# Annif tutorial @ SWIB19



THE NATIONAL LIBRARY OF FINLAND Introduction to Annif



Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics

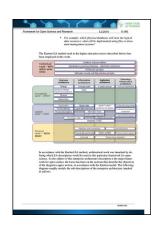
# Subject indexing

- a.k.a. topic indexing, topic assignment
- ~ multi-label classification
- ~ tagging









### Extrablad till Å**ro underrättelser**

n 2. Harris il decenter l

#### Finlands oavhängighet.

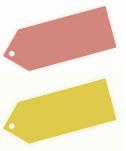
Lantdagen emfattar regeringens problamation om Finlands fullständiga oarhängigbet och andater sig till huvudprinsiperna i regeringens program för tryggandet av landets nya ställning. Beslutet fattades med 100 rönter enns 88 vilka tillfölie ett av socialdemokraterna formulærat förslag.

Ye newfells borrasses dest t Eviling for a welfe far.  Now Brein then vis please t day of 2 septes at av de houseles of at- moneyer transfel Krein old at- mineral transfel old at- mineral transfel old at- mineral transfel old at- mineral transfel old at- mineral colorated.  Nowed to common old attack.	Virual appliance of the strong statement between the statement of the stat	Detects white.  Med Htt river ster to problem.  Med Htt river ster to problem.  Medians to beganter properties.  VA'daddd av arbeitare.  Frenz, Author & Domprios she-	Aterupptagandet av trafiken Abo-Stockholm.
the meg finding till ar requessioner, villen to byggd of day mend, att Polond in on obsessed requiries behinder handlen doorse trackers to day to be an higher entendere on the day following, doors princip mit and films films and the or of the second or or of the second of the secon	Sunderspin borio oppino pinnoli indei piong lis columbia se finitali piong lis columbia se finitali pione indei Pidena se di Protesso Dave darkej broks suderpino tida untersas nabi-filiarente indeis sundeis protectione delle distribution delle distribution delle distribution delle distribution delle distribution di distribution distribution distribution di distribution di distribution di dist	for onto sembrana, or heatenge- mender i fig altima or indicate- riams, celt institude i altimate or finance. Superiore military. Vag sembra military. I den abest since bestea 2 on finance orbitation from the first indicate indicate and the first indicate and the control of the control of the many of the minimum and the minimum and the minimum and the minimum and the minimum and the minimum and the minimum and the minimum and the minimum and the minimum and the minimum and the minimum and the minimum and the minimum and the minimum and the minimum and the minimum and the minimum and the minimum and the	Season Hibilitie ett uthatt.  Enten nich vi erhet har ornans bebender frage, sir appaapski av author sit here for-Docketen dere Under er sicher for topen, blanding, Underen sitt great blanding, Underen sit jurar blatt er er sicher sit har an ensistent oppning sachebunder oppning sachebunder sowiester blatt er er derbeit in der er er blatt nach har er er er er er er er blatt nach har er er er er er er er er er er e
edite hive oblind vickers de hi- ghter reparages medicht ware	often op skerweit result oft respire and des Perrodoscul- ter for explosiones cellusis. For	white anyon I upto an item to believe to be believed to the first transfer.	Charles Serie sales les lons ill et yolder six, abre soules series traines for religiousle so

Pris 25 penal.

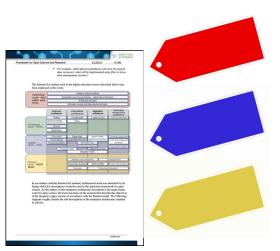












### Extrablad till Åbo underrättelser

Marrie S

#### Finlands oavhängighet.

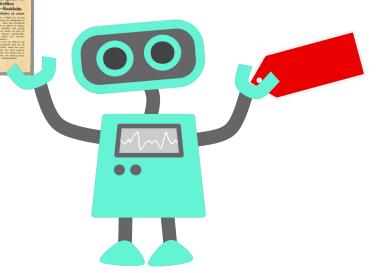
Landagen omfattar regeringens produmation om finlands fullständiga oarhängigbet och ansluter sig till havudprinsjerna i regeringens program för tryggandet av landets nya ställning. Beslutte fattades med 100 rönter omst 85 vilka tillfölle ett av socialdemskrateran formåerat förslag.

Var amartalla kontranta.	A Narial Samakraterra.	ivelens ti intono palo-	Advert till tellar heldstene en
ent t Boternatere medce-	vagnar upplänten ett	sands follow landages tibitis etc.	do not been use all places to
S reberer on the de production	with the alliture Philosophis beating prin-	settion is beauty grape on	Aterupotavandet

continuentes. Il re-controlle de l'acceptant de l'a

Widelid or wieders,
The second of the second

Pris 25 pensi



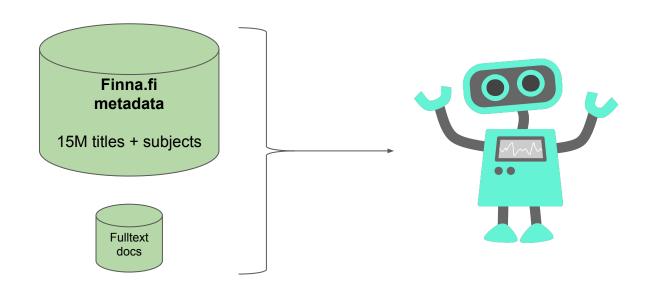
# **Idea of Annif**



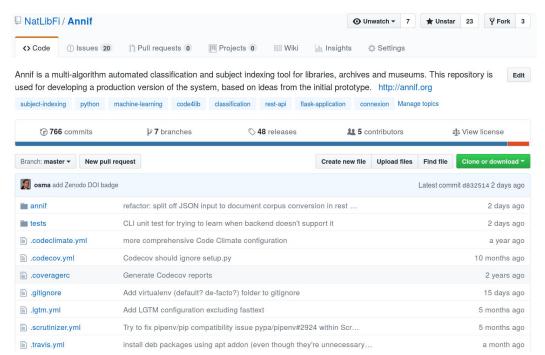


We have a lot of LAM metadata, e.g. 15M records in Finna.fi discovery service

# Machine learning using library data



#### **Annif Architecture** Short text / metadata training data training data Any metadata / document management Indexed system full text Flask/ documents nn\_ensemble tfidf Connexion **REST API** web app training fasttext pav data Configuration Mobile module vw ensemble vw multi app Analyzer modules ensemble **REST API** maui Maui Server Vocabulary regular backends fusion backends module O¢R Microservice around Maui NEWS ■ 1 5 Project Subject configuration vocabulary



### **Annif on GitHub**

Python 3.5+ code base Apache License 2.0

Fully unit tested (99% coverage)
PEP8 style guide compliant
Usage <u>documentation</u> in the wiki

https://github.com/NatLibFi/Annif

## Accessing Annif

- **Command line interface** setup and administration
  - training models
  - testing and evaluating models
  - bulk indexing of documents

### Web user interface

- interactive testing of models

### REST API

- integrating Annif services to other systems

# **Apply Annif on your own data!**

Choose an indexing vocabulary

Prepare a corpus from your existing metadata

Train a model using the corpus

Use it to index new documents