

Annif, Finto AI ja HPD-hanke

Osma Suominen

14.9.2020

Tiedonhallinnan ohjausryhmä

Extrablad till
ÅBO UNDERRÄTTELSE

N:o 2. Åbo, den 6 Mars 1917.

Finlands oavhängighet.

Landtagen omfattar regeringens proposition om Finlands fullständiga oavhängighet och ansluter sig till huvudsakligen i regeringens program för tryggandet av landets nya ställning. Beslutet fattades med 100 rösterna mot 28 vilka till följande ett av socialdemokraternas förordnande förslag.

Finlandens ställning.
Finland har varit ett självständigt och oavhängigt land sedan 1809. Dess ställning som självständigt och oavhängigt land har varit en förutsättning för dess utveckling och förvärdning. Detta har varit en förutsättning för dess utveckling och förvärdning. Detta har varit en förutsättning för dess utveckling och förvärdning.

Finlandens ställning.
Finland har varit ett självständigt och oavhängigt land sedan 1809. Dess ställning som självständigt och oavhängigt land har varit en förutsättning för dess utveckling och förvärdning. Detta har varit en förutsättning för dess utveckling och förvärdning.

Pris 25 penn.

Journal of
Learning Sciences and Research
3, 2 (2015) 1-8 (8)

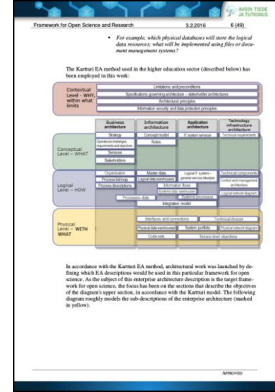
• For example, which physical disciplines will cover the logical skills necessary to deal with the implementation using EA or document management systems?

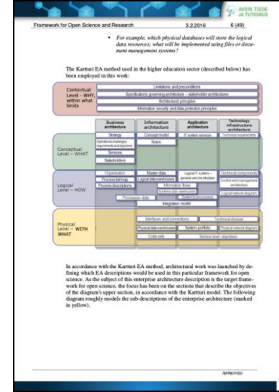
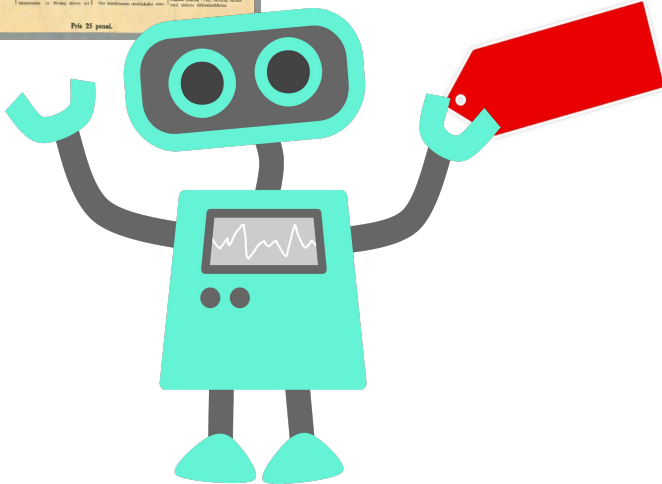
The Kartus EA method used in the higher education sector (described below) has been reported in the work:

Conceptual Level - WET method used	Knowledge of mathematics		Technology applications
	Mathematical modelling	Mathematical problem solving	
Conceptual Level - POW	Mathematical modelling	Mathematical problem solving	Technology applications
	Mathematical modelling	Mathematical problem solving	
Practical Level - WET method used	Mathematical modelling	Mathematical problem solving	Technology applications
	Mathematical modelling	Mathematical problem solving	

In accordance with the Kartus EA method, additional work was conducted by the Faculty EA description would be used in this particular framework for open source. An objective of this enterprise architecture description is to design these work for open source. The focus has been on the activities that describe the objective of the enterprise architecture, in accordance with the Kartus model. The following diagram simply models the sub-structure of the enterprise architecture (based on factors).

www.kartus.com







OPEN
CALAIS



THOMSON REUTERS

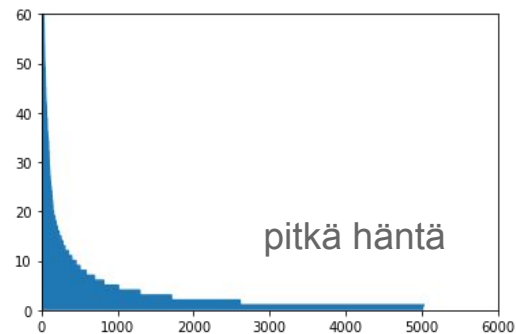




~~YSA YSO
Allärs KOKO~~

€ £ \$

Sisällönkuvailu on vaikeaa



ihmisille:

- **Subjektivisuus:** kun kaksi eri ihmistä kuvailee saman dokumentin, vain $\sim\frac{1}{3}$ aiheista on samoja
- **Paljon käsitteitä:** kymmeniä tuhansia mahdollisia aiheita joista valita
- **Sanasto elää:** uusia käsitteitä lisätään, vanhoja nimetään ja määritellään uudelleen

tekoälylle:

- **Pitkän hännän jakauma:** useimpien aiheiden käytöstä vain vähän esimerkkejä
- **Paljon käsitteitä:** vaatii monimutkaisia ja laskennallisesti raskaita malleja
- **Vaikea arvioida:** hyvien ehdotusten erottelu huonoista vaatii ihmistyötä
- **Sanasto elää:** malleja täytyy kouluttaa uudelleen



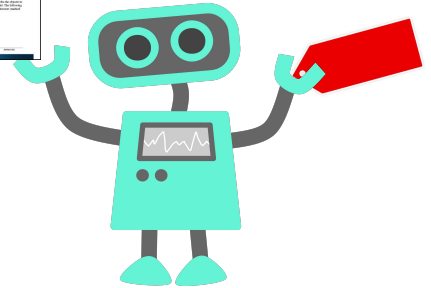
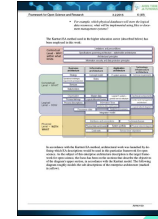
Tuottaa sisällönkuvailua (asiasanoitus ja luokitus) tekstin perusteella

Kansalliskirjastossa kehitetty työkalu (prototyyppi 2017), avointa koodia

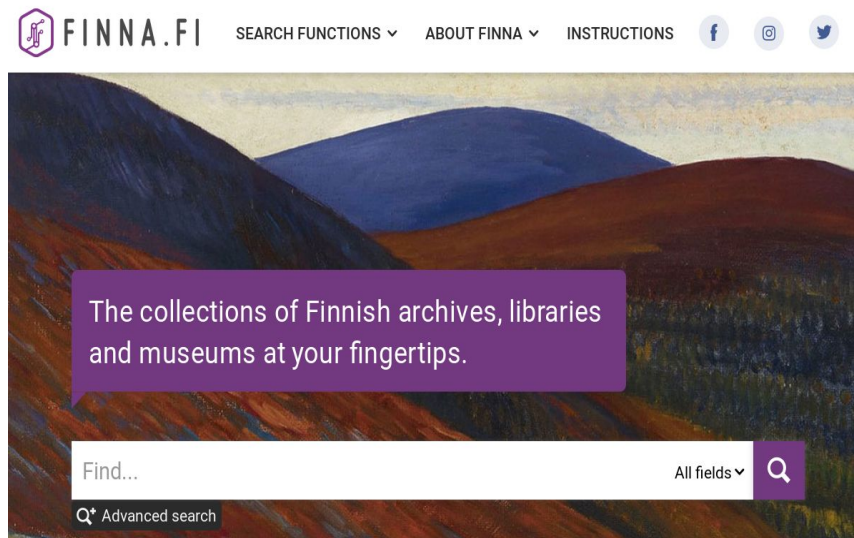
Soveltuu parhaiten asiateksteille, mutta testataan myös muilla aineistoilla

Perustuu **kieliteknologiaan** ja **koneoppimiseen**

Voidaan käyttää **rajapinnan kautta** tai **itsenäisenä työkaluna**



Annif oppii aineistosta, miten tunnistaa aiheita



annif

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



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> Edit

[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
 .lgm.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

Kattavat yksikkötestit (99%)

PEP8 tyylisäännöt

[Tekninen dokumentaatio](#) wikissä

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

Annifin laatu ihmisten arvioitavana

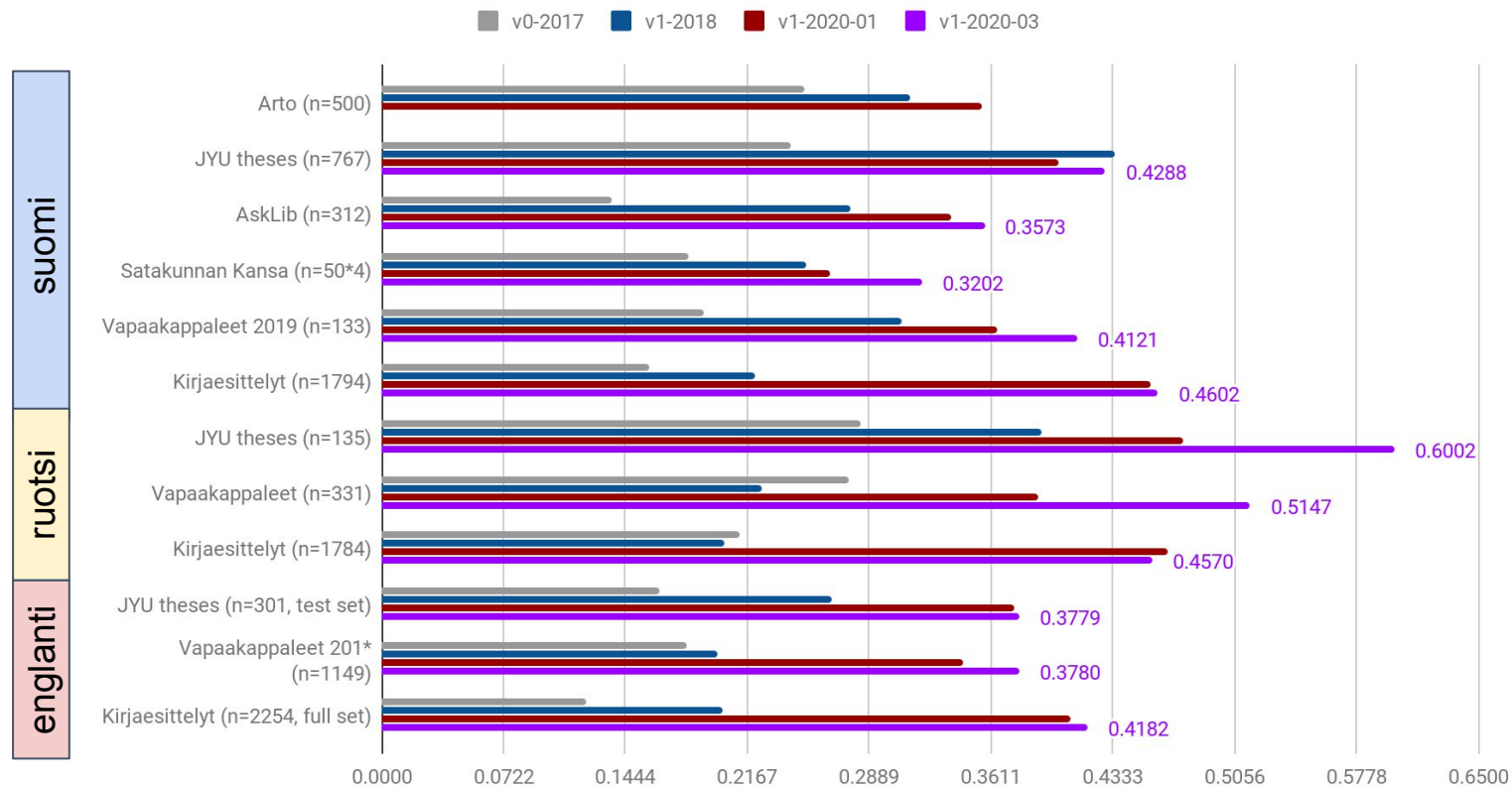


Kuva: Mikko Lappalainen.

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>

Mallipäivitys maaliskuuta 2020: YSO Diotima (2020.1), uusi Finna-koulutusdata, neuroverkko-ensemble -mallit

F1@5 laatuarviot eri aineistoille ja Annif-rajapintaversioille



Automaattisen kuvailun kehittäminen Kansalliskirjastolla 2019-2020

1. Kehitetään Annif-työkalua erityisesti sisällönkuvailun laadun näkökulmasta.
2. **Julkaistaan avoin rajapintapalvelu automaattiseen sisällönkuvailuun Finto-palvelun yhteyteen (Finto AI) ja tuetaan sen käyttöönottoa eri organisaatioissa.**
3. Edistetään Annif-avusteisen kuvailun käyttöä Kansalliskirjaston omissa järjestelmissä (julkaisuarkistot, e-vapaakappaleet...).
4. Tutkitaan automaattisen bibliografisen kuvailun mahdollisuuksia kuvailuprosessien kehittämisessä.
5. Viestintä ja yhteistyö.



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



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

Finto AI julkaistiin 28.5.2020

Automaattinen sisällönkuvailu osaksi Finton palveluja

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

Jaa



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 opinnäytetö- ja tietokirja-

<https://www.sttinfo.fi/tiedote/automaattinen-sisallonnkuvailu-osaksi-finton-palveluja?publisherId=3747&releaseId=69882342>

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

HPD-hanke

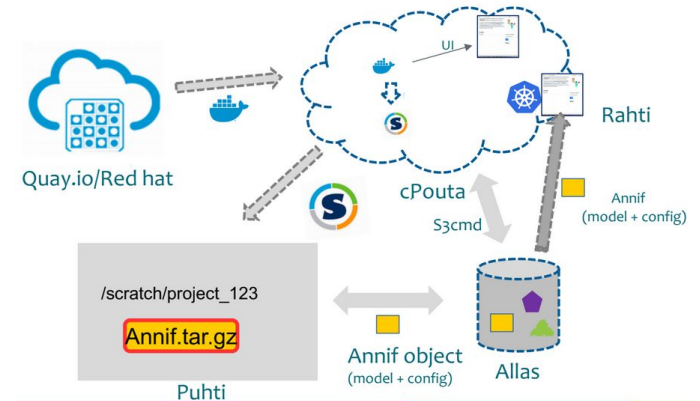
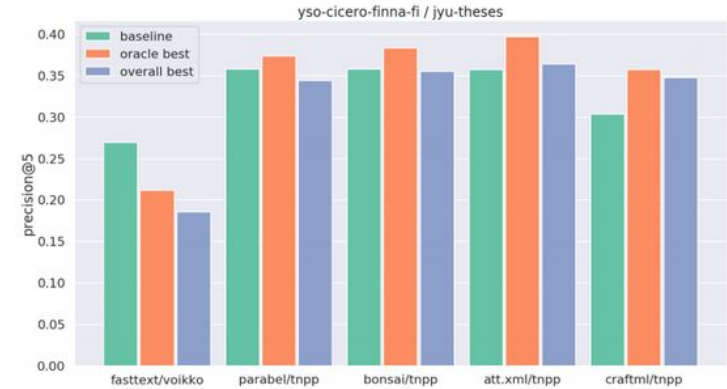
Tekoäly rikastaa muistiorganisaatioiden digitaalisia aineistoja

CSC:n, Kansalliskirjaston ja Kansallisarkiston *High-Performance Digitisation* -hankkeessa luodaan älykäs annotaatioputki arkistoitujen aineistojen käsittelyyn ja rikastukseen. Aineistot voivat olla esimerkiksi skannattuja sanomalehtiä, kirjoja, kuvia ja virallisia asiakirjoja. Tekoälyn ja koneoppimisen avulla data laatua voidaan parantaa ja dataa voidaan hyödyntää tehokkaammin.

Ajalle 9/2019 - 9/2020, mutta myönnetty jatkoaikaa vuoden 2020 loppuun

HPD-hanke käytännössä

- Kansalliskirjasto toimittaa koeaineistoja CSC:n asiantuntijoille
- CSC testaa eri automaattisen luokittelun ja sisällönkuvailun algoritmeja suurteholaskentaympäristössä
 - Lupaavimmat algoritmit (mm. Omikujii) integroidaan Annifiin
- Lisäksi suunnitellaan automaattisen sisällönkuvailun teknistä infrastruktuuria sekä integraatioita muihin järjestelmiin



Annif / Finto AI käytössä

Jyväskylän yliopiston JYX-julkaisuarkisto

Opiskelijat lataavat omat gradunsa, Annif / Finto AI ehdottaa niille aiheita

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

Sama ajatus kuin JYXissä: Finto AI ehdottaa aiheita uusille dokumenteille

Pilottikäytössä
2.3.2020 alkaen,
toteutus
Anis Moubarik

Asiasanat:

Annif-ehdotukset

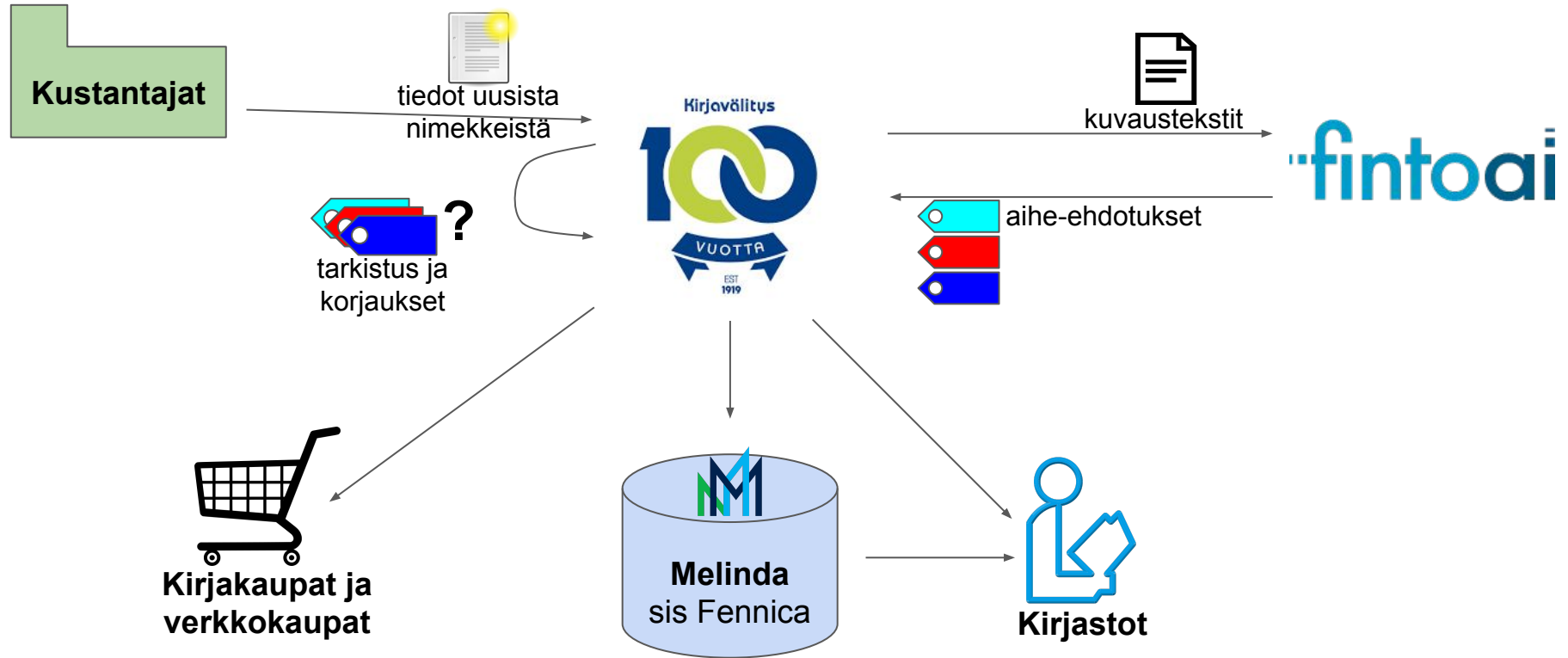
- | | |
|--|--|
| <input type="checkbox"/> working abroad | <input type="checkbox"/> families (groups) |
| <input type="checkbox"/> career development | <input type="checkbox"/> managers and executives |
| <input type="checkbox"/> career | <input type="checkbox"/> human resources |
| <input type="checkbox"/> 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.

Kirjavälitys Oy käyttää Finto AI:tä



Käyttöönotto Kansalliskirjaston järjestelmissä

The screenshot shows two overlapping web interfaces. The top interface is the LUT University Doria system, featuring a search bar with the text 'LUTPubin haku- ja käyttöohje' and a 'HAE' button. Below it, the Julkari system is visible, with a search bar and a 'Haku' button. The Julkari interface includes a description: 'Julkari on Sosiaali- ja terveysministeriön julkaisuista on saatavilla sähköinen versi laitoksen osalta myös julkaisurekisterinä. Julkarin julkaisuja voi hakea mm. julkaist. järjestetty kokoelmittain, joten sanahaun Julkari on vapaasti käytettävissä. Julkais...

The screenshot shows the Kansalliskirjasto digital publishing system interface. The header includes the logo and the text 'Digitaalisten julkaisujen arkistointi'. Below the header, there are several sections: 'Luovuttajan tiedot', 'Julkaisun tiedot', and 'Julkaisun tekemäärä'. The 'Luovuttajan tiedot' section includes fields for 'Yhteystiedot', 'Sähköpostiosoite', 'Puhelinnumero', and 'Organisaatio'. The 'Julkaisun tiedot' section includes fields for 'Julkaisun tyyppi' and 'Päättämätiedot'. The 'Julkaisun tekemäärä' section includes a field for 'Luovuttajan tiedot'.

E-vapaakappaleiden vastaanotto ja käsittely

DSpace-pohjaiset julkaisuarkistot: Osuva, Theseus ...

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

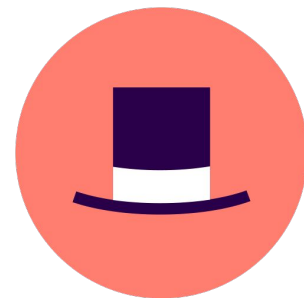


KIRJASTOT.fi



Suomi.fi

Annif-verkkokurssi syksyllä 2020



21.9. [Introduction to Annif and automated subject indexing](#)

Esitys DCMI Virtual -konferenssissä

Osma Suominen, Koraljka Golub, Annemieke Romein & Sara Veldhoen

[Itsenäisesti suoritettava verkkokurssi](#) sis. videoita ja harjoitustehtäviä

9.10. ja 21.10. verkkokurssin vuorovaikutteiset osuudet tukea tarvitseville

Kiitos!

Juho Inkinen, Mona Lehtinen, Osma Suominen

