesta voi lukea lisää mm. tästä [blogitekstistä](#).

Kirjoitukset

- Blogikirjoitus ([suomeksi](#) ja [englanniksi](#))
- Sisällönkuvailun automatisoinnin haasteita ja ratkaisuja kulttuuriperintöorganisaatiossa
 - Whitepaper, tekninen kuvaus automaattisen kuvailun palvelusta:
<http://urn.fi/URN:ISBN:978-951-51-7233-4>
- Automaattisen kuvailun palvelun integroiminen Kansalliskirjaston järjestelmäkokonaisuuteen – tietovirrat ja prosessit
 - Proof of Concept, käsittelee automaattisen kuvailun käyttöönottoa Kansalliskirjastossa:
<http://urn.fi/URN:ISBN:978-951-51-6986-0>

Verkostot ja yhteistyö: suunnitelmia

- Yhteistyö Linné-yliopiston ja Ruotsin kansalliskirjaston kanssa
- Annif-tutoriaali Uumajassa, Baltian maiden kansalliskirjastoille
- Pohjoismaiseen yhteistyöprojektiin 2022–2023 haettu rahoitusta NeIC:ltä (Nordic e-Infrastructure Collaboration)

Annifiin liittyvää koulutusta tulossa toukokuussa

- 19.5.
 - Ilmoittautuminen ja ohjelma: <https://www.kiwi.fi/x/FwW5D>
- Perustuu aikaisempaan [tutoriaaliin](#)
- Miten Annifin saa käyttöön omalle aineistolle?
- Automaattisen kuvailun verkoston s-postilistalla tiedotusta mm. tästä aiheesta (liittymisohjeet verkoston [sivulla](#))

www.kansalliskirjasto.fi

Kiitos!