

annif

fintoai

(Annif ja) Finto AI sisällönkuvailun apuvälineenä

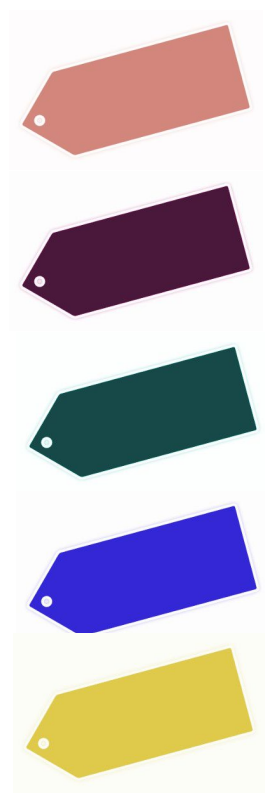
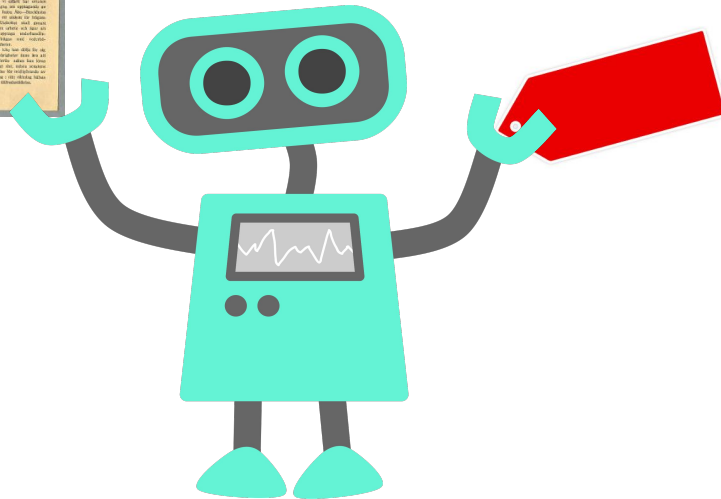
Mona Lehtinen, Osma Suominen, Juho Inkinen
Melindan syyswebinaari



Sisältö

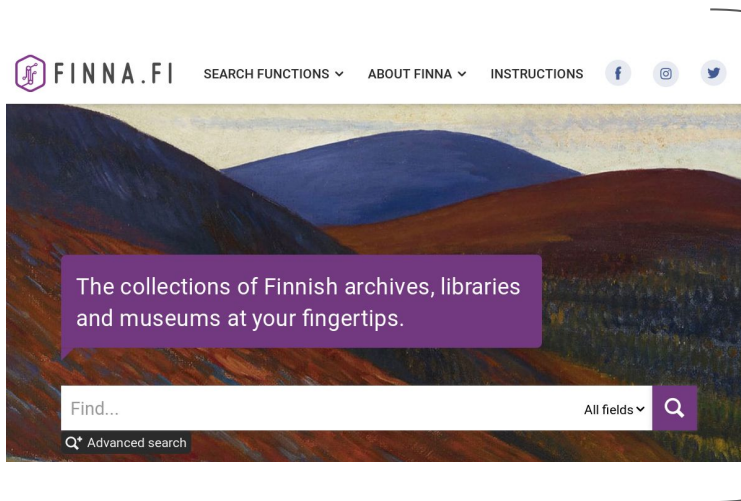
1. Annifin ja Finto AI:n kehitys
2. Automaattisen asiasanoituksen laatu
3. Yhteisö
4. Käyttöönottoja
5. Opittua

1. Annifin ja Finto Al:n kehitys



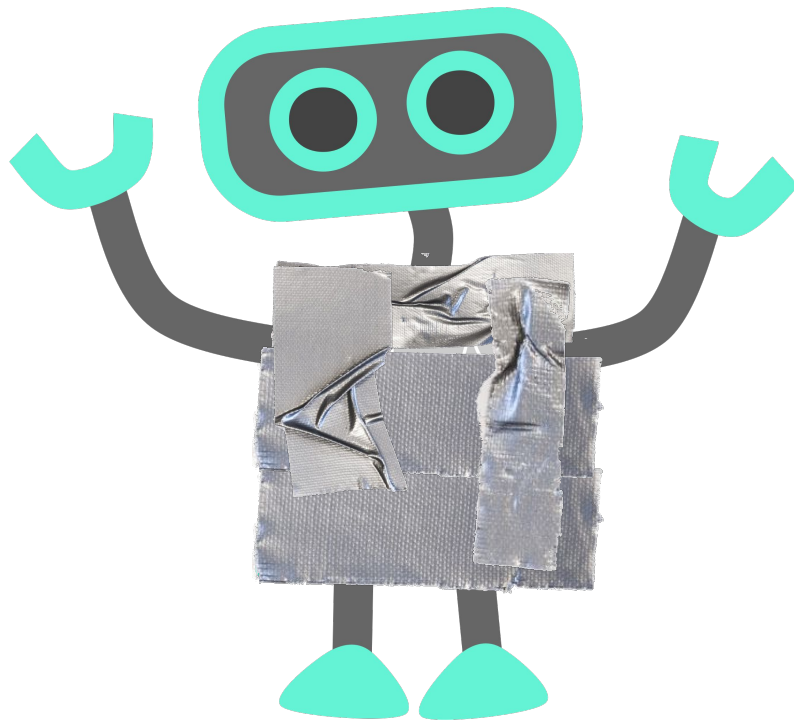
YSO, General Finnish Ontology
with 30,000+ subjects

Annif oppii aineistosta, miten tunnistaa aiheita



annif

Ensimmäinen prototyyppi (2017)



Annifin kehityksen lähtökohtia (2018 →)

1. Monikielisyys ja sanastoriippumattomuus
2. Erilaisten automaattisen kuvailun ratkaisujen / algoritmien tukeminen
3. Toteutus avoimen lähdekoodin hengessä komentorivillä, REST-rajapinnan sekä verkkolomakkeen kautta



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

Annif GitHubissa

Python 3.6+

Apache License 2.0

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



annif 0.49.0

```
pip install annif
```

Automated subject indexing and classification tool

Python paketti PyPI:ssä



natlibfi / annif

Docker-kuvia Quay.io:ssa

Nettisivu ja testilomake annif.org

INPUT TEXT

SWIB focuses on Linked Open Data (LOD) in libraries and related organizations. It is well established as an event where IT staff, developers, librarians, and researchers from over the world meet and mingle and learn from each other. The topics of talks and workshops at SWIB revolve around opening data, linking data and creating tools and software for LOD production scenarios. These areas of focus are supplemented by presentations of research projects in applied sciences, industry applications, and LOD activities in other areas.

As usual, SWIB20 will be organized by the ZBW - Leibniz Information Centre for Economics and the North Rhine-Westphalian Library Service Centre (hbz). The conference language is English.

Would you like to share your experiences working on an interesting service, research topic or project – not just what you did, but also how you did it?

For this SWIB rendition we adjusted the formats to the online environment:

Presentations (15 minutes plus 5 q&a)

Practical workshops or tutorials (maximum 120 min)

PROJECT (VOCABULARY AND LANGUAGE)

YSO NN Ensemble English ▼

MAX # OF SUGGESTIONS

10 15 20

Get suggestions →

annif

SUGGESTED SUBJECTS

- data acquisition
- libraries
- linked open data
- information retrieval
- metadata
- data mining
- Finland
- information and communications technology
- information technology
- social media

Finto AI



[About](#) [Feedback](#)

[suomeksi](#) [på svenska](#)

Finto AI suggests subjects for a given text. It's based on Annif, a tool for automated subject indexing. [Read more...](#)

API service

Finto AI is also an API service that can be integrated to other systems.

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

Enter text to be indexed

In computer science, artificial intelligence (AI), sometimes called machine intelligence, is intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans and animals. Leading AI textbooks define the field as the study of "intelligent agents": any device that perceives its environment and takes actions that maximize its chance of successfully achieving its goals [1] Colloquially, the term "artificial intelligence" is often used to describe machines (or computers) that simulate human intelligence, or that associate with the human mind, such as expert systems.

As machines become increasingly capable, some have argued that they should be removed from the definition of intelligence. The Turing Test, proposed by Alan Turing, is often cited in the context of AI. In Tesler's Theorem says "AI is whatever we can't get a computer to do". AI recognition is frequently equated with the ability to perform tasks that were previously routine technology.[6] Modern AI research includes the development of systems that successfully understand human speech,[7] competing at the highest level in strategic game systems (such as chess and Go),[8] autonomously operating cars, intelligent routing in content delivery networks, and military simulations.

**Avattu
toukokuussa
2020**

ai.finto.fi

Subject indexing

Vocabulary and text language

YSO English

Maximum # of suggestions

10

15

20

Get subject suggestions

Suggestions

- [artificial intelligence](#)
- [machine learning](#)
- [intelligence \(mental properties\)](#)
- [information technology](#)
- [computational science](#)
- [computer science](#)
- [computers](#)
- [computer-assisted teaching](#)
- [learning](#)
- [automation](#)

2. Automaattisen asiasanoituksen laatu

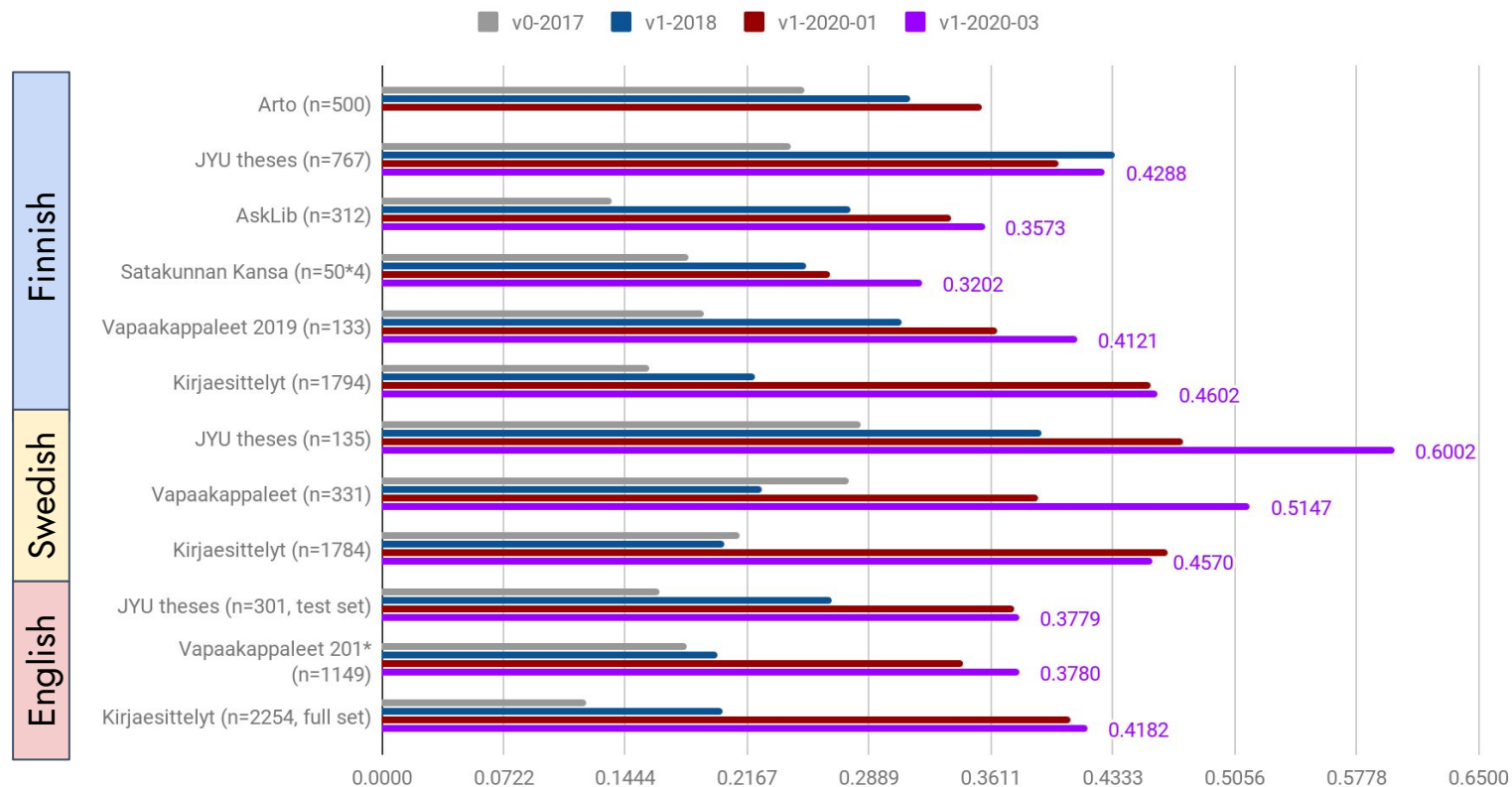
Koulutus- ja arviointiaineistoa

1. Metadata-aineistoa Finnasta
2. Kysy kirjastonhoitajalta -aineistoa
3. Jyväskylän yliopiston opinnäytteitä
4. Kirjojen esittelytekstejä
5. E-aineistoa Kansalliskirjastosta
6. ...

[Annif-corpora](#) GitHubissa

Vertailu ihmisen tekemiin asiasanoituksiin

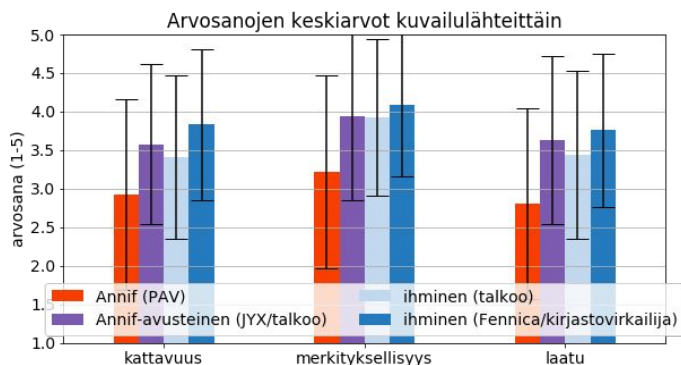
F1@5 pisteet eri aineistoilla ja versioilla



Annif ihmisten arvioitavana

Vuoden 2019 kirjastoverkkopäivien työpajassa, **48 osalistujaa** arvioi **50 dokumentin** asiasanoja. Asiasanoituksen oli tehnyt joko ihminen tai Annif.

Paraskaan algoritmi (pun.) ei saavuttanut ihmiskuvailijoiden tasoa laatuarvioissa (vas.) tai poistettaviksi ehdotettujen asiasanojen määrässä (oik).



Poistettujen asiasanojen keskimääräiset %-osuudet kuvailulähteittäin

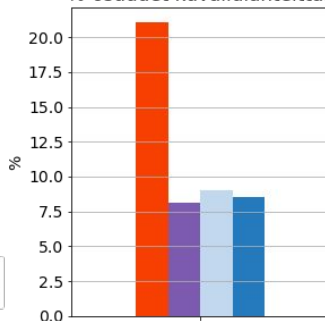


Photo: Mikko Lappalainen.

Lehtinen M., Inkinen J. & Suominen O. (2019). Aaveita koneessa: Automaattisen sisällönkuvailun arviointia Kirjastoverkkopäivillä 2019. [Tietolinja, 2019\(2\)](http://urn.fi/URN:NBN:fi-fe2019120445612). <http://urn.fi/URN:NBN:fi-fe2019120445612>

Annif-Leiki -vertailu Ylessä

- Annif vs Leiki 28:n ihmisen arvioitavana
- Noin 100 suomen- ja ruotsinkielistä artikkelia avainsanoineen
 - liike-elämä, tiede, kulttuuri, urheilu

Suomi: Annif **hieman parempi** kuin Leiki

Ruotsi: Annif **huomattavasti parempi** kuin Leiki

Arvioi asiasanojen osuvuutta

Voit katsoa asiasanan kuvauksen käymällä osoitteessa "https://meta.api.yle.fi/v1/concepts.json?app_id=HIEKKALAATIKKO&app_key=HIEKKALAATIKKO&yle_id=YLE_ID" ottamalla YLE_ID:n asiasanan suluista

	essential keskeinen	ok ok	non relevant epärelevantti	wrong väärä
professorit (18-211223)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
aurinkokunnat (18-8786)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Markku Wilenius (18-295234)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
fossiiliset polttoaineet (18-2396)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
ilmasto (18-215534)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
tulevaisuus (18-211385)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Arvosteluasteikko

Avainsanat

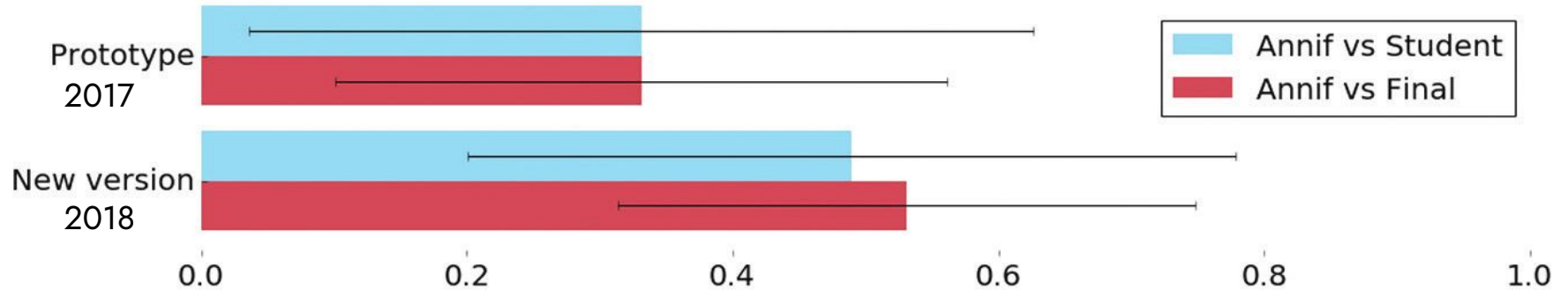
Käytännön arviointia

JYX julkaisuarkisto, Jyväskylän yliopisto:

Samankaltaisuus (F1)

a) opiskelijan valitsemien ja Annifin ehdottamien välillä (sin)

b) kirjastovirkailijan / informaatikon vahvistamien ja Annifin ehdotusten välillä (pun)



Suominen, O., 2019. Annif: DIY automated subject indexing using multiple algorithms. *LIBER Quarterly*, 29(1), pp.1–25. DOI: <http://doi.org/10.18352/lq.10285>

3. Yhteisö

Yhteistyökumppaneita



KB } nationale
bibliotheek



KANSALLISARKISTO



University of Jyväskylä | JYX Digital Repository

KANSALLINEN AUDIOVISUAALINEN INSTITUUTTI
NATIONELLA AUDIOVISUELLA INSTITUTET
NATIONAL AUDIOVISUAL INSTITUTE



ZBW

Leibniz-Informationszentrum
Wirtschaft
Leibniz Information Centre
for Economics



KIRJASTOT.fi



Suomi.fi

Wiki: ohjeet ja dokumentaatio

GitHubissa

- issues
- pull requests

Welcome to the Annif wiki!

- [Getting started](#)
- [System requirements](#)
- [Optional features and dependencies](#)
- [Usage with Docker](#)
- [Architecture](#)
- [Commands](#)
- [Web user interface](#)
- [Corpus formats](#)
 - [Document corpus formats](#)
 - [Subject vocabulary formats](#)
- [Project configuration](#)
- [Analyzers](#)
- [Achieving good results](#)
- [Reusing preprocessed training data](#)
- [Running as a WSGI service](#)
- [Backends/Algorithms supported by Annif](#)
 - Regular backends for automated subject indexing and classification
 - [Backend: TF-IDF](#)
 - [Backend: fastText](#)
 - [Backend: Omikuji](#)
 - [Backend: Maui](#)
 - [Backend: vw_multi](#)
 - Fusion/Ensemble backends that combine results from other backends
 - [Backend: Ensemble](#)

Welcome to the [Annif](#) users' mailing list / web forum! This list can be used for

- general discussion about Annif, its features and usage scenarios
- asking for help with installing or running Annif
- future directions for Annif
- announcements for new versions and other Annif-related news

Annif-users ([Google group](#))



- | | | | | |
|--|---------------------------------|---|--------------|---|
| | Remi Malessa 2 | Loading OCLC FAST Vocabulary – Actually, my server has around 32MB of memory and it s | 14.55 | ☆ |
| | osma.s..... , sale...@g... 4 | Annif presentation and workshop at SWIB20 online conference – Thanks Osma for sharing | 10. marrask. | |
| | mona.l...@... , ... Annif... 13 | Annif tutorial at the 2020 DCMI Virtual Conference – Hi All, There is also a second online h | 7. lokak. | ☆ |
| | sara.v...@kb.nl | Paper on Annif for categorizing laws – Dear all, Some of you may have read my blogpost o | 8. syysk. | ☆ |
| | haig...@... , osma.s..... 2 | Running Annif as a WSGI service - connexion module error – Hi Thomas! It looks like the P | 7. syysk. | ☆ |
| | juho.k..... , osma.s..... 4 | Results with Annif in employment service context – Hi Juho! juho.k...@gmail.com kirjoitti 2 | 31. elok. | ☆ |
| | stephane5... , osma.s..... 2 | Matter of vocabulary – Hi Stephane, Since you have a SKOS vocabulary, you can just load it | 26. elok. | ☆ |
| | osma.s...@helsinki.fi | ANN: Annif 0.49 released – Annif 0.49 has been released! https://github.com/NatLibFi/Ann | 30. heinäk. | ☆ |

Automaattisen kuvailun verkosto

<https://www.kiwi.fi/x/qgrUBg>

Automaattisen kuvailun verkosto on avoin verkosto, jonka tavoitteena on jakaa tietoa erilaisista kuvailun automatisointiin liittyvistä ratkaisuista. Verkosto kokoontuu tarpeen mukaan muutaman kerran vuodessa.

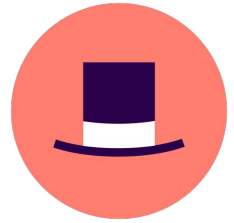
Verkostolle on myös perustettu s-postilista, jota tullaan käyttämään lähinnä tiedottamiseen. Listalle saa liittyä vapaasti lähettämällä sähköpostia osoitteeseen majordomo (ät) helsinki.fi. Huomaa, että:

1. Viestin otsikkokentän tulee olla tyhjä
2. Tekstiosassa tulee lukea pelkästään: subscribe automaattinen-kuvailu (Mahdollinen allekirjoituskin kannattaa siis poistaa viestistä.)

Listalta voi erota vastaavasti lähettämällä em. osoitteeseen tyhjällä otsikolla viesti unsubscribe automaattinen-kuvailu.

Annif tutoriaali

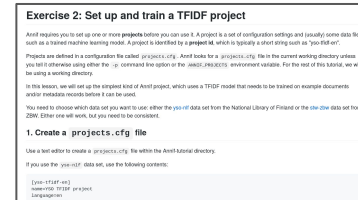
itseopiskeluun



SWIB19
Semantic Web in Libraries

DCMI Virtual, 2020
September 14th-25th, 2020

SWIB20
Semantic Web in Libraries



Videot ja tehtävät YouTubessa & GitHubissa!



4. Käyttöönottoja

JYX-julkaisuarkisto, Jyväskylän yliopisto

Opiskelijat tallentavat väitöskirjoja ja graduja, Annif / Finto AI ehdottaa aiheita*

Keywords

<p>Keyword suggestions Choose valid keywords by clicking</p>	<ul style="list-style-type: none"><input type="checkbox"/> information management systems [YSO]<input type="checkbox"/> metadata [YSO]<input type="checkbox"/> connections (technical systems) [YSO]<input type="checkbox"/> content management [YSO]<input type="checkbox"/> multimedia (information technology) [YSO]<input type="checkbox"/> digital libraries [YSO]<input type="checkbox"/> XML [YSO]<input type="checkbox"/> semantic web [YSO]<input type="checkbox"/> open source code [YSO]<input type="checkbox"/> open data [YSO]<input type="checkbox"/> user-centeredness [YSO]<input type="checkbox"/> archives (memory organisations) [YSO]<input type="checkbox"/> seeking [YSO]<input type="checkbox"/> Works [YSO]<input type="checkbox"/> cloud services [YSO]<input type="checkbox"/> electronic publications [YSO]
<p>Your own keywords Comma separated list</p>	<input type="text" value="keyword 1, keyword 2"/>

Toteutus (DSpace & [GLAMpipe](#)): Ari Häyrinen

*käytössä YSO

Osuva / Vaasan yliopisto

Trepo / Tampereen yliopisto

Theseus / Ammattikorkeakoulut

Sama idea kuin JYX: issä; opiskelijat tallentavat opinnäytteensä,
Annif / Finto AI ehdottaa asiasanoja

Asiasanat:

Pilotti Osuvassa maaliskuussa
2020, muut myöhemmin.

DSpace-integraation tekninen
toteutus: Anis Moubarik.

Annif-ehdotukset

<input type="checkbox"/> working abroad	<input type="checkbox"/> families (groups)
<input type="checkbox"/> career development	<input type="checkbox"/> managers and executives
<input type="checkbox"/> career	<input type="checkbox"/> human resources
<input type="checkbox"/> adaptation (change)	<input type="checkbox"/> work
<input type="checkbox"/> expatriates	<input type="checkbox"/> 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.

E-vapaakappaleet

Marraskuussa 2020 **Finto AI** käyttöön e-vapaakappaleiden luovutuslomakkeella.

Toteutus: Erik Lindgren, Mikko Merioksa, Satu Niininen



1640
KANSALLIS
KIRJASTO

fi sv en

Digitaalisten julkaisujen arkistointi

* Pakollinen tieto

*Vähintään toinen (sähköpostiosoite tai puhelin) on pakollinen tieto.

Luovuttajan tiedot

Yhteyshenkilö * Tämä on pakollinen kenttä

Sähköpostiosoite *

Puhelinnumero *

Organisaatio

Muista nämä tiedot.

Julkaisun tiedot

Julkaisujen lukumäärä

Luovutan yhden julkaisun Luovutan useita julkaisuja

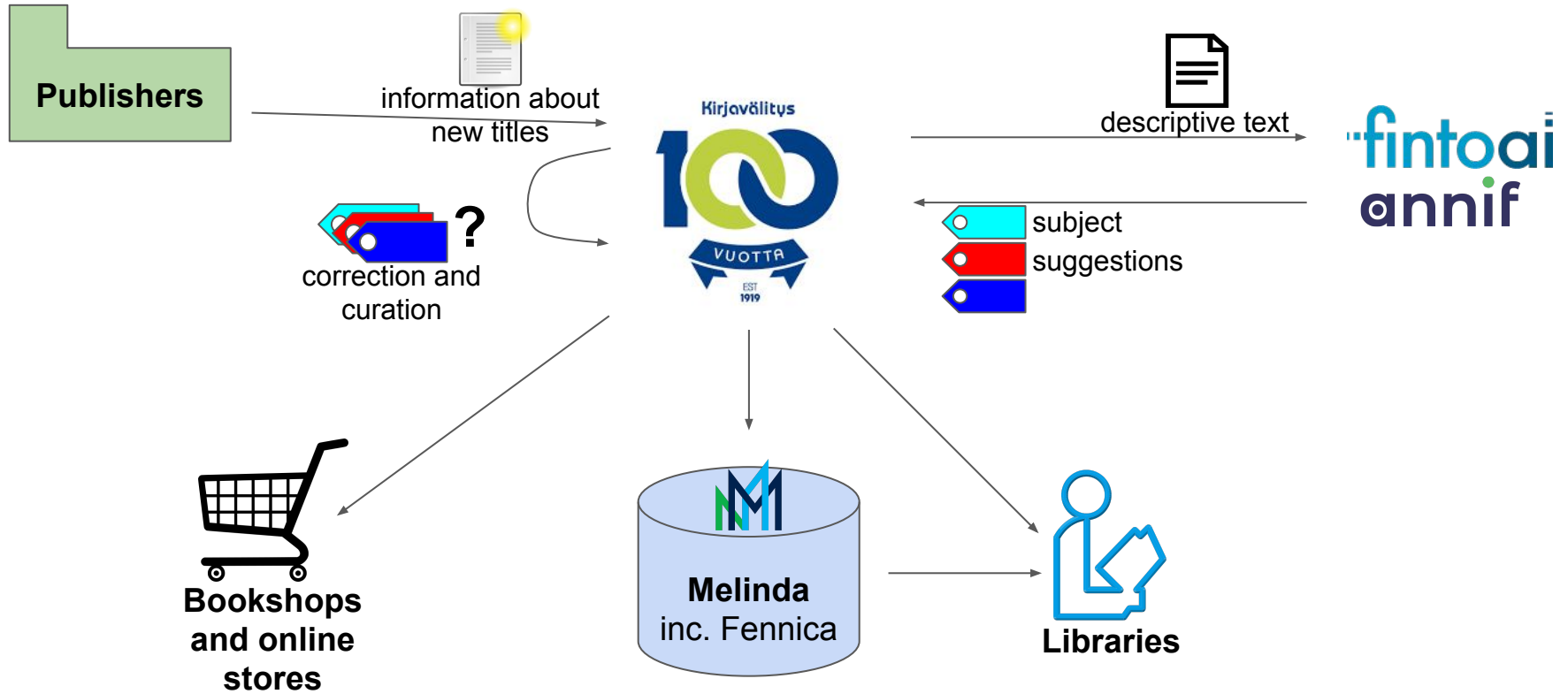
Julkaisun tyyppi

kirja muotti äänite muu

Perustiedot

ISBN (viivoilla)

Kirjavälitys Oy



5. Opittua

Opittua

- Asiasanoituksen haasteet
- Algoritmit toimivat lähes aina paremmin yhdessä kuin erikseen
- Arviointi on jatkuva ja laaja prosessi
- Vaiheittainen siirtyminen testauksen kautta tuotantoon
- Yhteistyössä on voimaa

Lueffavaa

Lehtonen, T., Piukkula, J. Automaattinen asiasanoitus Radio- ja televisio-ohjelmatietokanta Ritvassa. Informaatiotutkimus 39 (1), 2020. DOI: <https://doi.org/10.23978/inf.88107>

Martijn Kleppe, Sara Veldhoen, Meta van der Waal-Gentenaar, Brigitte den Oudsten, & Dorien Haagsma. (2019). Exploration possibilities Automated Generation of Metadata. DOI: <http://doi.org/10.5281/zenodo.3375192>

Romein, C.A., Gruijter, M.D., & Veldhoen, S. (2020). The Datafication of Early Modern Ordinances. DH Benelux Journal, issue 2, 2020. <https://journal.dhbenelux.org/journal/issues/002/article-23-romein/article-23-romein.html>

Kiitos!



Juho Inkinen



Mona Lehtinen



Osma Suominen