

# Aiheentunnistustyökalu Annif

Tutkimustietovaranto-hankkeen ohjausryhmä

OKM, 27.5.2019

Osma Suominen

# Extrablad till ÅBO UNDERRÄTTELSE

No 2.  
Finlands oavhängighet.

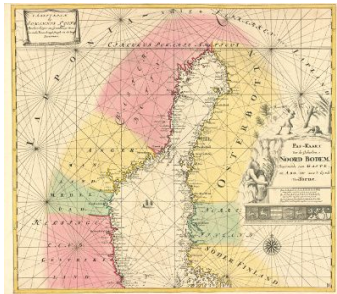
Landsdagen omfattar regeringens proklamation om Finlands fullständiga oavhängighet och ansluter sig till huvudsaken i regeringens program för tryggandet av landets nya ställning. Beslutet fattades med 100 rösterna mot 58 vilka till följande ett av socialdemokraternas förordnande förelag.

**Verkligheten beträffande**  
Finland har varit en del af Sverige sedan 1809. Sedan 1809 har Finland varit en del af Sverige sedan 1809. Sedan 1809 har Finland varit en del af Sverige sedan 1809.

**Årskiftet**  
Årskiftet har varit en del af Sverige sedan 1809. Årskiftet har varit en del af Sverige sedan 1809. Årskiftet har varit en del af Sverige sedan 1809.

**Årskiftet**  
Årskiftet har varit en del af Sverige sedan 1809. Årskiftet har varit en del af Sverige sedan 1809. Årskiftet har varit en del af Sverige sedan 1809.

Pris 25 penn.



JANIS 1856  
IN 1856

Framework for Open Science and Research      3.3.2019      6/196

For example, which physical disciplines will store the digital data resources, who will be responsible using files for data-management systems?

The Kartus EA method used in the higher education sector (described below) has been explained in this work.

Open Level	Open Level	Open Level	Open Level	Open Level
Level 1: <b>Open Access</b>	Level 2: <b>Open Access</b>	Level 3: <b>Open Access</b>	Level 4: <b>Open Access</b>	Level 5: <b>Open Access</b>
Level 6: <b>Open Access</b>	Level 7: <b>Open Access</b>	Level 8: <b>Open Access</b>	Level 9: <b>Open Access</b>	Level 10: <b>Open Access</b>
Level 11: <b>Open Access</b>	Level 12: <b>Open Access</b>	Level 13: <b>Open Access</b>	Level 14: <b>Open Access</b>	Level 15: <b>Open Access</b>
Level 16: <b>Open Access</b>	Level 17: <b>Open Access</b>	Level 18: <b>Open Access</b>	Level 19: <b>Open Access</b>	Level 20: <b>Open Access</b>
Level 21: <b>Open Access</b>	Level 22: <b>Open Access</b>	Level 23: <b>Open Access</b>	Level 24: <b>Open Access</b>	Level 25: <b>Open Access</b>
Level 26: <b>Open Access</b>	Level 27: <b>Open Access</b>	Level 28: <b>Open Access</b>	Level 29: <b>Open Access</b>	Level 30: <b>Open Access</b>
Level 31: <b>Open Access</b>	Level 32: <b>Open Access</b>	Level 33: <b>Open Access</b>	Level 34: <b>Open Access</b>	Level 35: <b>Open Access</b>
Level 36: <b>Open Access</b>	Level 37: <b>Open Access</b>	Level 38: <b>Open Access</b>	Level 39: <b>Open Access</b>	Level 40: <b>Open Access</b>
Level 41: <b>Open Access</b>	Level 42: <b>Open Access</b>	Level 43: <b>Open Access</b>	Level 44: <b>Open Access</b>	Level 45: <b>Open Access</b>
Level 46: <b>Open Access</b>	Level 47: <b>Open Access</b>	Level 48: <b>Open Access</b>	Level 49: <b>Open Access</b>	Level 50: <b>Open Access</b>

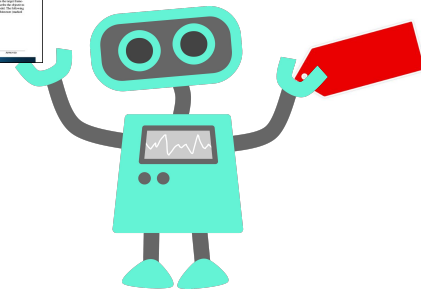
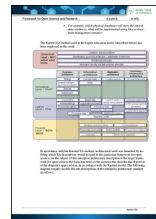
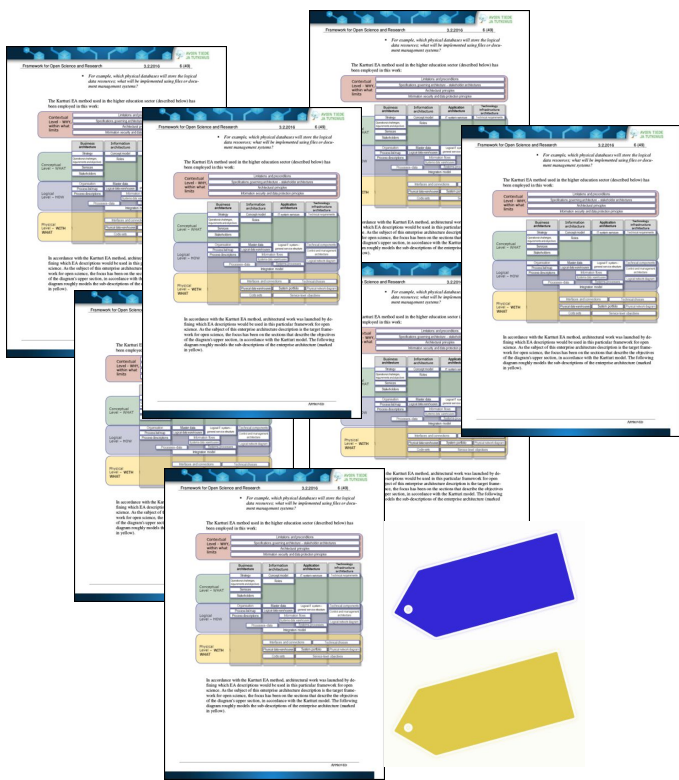
In accordance with the Kartus EA method, extensive work was conducted by the Kartus EA description model by using the physical disciplines for open science. As the objective of the enterprise architecture description is the right model work for open science, the focus has been on the activities that describe the objectives of the digital open science, in accordance with the Kartus model. The following diagram roughly models the sub-description of the enterprise architecture (enabled in others).











**Annif**

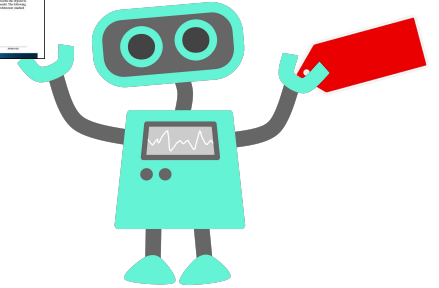
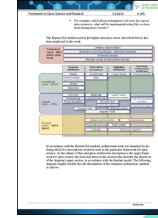
# Annif

**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...

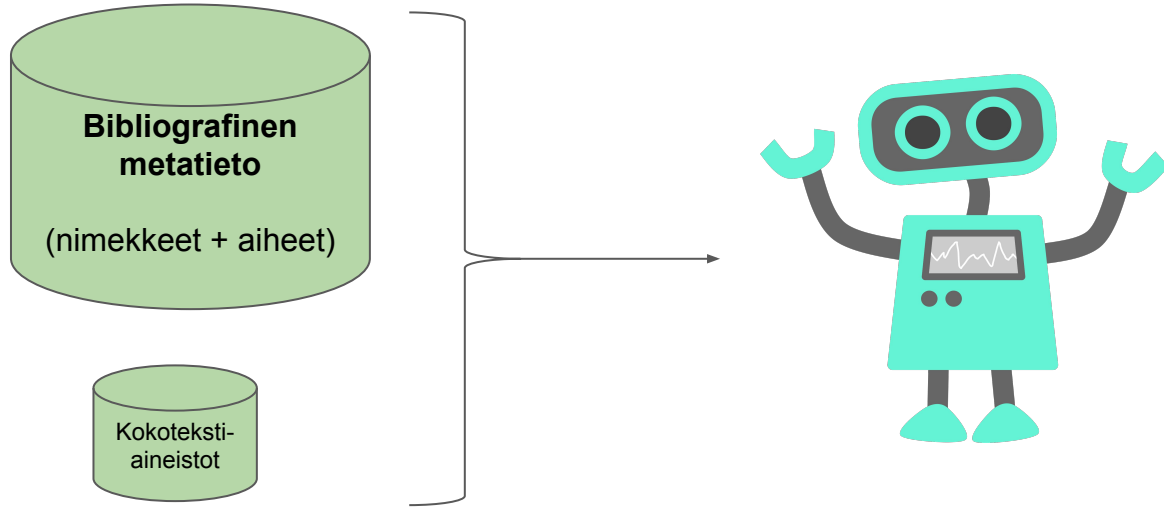
Kaikki osumat ▾



🔍 Tarkennettu haku

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

# Annif oppii aineistosta, miten tunnistaa aiheita



**Mitä Annifilla voi tehdä?**

# 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 Häyrinen

# Wikipedian kuvailu YSOlla

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)

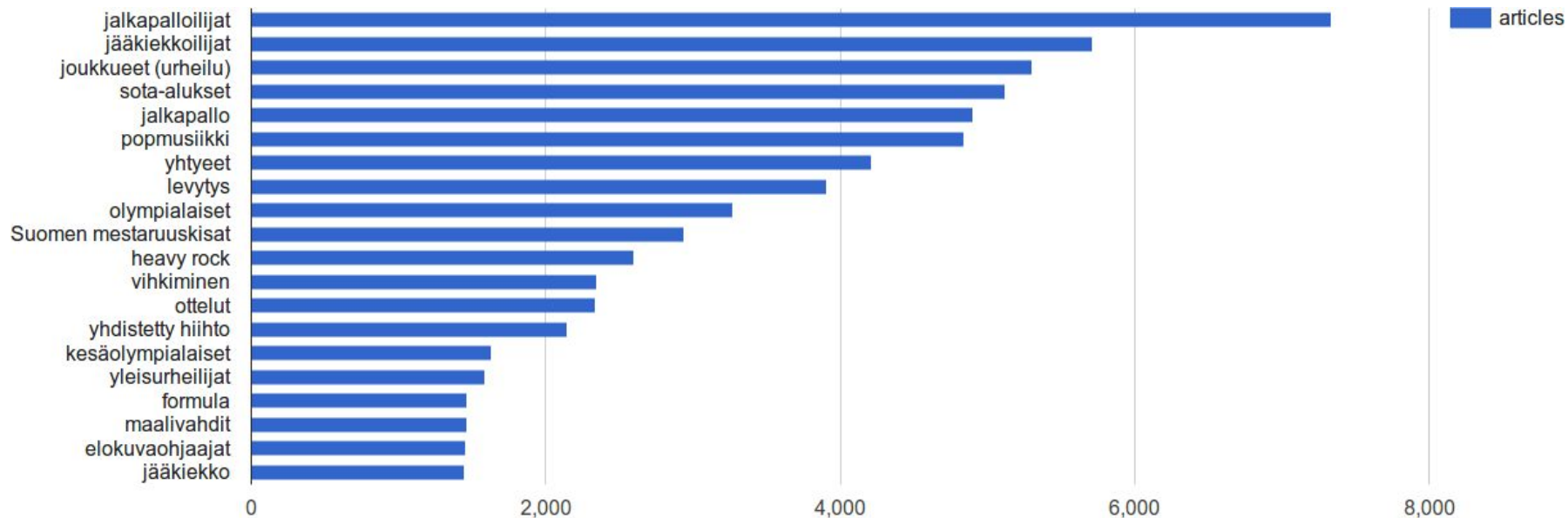
## Wikipedia-artikkeli

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

## YSO-aiheet

rannat  
kilpa-autoilijat, formula, mikroautoilu  
kansanrunoudentutkijat, 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:

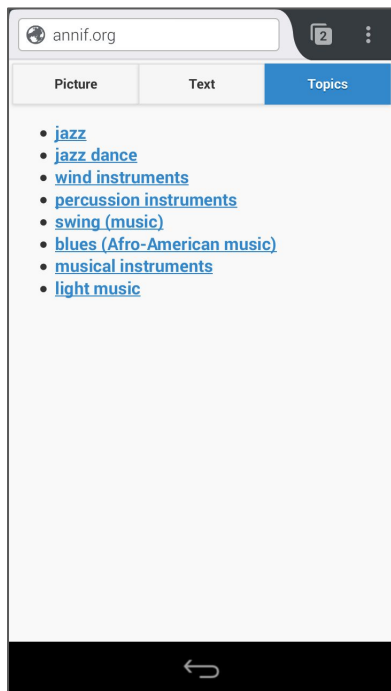
Petteri Lehtonen [CC BY-SA 3.0]

Hockeybroad/Cheryl Adams [CC BY-SA 3.0]

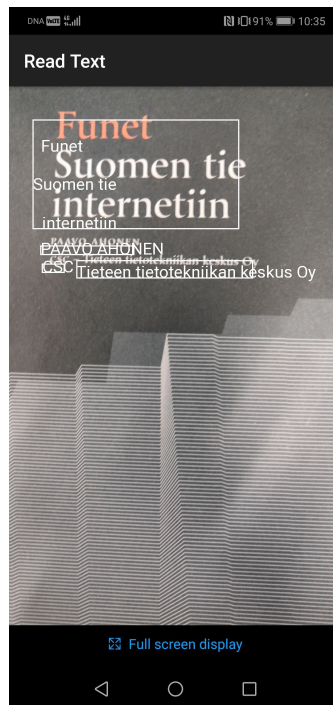
Tomisti [CC BY-SA 3.0]

Tuomas Vitikainen [CC BY-SA 3.0]

# Mobiilisovellukset



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

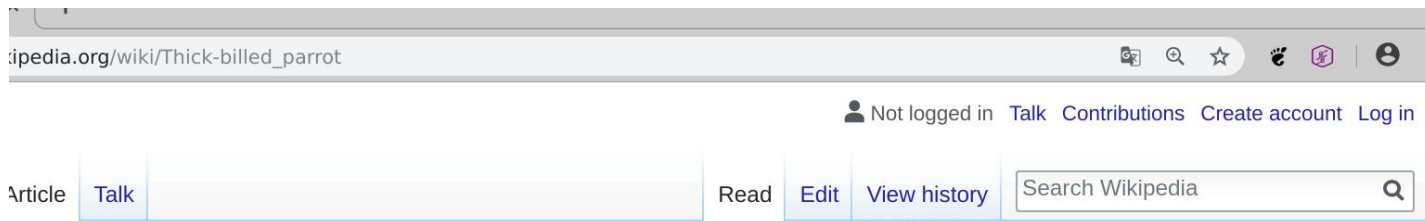


Prototyyppi Android-sovelluksesta, OCR laitteessa  
(tekijä Okko Vainonen)





# Finna Recommends -selainlaajennos Chromelle



## Thick-billed parrot

From Wikipedia, the free encyclopedia

*This page is about the species of parrot. For the genus of parrots, see [Rhynchopsitta](#).*

The **thick-billed parrot** (*Rhynchopsitta pachyrhyncha*) is a medium-sized green and red [parrot](#) found in Mexico, that formerly ranged into the [southwestern United States](#). Its position in parrot [phylogeny](#) is the subject of ongoing discussion; it is sometimes referred to as thick-billed [macaw](#) or thick-billed [conure](#). In Mexico, it is locally called *guacamaya* ("macaw") or *cotorra serrana* ("mountain parrot"). Classified internationally as Endangered through [IUCN](#),<sup>[1]</sup> the thick-billed parrot's decline has been central to multiple controversies over wildlife management.

### Contents [\[hide\]](#)

- [Taxonomy](#)
- [Description](#)



Analysoi  
valitun tekstin  
miltä tahansa sivulta  
Annifin APilla  
ja hakee  
**kirjasuosituksia**  
Finnasta

**WIDE-hackathon**  
Yazan Alhalabi  
Samuel Akangbe  
Steven Nebo

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

## Try Annif!

Text to analyze:

Digitalisoituvan tutkimuksen ja avoimen tieteen myötä erilaisten tutkimustuotosten määrät kasvavat ja tietoja niistä kerätään yhä enenevässä määrin erilaisiin tietojärjestelmiin. Tutkimustuotosten järkevä ja helppo hallinnointi vaatii tutkimusta kuvaavien tietojen yhdenmukaista määrittelyä ja sujuvaa käsittelyä. Tutkimusta kuvaavat tiedot ovatesimerkiksi julkaisuja, tutkimusaineistoja, tutkimusinfrastruktuureita, tutkijoita, hankkeita/projekteja ja tutkimusryhmiä koskevat metatiedot. Tutkimusta kuvaavia tietoja käytetään tutkimustyössä ja tutkimukseen liittyvässä hallinnollisissa prosesseissa, esim. tiedonhauissa, rahoitushauissa ja -raportoinnissa, julkaisutoiminnassa, tilastoinnissa ja muussa tutkimustiedon levittämisessä. Tietovirtojen järkeistäminen vähentää hallinnollista työtaakkaa sekä parantaa tietojen saatavuutta ja löydettävyyttä. Siitä hyötyvät niin tutkijat, tutkimusorganisaatiot ja rahoittajat kuin tietoon perustuvia päätöksiä tekevä hallinto. Lisäksi tutkimustieto on paremmin sitä hyödyntävien kansalaisten saatavilla.

Opetus- ja kulttuuriministeriön tavoitteena on, että tutkimusta kuvaavat tiedot ovat jatkassa sujuvasti hyödynnettävissä erilaisissa palveluissa, tiedot syötetään järjestelmiin vain kerran ja tutkimustietojen syöttämistä automatisoidaan. Tietovirtojen sujuvoittaminen edellyttää avoimia

Project (vocabulary and language):

YSO ensemble Finnish

Analyze

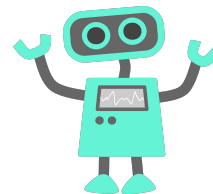
## Results



# 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"}
```

]

# Kiitos!

## Kysymyksiä?

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

Tämä esitys: <https://tinyurl.com/annif-okm>

**Annif: DIY automated subject indexing using multiple algorithms.**

Manuscript, submitted for review.

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