

# Mikä on Annif ja miksi tarvitset sitä?

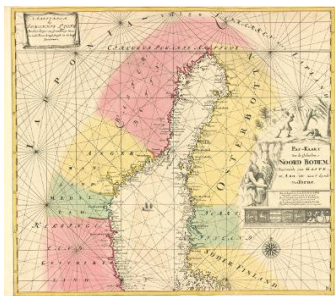
Missä mennään –webinaari  
20.9.2019

Mona Lehtinen (/Osma Suominen<sup>1</sup>)

**Mikä on Annif...**



Kirjastoissa hallinnoidaan suurta määrää meta-/kuvailutietoa eri dokumenteista



Framework for Open Services and Research - 2020 - 2021 - 2022 - 2023

For example, which physical architectures will allow the largest data resources, who will be responsible for setting up an advanced data resource system?

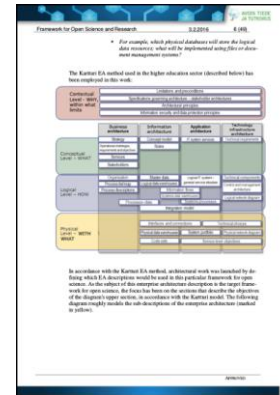
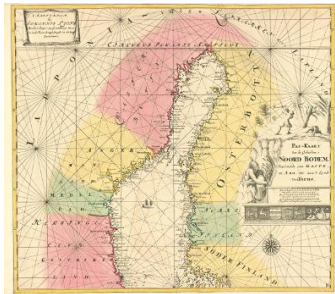
The Kartica EA method used in the higher education sector (described below) has been employed in this way.


Category	Architecture	Infrastructure	Software	Implementation
Open Services	...	...	...	...
Open Research	...	...	...	...
Open Data	...	...	...	...
Open Text	...	...	...	...
Open Video	...	...	...	...
Open Audio	...	...	...	...
Open Image	...	...	...	...

In accordance with the Kartica EA method, architectural work was conducted by the Kartica EA description would be used in the graphical framework for open services. An overview of the emerging architectural approaches in the open services work for open services. The focus has been on the architecture that describes the objectives of the digital open services, in accordance with the Kartica model. The following diagram roughly illustrates the main architecture described in the text.



Dokumentit kuvaillaan (asiasanoitetaan ja luokitellaan) kontrolloidun sanaston termeillä, jotta ne olisi helppo löytää

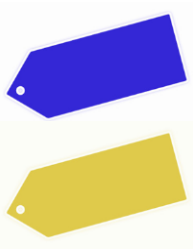



  
 Suomen Keskustiedustelu -palvelus

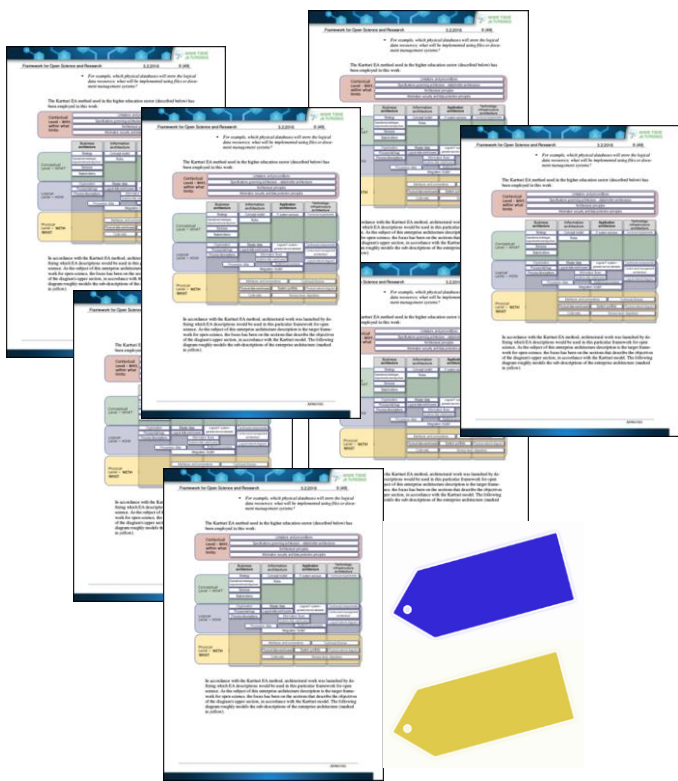
The Report is submitted to the Register when event identified within the time period of the week.

Event Name	Event Date	Event Location	Event Details	
			Event Type	Event Status
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...

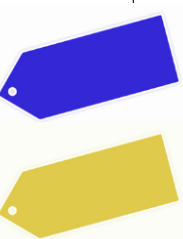
We understand that the Report is submitted and identified with the identified by the ...

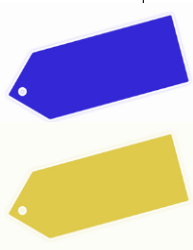
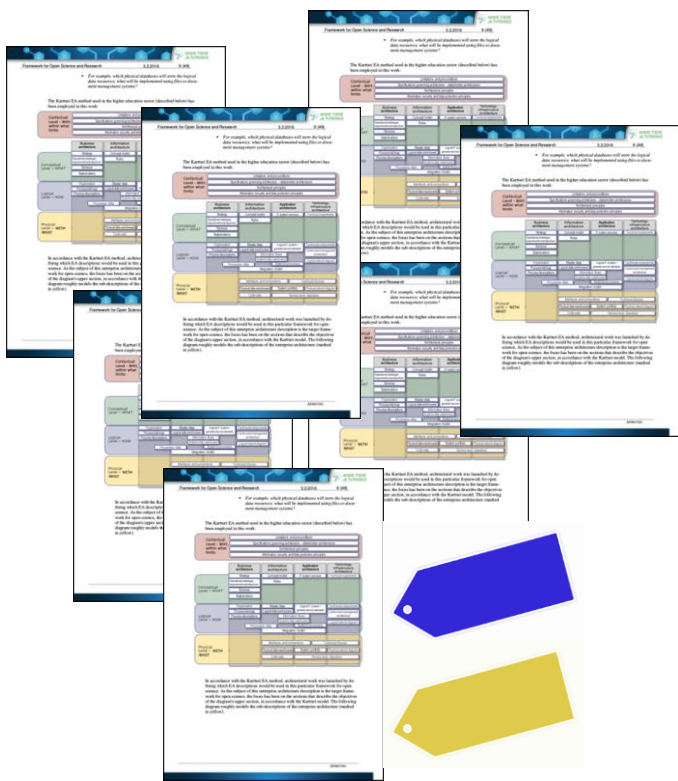


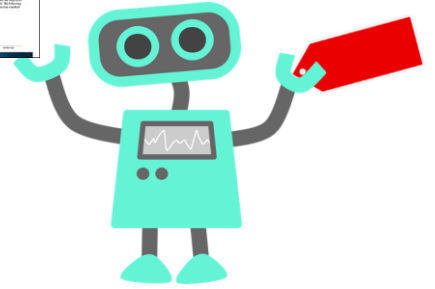
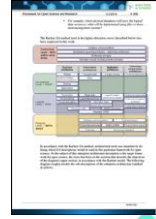
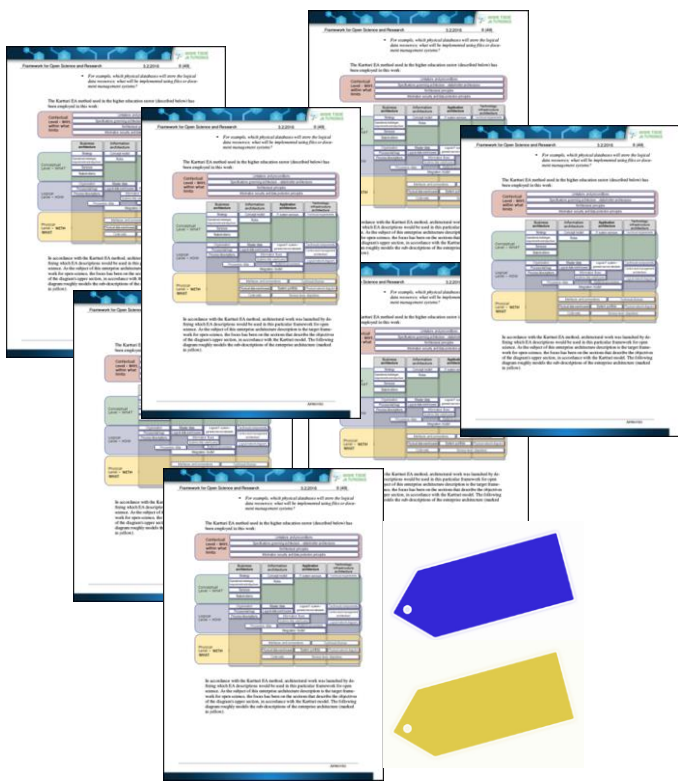

  
 Suomalainen  
 asiantuntija- ja  
 ontologiapalvelu



Entä kun aineiston määrä kasvaa?







**Annif**



# Annif

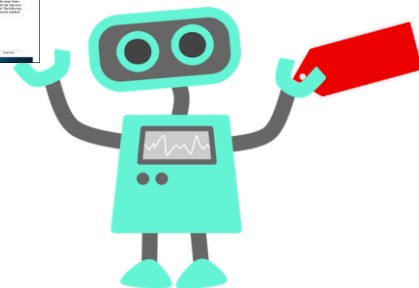
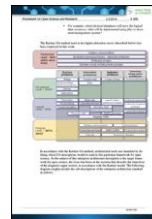
## Automaattinen asiasanoittaja

Tuottaa sisällönkuvailua (asiasanoitus ja luokitus) tekstin perusteella

Kansalliskirjastossa kehitetty työkalu, avointa koodia

Soveltuu parhaiten asiatekstille, mutta testataan myös muilla aineistoilla

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



Suomen arkistojen, kirjastojen ja museoiden aarteet samalla haulla.

Hae...

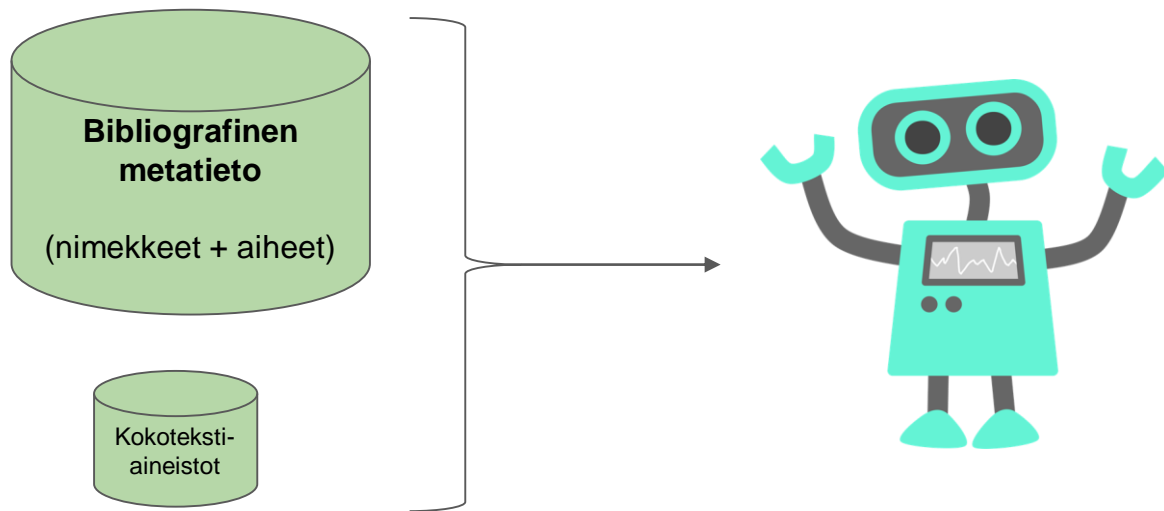
Kaikki osumat ▾



🔍 Tarkennettu haku

Paljon muistiorganisaatioiden metatietoa, mm. **15M tietuetta** [Finnassa](#)

# Annif oppii aineistosta, miten tunnistaa aiheita



# JYX-julkaisuarkisto, Jyväskylän yliopisto

Opiskelijat lataavat opinnäytteensä arkistoon, Annif ehdottaa sopivia 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

Toteutus  
DSpace &  
[GLAMpipe](#) -  
ohjelmistoilla,  
Ari Häyrynen

# Wikipedian kuvailu Annifilla

Suomenkielisessä Wikipediassa on 410 000 artikkelia (raakatekstinä 620 MB)

Automaattinen sisällönkuvailu kesti n. 6-7 tuntia

Aiheita annettiin 1-3 per artikkeli (keskimäärin n. 2)

**Esimerkkejä:** (satunnaisesti valittu)

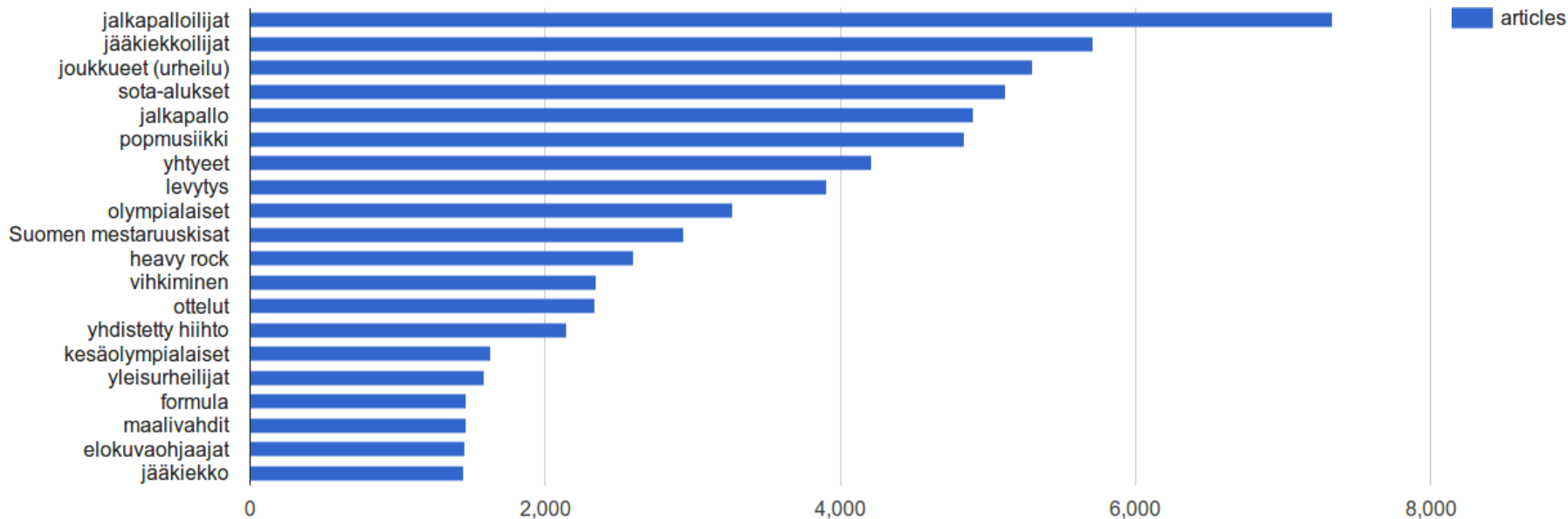
## Wikipedia-artikkeli

Ahvenuslammi (Urjala)  
Brazilian Grand Prix 2016  
Guy Topelius  
HMS Laforey  
Liigacup  
Pää Kii  
RT-21M Pioneer  
Runoja  
Sjur Røthe  
Veikko Lavi

## Ehdotetut aiheet yleisestä suomalaisesta ontologiasta ([YSO](#))

rannat  
kilpa-autoilijat, formula, mikroautoilu  
kansanrunoutentutkijat, sakariini  
sota-alukset  
jalkapallo, jalkapalloilijat  
yhtyeet, popmusiikki  
ohjukset  
popmusiikki, levytys, sävellykset  
hiihtäjät, hiihto, yhdistetty hiihto  
sanoittajat, kupletit

# Yleisimmät aiheet Wikipediassa



# Yleisimmät aiheet Wikipediassa



## Image credits:

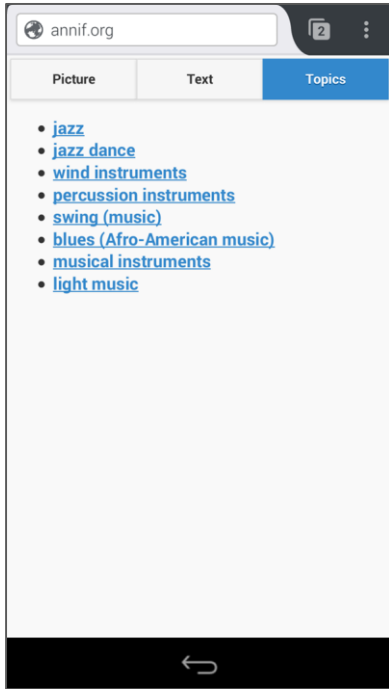
Petteri Lehtonen [CC BY-SA 3.0]

Hockeybroad/Cheryl Adams [CC BY-SA 3.0]

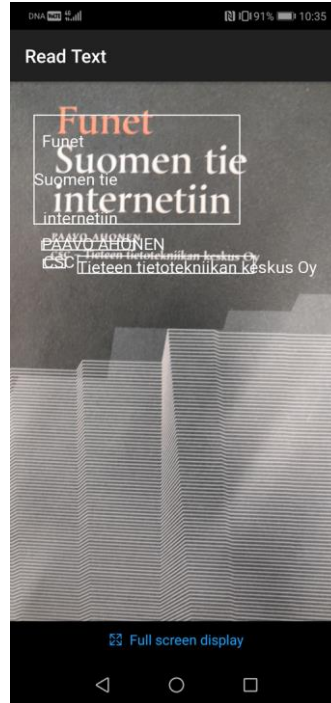
Tomisti [CC BY-SA 3.0]

Tuomas Vitikainen [CC BY-SA 3.0]

# Mobiilisovellukset



Prototyyppi web-sovellus  
OCR-pilvipalvelu ocr.space  
[m.annif.org](http://m.annif.org)



Prototyyppi Android-sovelluksesta, OCR laitteessa  
(tekijä Okko Vainonen)





# Lomake testausta varten: [annif.org](https://annif.org)

## Try Annif!

Text to analyze:

Rinnakkaisohjelma Toiminnallista ja osallistavaa

NÄYTTELYT AVOINNA klo 9–16

AALTO-SALI

9.00–9.20

Koneen kanssa. Keksijä, yrittäjä, säveltäjä Perttu Pölönen

9.20–10.00

Tekoölyn aika ja ihminen. Professori Mika Pantzar.

10.00–10.30

TAUKO, VERKOSTOITUMINEN, TUTUSTUMINEN NÄYTTELYIHIN

Project (vocabulary and language):

YSO ensemble Finnish

Analyze

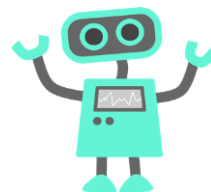
## Results



# 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"}
```

]

**... ja miksi tarvitset sitä?**

# Miksi automatisoida sisällönkuvailua?

- **Datan ja aineiston määrän kasvu**
  - Yleinen digitalisaatio = digitaalisen materiaalin kasvu
- **Metadatan kysynnän kasvu**
- **Henkilöstöressit**
  - Henkilötyövuodet vs. aineiston määrä?
  - Puoliautomasointi vapauttaa resursseja rutiinitehtävistä vaativampaan asiantuntijatyöhön

# Kiitos!

## Kysymyksiä?

[mona.lehtinen@helsinki.fi](mailto:mona.lehtinen@helsinki.fi)  
[osma.suominen@helsinki.fi](mailto:osma.suominen@helsinki.fi)  
<http://annif.org>

**Psst... Annif: DIY automated subject indexing using multiple algorithms:**  
<http://doi.org/10.18352/lq.10285>

<sup>1</sup> Esitys pohjautuu Osma Suomisen esitykseen [Kirjastopäivillä 2019](#)