



Annif tutorial @ SWIB19



Introduction to Annif



Leibniz-Informationszentrum
Wirtschaft
Leibniz Information Centre
for Economics

Subject indexing

a.k.a. topic indexing, topic assignment

~ multi-label classification


~ tagging

Idea of Annif



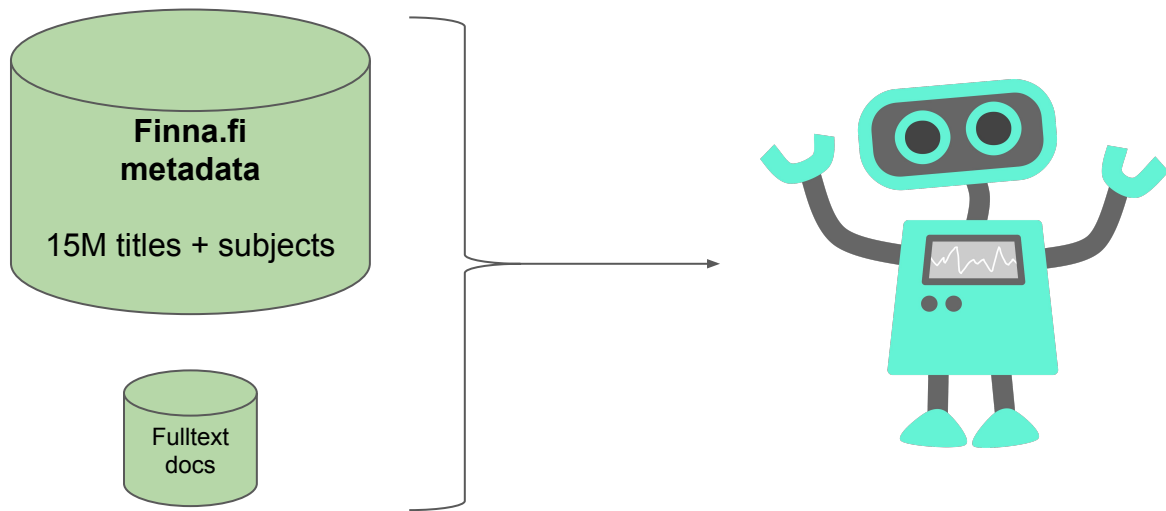
The collections of Finnish archives, libraries
and museums at your fingertips.

All fields ▾

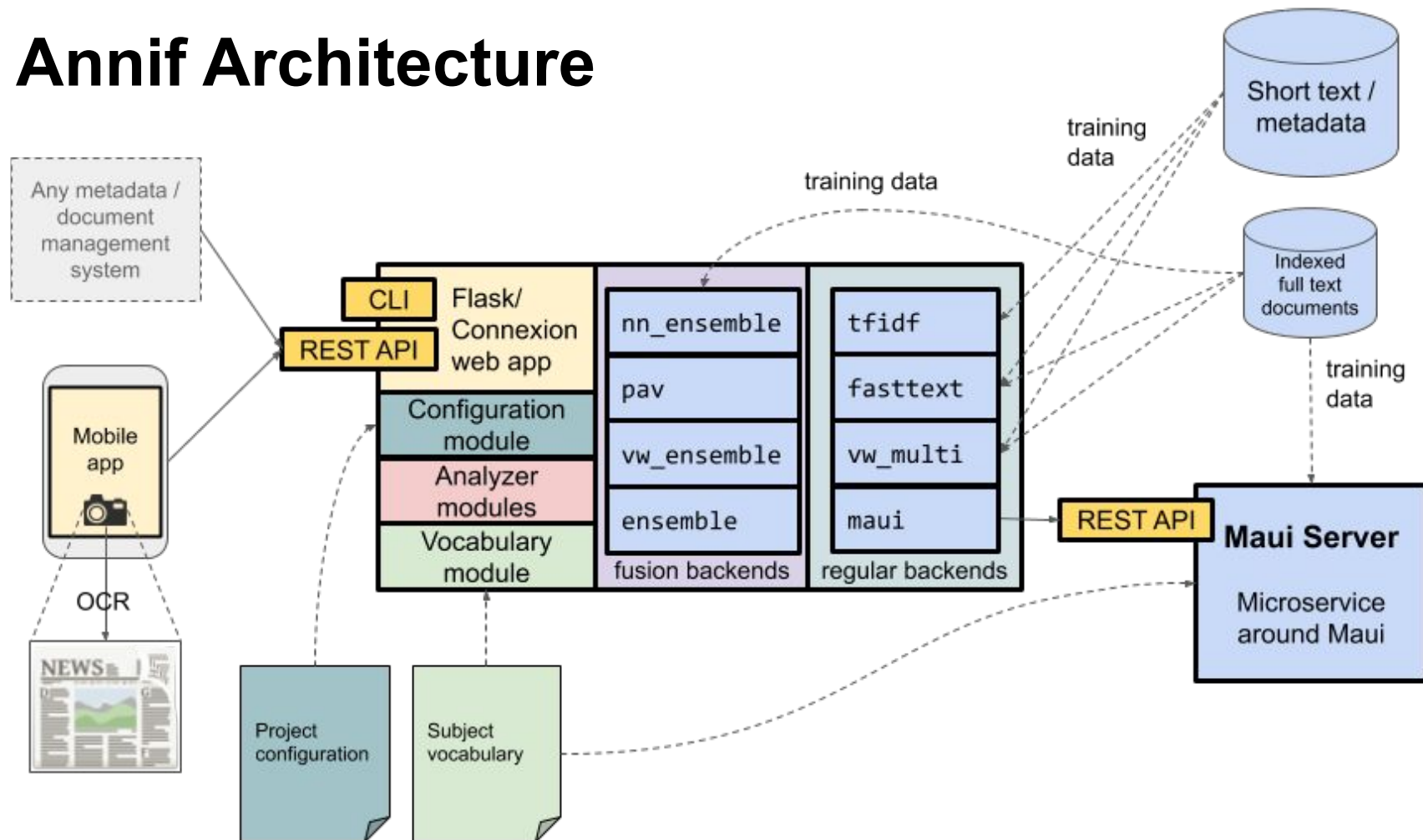
 Advanced search

We have **a lot** of LAM metadata, e.g. **15M records** in [Finna.fi](https://finna.fi) discovery service

Machine learning using library data



Annif Architecture



NatLibFi / Annif

Unwatch 7 Unstar 23 Fork 3

Code Issues 20 Pull requests 0 Projects 0 Wiki Insights Settings

Annif is a multi-algorithm automated classification and subject indexing tool for libraries, archives and museums. This repository is used for developing a production version of the system, based on ideas from the initial prototype. <http://annif.org>

subject-indexing python machine-learning code4lib classification rest-api flask-application connexion Manage topics

766 commits 7 branches 48 releases 5 contributors View license

Branch: master New pull request Create new file Upload files Find file Clone or download

osma add Zenodo DOI badge Latest commit d832514 2 days ago

annif	refactor: split off JSON input to document corpus conversion in rest ...	2 days ago
tests	CLI unit test for trying to learn when backend doesn't support it	2 days ago
.codeclimate.yml	more comprehensive Code Climate configuration	a year ago
.codecov.yml	Codecov should ignore setup.py	10 months ago
.coveragerc	Generate Codecov reports	2 years ago
.gitignore	Add virtualenv (default? de-facto?) folder to gitignore	15 days ago
.lgtn.yml	Add LGTM configuration excluding fasttext	5 months ago
.scrutinizer.yml	Try to fix pipenv/pip compatibility issue pypa/pipenv#2924 within Scr...	5 months ago
.travis.yml	install deb packages using apt addon (even though they're unnecessary...	a month ago

Annif on GitHub

Python 3.5+ code base

Apache License 2.0

Fully unit tested (99% coverage)

PEP8 style guide compliant

Usage [documentation](#) 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

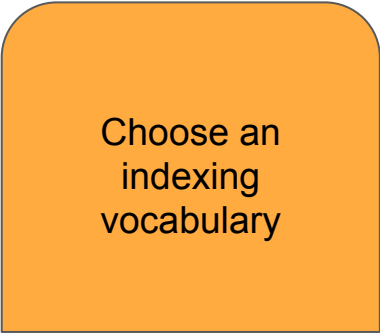
Web user interface

- interactive testing of models

REST API

- integrating Annif services to other systems

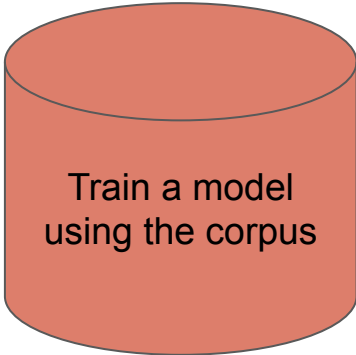
Apply Annif on your own data!




Choose an
indexing
vocabulary



Prepare a
corpus
from your
existing
metadata



Train a model
using the corpus



Use it to index
new documents