

Annif ja automaattisen kuvailun edistäminen

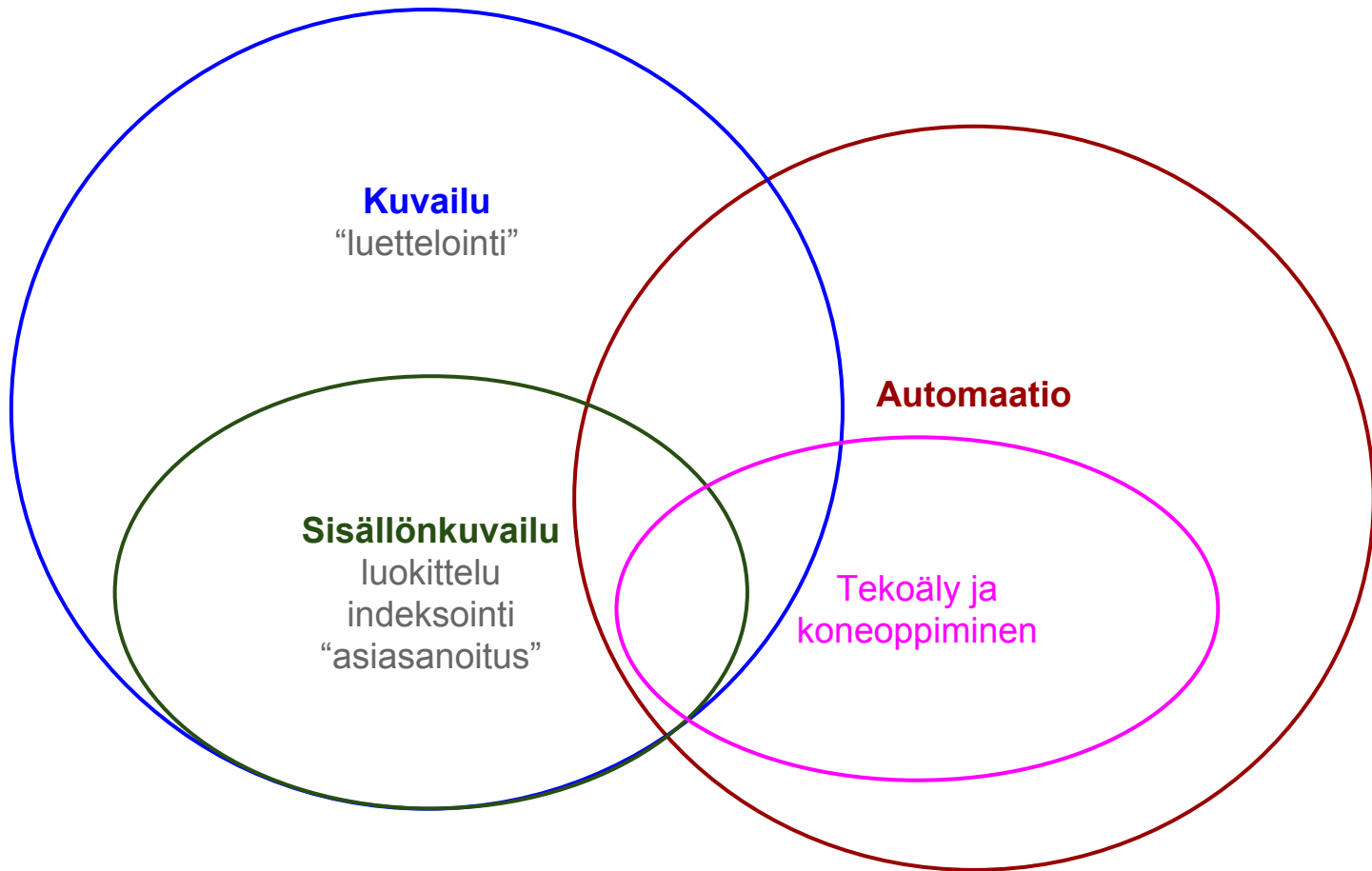
Osma Suominen

Finto-ohjausryhmä 27.2.2019

Kuvailu
"luettelointi"

The diagram consists of two overlapping circles. The larger, outer circle is blue and contains the text 'Kuvailu' and '"luettelointi"'. The smaller, inner circle is green and contains the text 'Sisällönkuvailu', 'luokittelu', 'indeksointi', and '"asiasanoitus"'. The green circle is positioned such that it is entirely contained within the blue circle, illustrating that 'Sisällönkuvailu' is a subset of 'Kuvailu'.

Sisällönkuvailu
luokittelu
indeksointi
"asiasanoitus"



Kuvailu

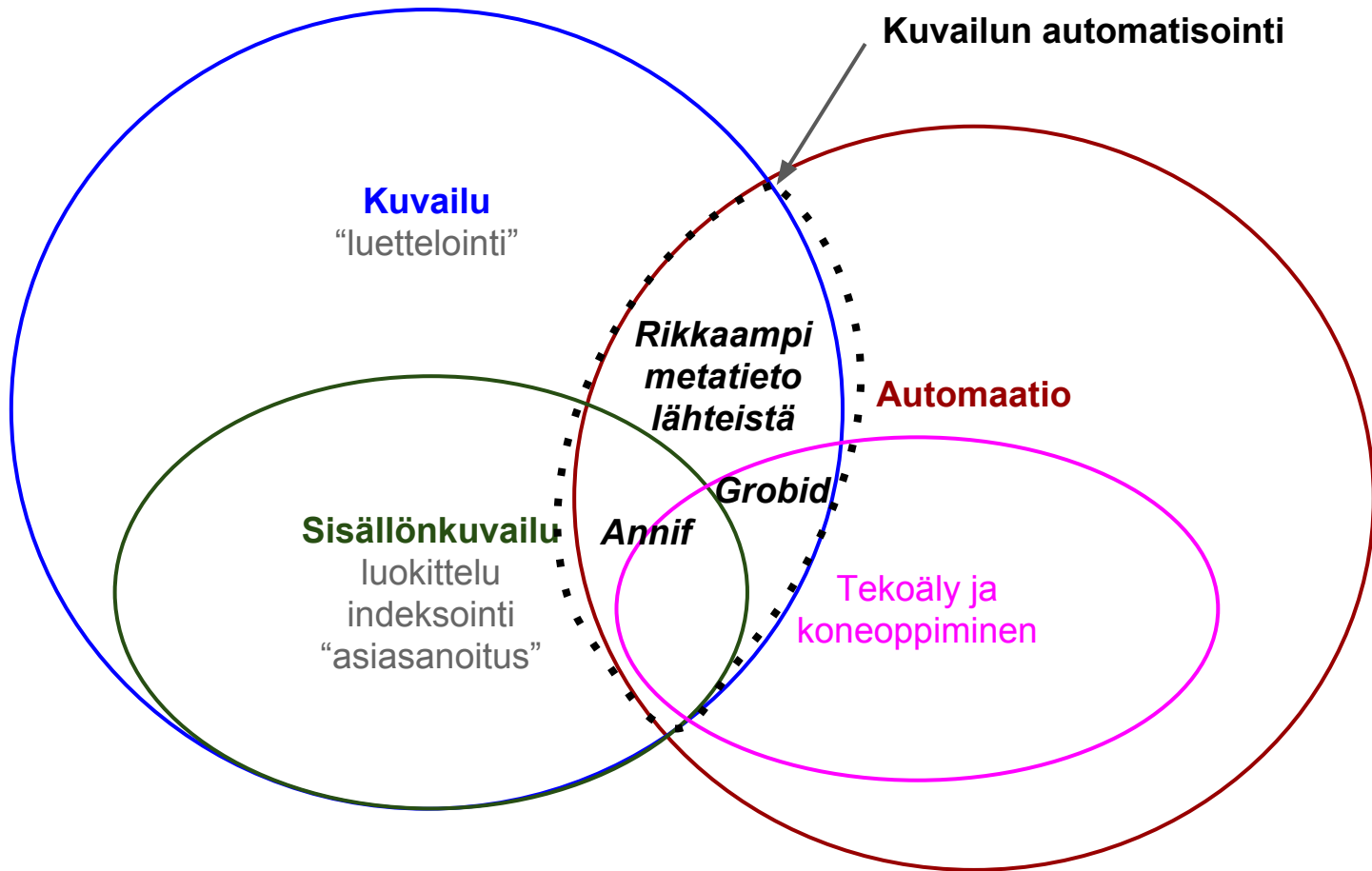
“luettelointi”

Sisällönkuvailu

luokittelu
indeksointi
“asiasanoitus”

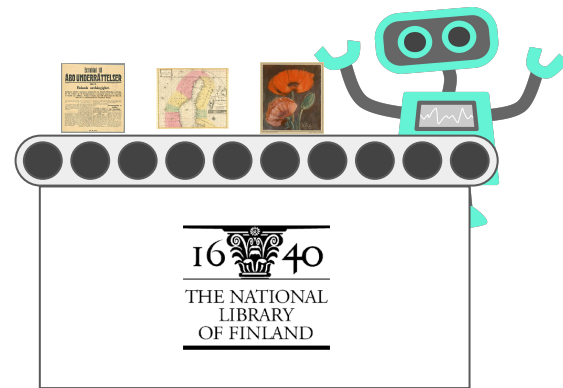
Automaatio

Tekoäly ja
koneoppiminen



Tavoitteet

1. Kuvailuprosessien tehostaminen:
asiantuntijatyö siihen mihin sitä erityisesti tarvitaan
2. Parempien palveluiden tarjoaminen yhteistökumppaneille
3. Kansainvälinen yhteistyö, näkyvyys ja maine

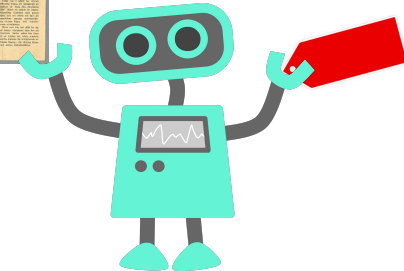


Työkalut

Grobid

Depth-resolved analysis of spontaneous phase separation in the growth of lattice-matched AlInN	title
A. Redondo-Cubero ^{1,2*} , K. Lorenz ³ , R. Gago ⁴ , N. Franco ¹ , M.-A. di Forte Poisson ⁵ , E. Alves ⁶ and E. Muñoz ¹	authors
¹ ISOM and Dpt. de Ingeniería Electrónica, ETSI Telecomunicación, Universidad Politécnica de Madrid, E-28040 Madrid, Spain.	affiliation
² Centro de Micro-Análisis de Materiales, Universidad Autónoma de Madrid, E-28049 Madrid, Spain.	
³ Instituto Tecnológico e Nuclear, Estrada nacional 10, 2686-953 Sacavém, Portugal.	
⁴ Instituto de Ciencia de Materiales de Madrid (ICMM), E-28049 Madrid, Spain.	
⁵ Thales Research & Technology/TIGER, 91461 Marcoussis Cedex, France.	
ABSTRACT:	abstract
We report the detection of phase separation of an Al _{1-x} In _x N/GaN heterojunction grown close to lattice matched conditions (x=0.18) by means of Rutherford backscattering spectrometry in channeling geometry and high resolution x-ray diffraction. An initial pseudomorphic growth of the film was found, with good single crystalline quality, the	

Annif



Tuottaa perusmetatietoja (nimeke, tekijät, julkaisuajankohta ...) PDF-tiedostoista

Parhaillaan testattavana Kansalliskirjastossa

Tarkoitettu tieteellisille artikkeleille, mutta testataan myös muuntyyppisillä julkaisuilla

Perustuu koneoppimismalleille, joita voitaisiin kouluttaa uudelleen suomalaisella aineistolla

Tuottaa sisällönkuvailua (asiasanoitus ja luokitus) tekstin perusteella

Kansalliskirjastossa kehitetty

Soveltuu parhaiten asiatekstille, mutta testataan myös muilla aineistoilla

Perustuu kieliteknologiaan ja koneoppimiseen

Sanastot, ontologiat, luokitukset...

YSO	Toimivat mallit olemassa ja jo käytössä mm. Jyväskylän yliopiston kirjastossa. Kehitys jatkuu.
YSO-paikat	Ei vielä aloitettu. Tarkoitus testata mm. Satakunnan Kansan artikkeleilla.
YKL	YKL julkaistu Fintossa 2019 alussa. Alustavia automaattisen YKL-luokituksen kokeiluja tehty mm. Makupalat-aineistolla ja Kirjavälityksen tuotetiedoilla.
UDK	Ei vielä aloitettu. Paljon UDK:ta käyttävää metatietoa saatavana mm. Fennicassa ja Melindassa.
Kaunokki	Kaunokki julkaistu Fintossa 2019 alussa. Automaatiokokeiluja ei vielä aloitettu. Tutkittava, missä määrin kaunokirjallisuuden sisällönkuvailua ylipäänsä voidaan automatisoida.

Aineistotyypit

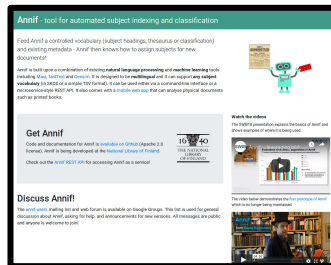
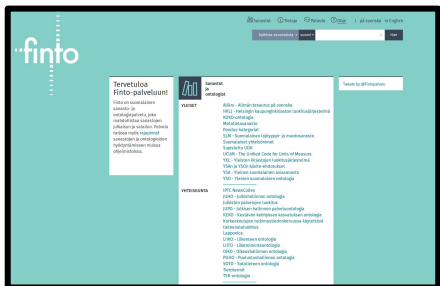
Tähän mennessä parhaita tuloksia saatu graduilla, väitöskirjoilla ja e-artikkeleilla. Yleisten kirjastojen kiinnostus kohdistuu myös muuntyyppisiin aineistoihin:

- e-kirjat
- kustantajien / Kirjavälityksen tuotekuvaukset
- lehtiartikkelit (mm. Satakunnan Kansan arkistot)
- kysymys-vastausparit (esim. Kysy kirjastonhoitajalta -palsta)
- linkkikokoelmat (mm. Makupalat)

Kaikkia em. aineistotyyppisiä on **jo saatu käyttöön** yleisiltä kirjastoilta ja Kirjavälitykseltä ja niillä **on jo tehty** automaattisen sisällönkuvailun kokeiluja.

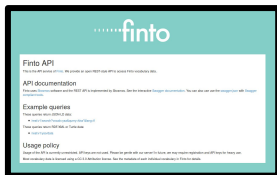
Automaattisen sisällönkuvailun palvelut

Finton palveluvalikoimaa laajennetaan automaattiseen sisällönkuvailuun
Dokumentaatio, käyttönoton tuki, palaute jne. Finton nykyisten kanavien kautta

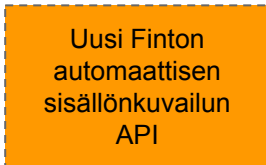


annif.org

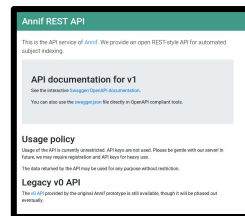
Annif-ohjelmiston käyttökorttisivusto



api.finto.fi/rest/
rajapinta sanastodataan
tuotantokäyttöä varten



api.finto.fi/xxx/
rajapinta autom. kuvailuun
tuotantokäyttöä varten

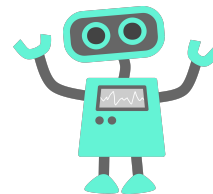


api.annif.org
demorajapinta automaattiseen
sisällönkuvailuun

Annif API esimerkki

“The quick brown fox jumped over the lazy dog.”

Analyze this!



Annif API

results=[

```
{uri="<http://www.yso.fi/onto/yso/p2228>", score=0.2595, label="red fox"},  
{uri="<http://www.yso.fi/onto/yso/p5319>", score=0.2039, label="dog"},  
{uri="<http://www.yso.fi/onto/yso/p8122>", score=0.1946, label="laziness"},  
{uri="<http://www.yso.fi/onto/yso/p25726>", score=0.1285, label="brown"},  
{uri="<http://www.yso.fi/onto/yso/p4760>", score=0.1220, label="triple jump"}
```

]

Automaattisen kuvailun verkosto Suomessa

- automaattisen kuvailun verkostossa mukana Yle, CSC, Kansallisarkisto, Xamk, JYU, Yleiset kirjastot / VAKE, VNK, Kirjavälitys ...
- Kirjavälitystä kiinnostaa Annifin hyödyntäminen metatiedon tuotannossa
 - Kehityshanke Kirjavälityksen kanssa 2019-2020: *“Metadatan tuotannon tehostaminen ja sen rikastaminen automaation avulla”*
- Yle ja Kavi haluavat asiasanoittaa tv-ohjelmia tekstityksen perusteella
- Museovirasto haluaa asiasanoittaa mm. rakennettujen ympäristön kuvauksia ja arkeologisia löytöjä
- Jyväskylän yliopisto käyttää jo Annifia JYX-julkaisuarkistossa (gradut ja vk:t)

Seuraavia askeleita

1. Rekrytointien käynnistäminen

- Tietoasiantuntija laatutyöhön ja verkostojen koordinointiin
- Tietojärjestelmäasiantuntija Annif- ja muuhun tekniseen kehitystyöhön

2. Finton palveluvalikoiman laajentaminen automaattiseen sisällönkuvailuun

- Finto tarjoaisi Annif API:n nykyisen Finto/Skosmos API:n rinnalla

3. Kansallisen yhteistyöverkoston aktivointi

- Automaattisen kuvailun verkoston seuraavan tapaamisen järjestäminen
- Työpajojen järjestäminen, mm. Kirjastoverkkopäivien yhteydessä

4. Kansainvälisen yhteistyön laajentaminen

- Annifille kehittymässä kansainvälinen käyttäjäyhteisö (mm. annif-users foorumi)
- DCMI Interest Groupin perustaminen automaattisen sisällönkuvailun ympärille

Kiitos!

Kysymyksiä?

osma.suominen@helsinki.fi - @OsmaSuominen

Tämä esitys: <https://tinyurl.com/kuvailun-automatisointi-finto>