annif tutorial

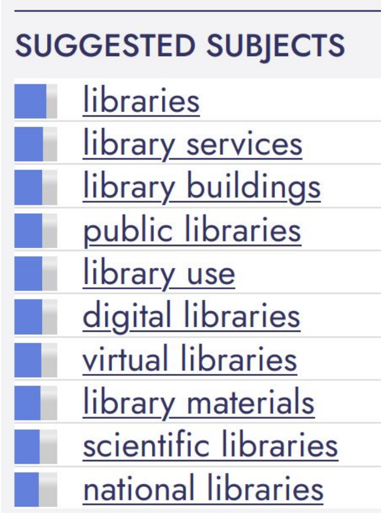
Metrics & evaluation

1640
THE NATIONAL
LIBRARY
OF FINLAND

ZBW
Leibniz-Informationszentrum
Wirtschaft
Leibniz Information Centre
for Economics

How well an Annif model works?

Manual grading

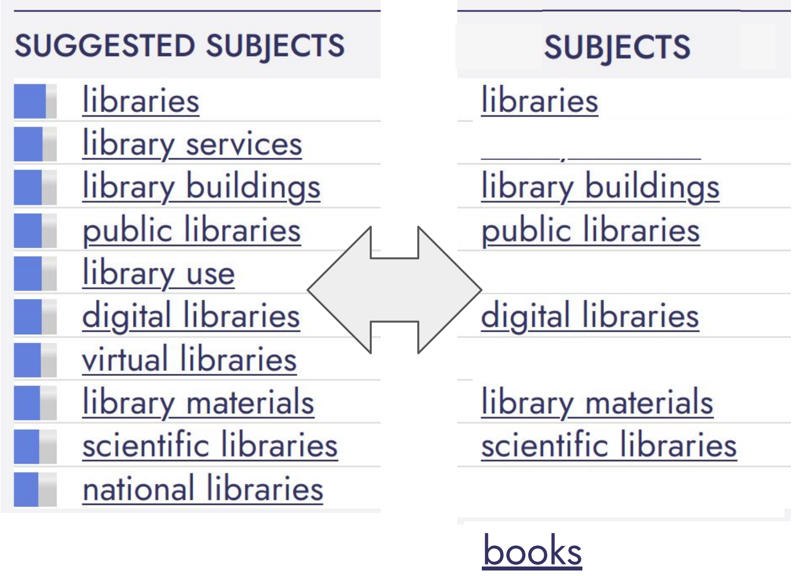


SUGGESTED SUBJECTS

- libraries
- library services
- library buildings
- public libraries
- library use
- digital libraries
- virtual libraries
- library materials
- scientific libraries
- national libraries

A-

Automatic comparison to gold-standard subjects by human



SUGGESTED SUBJECTS

- libraries
- library services
- library buildings
- public libraries
- library use
- digital libraries
- virtual libraries
- library materials
- scientific libraries
- national libraries

SUBJECTS

- libraries
- library buildings
- public libraries
- digital libraries
- library materials
- scientific libraries
- books

metric → **0.942**

Metrics in brief

Metrics provide numeric values that can be compared easily. We use metrics from machine learning / information retrieval.

In this tutorial, we will consider the following:

- precision & recall
- F1 score
- Normalized Discounted Cumulative Gain (NDCG)

These metrics take values between 0.0 – *worst*, and 1.0 – *best*.

Precision, recall and F1 score

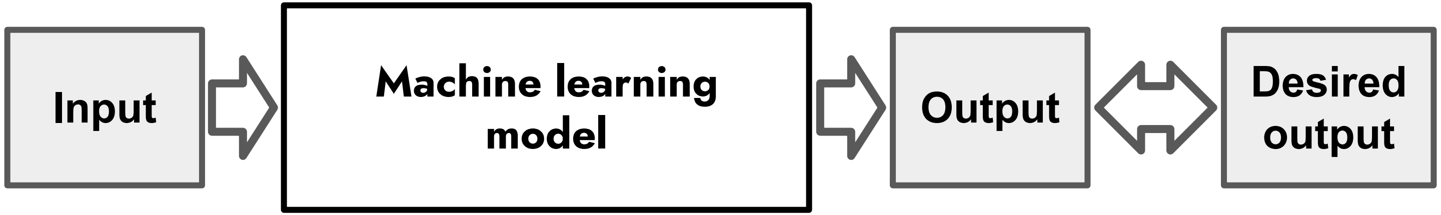
- **Precision:** fraction of the correct subjects among the subjects suggested
  “How many of the suggested subjects are actually correct?”
- **Recall:** fraction of all correct subjects that were actually suggested
  “How many of those subjects that should be suggested have actually been suggested?”
- The **F1 score** is the harmonic mean between precision and recall
  (i.e., a way of combining precision and recall values into one).

NDCG – Normalized Discounted Cumulative Gain

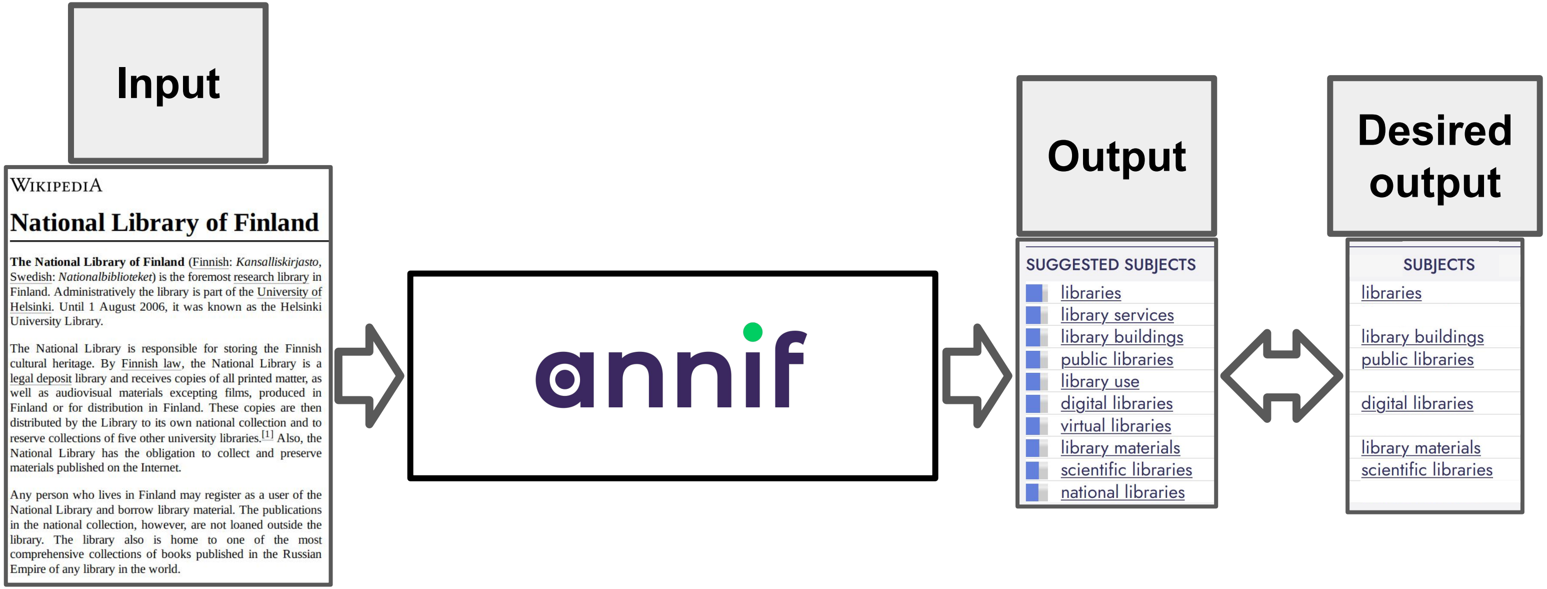
The **NDCG** is a ranking-based measure, i.e., the order of the subjects suggested is taken into account:

Getting the top ranked (highest score) result right will matter more than getting the 2nd or 3rd right.

Test, train and validation data



Test, train and validation data



Wikipedia

National Library of Finland

The National Library of Finland (Finnish: *Kansalliskirjasto*, Swedish: *Nationalbiblioteket*) is the foremost research library in Finland. Administratively the library is part of the University of Helsinki. Until 1 August 2006, it was known as the Helsinki University Library.

The National Library is responsible for storing the Finnish cultural heritage. By Finnish law, the National Library is a legal deposit library and receives copies of all printed matter, as well as audiovisual materials excepting films, produced in Finland or for distribution in Finland. These copies are then distributed by the Library to its own national collection and to reserve collections of five other university libraries.[1] Also, the National Library has the obligation to collect and preserve materials published on the Internet.

Any person who lives in Finland may register as a user of the National Library and borrow library material. The publications in the national collection, however, are not loaned outside the library. The library also is home to one of the most comprehensive collections of books published in the Russian Empire of any library in the world.

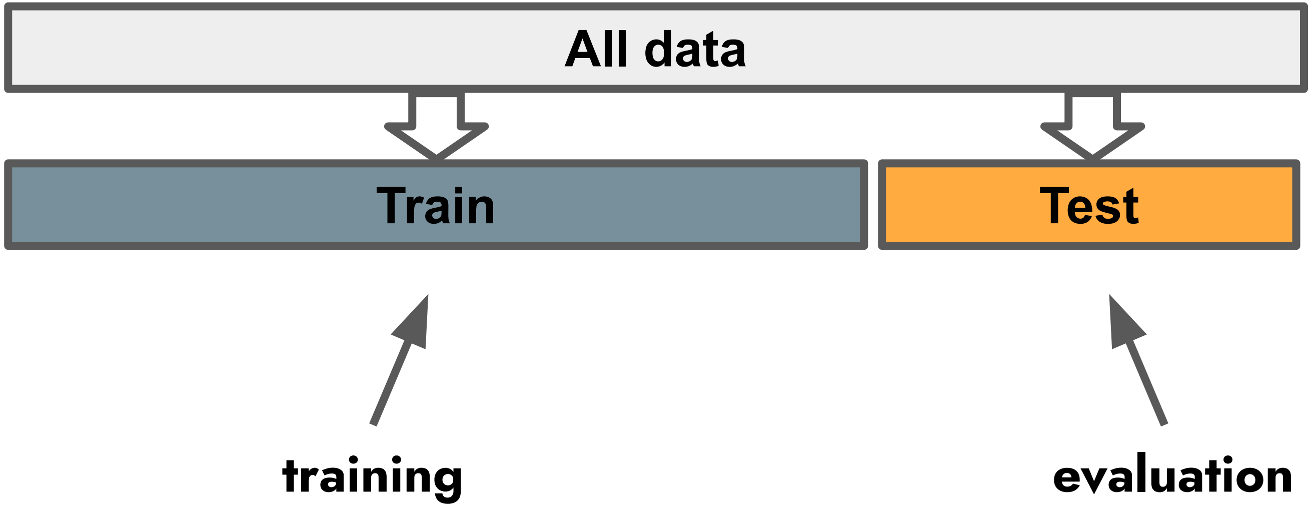
SUGGESTED SUBJECTS

- libraries
- library services
- library buildings
- public libraries
- library use
- digital libraries
- virtual libraries
- library materials
- scientific libraries
- national libraries

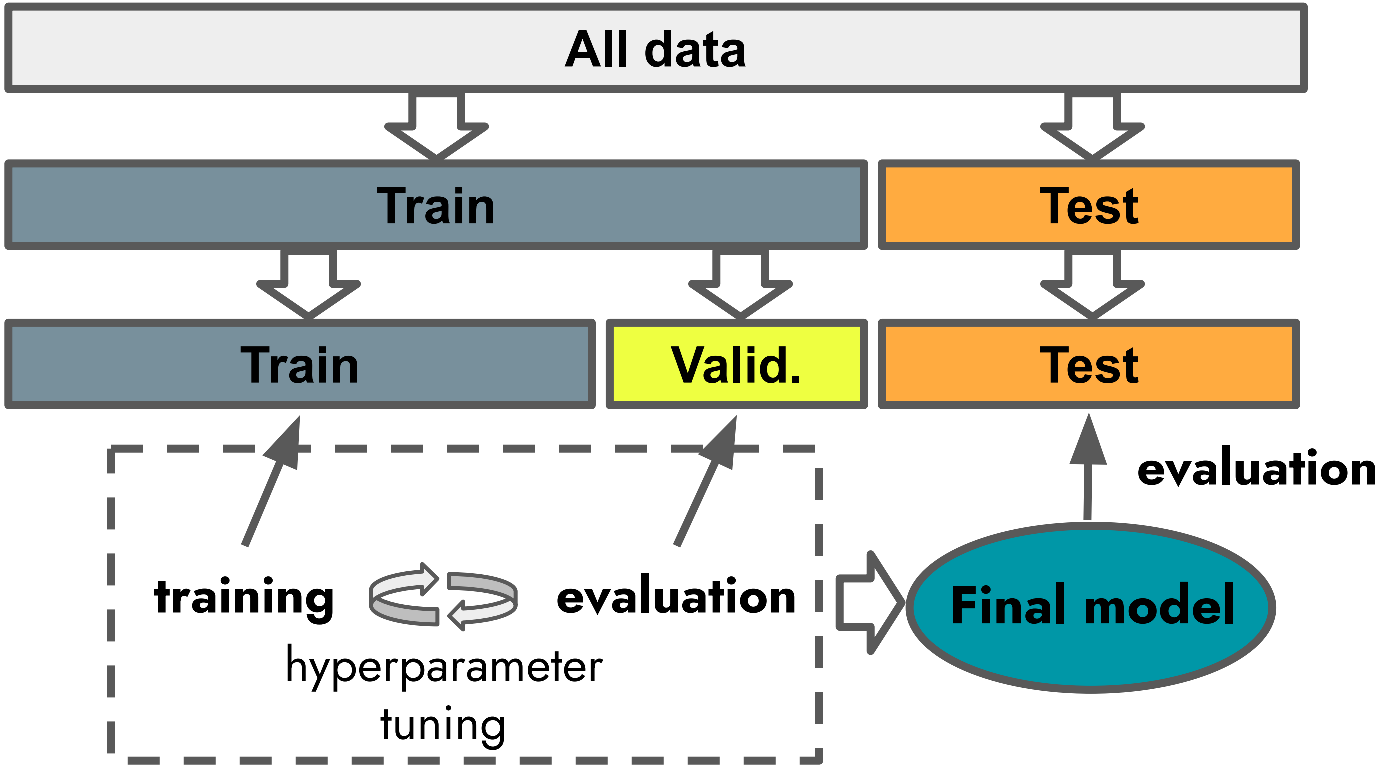
SUBJECTS

- libraries
- library buildings
- public libraries
- digital libraries
- library materials
- scientific libraries

Test, train and validation data



Test, train and validation data

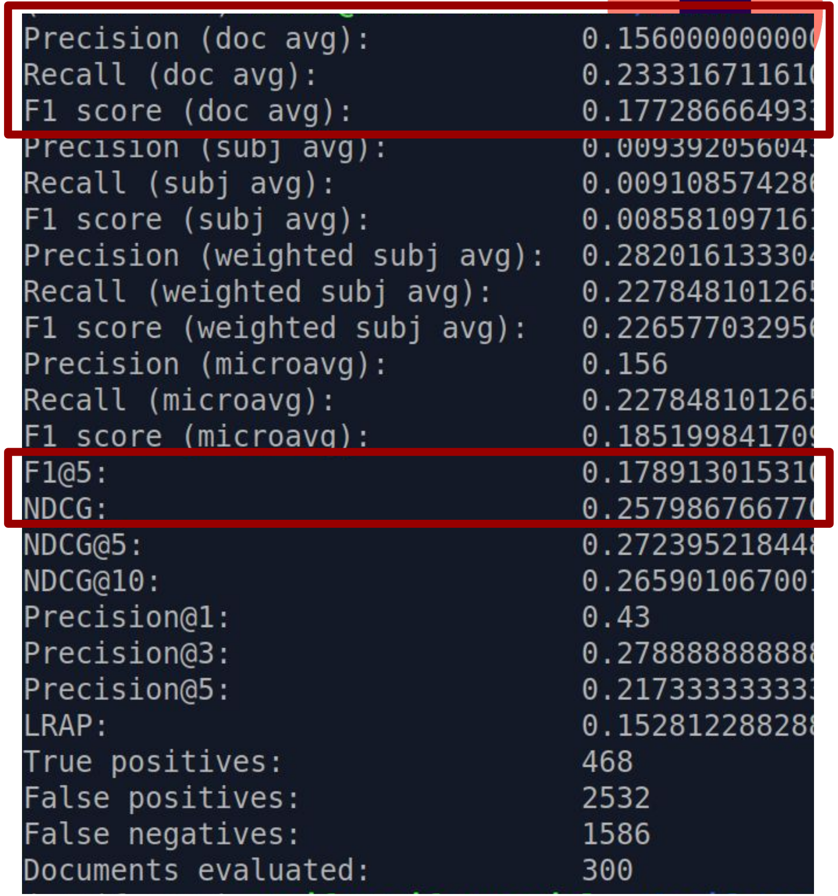


Evaluation in Annif

Annif has a built-in command for evaluation:

```
annif eval project_id path/to/docs
```

Output is a report with several metrics



```
Precision (doc avg):              0.156000000000
Recall (doc avg):                 0.233316711610
F1 score (doc avg):               0.177286664933
Precision (subj avg):             0.009392056045
Recall (subj avg):                0.009108574286
F1 score (subj avg):              0.008581097161
Precision (weighted subj avg):    0.282016133304
Recall (weighted subj avg):       0.227848101265
F1 score (weighted subj avg):     0.226577032956
Precision (microavg):             0.156
Recall (microavg):                0.227848101265
F1 score (microavg):              0.185199841709
F1@5:                             0.178913015310
NDCG:                             0.257986766770
NDCG@5:                           0.272395218448
NDCG@10:                          0.265901067001
Precision@1:                      0.43
Precision@3:                      0.278888888888
Precision@5:                      0.217333333333
LRAP:                             0.152812288288
True positives:                   468
False positives:                  2532
False negatives:                  1586
Documents evaluated:              300
```