



*"Notta aivan ei ehkä kannata ajatuksetta ottaa..."*

Kirjastoverkkopäivät 2019

# Annif oikeissa töissä

Miten ANNIFia käytetään JYU:n Avoimen tiedon keskuksessa

Ari Häyrinen



*"Teknisesti toimii, mutta olisiko kuitenkin vähän liian, hmmm.... "luova"? Vapaamielinen/laaja-alainen...?"*

*...*

*Mitä johtamisaiheiseen 'johtaminen ja sosiaalinen oikeudenmukaisuus' artikkeliin tulee, niin sille olisi tarjolla ortodoksista kirkkomusiikkia (orthodox church music), salaojitusta (subsurface drainage) ja kestopuovia (thermoplastic)"*

## TAUSTAA: JYX julkaisuarkiston uudistus

- kaikki työnkulut ulkoisia sovelluksia
- mahdollistaa millaiset työnkulut tahansa
  - Annifin lisäys helppoa



*"Laitoin myös oman artikkelini testatakseni asiasanoja, muutoin osuvat (englanninkielinen artikkeli) mutta joukossa kolme kummaa, kuten esim. 'ryijy', Mutta pääsääntöisesti asiasanat liittyivät aiheeseen ja niistä koko joukko sellaisia, joita olisin itsekin ko. artikkeliin asiasanoittanut"*

## Miksi Annif otettiin käyttöön?

- Asiasanoitus on vaikeaa ja aikaa vievää
- Annifin käyttöönotto oli tehty helpoksi!
  - "vakio-Annif" on yhden rajapintakutsun päässä



*"XML-tiedonhakua käsittelevän graduni  
tapauksessa kone suoriutui melkoisen hyvin:*

*kyselykielet  
tiedonhakujärjestelmät  
kuvauskielet  
XML"*

## Annif ja JYXin graduputki

- Opiskelija julkaisee gradunsa verkossa JYXissä
- Lomakkeella näytetään Annifin ehdotukset
- Informaatikko tarkastaa asiasanoituksen ennen julkaisua



Pääsivu > Lomakkeet > Opinnäytteen tallennus

## Opinnäytteen JYX-tallennus

Lähetä opinnäytteesi julkaistavaksi **JYX-julkaisuarkistoon** yhtenä PDF-tiedostona tällä lomakkeella. Liiteaineiston voit toimittaa meille tarvittaessa erikseen.

1. Lataa PDF-tiedosto

Valitse tiedosto

Browse... No file selected.

jatka

Ohjeita PDF-tiedoston tekoon: [PDF-tiedoston luontiohje](#)

Jos sinulla on opinnäytteeseesi liittyen muutakin materiaalia (esim. tutkimusdataa) tai törmäät ongelmiin esimerkiksi pdf-tiedoston luomisessa, ota yhteyttä meihin! [tutkielmat@library.jyu.fi](mailto:tutkielmat@library.jyu.fi) tai puh. 040 805 3785

Teknisiä ongelmia lomakkeen kanssa? Laita viestiä: [jyx-support@jyu.fi](mailto:jyx-support@jyu.fi)





**abstrakti suomeksi**  
suomenkielinen tai englanninkielinen  
vaaditaan

**abstrakti englanniksi**  
suomenkielinen tai englanninkielinen  
vaaditaan

## asiasanoitus

**asiasanaehdotukset**  
Valitse työhösi sopivat klikkaamalla

- ☐ documentation [YSO]
- ☐ users [YSO]
- ☐ query languages [YSO]
- ☐ user interfaces [YSO]
- ☐ cultural research [YSO]
- ☐ study of cultures [YSO]
- ☐ culture [YSO]
- ☐ connections (technical systems) [YSO]
- ☐ demands [YSO]
- ☐ alteration work [YSO]

**omat asiasanasi**



OSC voi purkaa sopimuksen välittömästi, mikäli osoittautuu, että tekijät ovat antaneet pykälän 4 osalta harhaanjohtavaa tai virheellistä tietoa. Sopimuksen purkaminen ei vapauta tekijöitä vastuustaan ja mahdollisista vahingonkorvausvelvollisuuksistaan. Tekijät voivat irtisanoa sopimuksen kirjallisesti päättymään kuukauden kuluttua irtisanomisesta. Sopimuksen irtisanomisen jälkeen OSC:lla ei ole oikeutta pitää julkaisua avoimessa tietoverkossa yleisesti saatavilla, mutta OSC saa pitää hallussaan teoksen elektronisen arkisto-kappaleen. Arkistokappaleet ovat käytettävissä muun elektronisen vapaakappaleaineiston tavoin vapaakappalekirjastoissa olevilta, tämän aineiston käyttöön varatuilta laitteilta. Mikäli sopimuksen purkaminen johtuu tekijöiden sopimusrikkomuksesta, ovat tekijät velvollisia korvaamaan OSC:lle purkamisesta aiheutuvat kustannukset ja välittömät vahingot. Kummallakin osapuolella on oikeus purkaa sopimus, jos sen ehtojen täyttäminen ylivoimaisen esteen jatkumisen vuoksi tulee mahdottomaksi. Ylivoimaiseksi esteeksi katsotaan sellainen sopimuksen täyttämisen estävä ja sopimuksen syntymisen jälkeen sattunut epätavallinen ja asiaan vaikuttava tapahtuma, jota osapuolten ei ole ollut syytä ottaa huomioon sopimusta tehtäessä, joka on osapuolista riippumaton ja jonka estävää vaikutusta ei voida poistaa ilman kohtuuttomaksi katsottavia lisäkustannuksia tai kohtuuttomaksi katsottavaa ajanhukkaa.

## 9. RIITOJEN RATKAISU

Tämän sopimuksen tulkinnasta aiheutuvat erimielisyydet ratkaistaan ensisijaisesti neuvotteluin. Mikäli näin ei päästä ratkaisuun, riita saatetaan Jyväskylän käräjäoikeuden ratkaistavaksi.

Hyväksyn sopimuksen ja sitoudun noudattamaan sen ehtoja. Luovutan laatimani materiaalin ylläpidettäväksi ja jaeltavaksi yliopiston aineistopalvelimelle. Vakuutan, ettei materiaalin käyttö sopimuksen mukaisesti edellytä minkään kolmannen osapuolen lupaa.

2019-10-21

- Häyrinen, Ari

### julkaisulupa

- ☒ Hyväksyn julkaisusopimuksen ja opinnäytteen saa julkaista avoimesti JYX-julkaisuarkistossa.
- ☐ Kiellän opinnäytteeni avoimen julkaisun JYX-julkaisuarkistossa.

4. Lähetä JYXiin

[palaa takaisin](#) [Lähetä työ julkaistavaksi](#)



JYX-admin > opinnäytejulkaisut

## Lomakejulkaisut (gradut)

Tallentamatta: 5, talletettu: 1995

pvm	otsikko	lähettäjä	tyyppi	lupa avoimeen julkaisuun	tila	JYX	poista
2019-10-21	ARIN TESTI - ignore me!	ari.hayrinen@ju.fi	gradu	Häyrinen, Ari <b>OK</b>	ota käsittelyyn	ei vielä JYXissä	poista
2019-10-11	Arin testi - ignore me!	ari.hayrinen@ju.fi	gradu	Häyrinen, Ari <b>Informaatio vapautettu</b>	ota käsittelyyn	ei vielä JYXissä	poista
2019-10-11	Arin testi - ignore me!	ari.hayrinen@ju.fi	gradu	Häyrinen, Ari <b>Informaatio vapautettu</b>	ota käsittelyyn	ei vielä JYXissä	poista
2019-10-11	Arin testi - ignore me!	ari.hayrinen@ju.fi	gradu	Häyrinen, Ari <b>Informaatio vapautettu</b>	ota käsittelyyn	ei vielä JYXissä	poista





## asiasanoitus

### vapaat asiasanat

Pilkulla erotettuna.

### Kaikki YSO-asiasanat

Vihreällä merkityt ovat opiskelijan  
valitsemia.

### YSO-haku

haku enterillä, esc nollaa

- ☒ dokumentointi : documentation [YSO]
- ☐ käyttäjät : users [YSO]
- ☒ kyselykielet : query languages [YSO]
- ☒ käyttöliittymät : user interfaces [YSO]
- ☐ kulttuurintutkimus : cultural research [YSO]
- ☐ kulttuurien tutkimus : study of cultures [YSO]
- ☒ kulttuuri : culture [YSO]
- ☐ liittymät : connections (technical systems) [YSO]
- ☐ vaatimukset : demands [YSO]
- ☐ muutostyöt : alteration work [YSO]



*"Osaako kone asiasanoittaa?"*

*Nyt voit itse kokeilla ja samalla voit auttaa uuden JYXin opinnäytelomakkeen testaamisessa. Lomake käyttää Kansalliskirjaston Annif -nimistä asiasanoitussysteemiä, joka perustuu Finnan tietoihin ja se tarjoaa lomakkeen käyttäjälle mahdollisuuden valita haluamansa asisanat Annifin ehdotuksista."*

## Sisäinen testaus

- Automaattinen asiasanoitus uusi asia (meille)
- Annifiin tutustuminen aloitettiin sisäisellä testauksella
- Yläosan lainaukset ovat sisäiseltä keskustelufoorumilta



*"Oman kirjallisuudentutkimuksen graduni kanssa asiasanoitus tosin ei onnistunut sinne päinkään. Erikoista miten annif kaiveli useita sanoja erilaisista piharakennuksista, vaikka gradussani ei niitä juurikaan mainita. Romaanikin tainnut muuttua asisanoissa romaniksi..."*

*arot  
huvimajat  
pyhäkoulu  
romanikulttuuri  
leikkimökit  
piharakennukset"*



## Sisäinen testaus

- Annifin proto -versio antoi paljon huteja



*”\*\*\*\* onnistui hämmentämään edellisen ANNIFin gradullaan, joka ANNIFin mielestä käsitteli leikkimökkejä, aroja ja pyhäkouluja. Näistä asioistahan Strimberg Strindberg jatkuvasti kirjoitti.*



*However, ANNIF 2 on toista mieltä:*

- identiteetti*
- postmoderni*
- subjektiviteetti*
- romaani*
- subjekti (lauseenjäsen)*
- subjekti (filosofia)*
- tunteet*
- ihmiskuva*
- taiteilijuus*
- dialogisuus”*

## Sisäinen testaus

- Seuraava versio paransi **tuloksia huomattavasti!**

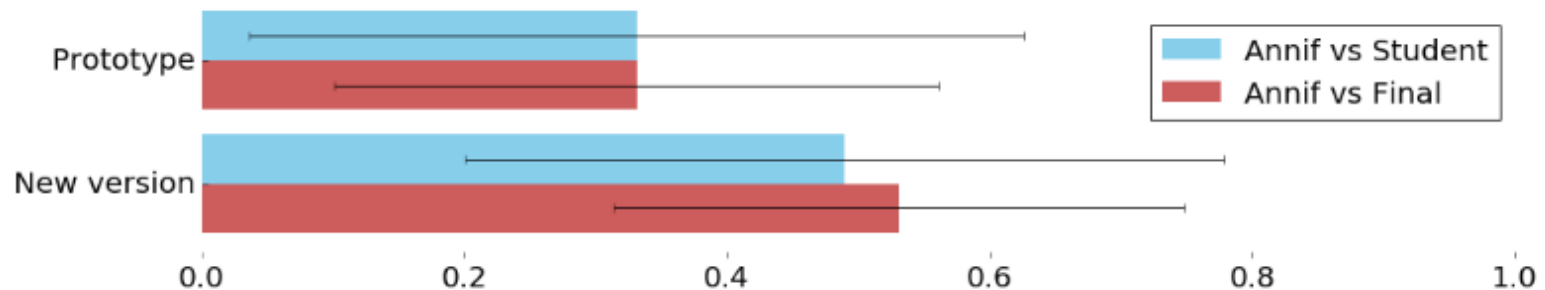


# Annif: DIY automated subject indexing using multiple algorithms

Osma Suominen 2019

<https://www.doria.fi/handle/10024/169004>

**Figure 4.** F1 similarity between Annif suggestions, student-selected subjects and final subjects in JYX, for the Annif prototype and new version



The similarity scores for the new version are analyzed broken down by university department in Figure 5. Due to the relatively small number of documents and the high variation in F1 scores, we cannot draw any firm conclusions, but it appears that the best results are obtained in the humanities, while results are not as good in mathematics, science and technology. This pattern may be due to differences in granularity of the subject vocabulary in different topical areas, as well as the different nature of concepts in different fields: in the humanities, concepts may often be broader and fuzzier, whereas in more technical fields they can be more specific and strictly bounded.

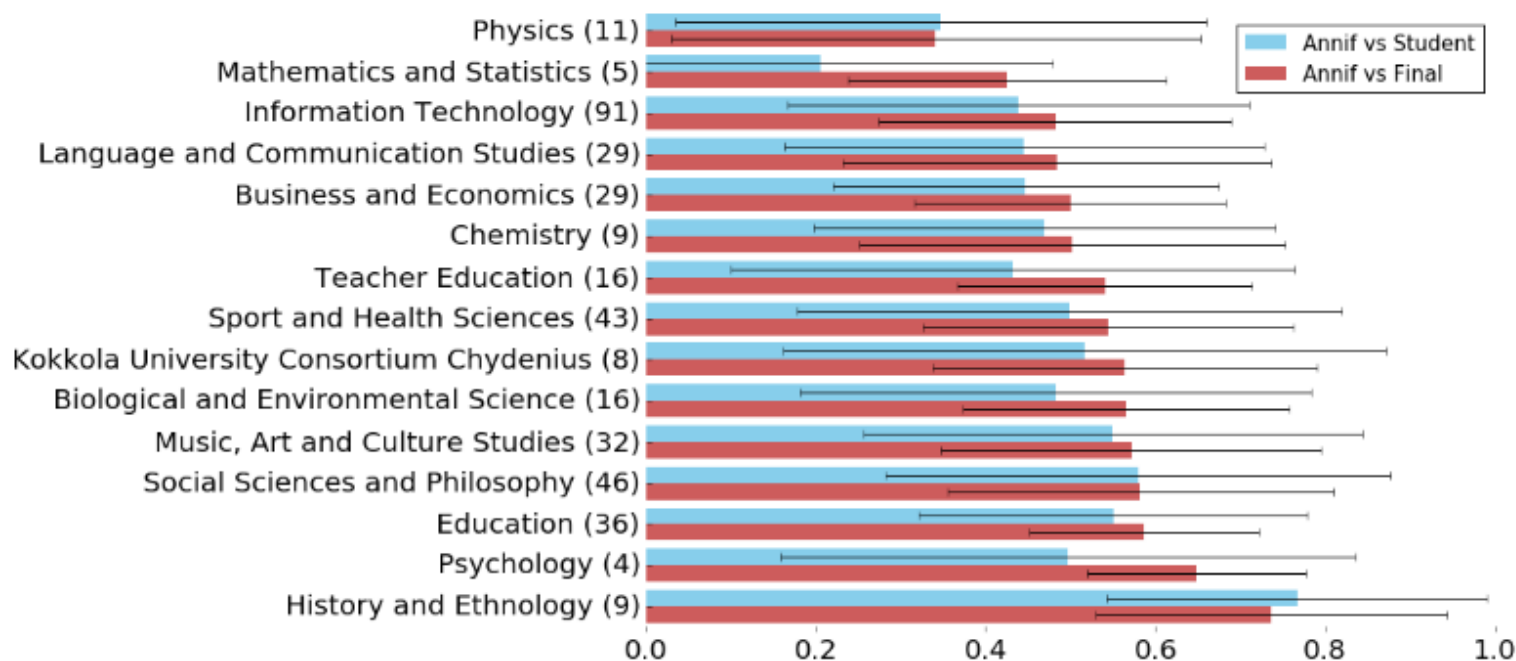


# Annif: DIY automated subject indexing using multiple algorithms

Osma Suominen 2019

<https://www.doria.fi/handle/10024/169004>

**Figure 5.** F1 similarity between Annif (new version) suggestions, student-selected subjects and final subjects in JYX, by department





*"ANNIFin uudempi versio on ollut gradujen asiasanoituksen apuna marraskuusta lähtien. ANNIFin kehittäjä Osma Suominen kyseli datan perään eli onko Annif kehittynyt vai ei.*



*Graduja on tullut uuden Annifin läpi vähän alle neljäsataa. ja nyt voi todeta, että onhan se kehittynyt.*

*Noin 85 prosenttia opiskelijoista kelpuutti ainakin yhden Annifin ehdotuksen. Aiemmin tämä oli luokkaa 65 -70 prosenttia.*

*Yli kolmasosa opiskelijoista valitsi yli puolet Annifin tarjoamista termeistä. Tätä voi pitää jo aika erinomaisena tuloksena.*

*Kaikki Annifin tarjoamat termit kelpasivat 14 opiskelijalle."*

## Kannattiko?

- KYLLÄ!



*"Olen huomannut saman, Annif on kehittynyt uuden version myötä huomattavasti."*



## Mitä seuraavaksi?

- Uusien väitöskirjojen asiasanoitus
- Kokeiluja museoaineistolla
  - uusi näkökulma aineistoon





*"No niin. ANNIF 2.0 ymmärtää jo mestariteoksia."*



# Kiitos

ari.hayrinen@jyu.fi



# JYX-LOMAKKEET

